# EXÁMENES

## Curso 2011/12

**Cuestiones** (45 minutos, 5 puntos)

1. Se ha realizado una encuesta entre los estudiantes de grado del MIT (Massachusetts Institute of Technology) para conocer sus preferencias tecnológicas. El 35% de los entrevistados tienen un iPhone y un iPad, el 80% tiene al menos uno de estos dispositivos y el 60% no tiene iPad. Para un estudiante elegido al azar, calcula la probabilidad de que:

   a. Disponga de iPhone y no de iPad.

   b. Tenga un iPad pero no un iPhone.

   c. Tenga únicamente uno de los dos dispositivos.

   d. No disponga de ninguno de los dos.

2. Un profesor tiene las notas de un examen de sus 100 alumnos en una hoja de cálculo. La primera columna corresponden a la nota del primer ejercicio ($x_i$), la segunda columna corresponde a las notas del segundo ejercicio ($y_i$) y en la tercera columna se encuentra el promedio de las notas de los dos ejercicios:

$$z_i = \frac{1}{2}x_i + \frac{1}{2}y_i$$

   La media del primer ejercicio ha sido $\bar{x} = 7.2$ con desviación típica $s_x = 1.2$ y la nota media del segundo ejercicio es $\bar{y} = 5.4$ con deviación típica $s_y = 0.81$. El profesor ha calculado el coeficiente de correlación entre las notas del primer y segundo ejercicio y ha resultado igual a 0.75.

   (a) Demuestra que:

$$s_z^2 = \frac{1}{4}s_x^2 + \frac{1}{4}s_y^2 + \frac{1}{2}s_{xy}$$

   (b) La nota del examen la ha obtenido dando pesos iguales a los dos ejercicios, está considerando la posibilidad de utilizar como nota final

$$u_i = a\,x_i + (1-a)\,y_i$$

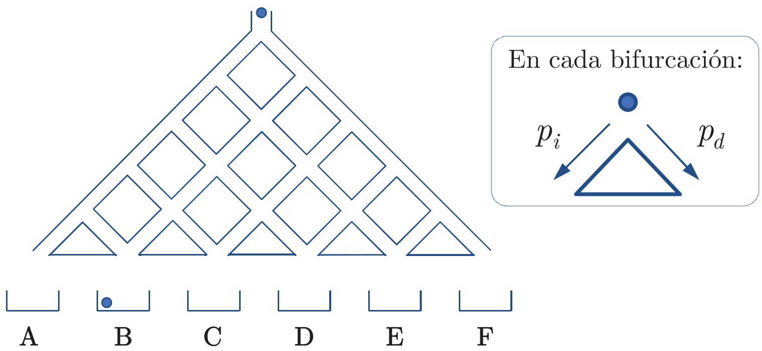
   utilizando el valor $a$, tal que $0 \leq a \leq 1$, que hace que la varianza $s_u^2$ sea lo más grande posible. Calcular $a$.

3. Disponemos de un juego de canicas como el que se muestra en la figura. La canica se desplaza en sentido descendente, pudiendo ir hacia la derecha o hacia la izquierda en cada bifurcación, con probabilidades $p_d$ y $p_i$ respectivamente.

   Se pide calcular la probabilidad de que la canica caiga en el cesto **B**:

   (a) Suponiendo que $p_d = p_i = 0.5$

   (b) Suponiendo que $p_d = 0.4$ y $p_i = 0.6$



**Problema**

(45 minutos, 5 puntos)

Desde un restaurante de comida rápida se envían los pedidos en moto. La función de densidad de la distancia recorrida a lo largo de la carretera, $X$, es

$$f_X(x) = ke^{-x}, \quad 0 < x < \infty$$

1. Calcular el valor de $k$, así como la función de distribución de $X$.
2. ¿Cuál es la probabilidad de que en un viaje tenga que recorrer más de 2km?

   Si realiza 10 viajes, ¿cuál es la probabilidad de que en el viaje más largo recorra más de 2km?
3. Se han añadido al negocio dos restaurantes en otras dos carreteras. Las variables aleatorias $Y$ y $Z$ correspondientes a las distancias recorridas en cada uno tienen las siguientes funciones de densidad:

   $$f_Y(y) = 2e^{-2y}, \quad 0 < y < \infty$$

   y

   $$f_Z(z) = 3e^{-3z}, \quad 0 < z < \infty$$

   Las probabilidades de que a un motorista se le asigne trabajar en las 3 carreteras son $p_X = 0.4$, $p_Y = 0.3$ y $p_Z = 0.3$ y en el último reparto recorrió más de 2 km, ¿cuál es la probabilidad de que lo hiciese en la segunda carretera?

**Cuestiones** (45 minutos, 5 puntos)

1. Se ha realizado una encuesta entre los estudiantes de grado del MIT (Massachusetts Institute of Technology) para conocer sus preferencias tecnológicas. El 35 % de los entrevistados tienen un iPhone y un iPad, el 80 % tiene al menos uno de estos dispositivos y el 60 % no tiene iPad. Para un estudiante elegido al azar, calcula la probabilidad de que:

   a. Disponga de iPhone y no de iPad.

   b. Tenga un iPad pero no un iPhone.

   c. Tenga únicamente uno de los dos dispositivos.

   d. No disponga de ninguno de los dos.

   *Solución:*

   Suceso A = "Tiene iPhone".
   Suceso B = "Tiene iPad".

   Datos:
   $P(A \cap B) = 0{,}35$.
   $P(A \cup B) = 0{,}80$.
   $P(\overline{B}) = 0{,}60$.

   a) Deseamos calcular $P(A \cap \overline{B})$, y sabemos

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$

   Entonces,

$$P(A \cap \overline{B}) = P(A) - P(A \cap B)$$

   Para calcular $P(A)$ utilizamos los datos que nos proporciona el enunciado:

$$\begin{aligned} P(A) &= P(A \cup B) - P(B) + P(A \cap B) \\ &= 0{,}80 - 0{,}40 + 0{,}35 \\ &= 0{,}75 \end{aligned}$$

   Por tanto,

$$\begin{aligned} P(A \cap \overline{B}) &= P(A) - P(A \cap B) \\ &= 0{,}75 - 0{,}35 \\ &= 0{,}40 \end{aligned}$$

   b) Deseamos calcular $P(B \cap \overline{A})$. Sabemos

$$P(B) = P(B \cap A) + P(B \cap \overline{A})$$

   Por tanto:

$$\begin{aligned} P(B \cap \overline{A}) &= P(B) - P(A \cap B) \\ &= 0{,}40 - 0{,}35 \\ &= 0{,}05 \end{aligned}$$

c) Ahora calculamos la probabilidad de que sólo tengan uno de estos dispositivos:

$$P(A \cap \overline{B}) + P(\overline{A} \cap B) \quad = \quad 0{,}45$$

d) Se pueden dar las siguientes situaciones:

$$\left.\begin{array}{lcc} A \cap B & \rightarrow & 0{,}35 \\ A \cap \overline{B} & \rightarrow & 0{,}40 \\ \overline{A} \cap B & \rightarrow & 0{,}05 \\ \overline{A} \cap \overline{B} & \rightarrow & ? \end{array}\right\} \quad P(\overline{A} \cap \overline{B}) = 1 - 0{,}8 = 0{,}2$$

2. Un profesor tiene las notas de un examen de sus 100 alumnos en una hoja de cálculo. La primera columna corresponden a la nota del primer ejercicio ($x_i$), la segunda columna corresponde a las notas del segundo ejercicio ($y_i$) y en la tercera columna se encuentra el promedio de las notas de los dos ejercicios:

$$z_i=\frac{1}{2}x_i+\frac{1}{2}y_i$$

La media del primer ejercicio ha sido $\bar{x} = 7{,}2$ con desviación típica $s_x = 1{,}2$ y la nota media del segundo ejercicio es $\bar{y} = 5{,}4$ con deviación típica $s_y = 0{,}81$. El profesor ha calculado el coeficiente de correlación entre las notas del primer y segundo ejercicio y ha resultado igual a 0.5.

*a*) Demuestra que:

$$s_z^2=\frac{1}{4}s_x^2+\frac{1}{4}s_y^2+\frac{1}{2}s_{xy}$$

*b*) La nota del examen la ha obtenido dando pesos iguales a los dos ejercicios, está considerando la posibilidad de utilizar como nota final

$$u_i = a\,x_i + (1-a)\,y_i$$

utilizando el valor $a$, tal que $0 \leq a \leq 1$, que hace que la varianza $s_u^2$ sea lo más grande posible. Calcular $a$.

*Solución:*

Tenemos:

$$z_i = \alpha x_i + \beta y_i$$

con $\alpha = 1/2$ y $\beta = 1/2$.

Por tanto:

$$\begin{aligned}
\overline{z} &= \frac{\sum z_i}{n} \\
&= \frac{\sum(\alpha x_i + \beta y_i)}{n} \\
&= \alpha\frac{\sum x_i}{n} + \beta\frac{\sum y_i}{n} \\
&= \alpha\overline{x} + \beta\overline{y}
\end{aligned}$$

$$\begin{aligned}
s_z^2 &= \frac{\sum\left(z_i - \overline{z}\right)^2}{n} \\
&= \frac{\sum\left(\alpha x_i + \beta y_i - \alpha\overline{x} - \beta\overline{y}\right)^2}{n} \\
&= \frac{\sum\left(\alpha\left(x_i - \overline{x}\right) + \beta\left(y_i - \overline{y}\right)\right)^2}{n} \\
&= \alpha^2\frac{\sum\left(x_i - \overline{x}\right)^2}{n} + \beta^2\frac{\sum\left(y_i - \overline{y}\right)^2}{n} + 2\alpha\beta\frac{\sum\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{n} \\
&= \alpha^2 s_x^2 + \beta^2 s_y^2 + 2\alpha\beta s_{xy}
\end{aligned}$$

Si ahora substituyo $\alpha = 1/2$ y $\beta = 1/2$, obtenemos lo que me piden:

$$s_z^2 = \frac{1}{4}s_x^2 + \frac{1}{4}s_y^2 + \frac{1}{2}s_{xy}$$

Ahora, substituyendo con los valores del re3sultado, obtengo lo que me piden:

$$s_z^2 = \frac{1}{4}1{,}2^2 + \frac{1}{4}0{,}81^2 + \frac{1}{2}0{,}5 \cdot 1{,}2 \cdot 0{,}81 = 0{,}767$$

Ahora, para $\alpha = a$ y $\beta = 1 - a$, tenemos:

$$s_z^2 = a^2 s_x^2 + (1-a)^2 s_y^2 + 2a(1-a)s_{xy}$$

Para calcular el máximo con respecto a la variable $a$, igualamos a cero la primera derivada:

$$\frac{ds_z^2}{da} = s_z^2 = 2as_x^2 - 2(1-a)s_y^2 + 2a(1-2a)s_{xy} = 0$$

y obtenemos

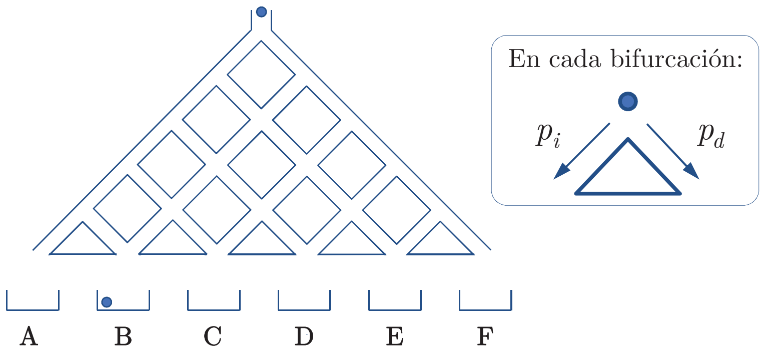
$$a = \frac{s_y^2 - s_{xy}}{s_x^2 + s_y^2 - 2s_{xy}} = 0{,}1513$$

3. Disponemos de un juego de canicas como el que se muestra en la figura. La canica se desplaza en sentido descendente, pudiendo ir hacia la derecha o hacia la izquierda en cada bifurcación, con probabilidades $p_d$ y $p_i$ respectivamente.

   Se pide calcular la probabilidad de que la canica caiga en el cesto **B**:

   *a*) Suponiendo que $p_d = p_i = 0{,}5$

   *b*) Suponiendo que $p_d = 0{,}4$ y $p_i = 0{,}6$




*Solución:*

Para que la canica llegue a las bandejas, ha de pasar por cinco bifurcaciones. En cada bifurcación, puede ir hacia la derecha (D) o hacia la izquierda (I). Por ejemplo, para caer en la celda A, en todas las bifurcaciones ha de ir hacia la izquierda:

$$I_1I_2I_3I_4I_5$$

Para caer en la celda B, en una de las bifurcaciones ha de ir a la derecha, y las cuatro restantes a la izquierda. El salto a la derecha lo puede dar en cualquiera de las cinco bifurcaciones, es decir, que se pueden dar estos cinco casos posibles:

$$\begin{array}{c} D_1I_2I_3I_4I_5 \\ I_1D_2I_3I_4I_5 \\ I_1I_2D_3I_4I_5 \\ I_1I_2I_3D_4I_5 \\ I_1I_2I_3I_4D_5 \end{array}$$

Como se puede observar, hay $\begin{pmatrix} 5 \\ 1 \end{pmatrix}$ formas diferentes de colocar “el salto a la derecha” en un conjunto de cinco saltos.

La probabilidad de cualquiera de estos sucesos es $p_I^4 p_D$, pues son independientes.

La probabilidad que se pide, por tanto, es:

$$P(B) = 5p_I^4 p_D = \begin{cases} a) & 5 \cdot 0{,}5^4 \cdot 0{,}5 = 0{,}1563 \\ b) & 5 \cdot 0{,}6^4 \cdot 0{,}4 = 0{,}2592 \end{cases}$$

**Problema**
(45 minutos, 5 puntos)

Desde un restaurante de comida rápida se envían los pedidos en moto. La función de densidad de la distancia recorrida a lo largo de la carretera, $X$, es

$$f_X(x) = ke^{-x}, \quad 0 < x < \infty$$

1. Calcular el valor de $k$, así como la función de distribución de $X$.
2. ¿Cuál es la probabilidad de que en un viaje tenga que recorrer más de 2km?
   Si realiza 10 viajes, ¿cuál es la probabilidad de que en el viaje más largo recorra más de 2km?
3. Se han añadido al negocio dos restaurantes en otras dos carreteras. Las variables aleatorias $Y$ y $Z$ correspondientes a las distancias recorridas en cada uno tienen las siguientes funciones de densidad:
$$f_Y(y) = 2e^{-2y}, \quad 0 < y < \infty$$
   y
$$f_Z(z) = 3e^{-3z}, \quad 0 < z < \infty$$
   Las probabilidades de que a un motorista se le asigne trabajar en las 3 carreteras son $p_X = 0{,}4$, $p_Y = 0{,}3$ y $p_Z = 0{,}3$ y en el último reparto recorrió más de 2 km, ¿cuál es la probabilidad de que lo hiciese en la segunda carretera?

*Solución:*

1. Una de las propiedades de cualquier función de distribución es que su área sea uno. Por tanto, hacemos que la función $F_X(x)$ cumpla esta propiedad:
$$\int_0^\infty ke^{-x}dx = 1 \quad \Rightarrow \quad k = 1$$
   La función de distribución de X se calcula de la siguiente manera:
$$F_X(x) = P(X \leq x) = \int_0^x f_X(x)dx = 1 - e^{-x} \ , \ x \geq 0$$
2. La probabilidad de que un viaje sea mayor que dos kilómetros, se calcula como sigue:
$$P(X > 2) = \int_2^\infty f_X(x)dx = e^{-2} = 0{,}1353$$
   Para que de 10 viajes el más largo sea como máximo de dos kilómetros, implica que todos los viajes tienen que ser menores de dos kilómetros. Suponiendo que son independientes:
$$P(\text{Todos menores de 2 kilómetros}) = (1 - e^{-2})^{10} = 0{,}2336$$
   Para calcular la probabilidad de que en un viaje tenga que recorrer más de dos kilómetros, utilizamos la propiedad de que es la probabilidad complementaria a la calculada anteriormente. Es decir:
$$P(\text{Alguno mayor de 2 kilómetros}) = 1 - P(\text{Todos menores de 2 kilómetros})$$
   y, por tanto,
$$P(\text{Alguno mayor de 2 kilómetros}) = 1 - 0{,}2336 = 0{,}7664$$

3. Es directo calcular que:

$$P(Y > 2) = e^{-2 \cdot 2} = e^{-4}$$
$$P(Z > 2) = e^{-3 \cdot 2} = e^{-6}$$

Definimos $A_i$=“Reparto ha sido en la $i$-ésima carretera”

$$P(A_2|\text{Viaje} > 2) = \frac{P(Y > 2) \cdot P(A_2)}{P(X > 2)P(A_1) + P(Y > 2)P(A_2) + P(Z > 2)P(A_3)}$$

donde:
$P(\text{Viaje} > 2|A_1) = P(X > 2)$
$P(\text{Viaje} > 2|A_2) = P(Y > 2)$
$P(\text{Viaje} > 2|A_3) = P(Z > 2)$

Por tanto:

$$P(A_2|\text{Viaje} > 2) = \frac{0{,}3 \cdot e^{-4}}{0{,}4 \cdot e^{-2} + 0{,}3 \cdot e^{-4} + 0{,}3 \cdot e^{-6}} = 0{,}091$$

**Cuestiones** (45 minutos, 5 puntos)

1. De un lote de 20000 piezas se extraen 250 al azar y se rechaza el lote cuando el número de piezas defectuosas es mayor que $c$.

   Calcular $c$ para que la probabilidad de rechazar un lote con un 3 % de defectuosas sea 0.05.

2. El estroncio-90 es un radionucleido que se desintegra de forma aleatoria emitiendo una partícula $\beta^-$ según un proceso de Poisson con parámetro (constante de desintegración) $\lambda = 0{,}0239843$ desintegraciones/año.

   a) ¿Cuál es la probabilidad de que transcurran 100 años antes de que un radionucleido emita una partícula?

   Si se dispone de $1.3398\times\ 10^{16}$ radionucleidos de estroncio-90 ($2\mu g$),

   b) ¿Cuál será el número esperado de radionucleidos sin desintegrar al cabo de 100 años?

3. Una tribu indígena del Amazonas construye puentes. Sabemos que el peso que soportan estos puentes sigue una distribución normal, con media 300 kg. y desviación típica 10 kg. Ahora mismo están cruzando el puente dos indígenas y tres monos, y me gustaría saber si es fiable que yo también me suba. Sabiendo que mi peso es de 75 kg., ¿cuál es la probabilidad de que el puente no se destruya al subirme?

   *Datos*: El peso de un indígena sigue una distribución normal de media 70 kg. y varianza 25 kg. El peso de un mono sigue una distribución normal de media 20 kg. y varianza 9 kg. Considerad que el peso de cada persona o animal son independientes entre sí.

**Problema** (45 minutos, 5 puntos)

Las llamadas telefónicas que llegan a la centralita de una compañía es un proceso de Poisson con parámetro $\lambda$. Empezando en un punto arbitrario, que denominamos $t = 0$, sea $T_1$ la variable aleatoria "instante en el que se produce la primera llamada". El tiempo $Y$ que transcurre desde el tiempo observado $t_1$ de la primera llamada hasta que llega la siguiente llamada es una variable aleatoria exponencial de parámetro $\lambda$. Si $t_2$ es el tiempo de ocurrencia de la segunda llamada medido desde el origen $t = 0$, entonces $T_2 = t_1 + Y$, siendo $Y = T_2 - t_1$.

La función de densidad condicionada de $T_2$ dado que $T_1 = t_1$ es:

$$f_{T_2|T_1}(t_2|t_1) = \lambda \cdot e^{-\lambda(t_2 - t_1)} \quad , \quad 0 < t_1 < t_2$$

1. Calcula la función de densidad de la variable aleatoria $T_1$ y, a continuación, la función de densidad conjunta de $T_1$ y $T_2$.

2. Calcula la función de densidad marginal de $T_2$.

3. Calcula la función de densidad condicionada de $T_1$ dado que $T_2 = t_2$.

**Cuestiones** (45 minutos, 5 puntos)

1. De un lote de 20000 piezas se extraen 250 al azar y se rechaza el lote cuando el número de piezas defectuosas es mayor que $c$.

   Calcular $c$ para que la probabilidad de rechazar un lote con un 3 % de defectuosas sea 0.05.

***Solución:***

Tenemos una muestra de 250 elementos, cuyos elementos tienen una probabilidad de 0,03 de ser defectuosos. Por tanto, la distribución del número de elementos defectuosos sigue una distribución binomial:

$$X \to B(n = 250, p = 0{,}03)$$

Se rechaza el lote cuando $X > c$, y deseamos que $P(X > c) = 0{,}05$. Por tanto, utilizando la aproximación a la normal:

$$\begin{aligned} P(X > c | X \to B(n = 250, p = 0{,}03)) &\simeq P(X \geq c + 0{,}5 | X \to N(np, \sqrt{np(1-p)}\ )) \\ &\simeq P(X \geq c + 0{,}5 | X \to N(7{,}5,\ 2{,}697)) \end{aligned}$$

y, por tanto:

$$\begin{aligned} P(X > c) &= P(\frac{X-7}{2{,}697} \geq \frac{c-7}{2{,}697}) \\ &= P(Z \geq \frac{c-7}{2{,}697}) \\ &= 1 - P(Z < \frac{c-7}{2{,}697}) \\ &= 0{,}05 \end{aligned}$$

De la expresión anterior y consultando las tablas de la distribución N(0,1), tenemos:

$$1{,}645 = \frac{c-7}{2{,}697}$$

De donde se deduce: $c = 11{,}43 \simeq 12$.

2. El estroncio-90 es un radionucleido que se desintegra de forma aleatoria emitiendo una partícula $\beta^-$ según un proceso de Poisson con parámetro (constante de desintegración) $\lambda = 0{,}0239843$ desintegraciones/año.

   a) ¿Cuál es la probabilidad de que transcurran 100 años antes de que un radionucleido emita una partícula?

   Si se dispone de 1.3398× $10^{16}$ radionucleidos de estroncio 90 ($2\mu g$),

   b) ¿Cuál será el número esperado de radionucleidos sin desintegrar al cabo de 100 años?

***Solución:***

Definimos la variable aleatoria X como el "número de radionucleidos sin desintegrar al cabo de 100 años". Se observa que sigue una distribución Binomial:

$$B(n = 1{,}3398 \cdot 10^6, p = P(T > 100 \text{ años}))$$

El tiempo de desintegración T se modela como una variable aleatoria con distribución exponencial, cuya función de densidad es $f_T(t) = \lambda e^{-\lambda t}$.

Calculamos la probabilidad de que un radionucleido esté 100 años sin desintegrarse:

$$\begin{aligned} P(T > 100 \text{ años}) &= 1 - F_T(t = 100) \\ &= e^{-100\lambda} \\ &= 0{,}0909 \end{aligned}$$

Y ahora obtenemos el número esperado de radionucleidos sin desintegrar al cabo de 100 años:

$$E[X] = np = 1{,}3398 \cdot 10^{16} \cdot 0{,}0909 = 1{,}2179 \cdot 10^{15}$$

3. Una tribu indígena del Amazonas construye puentes. Sabemos que el peso que soportan estos puentes sigue una distribución normal, con media 300 kg. y desviación típica 10 kg. Ahora mismo están cruzando el puente dos indígenas y tres monos, y me gustaría saber si es fiable que yo también me suba. Sabiendo que mi peso es de 75 kg., ¿cuál es la probabilidad de que el puente no se destruya al subirme?

   *Datos*: El peso de un indígena sigue una distribución normal de media 70 kg. y varianza 25 kg$^2$. El peso de un mono sigue una distribución normal de media 20 kg. y varianza 9 kg$^2$. Considerad que el peso de cada persona o animal son independientes entre sí.

***Solución:***

Sabemos que la capacidad $C$ del puente, el peso $P$ de las personas y el peso $M$ de los monos se distribuyen según normales independientes:

$$C \to N(300, 10)$$
$$P \to N(70, 5)$$
$$M \to N(20, 3)$$

La capacidad disponible del puente, en el caso de que esté manteniendo a dos personas y tres monos, se calcula como $D = C - P_1 - P_2 - M_1 - M_2 - M_3$. La probabilidad que se pide, por tanto, es si esta capacidad disponible es mayor que 75 kg.

La variable aleatoria D es normal, ya que se calcula como combinación lineal de variables normales. Calculamos la media y varianza:

$$\mu_D = 300 - 2 \cdot 70 - 3 \cdot 20 = 100$$

$$\sigma_D^2 = 100 + 2 \cdot 25 + 3 \cdot 9 = 177$$

Por tanto:

$$\begin{aligned} P(D > 75) &= P\left(\frac{D - 100}{\sqrt{177}} > \frac{75 - 100}{\sqrt{177}}\right) \\ &= P\left(Z > \frac{75 - 100}{\sqrt{177}}\right) \\ &= P(Z > -1{,}879) \\ &= 0{,}9699 \end{aligned}$$

**Problema** (45 minutos, 5 puntos)

Las llamadas telefónicas que llegan a la centralita de una compañía es un proceso de Poisson con parámetro $\lambda$. Empezando en un punto arbitrario, que denominamos $t = 0$, sea $T_1$ la variable aleatoria "instante en el que se produce la primera llamada". El tiempo $Y$ que transcurre desde el tiempo observado $t_1$ de la primera llamada hasta que llega la siguiente llamada es una variable aleatoria exponencial de parámetro $\lambda$. Si $t_2$ es el tiempo de ocurrencia de la segunda llamada medido desde el origen $t = 0$, entonces $T_2 = t_1 + Y$, siendo $Y = T_2 - t_1$.

La función de densidad condicionada de $T_2$ dado que $T_1 = t_1$ es:

$$f_{T_2|T_1}(t_2|t_1) = \lambda \cdot e^{-\lambda(t_2-t_1)} \quad , \quad 0 < t_1 < t_2$$

1. Calcula la función de densidad de la variable aleatoria $T_1$ y, a continuación, la función de densidad conjunta de $T_1$ y $T_2$.
2. Calcula la función de densidad marginal de $T_2$.
3. Calcula la función de densidad condicionada de $T_1$ dado que $T_2 = t_2$.

***Solución:***

1. La función de densidad de la variable aleatoria $T_1$ es:

$$f_{T_1}(t_1) = \lambda e^{-\lambda t_1} \quad , \quad 0 < t_1$$

La función de densidad conjunta de $T_1$ y $T_2$ es:

$$\begin{aligned} f_{T_1,T_2}(t_1,t_2) &= f_{T_1}(t_1) \cdot f_{T_2|T_1}(t_2|t_1) \\ &= \lambda e^{-\lambda t_1} \cdot \lambda e^{-\lambda(t_2-t_1)} \\ &= \lambda^2 e^{-\lambda t_2} \quad , \quad 0 < t_1 < t_2 \end{aligned}$$

2. La marginal de $T_2$ es:

$$\begin{aligned} f_{T_2}(t_2) &= \int_{-\infty}^{\infty} f_{T_1,T_2}(t_1,t_2) dt_1 \\ &= \int_0^{t_2} \lambda^2 e^{-\lambda t_2} dt_1 \\ &= \lambda^2 t_2 e^{-\lambda t_2} \quad , \quad t_2 > 0 \end{aligned}$$

3. La función de densidad condicionada de $T_1$ dado que $T_2 = t_2$ es:

$$\begin{aligned} f_{T_1|T_2}(t_1|t_2) &= \frac{f_{T_1,T_2}(t_1,t_2)}{f_{T_2}(t_2)} \\ &= \frac{\lambda^2 e^{-\lambda t2}}{\lambda^2 t_2 e^{-\lambda t_2}} \\ &= \frac{1}{t_2} \quad , \quad 0 < t_1 < t_2 \end{aligned}$$

Se observa que esta función es uniforme en el intervalo $(0, t_2)$.

**Cuestiones** (45 minutos, 5 puntos)

1. Un alumno de Estadística desea conocer la proporción $p$ de alumnos que aprobaron el Segundo Examen Parcial de la asignatura. Como el alumno no tiene tiempo para encuestar a todos los alumnos de la asignatura, realiza un muestreo aleatorio simple, preguntando a un cierto número de alumnos si aprobaron el examen. En base a la información recogida, el alumno calcula el intervalo de confianza para el parámetro $p$, obteniendo que se encuentra en los límites (0,5; 0,8) con una confianza del 95 %.

   Calcula el número total de personas que fueron encuestadas por el alumno, y el número de ellas que contestaron afirmativamente.

2. Se ha realizado un experimento con 75 bombillas incandescentes para cuantificar su duración. Se han obtenido los resultados siguientes:

| Duración (horas) | Nº de bombillas |
|---|---|
| $< 100$ | 25 |
| $[100, 200)$ | 15 |
| $[200, 300)$ | 15 |
| $[300, 400)$ | 8 |
| $\geq 400$ | 12 |
| Total | 75 |

   Contraste con un nivel de significación $\alpha = 0,05$ que la variable aleatoria $T$: *duración de la bombilla incandescente,* es una variable aleatoria exponencial con media 200 horas.

3. Sea una variable aleatoria discreta con distibución uniforme que toma los valores $i = 1, .., N$. Estimar $N$ por el método de los momentos, calcular el sesgo del estimador y su error cuadrático medio sabiendo que:

$$\sum_{i=1}^{N} i^2 = N(N+1)(2N+1)/6 \ .$$

**Problema** (45 minutos, 5 puntos)

Una empresa suiza fabrica piezas de alta precisión para relojes de pulsera de distintos fabricantes. Su sistema de fabricación tiene un error con una desviación típica igual a 0.010 mm. Uno de los fabricantes le ha encargado tornillos de 6.24 mm.

Para controlar el proceso de fabricación, cada hora se toma una muestra de 25 tornillos y se calcula su media. Si la media obtenida es 6.245 mm o mayor, se detiene el proceso y se hace un reajuste del mismo. Cuando el sistema se desajusta, la línea produce tornillos más largos de lo que debería. Nunca se producen errores en el otro sentido.

1. Calcular cuál es la probabilidad que el procedimiento indique que hay que hacer un ajuste cuando el proceso está fabricando correctamente con media 6.24 mm.

2. El proceso se ha desajustado y empieza a fabricar con media igual a 6.25 mm, ¿cuál es la probabilidad de detectarlo el control?

3. Los datos de las últimas diez medidas obtenidas son:

$$6.2411,\ 6.2437,\ 6.2355,\ 6.2406,\ 6.2374,$$
$$6.2455,\ 6.2472,\ 6.2407,\ 6.2391,\ 6.2471$$

   La desviación típica corregida de estos valores es de 0.004 mm. (Atención: no se disponen de las medidas individuales. Sólo de las medias). Si las condiciones de fabricación no han cambiado en las últimas 10 horas, contrasta si ha habido un aumento en la varianza del proceso.

**Cuestiones** (45 minutos, 5 puntos)

1. Un alumno de Estadística desea conocer la proporción $p$ de alumnos que aprobaron el Segundo Examen Parcial de la asignatura. Como el alumno no tiene tiempo para encuestar a todos los alumnos de la asignatura, realiza un muestreo aleatorio simple, preguntando a un cierto número de alumnos si aprobaron el examen. En base a la información recogida, el alumno calcula el intervalo de confianza para el parámetro $p$, obteniendo que se encuentra en los límites (0,5; 0,8) con una confianza del 95 %.

   Calcula el número total de personas que fueron encuestadas por el alumno, y el número de ellas que contestaron afirmativamente.

   *Solución*:

   El intervalo de confianza se calcula como:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0{,}5$$
$$\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0{,}8$$

   Resolviendo el anterior sistema de ecuaciones, con $\alpha = 0{,}05$, se obtiene:

$$\begin{cases} \hat{p} = 0{,}65 \\ n = 38,84 \sim 39 \end{cases}$$

2. Se ha realizado un experimento con 75 bombillas incandescentes para cuantificar su duración. Se han obtenido los resultados siguientes:

| Duración (horas) | Nº de bombillas |
|---|---|
| $< 100$ | 25 |
| $[100, 200)$ | 15 |
| $[200, 300)$ | 15 |
| $[300, 400)$ | 8 |
| $\geq 400$ | 12 |
| Total | 75 |

Contraste con un nivel de significación $\alpha = 0,05$ que la variable aleatoria $T$: *duración de la bombilla incandescente,* es una variable aleatoria exponencial con media 200 horas.

*Solución*:

La funcion de densidad: $f_T(t) = \frac{1}{200}\exp(-t/200)$

La función de distribución :$F_T(t) = 1 - \exp(-t/200)$

No es necesario estimar ningún parámetro para calcular la probabilidad.

| Clase | $O_i$ | $E_i = n \times p_i$ | $\frac{(Oi - \text{Ei})^2}{E_i}$ |
|---|---|---|---|
| $< 100$ | 25 | 29.51 | 0.6893 |
| $[100, 200)$ | 15 | 17.90 | 0.4698 |
| $[200, 300)$ | 15 | 10.86 | 1.5782 |
| $[300, 400)$ | 8 | 6.58 | 0.3064 |
| $\geq 400$ | 12 | 10.15 | 0.3372 |
| Total | 75 | 75 | 3.3809 |

$\chi^2_{5-0-1;\alpha=0,05} = 9{,}488$

Como $X^2_M = 3{,}3809 \leq 9{,}488$ no se puede rechazar $H_o$

3. Sea una variable aleatoria discreta con distibución uniforme que toma los valores $i = 1, .., N$. Estimar $N$ por el método de los momentos, calcular el sesgo del estimador y su error cuadrático medio sabiendo que:

$$\sum_{i=1}^{N} i^2 = N(N+1)(2N+1)/6 \ .$$

*Solución*:
Igualamos la esperanza de la población con la media de la muestra:

$$E(X) = \sum_{x=1}^{N} x \cdot P(x) = \sum_{x=1}^{N} x \cdot \frac{1}{N} = \frac{N+1}{2}$$

$$E(X) = \overline{x} \longrightarrow \widehat{N} = 2\overline{x} - 1$$

Ahora calculamos la esperanza del estimador:

$$E[\widehat{N}] = 2E[\overline{X}] - 1 = 2\frac{N+1}{2} - 1 = N$$

Calculamos la varianza del estimador:

$$var[\widehat{N}] = 4E[\overline{X}] = \frac{4var[X]}{n};$$

$$\begin{aligned} var[X] &= E[X^2] - ((1+N)/2)^2 \\ &= (1/N)N(N+1)(2N+1)/6 - ((1+N)/2)\hat{\ }2 \\ &= (N+1)[(2N+1)/6 - (1+N)/4] \\ &= \frac{(N+1)(N-1)}{12} \end{aligned}$$

$$var[\widehat{N}] = \frac{(N+1)(N-1)}{3n}$$

Finalmente, calculamos el error cuadrático medio del estimador:

$$\begin{aligned} ECM[\widehat{N}] &= sesgo^2 + var[\widehat{N}] \\ &= 0 + var[\widehat{N}] \\ &= \frac{(N+1)(N-1)}{3n} \end{aligned}$$

**Problema** (45 minutos, 5 puntos)

Una empresa suiza fabrica piezas de alta precisión para relojes de pulsera de distintos fabricantes. Su sistema de fabricación tiene un error con una desviación típica igual a 0.010 mm. Uno de los fabricantes le ha encargado tornillos de 6.24 mm.

Para controlar el proceso de fabricación, cada hora se toma una muestra de 25 tornillos y se calcula su media. Si la media obtenida es 6.245 mm o mayor, se detiene el proceso y se hace un reajuste del mismo. Cuando el sistema se desajusta, la línea produce tornillos más largos de lo que debería. Nunca se producen errores en el otro sentido.

1. Calcular cuál es la probabilidad que el procedimiento indique que hay que hacer un ajuste cuando el proceso está fabricando correctamente con media 6.24 mm.

2. El proceso se ha desajustado y empieza a fabricar con media igual a 6.25 mm, ¿cuál es la probabilidad de detectarlo el control?

3. Los datos de las últimas diez medidas obtenidas son:

   6.2411, 6.2437, 6.2355, 6.2406, 6.2374, 6.2455, 6.2472, 6.2407, 6.2391, 6.2471

   La desviación típica corregida de estos valores es de 0.004 mm. (Atención: no se disponen de las medidas individuales. Sólo de las medias). Si las condiciones de fabricación no han cambiado en las últimas 10 horas, contrasta si ha habido un aumento en la varianza del proceso.

*Solución*:

**1)** Si el proceso fabrica correctamente, $E[X] = \mu = 6{,}24$ .

En estas condiciones $E[\overline{X}] = 6{,}24; Var[\overline{X}] = \dfrac{(0{,}010)^2}{25}$.

P(ajuste cuando el proceso fabrica correctamente) $= P(\overline{X} > 6{,}245 | \overline{X} \rightsquigarrow N(6{,}24, \dfrac{0{,}010}{5})) =$

$= P(Z > \dfrac{6{,}245 - 6{,}24}{0{,}010/5}) = P(Z > 2.5) = 1 - \phi(2{,}5) = 0{,}006.$

**2)** P(detectar el desajuste) $= P(\overline{X} > 6{,}245 | \overline{X} \rightsquigarrow N(6{,}25, \dfrac{0{,}010}{5}) =$

$= P(Z > \dfrac{6{,}245 - 6{,}25}{0{,}010/5}) = P(Z > -2.5) = \phi(2{,}5) = 0{,}994.$

**3)** Los datos corresponden a las medias de 25 datos obtenidas en las últimas 10 horas.

Si el proceso funciona bien $Var[\overline{X}] = \dfrac{\sigma^2}{25} = \dfrac{(0{,}010)^2}{25} = 4{,}0 \times 10^{-6}$

El contraste se realiza sobre la varianza de la media muestral.

$$\begin{cases} H_0 : \sigma^2_{\overline{X}} = 4{,}0 \times 10^{-6} \\ H_1 : \sigma^2_{\overline{X}} > 4{,}0 \times 10^{-6} \end{cases}$$

Si $H_0$ cierta $\dfrac{(n-1)\widehat{s}^2_{\overline{X}}}{\sigma^2_{\overline{X}}} \rightsquigarrow \chi^2_{n-1}$

Se rechaza $H_0$ si $\dfrac{9 \times \widehat{s}^2_{\overline{X}}}{4{,}0 \times 10^{-6}} > \chi^2_{0{,}95;9} = 16{,}92.$

Con los datos $X^2 = \dfrac{9 \times (0{,}004)^2}{4{,}0 \times 10^{-6}} = 36{,}007 > \chi^2_{0{,}95;9} = 16{,}92 \Longrightarrow$ Se rechaza $H_0$.Las condiciones de fabricación han cambiado, ha habido un aumento de la varianza del proceso.

**Problema - Primera Parte**
(60 minutos, 10 puntos)

Entre los estudiantes de bachillerato de un país se ha realizado un test de gramática, el tiempo (en horas) que tardan en terminarlo es una variable aleatoria cuya función de densidad es $f_T(t) = kt\exp(\square t^2/\alpha^2), t > 0, \alpha > 0$.

1) Calcular $k$ en función de $\alpha$ y la mediana de la distribución para $\alpha^2 = 2$

2) Si se eligen de forma independiente dos estudiantes al azar, ¿cuál es la probabilidad de que los dos tarden más de media hora en realizar el test? ($\alpha^2 = 2$).

3) Entre los estudiantes del país se hablan dos idiomas, A y B, de forma que todos los estudiantes hablan al menos uno de los dos. El 50 % de los estudiantes habla sólo A y el 20 % sólo B.

Supongase ahora que las distribuciones del tiempo en realizar el test son como la del primer apartado, con parámetros $\alpha^2 = 1$ para los que sólo hablan A, $\alpha^2 = 3$ para los que sólo hablan $B$ y $\alpha^2 = 2$ para los que hablan los dos idiomas. ¿Cuál es la probabilidad de que un estudiante elegido al azar tarde más de media hora en acabar el test?

4) Si un estudiante ha tardado más de 40 minutos, ¿cuál es la probabilidad de que hable los dos idiomas?

**Problema - Segunda Parte**
(60 minutos, 10 puntos)

Un Laboratorio de Termotecnia tiene que establecer un control para los fabricantes de termómetros de cámaras frigoríficas. En lotes de un gran número de unidades seleccionan al azar $n = 100$ termómetros, si el número de termómetros defectuosos en la muestra es menor o igual que $c = 3$ aceptan el lote y lo rechazan si el número de termómetros defectuosos en la muestra es mayor que $c = 3$.

1. ¿Cuál es la probabilidad $\alpha$ de rechazar un lote que tiene una proporción $p_A = 1\,\%$ de termómetros defectuosos?
2. ¿Cuál es la probabilidad $\beta$ de aceptar un lote que tiene una proporción $p_R = 7\,\%$ de termómetros defectuosos?
3. Calcular $n$ y $c$ si se desea que $\alpha = 0{,}05$ y $\beta = 0{,}05$. Utiliza la aproximación normal (no es necesario utilizar la corrección por continuidad).
4. Dibuja la curva que proporciona la probabilidad de aceptación en función de $p$ (la proporción de piezas defectuosas que contiene el lote) para los valores de $n$ y $c$ obtenidos en el apartado 3. Calcula los puntos para $p = 1\,\%, p = 3\,\%, p = 5\,\%$ y $p = 7\,\%$. Interpreta los resultados.

**Problema - Tercera Parte**
(60 minutos, 10 puntos)

Sergio, un alumno de la asignatura "*Estadística*", ha generado por simulación utilizando su ordenador una muestra de tamaño $n = 500$ de una variable aleatoria $X_1$ con distribución $N(\mu, \sigma)$. La media muestral, $\overline{x}_1$, es 1,9792 y la varianza muestral, $s_1^2 = 0{,}5347$.

1. Justifica si $\overline{x}_1$ y $s_1^2$ son estimadores centrados de $\mu$ y $\sigma^2$ respectivamente, calculando el sesgo correspondiente en caso de que no lo sean.

2. Sergio construye también por simulación una muestra de tamaño $n = 500$ de la variable aleatoria $Y$, que se define como $Y = \frac{X_1+...+X_{10}}{10}$, donde las $X_i$, $i = 1, ..., 10$ son variables aleatorias $N(\mu, \sigma)$ independientes.

   En la Figura 1 se muestra el histograma para esta muestra de $Y$ de tamaño 500 que Sergio ha obtenido por simulación. La media muestral que obtiene para $Y$ es $\overline{y}_S = 2{,}0062$, y la varianza muestral $s_S^2 = 0{,}0518$.

   Indica qué distribución (con sus parámetros correspondientes) cabe esperar para $Y$, justificando el resultado y calcula un intervalo de confianza para la media de $Y$.

3. Julio, amigo de Sergio y matriculado también en la asignatura "*Estadística*", ha intentado realizar un ejercicio de simulación como el del apartado 2, obteniendo una muestra también de tamaño 500 cuyo histograma se muestra en la Figura 2. A la vista de ambos histogramas, Sergio intuye que Julio se ha equivocado al realizar su ejercicio de simulación y decide hacer algunas comprobaciones, pues además sabe que los valores obtenidos por Julio para la media y varianza de su muestra han sido, respectivamente, $\overline{y}_J = 2{,}4038$ y $s_J^2 = 0{,}0496$.

   ¿Han generado Sergio y Julio sus respectivas muestras a partir de distribuciones con la misma varianza? Plantea y realiza el contraste correspondiente. ($\alpha =$ 0,05).

4. ¿Han generado Sergio y Julio sus respectivas muestras a partir de distribuciones con la misma media? Plantea y realiza el contraste correspondiente. Sabiendo que Sergio ha realizado el ejercicio correctamente, justifica a la vista del resultado del contraste si Julio se ha equivocado. ($\alpha =$ 0,05).

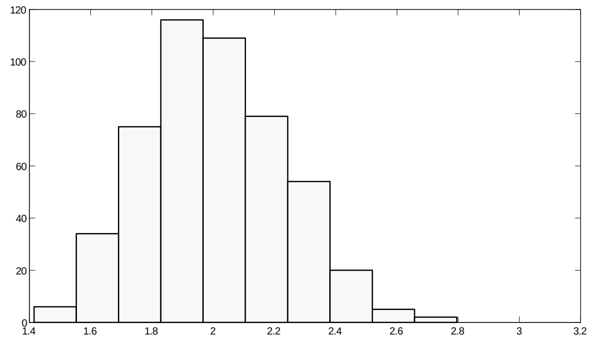



Figura 1: Histograma para la muestra de $Y$, de tamaño 500, obtenida por Sergio.

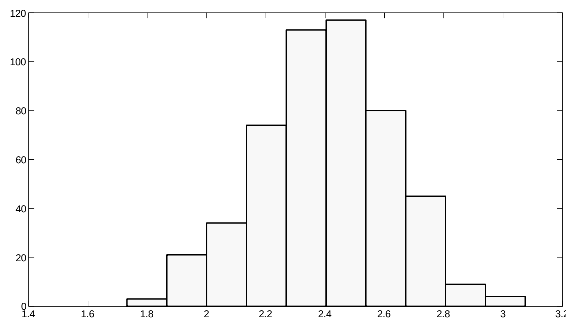



Figura 2: Histograma para la muestra de $Y$, de tamaño 500, obtenida por Julio.

**Solución de la primera parte**.

1.

Para calcular el valor de $k$, se impone que la integral de la función de densidad sea 1.

$\int\limits_0^\infty kt\exp(\square t^2/\alpha^2)dt = 1; \square k\frac{\alpha^2}{2}\int\limits_0^\infty \square\frac{2}{\alpha^2}t\exp(\square t^2/\alpha^2)dt = \square k\frac{\alpha^2}{2}[\exp(\square t^2/\alpha^2)]_0^\infty =$
$k\frac{\alpha^2}{2} = 1; k = \frac{2}{\alpha^2}$

Mediana

Se define como el valor $m$ tal que $F(m) = 0,5$ siendo $F$ la función de distribución.

$\int\limits_0^m t\exp(\square t^2/2)dt = 0,5; m = 1,17$

2.

Por independencia de los dos sucesos, la probabilidad conjunta es el producto de las probabilidades marginales.

$P(A$ y $B$ tardan más de 0,5) $=$ (*independencia*) $= P(A\ tarda\ más\ de\ 0,5)P(B\ tarda\ más\ de\ 0,5) = (P(A\ tarda\ más\ de\ 0,5))^2 = [\int\limits_{0,5}^\infty t\exp(\square t^2/2)dt]^2 =$
$0,778$

3

Teorema de la probabilidad total

$P(X > 0,5) = P(soloA)P(X > 0,5|soloA) + P(soloB)P(X > 0,5|soloB) + P(AyB)P(X > 0,5|AyB)$

$P(soloA) = 0,5; P(soloB) = 0,2; P(AyB) = 1 \square P(soloA) \square P(soloB) = 0,3$

$P(X > 0,5|soloA) = \int\limits_{0,5}^\infty 2t\exp(\square t^2)dt = 0,778$

$P(X > 0,5|soloB) = \int\limits_{0,5}^\infty \frac{2}{3}t\exp(\square t^2/3)dt = 0,92$

$P(X > 0,5|AyB) = \int\limits_{0,5}^\infty t\exp(\square t^2/2)dt = 0,882$

$P(X > 0,5) = 0,5 \times 0,778 + 0,2 \times 0,92 + 0,3 \times 0,882 = 0,838$

4

Teorema de Bayes

$P(AyB|X > 2/3) = \frac{P(AyB)P(X>0,5|AyB)}{P(AyB)P(X>0,5|AyB)+P(soloA)P(X>0,5|soloA)+P(soloB)P(X>0,5|soloB)}$

$$P(X > 2/3|soloA) = \int_{2/3}^{\infty} 2t \exp(\square t^2) dt = 0,64$$

$$P(X > 2/3|soloB) = \int_{2/3}^{\infty} \tfrac{2}{3} t \exp(\square t^2/3) dt = 0,865$$

$$P(X > 2/3|AyB) = \int_{2/3}^{\infty} t \exp(\square t^2/2) dt = 0,8$$

$$P(AyB|X > 2/3) = 0,327$$

**Solución Problema - Segunda Parte**

Sea X: v.a n$^o$ de termómetros defectuosos al observar 100. X$\rightarrow B(n = 100, p)$
Si $X \leq 3 \Longrightarrow$ Se Acepta el lote
Si $X > 3 \Longrightarrow$ Se rechaza el lote

**1.** $\alpha = P(X > 3 | X \rightarrow B(n = 100, p_A = 0.01)) = 1 - P(X \leq 3) =$

$$= 1 - \left[ \sum_{x=0}^{3} \binom{100}{x} 0.01^x 0.99^{100-x} \right] = 1 - 0.9816 = 0.0184.$$

**2.** $\beta = P(X \leq 3 | X \rightarrow B(n = 100, p_R = 0.07)) =$

$$= \left[ \sum_{x=0}^{3} \binom{100}{x} 0.07^x 0.93^{100-x} \right] = 0.0751.$$

**3.** $\alpha = P(X > c | X \sim N(np_A; \sqrt{np_A(1-p_A)})) = 0.05 \Longrightarrow$

$$\alpha = P\left( \frac{X - np_A}{\sqrt{np_A(1-p_A)}} > \frac{c - np_A}{\sqrt{np_A(1-p_A)}} \right) = 0.05 \Longrightarrow$$

$$\boxed{\frac{c - n0.01}{\sqrt{n0.01(1-0.01)}} = 1.64} \text{ Ec. 1}$$

$\beta = P(X \leq c | X \sim N(np_R; \sqrt{np_R(1-p_R)})) = 0.05 \Longrightarrow$

$$\beta = P\left( \frac{X - np_R}{\sqrt{np_R(1-p_R)}} \leq \frac{c - np_R}{\sqrt{np_R(1-p_R)}} \right) = 0.05 \Longrightarrow$$

$$\boxed{\frac{c - n0.07}{\sqrt{n0.07(1-0.07)}} = -1.64} \text{ Ec.2}$$

Resolviendo el sistema de dos ecuaciones con dos incógnitas se obtiene

$$n = 93.96 \approx 94$$
$$c = 2.5221 \approx 3$$

**4.** Considerando el plan $n = 94; c = 3$

$$P(X \leq 3 | X \sim N(94p; \sqrt{94p(1-p)})) = P\left( \frac{X - 94p}{\sqrt{94p(1-p)}} \leq \frac{3 - 94p}{\sqrt{94p(1-p)}} \right) =$$

$$P\left( Z \leq \frac{3 - 94p}{\sqrt{94p(1-p)}} \right) = \Phi\left( \frac{3 - 94p}{\sqrt{94p(1-p)}} \right)$$

| $p$ | $\Phi\left(\frac{3 - 94p}{\sqrt{94p(1-p)}}\right)$ | *P(aceptar)* |
|---|---|---|
| 0.01 | $\Phi(2.13)$ | 0.9834 |
| 0.03 | $\Phi(0.11)$ | 0.5438 |
| 0.05 | $\Phi(-0.80)$ | 0.2119 |
| 0.07 | $\Phi(-1.45)$ | 0.0735 |

A medida que aumenta la proporción de piezas defectuosas que contiene el lote, disminuye la probabilidad de aceptación, sin embargo estas probabilidades siguen siendo altas, por ejemplo la proporción de lotes que se aceptan con un 3% de piezas defectuosas (que triplica la $p_A = 1\%$) es superior al 50 %. Un razonamiento similar podría realizarse para los lotes con un 5% de piezas defectuosas.

# Problema (Parte 3)

**Sergio, un alumno de la asignatura "Estadística", ha generado por simulación utilizando su ordenador una muestra de tamaño** $n = 500$ **de una variable aleatoria** $X_1$ **con distribución** $N(\mu, \sigma)$**. La media muestral,** $\overline{x}_1$**, es** $1.9792$ **y la varianza muestral,** $s_1^2 = 0.5347$.

1. Justifica si $\overline{x}_1$ y $s_1^2$ son estimadores centrados de $\mu$ y $\sigma^2$ respectivamente, calculando el sesgo correspondiente en caso de que no lo sean.

   $E[\overline{x}_1] = E\left[\frac{x_{11}+...+x_{1n}}{n}\right] = \frac{n\mu}{n} = \mu$. Como $E[\widehat{\mu}] = E[\overline{x}_1]$ entonces $\widehat{\mu} = \overline{x}_1$ es un estimador centrado de $\mu$.

   Sin embargo, la varianza muestral, $s_1^2$ no es un estimador centrado, sí lo es $\widehat{s}_1^2$.

   $E[\widehat{s}_1^2] = \sigma^2$ y $(n \square 1)E[\widehat{s}_1^2] = nE[s_1^2] \rightarrow E[s_1^2] = \frac{(n \square 1)}{n}\sigma^2$.

   $\text{sesgo}(s_1^2) = E[s_1^2] \square \sigma^2 = \frac{(n \square 1)}{n}\sigma^2 \square \sigma^2 = \frac{\square \sigma^2}{n}$.

2. **Sergio construye también por simulación una muestra de tamaño** $n = 500$ **de la variable aleatoria** $Y$**, que se define como** $Y = \frac{X_1+...+X_{10}}{10}$**, donde las** $X_i$, $i = 1, ..., 10$ **son variables aleatorias** $N(\mu, \sigma)$ **independientes.**

   **En la Figura 1 se muestra el histograma para esta muestra de** $Y$ **de tamaño 500 que Sergio ha obtenido por simulación. La media muestral que obtiene para** $Y$ **es** $\overline{y}_S = 2.0062$**, y la varianza muestral** $s_S^2 = 0.0518$.

   **Indica qué distribución (con sus parámetros correspondientes) cabe esperar para** $Y$**, justificando el resultado y calcula un intervalo de confianza para la media de** $Y$.

   $Y \sim N(\mu, \frac{\sigma}{\sqrt{N}} = \frac{\sigma}{\sqrt{10}})$, que se puede justificar tanto por el Teorema Central del Límite como por el hecho de que la combinación lineal de v.a. normales es también una normal.

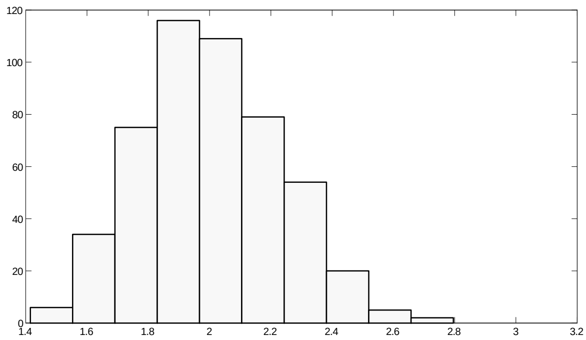



Figure 1: Histograma para la muestra de $Y$, de tamaño 500, obtenida por Sergio.

Obtención de la media y varianza de $Y$.

$E[Y] = E\left[\frac{X_1+...+X_{10}}{10}\right] = \frac{1}{10}\sum_{i=1}^{10} E(X_i) = \mu.$

$var[Y] = var\left[\frac{X_1+...+X_{10}}{10}\right] = \frac{1}{10^2}\sum_{i=1}^{10} var(X_i) = \frac{\sigma^2}{10}.$

Se pide calcular un intervalo de confianza para la media de $Y$, es decir, para $\mu$ :

$\mu \in \overline{y}_S \pm t_{500\square 1;\alpha/2}\frac{\widehat{s}_S}{\sqrt{500}} \rightarrow \mu \in 2.0062 \pm 1.96\frac{0.2278}{\sqrt{500}}$. Entonces $\mu \in 2.0062 \pm 0.0200$, $\mu \in [1.9862, 2.0262]$.

$(500 \square 1)\widehat{s}_S^2 = 500 s_S^2 \rightarrow \widehat{s}_S^2 = 0.0519$ y $\widehat{s}_S = 0.2278$.

3. **Julio, amigo de Sergio y matriculado también en la asignatura "Estadística", ha intentado realizar un ejercicio de simulación como el del apartado 2, obteniendo una muestra también de tamaño 500 cuyo histograma se muestra en la Figura 2. A la vista de ambos histogramas, Sergio intuye que Julio se ha equivocado al realizar su ejercicio de simulación y decide hacer algunas comprobaciones, pues además sabe que los valores obtenidos por Julio para la media y varianza de su muestra han sido, respectivamente, $\overline{y}_J = 2.4038$ y $s_J^2 = 0.0496$.**

   **¿Han generado Sergio y Julio sus respectivas muestras a partir de distribuciones con la misma varianza? Plantea y realiza el contraste correspondiente.**

1. En el enunciado se nos pide comparar las varianzas poblacionales $\sigma_S^2$ y $\sigma_J^2$, luego se plantea el contraste:

   $H_0 : \sigma_S^2/\sigma_J^2 = 1,$

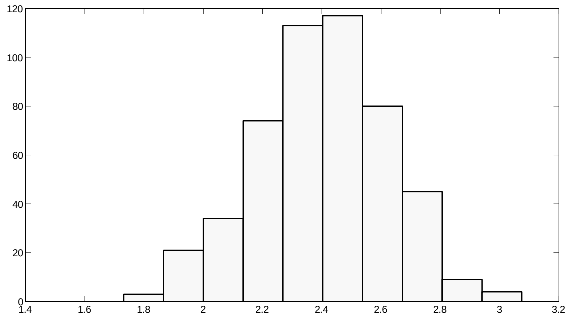


Figure 2: Histograma para la muestra de $Y$, de tamaño 500, obtenida por Julio.

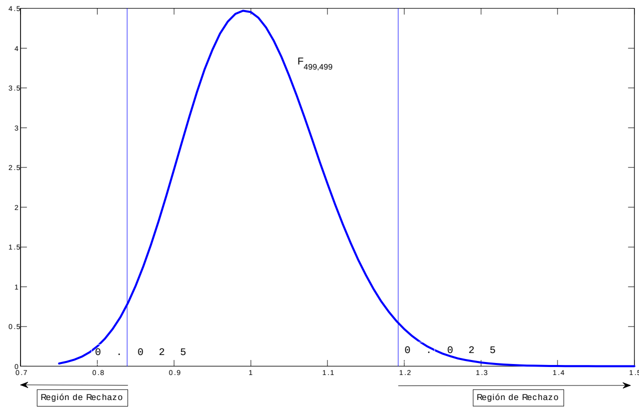



Figure 3: Regiones de aceptación y rechazo para el contraste de igualdad de varianzas

$H_1 : \sigma_S^2/\sigma_J^2 \neq 1.$

Se utiliza que $\frac{\frac{\widehat{s}_S^2}{\sigma_S^2}}{\frac{\widehat{s}_J^2}{\sigma_J^2}} \sim F_{500\square 1,500\square 1}$

Bajo la hipótesis nula $\sigma_S^2/\sigma_J^2 = 1$, luego $F_0 = \frac{\widehat{s}_S^2}{\widehat{s}_J^2} = \frac{0.0519}{0.0497} = 1.0443,$

ya que $(500 \square 1)\widehat{s}_J^2 = 500 s_J^2 \rightarrow \widehat{s}_J^2 = 0.0497$ y $\widehat{s}_J = 0.2229.$

Y de acuerdo con la Figura 3 y teniendo en cuenta que $F_0 = 1.0443$, no se rechaza $H_0$.

4. **¿Han generado Sergio y Julio sus respectivas muestras a partir de distribuciones con la misma media? Plantea y realiza el contraste correspondiente. Sabiendo que Sergio ha realizado el ejercicio correctamente, justifica a la vista del resultado del contraste si Julio se ha equivocado.**

Se pide un contraste de igualdad de medias (tenemos del apartado anterior que sí hay homocedasticidad).

El contraste a realizar es:

$H_0 : \mu_S - \mu_J = 0,$

$H_1 : \mu_S - \mu_J \neq 0.$

Se utiliza que $\frac{(\overline{y}_S - \overline{y}_J) - (\mu_S - \mu_J)}{\widehat{s}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim F_{500-1,500-1}$ y $\widehat{s}^2_R = \frac{n_1 - 1}{n_1 + n_2 - 2}\widehat{s}^2_S + \frac{n_1 - 1}{n_1 + n_2 - 2}\widehat{s}^2_J$,

$\widehat{s}^2_R = \frac{499}{998}0.0519 + \frac{499}{998}0.0497 = 0.0508$ y $\widehat{s}_R = 0.2254$.

$t_0 = \frac{(\overline{y}_S - \overline{y}_J)}{\widehat{s}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(2.0062 - 2.4038)}{0.2254\sqrt{\frac{1}{500} + \frac{1}{500}}} = -27.8909$. Y como la región de aceptación, para $\alpha = 0.05$ y considerando una $t_{499}$ estaría entre -1.96 y 1.96, se rechaza claramente la hipótesis nula en favor de la alternativa: Medias distintas, Julio se ha equivocado.

# EXÁMENES

## Curso 2012/13

**Cuestiones** (45 minutos – 5 puntos)

1. En la tabla se proporcionan los datos de déficit y deuda de los 27 países de la Unión Europea.
- Realiza los diagramas de caja (*bloxplot)* de la variable déficit, **indicando todos los valores numéricos necesarios para la construcción del *boxplot*.** Indicar también que países son atípicos.
- Comprueba gráficamente y numéricamente que las variables déficit y deuda están relacionadas.

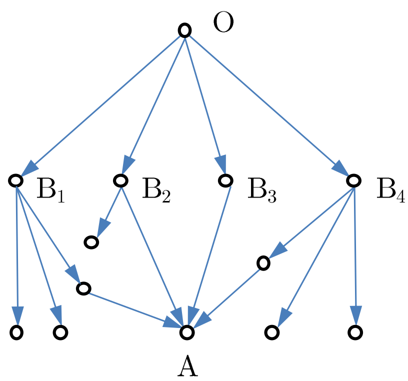
| Pais | Déficit | Deuda |
|---|---|---|
| Belgium | 3.7 | 98.0 |
| Bulgaria | 2.1 | 16.3 |
| Czech Republic | 3.1 | 41.2 |
| Denmark | 1.8 | 46.5 |
| Germany | 1.0 | 81.2 |
| Estonia | –1.0 | 6.0 |
| Ireland | 13.1 | 108.2 |
| Greece | 9.1 | 165.3 |
| Spain | 8.5 | 68.5 |
| France | 5.2 | 85.8 |
| Italy | 3.9 | 120.1 |
| Cyprus | 6.3 | 71.6 |
| Latvia | 3.5 | 42.6 |
| Lithuania | 5.5 | 38.5 |
| Luxembourg | 0.6 | 18.2 |
| Hungary | –4.3 | 80.6 |
| Malta | 2.7 | 72.0 |
| Netherlands | 4.7 | 65.2 |
| Austria | 2.6 | 72.2 |
| Poland | 5.1 | 56.3 |
| Portugal | 4.2 | 107.8 |
| Romania | 5.2 | 33.3 |
| Slovenia | 6.4 | 47.6 |
| Slovakia | 4.8 | 43.3 |
| Finland | 0.5 | 48.6 |
| Sweden | –0.3 | 38.4 |
| U. Kingdom | 8.3 | 85.7 |

Nota:

$$\mathrm{cov}\left([\text{Déficit Deuda}]\right) = \begin{pmatrix} 12.0 & 55.2 \\ 55.2 & 1199.7 \end{pmatrix}$$

2. Dos amigos, Antonio (A) y Basilio (B), deciden resolver una apuesta lanzando tiros triples a una canasta de baloncesto. El porcentaje de aciertos en tiros triples de A es 40% y el porcentaje de aciertos de B es 50%. Deciden lanzar de manera alterna A-B-A-B-A-.... hasta que uno enceste.
   - Si comienza lanzando A, ¿cuál es la probabilidad de que gane A?
   - ¿Cuál es la probabilidad de que gane A si deciden sortear previamente a cara y cruz quien empieza la secuencia?

3. Un esquiador sale del punto O y elige una de las pistas $OB_1$, $OB_2$, $OB_3$ y $OB_{44}$ al azar. En cualquiera de los cruces que se encuentre, elige también el camino al azar, de manera equiprobable.
¿Cuál es la probabilidad de que llegue al punto A?
(Las pistas sólo se recorren en la dirección de descenso que se indica)



**Problema** (45 minutos – 5 puntos)

Un fabricante recibe lotes con 100 piezas de un proveedor. El lote se considera inaceptable si contiene 5 o más piezas defectuosas. El fabricante no puede comprobar todas las piezas del lote y decide hacer una inspección tomando una muestra al azar (**sin reposición**).

1. Decide tomar una muestra de tamaño $m = 5$ y aceptar el lote si todas las piezas de la muestra son aceptables. ¿Cuál es la probabilidad de aceptar el lote si éste contiene exactamente 5 piezas defectuosas?

2. Decide tomar una muestra de tamaño $m = 10$, y aceptar el lote si en la muestra hay cero o una defectuosa. ¿Cuál es la probabilidad de aceptar el lote si éste contiene exactamente 5 piezas defectuosas?

3. Obtener la distribución de probabilidad de la variable aleatoria "número de defectuosas en la muestra de tamaño $m$", si el número de piezas en el lote es 100 y el lote contiene $K$ piezas defectuosas.

**Nota:** Observad que no es un proceso de Bernoulli (ya que no hay reposición): no se puede aplicar la Binomial.

**Solución Cuestión 1:**

*Apartado a*

n = 27

p = (n+1)/2 = 14

r = 7

s = n – r + 1 = 21

Mediana = 3.9

Q1 = 1.8

Q3 = 5.5

RI = 5.5 – 1.8 = 3.7

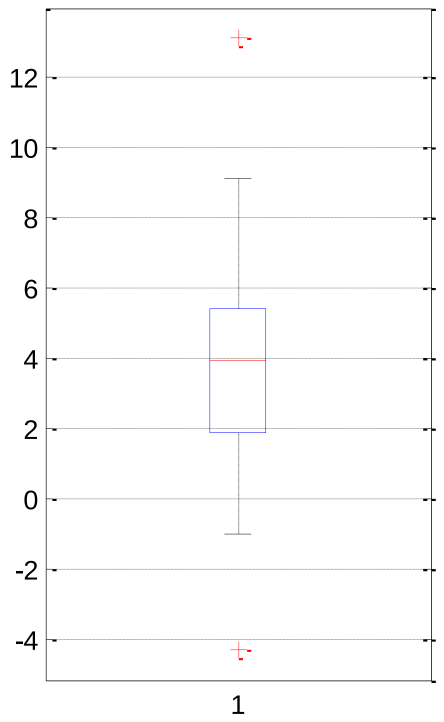
LI = 1.8 – 1.5 · 3.7 = –3.75

LS = 5.5 + 1.5 · 3.7 = 11.05

min{xi | xi >= LI} = –1

max{xi | xi <= LS} = 9.1

Atípicos: *Irlanda y Hungaria*



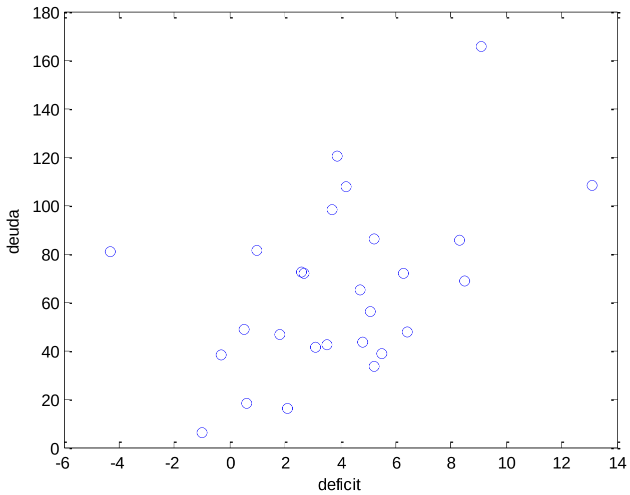

*Apartado b*

La relación entre ambas variables se muestra gráficamente el gráfico de dispersión.

Por otra parte, la relación lineal se muestra mediante la correlación, que se puede calcular directamente a partir de la matriz de varianzas-covarianzas proporcionada en el enunciado.

corr([Déficit Deuda]) = 0.4599

Lo cual denota que hay una cierta correlación.




Por tanto, se ha podido comprobar –gráficamente y numéricamente– que ambas variables están correlacionadas.

**Solución Cuestión 2:**

Definimos los siguientes sucesos:

$G_A$ : Gana A la apuesta.

$A_i$ : A marca canasta en el tiro $i$-ésimo

$B_i$ : B marca canasta en el tiro $i$-ésimo

*Apartado A*

Del enunciado, se sabe que A gana la apuesta si acierta antes que B. Por tanto, sabiendo que empieza sacando A, gana la apuesta en las siguientes ocasiones:

- A marca en el primer tiro: $P(A_1)$
- A y B fallan los primeros tiros, y A marca en el tercero: $P(\overline{A_1}\,\overline{B_2}A_3)$
- A y B fallan los primeros tiros, y A marca en el quinto: $P(\overline{A_1}\,\overline{B_2}\,\overline{A_3}\,\overline{B_4}A_5)$
- A y B fallan los primeros tiros, y A marca en el séptimo: $P(\overline{A_1}\,\overline{B_2}\,\overline{A_3}\,\overline{B_4}\,\overline{A_5}\,\overline{B_6}A_7)$
- ...y así sucesivamente.

$$
\begin{aligned}
P(G_A \mid saca\ A) &= P(A_1)+P(\overline{A_1}\,\overline{B_2}A_3)+P(\overline{A_1}\,\overline{B_2}\,\overline{A_3}\,\overline{B_4}A_5)+P(\overline{A_1}\,\overline{B_2}\,\overline{A_3}\,\overline{B_4}\,\overline{A_5}\,\overline{B_6}A_7)+\ldots \\
&= 0.4+0.6\cdot 0.5\cdot 0.4+(0.6\cdot 0.5)^2\cdot 0.4+(0.6\cdot 0.5)^3\cdot 0.4+\ldots \\
&= 0.4\cdot\left(1+0.6\cdot 0.5+(0.6\cdot 0.5)^2+(0.6\cdot 0.5)^3+\ldots\right) \\
&= 0.4\cdot\left(\sum_{k=0}^{\infty}(0.6\cdot 0.5)^k\right) \\
&= 0.4\cdot\frac{1}{1-(0.6\cdot 0.5)} \qquad \text{(suma de sucesión geométrica)} \\
&= 0.5714
\end{aligned}
$$

*Apartado B*

Si deciden sortear previamente a cara y cruz quien empieza la secuencia:

$$
\begin{aligned}
P(G_A) &= P(G_A \mid saca\ A)P(saca\ A)+P(G_A \mid saca\ B)P(saca\ B) \\
&= 0.5714\cdot 0.5+0.2857\cdot 0.5 \\
&= 0.4286
\end{aligned}
$$

donde:

$$
\begin{aligned}
P(G_A \mid saca\ B) &= P(\overline{B_1}A_2)+P(\overline{B_1}\,\overline{A_2}\,\overline{B_3}A_4)+P(\overline{B_1}\,\overline{A_2}\,\overline{B_3}\,\overline{A_4}\,\overline{B_5}A_6)+\ldots \\
&= 0.5\cdot 0.4+0.5\cdot(0.6\cdot 0.5)\cdot 0.4+0.5\cdot(0.6\cdot 0.5)^2\cdot 0.4+\ldots \\
&= 0.5\cdot 0.4\cdot\left(1+0.6\cdot 0.5+(0.6\cdot 0.5)^2+(0.6\cdot 0.5)^3+\ldots\right) \\
&= 0.5\cdot 0.4\cdot\left(\sum_{k=0}^{\infty}(0.6\cdot 0.5)^k\right) \\
&= 0.5\cdot 0.4\cdot\frac{1}{1-(0.6\cdot 0.5)} \\
&= 0.2857
\end{aligned}
$$

**Solución Cuestión 3:**

$$\begin{aligned}
\mathrm{P}(OA) &= \mathrm{P}(OA \mid OB_1)\cdot P\left(OB_1\right) + \mathrm{P}(OA \mid OB_2)\cdot P\left(OB_2\right) + \mathrm{P}(OA \mid OB_3)\cdot P\left(OB_3\right) + \mathrm{P}(OA \mid OB_4)\cdot P\left(OB_4\right) \\
&= \frac{1}{3}\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{1}{4} + \frac{1}{1}\cdot\frac{1}{4} + \frac{1}{3}\cdot\frac{1}{4} \\
&= 0.5417
\end{aligned}$$

**Solución Problema:**

*Apartado 1*

Se puede calcular como "casos favorables dividido entre casos posibles":

$$P(\text{Aceptar}) = P(\text{Ninguna defectuosa}) = \frac{\binom{95}{5}}{\binom{100}{5}} = 0.7696$$

O bien, mediante probabilidades condicionadas:

$$\begin{aligned} P(Aceptar) &= P(1^{a}\text{ buena})\cdot P(2^{a}\text{ buena}|1^{a}\text{ buena})\cdot P(3^{a}\text{ buena}|1^{a}\,2^{a}\text{ buenas})\cdot \\ &\quad \cdot P(4^{a}\text{ buena}|1^{a}\,2^{a}\,3^{a}\text{ buenas})\cdot P(5^{a}\text{ buena}|1^{a}\,2^{a}\,3^{a}\,4^{a}\text{ buenas}) \\ &= \frac{95}{100}\frac{94}{99}\frac{93}{98}\frac{92}{97}\frac{91}{96} = 0.7696 \end{aligned}$$

*Apartado 2*

$$\begin{aligned} P(\text{Aceptar}) &= P(\text{Ninguna defectuosa}) + P(1\text{ defectuosa}) \\ &= \frac{\binom{95}{10}}{\binom{100}{10}} + \frac{\binom{5}{1}\binom{95}{9}}{\binom{100}{10}} \\ &= 0.9231 \end{aligned}$$

*Apartado 3*

$$P(X=0) = \frac{\binom{100-K}{m}}{\binom{100}{m}}$$

$$P(X=1) = \frac{\binom{K}{1}\binom{100-K}{m-1}}{\binom{100}{m}}$$

Fórmula general:

$$P(X=x) = \frac{\binom{K}{x}\binom{100-K}{m-x}}{\binom{100}{m}} \qquad \max(0, m-N+K) \le x \le \min(K, m)$$

**Cuestiones** (45 minutos, 5 puntos)

1. **La empresa EMUS, S.L. adquiere cojinetes de fricción en lotes de 1000 unidades, acordando con el suministrador un nivel de calidad aceptable (AQL) de los lotes del 2,5 % de unidades defectuosas, siendo el nivel de calidad rechazable (RQL) igual al 10 %. Para reducir el número de piezas muestreadas define el siguiente plan de recepción en dos etapas. Toma 30 cojinetes, si en la muestra no hay defectuosas acepta el lote, si hay dos o más defectuosas rechaza el lote, pero si hay un cojinete defectuoso toma otros 30 cojinetes, de tal manera que si en el total de las dos muestras hay como máximo dos defectuosas se acepta el lote, rechazándolo en caso contrario. Calcule el riesgo del vendedor y del comprador según el plan de recepción propuesto en dos etapas.**

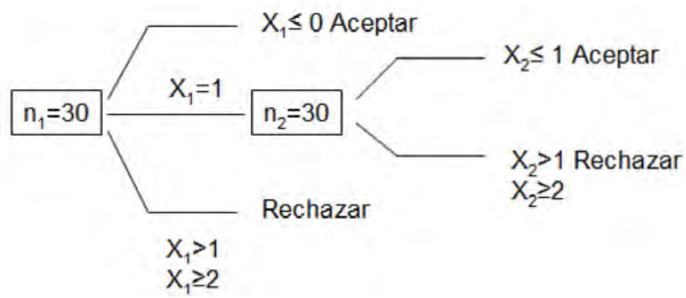



Figura 1: Esquema del plan de muestreo de la cuestion 1

2. **El número de defectos en hilos conductores de cobre sigue una distribución de Poisson de parámetro $\lambda$ defectos/km. Se toma un hilo conductor de 1 km y se observan 5 defectos, ¿cuál es la probabilidad de que los 5 defectos estén en la longitud A que se muestra en la figura?**

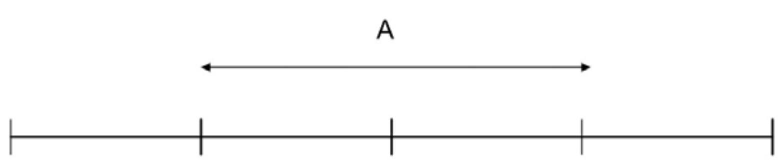



Figura 2: Esquema cuestion 2

3. **Sean $X_1$ y $X_2$ dos variables aleatorias con la misma varianza. Pruebe que las variables $W = X_1 + X_2$ y $V = X_1 - X_2$ están incorrelacionadas.**

**Problema** (45 minutos, 5 puntos)

**En un material se analizan tres variables aleatorias, $X$, $Y$, $Z$, que representan criterios de calidad del material. Las medias de esas tres variables son $10$, $20$ y $30$ respectivamente, y la matriz de varianzas y covarianzas**

$$\begin{bmatrix} 8 & 4 & 0 \\ 4 & 4 & 4 \\ 0 & 4 & 8 \end{bmatrix}$$

**La distribución conjunta es normal multivariante.**

**a) Si se toman $5$ piezas al azar de forma independiente ¿cual es la probabilidad de que el promedio de las mediciones de $X$ sea mayor que $12$?**

**b) Se han construido dos índices de calidad conjunta $Q_1 = 3X + 2Y + Z$ y $Q_2 = X + 2Y - Z$, ¿cuál es la probabilidad de que $Q_1 > Q_2 + 90$?**

**c) Si se toman dos piezas al azar de forma independiente ¿cuál es la probabilidad de que el mínimo de los dos valores de $Z$ sea mayor que $34$?**

**Cuestiones** (45 minutos, 5 puntos)

1. **La empresa EMUS, S.L. adquiere cojinetes de fricción en lotes de 1000 unidades, acordando con el suministrador un nivel de calidad aceptable (AQL) de los lotes del $2{,}5\,\%$ de unidades defectuosas, siendo el nivel de calidad rechazable (RQL) igual al $10\,\%$. Para reducir el número de piezas muestreadas define el siguiente plan de recepción en dos etapas. Toma 30 cojinetes, si en la muestra no hay defectuosas acepta el lote, si hay dos o más defectuosas rechaza el lote, pero si hay un cojinete defectuoso toma otros 30 cojinetes, de tal manera que si en el total de las dos muestras hay como máximo dos defectuosas se acepta el lote, rechazándolo en caso contrario. Calcule el riesgo del vendedor y del comprador según el plan de recepción propuesto en dos etapas.**

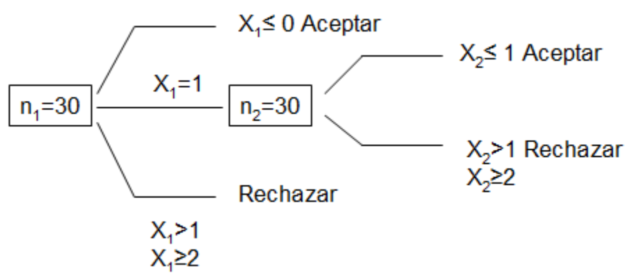



Figura 1: Esquema del plan de muestreo de la primera cuestión.

$p_A = 0{,}025$ y $p_R = 0{,}1$.

$\alpha =$ Riesgo del vendedor $= P(\text{rechazar el lote} \mid p = p_A = 0{,}025) = P(X_1 > 1 | X_1 \sim B(30,\, 0{,}025)) +$

$+P(X_1 = 1 | X_1 \sim B(30,\, 0{,}025)) \cdot P(X_2 \le 1 | X_2 \sim B(30,\, 0{,}025))$

Calculando $P(X_1 > 1 | X_1 \sim B(30,\, 0{,}025)) = 1 - P(X_1 \le 1 | X_1 \sim B(30,\, 0{,}025))$ quedaría:

$$1 - \left[\binom{30}{0} 0{,}025^0 \cdot 0{,}975^{30} + \binom{30}{1} 0{,}025^1 \cdot 0{,}975^{29}\right] +$$

$$+\left(\binom{30}{1} 0{,}025^1 \cdot 0{,}975^{29}\right) \cdot \left[\binom{30}{0} 0{,}025^0 \cdot 0{,}975^{30} + \binom{30}{1} 0{,}025^1 \cdot 0{,}975^{29}\right] = 0{,}234$$

$\beta =$ Riesgo del comprador $= P(\text{aceptar el lote} \mid p = p_R = 0{,}1) = P(X_1 = 0 | X_1 \sim B(30,\, 0{,}1)) +$

$+P(X_1 = 1 | X_1 \sim B(30,\, 0{,}1)) \cdot P(X_2 \le 1 | X_2 \sim B(30,\, 0{,}1)) =$

$$= \binom{30}{0} 0{,}1^0 \cdot 0{,}9^{30} + \left[\left(\binom{30}{1} 0{,}1^1 \cdot 0{,}9^{29}\right) \cdot \left(\binom{30}{0} 0{,}1^0 \cdot 0{,}9^{30} + \binom{30}{1} 0{,}1^1 \cdot 0{,}9^{19}\right)\right] = 0{,}068$$

2. **El número de defectos en hilos conductores de cobre sigue una distribución de Poisson de parámetro $\lambda$ defectos/km. Se toma un hilo conductor de 1 km y se observan 5 defectos, ¿cuál es la probabilidad de que los 5 defectos estén en la longitud A que se muestra en la figura?**

$Y =$ número de defectos en 1 Km $\sim Poisson(\lambda$ defectos/km$)$

$A =$ número de defectos en 1/2 Km $\sim Poisson(\lambda/2$ defectos en 1/2 km$)$

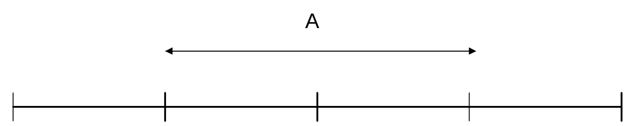



Figura 2: Esquema Cuestión 2.

$A-$ y $A+$: número de defectos en 1/4 Km, el primero y cuarto respectivamente de los de la Figura 2, tanto $A-$ como $A+ \sim$ Poisson($\lambda/4$ defectos en 1/4 km).

$$P(A=5|Y=5) = \frac{P(A{=}5 \cap Y{=}5)}{P(Y{=}5)} = \frac{P(A{-}{=}0 \cap A{=}5 \cap A{+}{=}0)}{P(Y{=}5)} = \frac{\frac{(\lambda/4)^0}{0!}e^{-\lambda/4} . \frac{(\lambda/2)^5}{5!}e^{-\lambda/2} . \frac{(\lambda/4)^0}{0!}e^{-\lambda/4}}{\frac{\lambda^5}{5!}e^{-\lambda}} = \frac{1}{2^5} = \frac{1}{32}$$

3. **Sean $X_1$ y $X_2$ dos variables aleatorias con la misma varianza. Pruebe que las variables $W = X_1 + X_2$ y $V = X_1 - X_2$ están incorrelacionadas.**

$X_1, X_2$ y se sabe que $\sigma^2_{X_1} = \sigma^2_{X_2} = \sigma^2$.

Las dos nuevas variables que se definen son:

$W = X_1 + X_2$

$V = X_1 - X_2$, y en forma matricial:

$$\begin{pmatrix} W \\ V \end{pmatrix} = A \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

Sea $M_X$ la matriz de varianzas-cov de $X_1$ y $X_2$ se tiene:

$$\text{Var}\begin{pmatrix} W \\ V \end{pmatrix} = AM_XA' = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2(\sigma^2 + \rho\sigma^2) & \sigma^2 + \rho\sigma^2 - \rho\sigma^2 - \sigma^2 \\ \sigma^2 - \rho\sigma^2 + \rho\sigma^2 - \sigma^2 & \sigma^2 - \rho\sigma^2 - \rho\sigma^2 + \sigma^2 \end{pmatrix} =$$

$$= \begin{pmatrix} 2(\sigma^2 + \rho\sigma^2) & 0 \\ 0 & 2(\sigma^2 - \rho\sigma^2) \end{pmatrix}.$$

Como la $cov(W, V) = 0 \rightarrow \rho_{WV} = 0$

**Problema** (45 minutos, 5 puntos)

**En un material se analizan tres variables aleatorias, $X$, $Y$, $Z$, que representan criterios de calidad del material. Las medias de esas tres variables son $10$, $20$ y $30$ respectivamente, y la matriz de varianzas y covarianzas**

$$\begin{bmatrix} 8 & 4 & 0 \\ 4 & 4 & 4 \\ 0 & 4 & 8 \end{bmatrix}$$

**La distribución conjunta es normal multivariante.**

**a) Si se toman $5$ piezas al azar de forma independiente ¿cual es la probabilidad de que el promedio de las mediciones de $X$ sea mayor que $12$?**

$\overline{X} \sim N(\mu = 10, \sigma^2 = 8/5 = 1,6)$
$P(\overline{X} > 35) = 1 - \phi((12-10)/\sqrt{1,6}) = 1 - \phi(1,58) = 1 - 0,943 = 0,057$

**b) Se han construido dos índices de calidad conjunta $Q_1 = 3X + 2Y + Z$ y $Q_2 = X + 2Y - Z$, ¿cuál es la probabilidad de que $Q_1 > Q_2 + 90$?**

$$D = Q_1 - Q_2; \begin{bmatrix} E[Q_1] \\ E[Q_2] \end{bmatrix} = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \\ 30 \end{bmatrix} = \begin{bmatrix} 100 \\ 20 \end{bmatrix}$$

$$M_Q = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 8 & 4 & 0 \\ 4 & 4 & 4 \\ 0 & 4 & 8 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 32 & 24 & 16 \\ 16 & 8 & 0 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 160 & 64 \\ 64 & 32 \end{bmatrix}$$

$D \sim N(\mu = 100 - 20 = 80; \sigma^2 = 160 + 32 - 2 \times 64 = 64)$

$P(D > 90) = 1 - \phi((90-80)/\sqrt{64}) = 1 - \phi(1,25) = 1 - 0,89 = 0,11$

**c) Si se toman dos piezas al azar de forma independiente ¿cuál es la probabilidad de que el mínimo de los dos valores de $Z$ sea mayor que $34$?**

$P(\text{mín} > 34) = P(Z_1 > 34, Z_2 > 34) = (F_Z(34))^2 = (\phi((34-30)/\sqrt{8}))^2 = (1-\phi(1,41))^2 = (1-0,92)^2 = 0,08^2 = 0,0064$

**Cuestiones** (45 minutos, 5 puntos)

1. Sea $X_1, X_2$, ...., $X_n$ una muestra aleatoria simple de una variable aleatoria $N(\mu, \sigma)$.

   Se propone como estimador de la varianza $\widehat{\sigma}^2 = a \sum_{i=1}^{n} (X_i - \overline{X})^2$ . Calcule el valor de $a$ para que $\widehat{\sigma}^2$ sea un estimador centrado, así como el valor de $a$ para que tenga Error Cuadrático Medio (ECM) mínimo.

2. Un zoólogo desea conocer el número $n$ de ejemplares (individuos) de una especie en peligro de extinción en un Parque Natural. Para ello utiliza un sistema de emisión-detección electrónico. Este sistema tiene una probabilidad de error de no detección del 20 %. En 10 periodos de observación, el sistema ha detectado 5, 5, 6, 4, 6, 3, 5, 4, 4, 5 individuos. Estime por el método de los momentos el número de individuos en el parque. Teniendo en cuenta los datos y la estimación obtenida, razone si es un buen estimador.

3. Los doscientos primeros dígitos del numero $\pi$ tienen la siguiente frecuencia:

| Dígito | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frecuencia | **19** | **20** | **24** | **19** | **22** | **20** | **16** | **12** | **25** | **23** |

   Contraste si los diez dígitos son igualmente probables. ($\alpha = 0{,}05$).

**Problema** (45 minutos, 5 puntos)

La Red de Carreteras del Estado (RCE) está dividida en segmentos o tramos, con similares características de intensidad de tráfico, en donde se registran el número de accidentes ocurridos en un determinado periodo de tiempo. Cada uno de los tramos puede tener un longitud diferente. En la tabla se presentan los datos correspondientes a una muestra aleatoria de 10 tramos de la RCE. Para cada uno de los tramos se recogen el número de accidentes ocurridos en el año 2008, así como su longitud en kilómetros (km).

| Tramo | Longitud ($km$) | Accidentes | Tramo | Longitud ($km$) | Accidentes |
|---|---|---|---|---|---|
| 1 | 7 | 6 | 6 | 3 | 3 |
| 2 | 29 | 45 | 7 | 7 | 5 |
| 3 | 21 | 31 | 8 | 25 | 16 |
| 4 | 16 | 11 | 9 | 24 | 25 |
| 5 | 3 | 3 | 10 | 3 | 5 |

En los estudios de accidentología se considera que el número de accidentes anuales en un tramo es una variable aleatoria con distribución de Poisson, y que la siniestralidad en un tramo es independiente de la de cualquier otro tramo.

1. Con los datos de la tabla, estime el número medio de accidentes por km para 2008 y proporcione un intervalo de confianza 95 %.

2. En los informes oficiales aparece reflejado, para el año 2008, que el número medio accidentes cada km es $\lambda = 1$ accidentes/km.

   Indique los tramos que presentan un número de accidentes significativamente mayor que el esperado (para $\lambda = 1$ accidentes/km). Utilice un nivel de significación $\alpha = 0{,}05$ . A estos tramos se les denomina "tramos de concentración de accidentes". Indique para dichos tramos el número máximo de accidentes que deberían haberse producido para no considerarlos como tramos de concentración de accidentes.

3. Debido a las medidas implantadas, el número medio de accidentes cada km se ha reducido. En el año 2010, el número total de accidentes en los diez tramos reflejados en la tabla ha sido 126. Realice el contraste $\begin{cases} H_0 : \lambda = 0{,}8 \\ H_1 : \lambda > 0{,}8 \end{cases}$. ( Utilice $\alpha = 0{,}05$).

   Calcule el nivel crítico del contraste (p-valor) y la probabilidad de error tipo II cuando $\lambda = 1$ accidentes/km .

**Cuestiones** (45 minutos, 5 puntos)

1. **Sea $X_1, X_2$, ...., $X_n$ una muestra aleatoria simple de una variable aleatoria $N(\mu, \sigma)$.**

   **Se propone como estimador de la varianza $\widehat{\sigma}^2 = a\sum_{i=1}^{n}(X_i - \overline{X})^2$. Calcule el valor de $a$ para que $\widehat{\sigma}^2$ sea un estimador centrado, así como el valor de $a$ para que tenga Error Cuadrático Medio (ECM) mínimo.**

   a) Para que el estimador $\widehat{\sigma}^2 = a\sum_{i=1}^{n}(X_i - \overline{X})^2$ sea centrado basta con que $\widehat{\sigma}^2 = \widehat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, de donde $a = n - 1$.

   También se obtiene ese resultado imponiendo que $E(\widehat{\sigma}^2) = \sigma^2$, que es la condición para que el sesgo sea 0. En ese caso $E(\widehat{\sigma}^2) = E(a\sum_{i=1}^{n}(X_i - \overline{X})^2) = a{\cdot}n{\cdot}\sigma^2\frac{n-1}{n}$, obtenido utilizando que $X \to N(\mu, \sigma)$ y que $\overline{X} \to N(\mu, \sigma/\sqrt{n})$ y las propiedades del operador Esperanza ($E$).

   b) Para minimizar el ECM:

   $ECM(\widehat{\sigma}^2) = \left(sesgo(\widehat{\sigma}^2)\right)^2 + var(\widehat{\sigma}^2)$ utilizamos que $sesgo(\widehat{\sigma}^2) = E(\widehat{\sigma}^2) - \sigma^2$ y del apartado anterior se tiene que $sesgo(\widehat{\sigma}^2) = \sigma^2(an - 1 - 1)$.

   Para la $var(\widehat{\sigma}^2) = var\left[a\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = a^2 var\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right]$. Y utilizando que $\frac{ns^2}{\sigma^2} \to \chi^2_{n-1}$, o $\frac{(n-1)\widehat{s}^2}{\sigma^2} \to \chi^2_{n-1}$, y que $\widehat{s}^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$ y que la $var(\chi^2_m) = 2m$ se tiene que $var(\widehat{\sigma}^2) = a^2\sigma^4 2(n-1)$ y sustituyendo en $ECM(\widehat{\sigma}^2) = \left(sesgo(\widehat{\sigma}^2)\right)^2 + var(\widehat{\sigma}^2)$ el valor del sesgo y la varianza y operando queda la siguiente expresión en función de $a$:

   $$ECM(\widehat{\sigma}^2) = \sigma^4(-a^2 + 2a + a^2n^2 + 1 - 2an).$$

   Para minimizar derivamos respecto de $a$ e igualamos a 0 :

   $$\begin{aligned} \frac{d(ECM(a))}{da} &= \sigma^4(-2a + 2 + 2an^2 - 2n), \\ \sigma^4(-2a + 2 + 2an^2 - 2n) &= 0 \text{ cuando } a = \frac{1}{n+1}. \end{aligned}$$

2. **Un zoólogo desea conocer el número $n$ de ejemplares (individuos) de una especie en peligro de extinción en un Parque Natural. Para ello utiliza un sistema de emisión-detección electrónico. Este sistema tiene una probabilidad de error de no detección del 20 %. En 10 periodos de observación, el sistema ha detectado 5, 5, 6, 4, 6, 3, 5, 4, 4, 5 individuos. Estime por el método de los momentos el número de individuos en el parque. Teniendo en cuenta los datos y la estimación obtenida, razone si es un buen estimador.**

   $p(\textit{detección}) = 1 - 0{,}2 = 0{,}8$

   Somo sólo estimamos un parámetro sólo igualamos momento de orden 1 muestral y el teórico para el modelo, Binomial ($n$, 0.8) en este caso.

$\overline{x} = \frac{5+5+6+4+6+3+5+4+4+5}{10} = 4{,}7$

$\widehat{n}{\cdot}0{,}8 = \overline{x} = 4{,}7 \rightarrow \widehat{n} = \frac{4{,}7}{0{,}8} = 5{,}875$

Discrepancia entre el máximo valor observado y el estimador para $n$.

3. **Los doscientos primeros dígitos del numero $\pi$ tienen la siguiente frecuencia:**

| Dígito | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Frecuencia** | **19** | **20** | **24** | **19** | **22** | **20** | **16** | **12** | **25** | **23** |

**Contraste si los diez dígitos son igualmente probables.** ($\alpha = 0{,}05$).

Contraste de bondad de ajuste de la Chi-cuadrado.

Utilizamos que $\sum \frac{(O_i - E_i)^2}{E_i} \rightarrow \chi^2_{K-r-1}$, y en este caso $K = 10$ y $r = 0$ pues no se estima ningún parámetro. Contraste UNILATERAL.

| Dígito | $O_i$ | $E_i$ | $\frac{(O_i-E_i)^2}{E_i}$ |
|---|---|---|---|
| 0 | 19 | 20 | 0.05 |
| 1 | 20 | 20 | 0 |
| 2 | 24 | 20 | 0.8 |
| 3 | 19 | 20 | 0.05 |
| 4 | 22 | 20 | 0.2 |
| 5 | 20 | 20 | 0 |
| 6 | 16 | 20 | 0.8 |
| 7 | 12 | 20 | 3.25 |
| 8 | 25 | 20 | 1.25 |
| 9 | 23 | 20 | 0.45 |

Sumando la última columna: $\sum \frac{(O_i-E_i)^2}{E_i} = 6{,}8 < 16{,}92$, que es el valor obtenido en tablas para $\chi^2_{9;0{,}05}$. No se rechaza la hipótesis nula (todos los dígitos igualmente probables).

**Problema** (45 minutos, 5 puntos)

**La Red de Carreteras del Estado (RCE) está dividida en segmentos o tramos, con similares características de intensidad de tráfico, en donde se registran el número de accidentes ocurridos en un determinado periodo de tiempo. Cada uno de los tramos puede tener un longitud diferente. En la tabla se presentan los datos correspondientes a una muestra aleatoria de 10 tramos de la RCE. Para cada uno de los tramos se recogen el número de accidentes ocurridos en el año 2008, así como su longitud en kilómetros (km).**

| Tramo | Longitud ($km$) | Accidentes | Tramo | Longitud ($km$) | Accidentes |
|---|---|---|---|---|---|
| 1 | 7 | 6 | 6 | 3 | 3 |
| 2 | 29 | 45 | 7 | 7 | 5 |
| 3 | 21 | 31 | 8 | 25 | 16 |
| 4 | 16 | 11 | 9 | 24 | 25 |
| 5 | 3 | 3 | 10 | 3 | 5 |

**En los estudios de accidentología se considera que el número de accidentes anuales en un tramo es una variable aleatoria con distribución de Poisson, y que la siniestralidad en un tramo es independiente de la de cualquier otro tramo.**

1. **Con los datos de la tabla, estime el número medio de accidentes por km para 2008 y proporcione un intervalo de confianza 95 %.**

2. **En los informes oficiales aparece reflejado, para el año 2008, que el número medio accidentes cada km es $\lambda = 1$ accidentes/km.**

   **Indique los tramos que presentan un número de accidentes significativamente mayor que el esperado (para $\lambda = 1$ accidentes/km). Utilice un nivel de significación $\alpha = 0{,}05$. A estos tramos se les denomina "tramos de concentración de accidentes". Indique para dichos tramos el número máximo de accidentes que deberían haberse producido para no considerarlos como tramos de concentración de accidentes.**

3. **Debido a las medidas implantadas, el número medio de accidentes cada km se ha reducido. En el año 2010, el número total de accidentes en los diez tramos reflejados en la tabla ha sido 126. Realice el contraste** $\begin{cases} H_0 : \lambda = 0{,}8 \\ H_1 : \lambda > 0{,}8 \end{cases}$**. (Utilice** $\alpha = 0{,}05$).

   **Calcule el nivel crítico del contraste (p-valor) y la probabilidad de error tipo II cuando $\lambda = 1$ accidentes/km .**

**Apartado 1)**

Sean $a_i$ y $l_i$ el número de accidentes y la longitud del tramo $i$ respectivamente.

$$\widehat{\lambda} = \frac{\sum_{i=1}^{10} a_i}{\sum_{i=1}^{10} l_i} = \frac{150 \text{ (n}^{\text{o}} \text{ de accidentes)}}{138 \text{ (n}^{\text{o}} \text{ de km)}} = 1{,}09 \frac{accidentes}{km}$$

$\lambda \in \widehat{\lambda} \pm 1,96\sqrt{\widehat{\lambda}/138} \Longrightarrow \lambda \in (0,916; 1,264)$ con 95 % de confianza.

**Apartado 2)**

$X$ : $v.a$ nº de accidentes en el tramo $i \Longrightarrow Poisson(\lambda_i = \lambda \times l_i) \rightsquigarrow N(\lambda \times l_i, \sqrt{\lambda \times l_i})$

Para saber si un tramo es un *tramo de concentración de accidentes*, lo que implicaría $\lambda > 1\dfrac{accidentes}{km}$, se debe realizar el siguiente contraste

$$\begin{cases} H_0 : \lambda_i = l_i \\ H_1 : \lambda > l_i \end{cases}$$

El contraste para cada uno de los tramos sería:

$z = \dfrac{a_i - l_i}{\sqrt{l_i}} \Longrightarrow$Se rechazará $H_0$ si $\dfrac{a_i - l_i}{\sqrt{l_i}} > 1{,}64 \Longrightarrow$Se rechazará $H_0$ si $a_i > l_i + 1{,}64\sqrt{l_i}$

**Para el tramo 2**, $z_0 = (45 - 29)/\sqrt{29} = 2{,}97 > 1{,}64 \Longrightarrow$Se rechaza $H_0$
Número máximo $= (A_{máx} - 29)/\sqrt{29} = 1{,}64 \Longrightarrow A_{máx} = 37{,}83. \approx 37$ accidentes.

**Para el tramo 3**, $z_0 = (31 - 21)/\sqrt{21} = 2{,}18 > 1{,}64 \Longrightarrow$Se rechaza $H_0$
Número máximo $= (A_{máx} - 21)/\sqrt{21} = 1{,}64 \Longrightarrow A_{máx} = 28{,}51. \approx 28$ accidentes.

Estos son los dos únicos tramos de concentración de accidentes. En el resto de los tramos no se rechaza la $H_0$.

| Estadístico $z$ | nº máximo |
|---|---|
| -0.378 | 11.35 |
| 2.97 | 37.83 |
| 2.18 | 28.51 |
| -1.33 | 22.58 |
| 0 | 5.84 |
| 0 | 5.84 |
| 0 | 11.35 |
| -1.8 | 33.22 |
| 0.2 | 32.05 |
| 1.15 | 5.84 |

**Apartado 3)**

Con los nuevos datos

$\widehat{\lambda} = \dfrac{126 \text{ (nº de accidentes)}}{138 \text{ (nº de km)}} = 0{,}9138\dfrac{accidentes}{km}$

En el contraste planteado, se rechazará $H_0$ si $\dfrac{\widehat{\lambda} - 0{,}8}{\sqrt{\frac{0{,}8}{138}}} > 1{,}64$;

Con los datos $z_0 = \dfrac{0{,}9138 - 0{,}8}{\sqrt{\frac{0{,}8}{138}}} = 1{,}4841 < 1{,}64 \Longrightarrow$ No hay evidencia suficiente para rechazar la Hipótesis nula

También se rechazará $H_0$ si $\widehat{\lambda} > 0{,}8 + 1{,}64 \times \sqrt{\frac{0{,}8}{138}} = 0{,}9249$

$P - valor = P(Z > 1{,}4841) = 1 - \Phi(1{,}48) = 0{,}0694$

$P(error\ tipo\ II) = P(Aceptar H_0\ |H_0\ es\ falsa) = P(\widehat{\lambda} < 0{,}9249|\lambda = 1\dfrac{accidentes}{km}) = P\left(\dfrac{\widehat{\lambda} - 1}{\sqrt{\frac{1}{138}}} < \dfrac{0{,}9241 - 1}{\sqrt{\frac{1}{138}}}\right) =$

$P(Z < -0{,}88) = \Phi(-0{,}88) = 0{,}19.$

**Problema 1** (60 minutos – 10 puntos)

Una persona tiene cinco monedas en una bolsa, dos de las cuales tienen caras por ambos lados, otra tiene cruces por ambos lados y otras dos son normales.

a) Coge una moneda al azar y la lanza, ¿cuál es la probabilidad de obtener CARA? Observa que el lado superior de la moneda es CARA. Sabiendo esto, ¿cuál es la probabilidad de que el otro lado de la moneda sea también CARA?

b) Devuelve la moneda a la bolsa. ¿Cuál es la probabilidad de obtener dos caras en dos lanzamientos consecutivos en los casos siguientes?
   1. Toma una moneda al azar de la bolsa y la misma moneda la lanza dos veces.
   2. Toma una moneda al azar de la bolsa, la lanza, a continuación devuelve la moneda a la bolsa y después de mezclar repite la extracción al azar y lanza por segunda vez.

c) ¿Cuál es la probabilidad de obtener dos caras si en una única extracción saca dos monedas al azar de la bolsa y las lanza?

d) En el caso del apartado c, se han obtenido dos CARAS. Sabiendo esto, ¿cuál es la probabilidad de que las dos monedas lanzadas sean iguales (las dos del mismo tipo)?

**Problema 1** (60 minutos, 10 puntos)

a) Llamando $M_1, M_2$ y $M_3$ a las monedas de dos caras, dos cruces y normales, respectivamente, la probabilidad de cada una de ellas es

$$P(M_1) = 2/5; P(M_2) = 1/5; P(M_3) = 2/5$$

Aplicando el teorema de la probabilidad total se obtiene

1. Llamando C al suceso obtener cara,
   $P(C) = P(C|M_1)P(M_1) + P(C|M_2)P(M_2) + P(C|M_3)P(M_3)$

$P(C) = 1 \times \frac{2}{5} + 0 \times \frac{1}{5} + \frac{1}{2} \times \frac{2}{5} = \frac{3}{5}$

2. Si el otro lado es otra cara, la moneda ha de ser de tipo M1. Por tanto, aplicando la probabilidad condicionada se tiene que:

$$P(M_1|C) = \frac{P(M_1 \cap C)}{P(C)} = \frac{P(M_1)}{P(C)} = \frac{\frac{2}{5}}{\frac{3}{5}} = \frac{2}{3}$$

b) Llamamos $C_1$ a suceso "obtener cara en el primer lanzamiento͞z $C_2$ al suceso "obtener cara en el segundo lanzamiento".

1. Si es con la misma moneda, se tiene que aplicando el teorema de la probabilidad total

$$\begin{aligned} P(C_1 \cap C_2) &= P(C_1 \cap C_2|M1) \cdot P(M_1) + P(C_1 \cap C_2|M2) \cdot P(M_2) + P(C_1 \cap C_2|M3) \cdot P(M_3) \\ &= 1 \times \frac{2}{5} + 0 \times \frac{1}{5} + \left(\frac{1}{2}\right)^2 \times \frac{2}{5} = \frac{1}{2}. \end{aligned}$$

2. Si se extraen dos monedas, los sucesos son independientes y

$$P(C_1 \cap C_2) = P(C_1)P(C_2) = \frac{3}{5} \times \frac{3}{5} = \frac{9}{25}.$$

c) Los posibles resultados y sus probabilidades para el tipo de las dos monedas extraídas son (el orden es irrelevante):

$$P(M_1 \cap M_1) = \tfrac{1}{10} \quad P(M_1 \cap M_2) = \tfrac{2}{10} \quad P(M_1 \cap M_3) = \tfrac{4}{10} \quad P(M_3 \cap M_3) = \tfrac{1}{10} \quad P(M_2 \cap M_3) = \tfrac{2}{10}$$

La probabilidad del suceso $C_1 \cap C_2$ se obtiene mediante

$$\begin{aligned} P(C_1 \cap C_2) &= P(C_1 \cap C_2|M_1 \cap M_1)P(M_1 \cap M_1) + P(C_1 \cap C_2|M_1 \cap M_2)P(M_1 \cap M_2) + \\ &\quad P(C_1 \cap C_2|M_1 \cap M_3)P(M_1 \cap M_3) + P(C_1 \cap C_2|M_3 \cap M_3)P(M_3 \cap M_3) + \\ &\quad P(C_1 \cap C_2|M_2 \cap M_3)P(M_2 \cap M_3) \\ &= 1 \times \frac{1}{10} + 0 \times \frac{2}{10} + \frac{1}{2} \times \frac{4}{10} + \frac{1}{4} \times \frac{1}{10} + 0 \times \frac{2}{10} = \frac{13}{40} \end{aligned}$$

d) La probabilidad de $M_1 \cap M_1$ si hemos observado $C_1 \cap C_2$ es

$$P(M_1 \cap M_1 | C_1 \cap C_2) = \frac{P(C_1 \cap C_2 | M_1 \cap M_1) P(M_1 \cap M_1)}{P(C_1 \cap C_2)} = \frac{1}{10} / \frac{13}{40} = \frac{4}{13}$$

y la probabilidad de $M_3 \cap M_3$ si hemos observado $C_1 \cap C_2$ es

$$P(M_3 \cap M_3 | C_1 \cap C_2) = \frac{P(C_1 \cap C_2 | M_3 \cap M_3) P(M_3 \cap M_3)}{P(C_1 \cap C_2)} = \frac{1}{40} / \frac{13}{40} = \frac{1}{13}$$

por tanto la probabilidad pedida es

$$\frac{4}{13} + \frac{1}{13} = \frac{5}{13}.$$

**Problema 2** (45 minutos – 10 puntos)

Un profesor de Educación Física sabe que la velocidad en carrera de sus alumnas de 4º ESO en larga distancia, X, sigue una distribución normal de media 7,5 Km/h y desviación típica 1.5 Km/h. Y sus alumnos del mismo curso una velocidad Y que sigue una distribución de media 9 Km/h con una desviación típica de 1 Km/k.

a) ¿Cuál es la probabilidad de que una alumna venza a un alumno?
b) El profesor está formando equipos de relevos (de larga distancia de dos alumnos, en donde se mantienen las velocidades X e Y anteriores). Para ello se plantea dos tipos de equipos, el primero, V, formado por un alumno y una alumna de 4º de la ESO y el segundo, W, formado por dos alumnos. Obtener la función de densidad conjunta de la variable aleatoria bidimensional $[V\ W]^T$ razonando la respuesta.
c) ¿Cuál es la probabilidad de que un equipo V venza a uno W?
d) El profesor de educación física se plantea la siguiente competición con sus alumnos: forma un equipo sólo con alumnos y otro sólo con alumnas y los pone a competir, es decir, una alumna compite siempre contra un alumno. Decide que las competiciones pararan cuando una alumna gane a un alumno. ¿Cuántas competiciones se realizarán por término medio?

**Problema 2** (60 minutos, 10 puntos)

La velocidad de las alumnas de $4^o$ de ESO ($X$) sigue una distribución $Niid(7,5, 1,5)$ y la de los alumnos, ($Y$) $Niid(9, 1)$. La velocidad de los alumnos y las alumnas de este curso son independientes.

a) La probabilidad de que una alumna venza a un alumno es:

$$P(X > Y) = P(X - Y > 0)$$

donde $X - Y$ sigue una distribución normal de media $7{,}5 - 9$ y de desviación típica $\sqrt{1{,}5^2 + 1^2}$, así $X - Y$ es $Niid(-1{,}5, \sqrt{3{,}25})$.Por lo tanto la probabilidad pedida es:

$$P(X > Y) = P(X - Y > 0) = P(Z > \frac{1{,}5}{\sqrt{3{,}25}}) = 0{,}2033$$

b) El profesor plantea dos equipos de la siguiente manera:

$$V = X_1 + Y_2$$

$$W = Y_3 + Y_4$$

donde las cuatro variables originales son independientes. Así:

$$E\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} 16{,}5 \\ 18 \end{bmatrix} \; y \; M_{VW} = \begin{bmatrix} 1{,}5^2 + 1^2 & 0 \\ 0 & 1^2 + 1^2 \end{bmatrix} = \begin{bmatrix} 3{,}25 & 0 \\ 0 & 2 \end{bmatrix}$$

Como $V$ y $W$ son independientes, la función de densidad conjunta se puede obtener como el producto de las funciones de densidad marginales, según:

$$f_{VW}(v, w) = f_V(v) f_W(w)$$

obteniendo:

$$f_{VW}(v, w) = \frac{1}{2\pi\sqrt{3{,}25}\sqrt{2}} \exp\left[-\frac{1}{2}\left[\left(\frac{v - 16{,}5}{\sqrt{3{,}25}}\right)^2 + \left(\frac{w - 18}{\sqrt{2}}\right)^2\right]\right]$$

c) La probabilidad de que un equipo $V$ venza a uno $W$ es:

$$P(V > W) = P(V - W > 0)$$

donde $V - W$ sigue una distribución normal de media $16{,}5 - 18$ y de desviación típica $\sqrt{3{,}25 + 2}$, así $V - W$ es $Niid(-1{,}5, \sqrt{5{,}25})$. Por lo tanto la probabilidad pedida es:

$$P(V > W) = P(V - W > 0) = P(Z > \frac{1{,}5}{\sqrt{5{,}25}}) = 0{,}2578$$

d) El número de competiciones, $C$, que se tendrán que realizar hasta que una alumna gane a un alumno sigue una distribución geométrica. Como nos piden el número medio de ellas, debemos obtener la esperanza de la variable geométrica $C$. Por lo tanto, $E[C] = \frac{1}{p}$, siendo $p$ la probabilidad de que una alumna gane a un alumno, que se ha calculado en el primer apartado de este problema. Así:

$$E[C] = \frac{1}{p} = \frac{1}{0{,}2033} = 4{,}9188 \simeq 5 \text{ competiciones}$$

**Problema 3** (60 minutos – 10 puntos)

La fábrica manchega *IDEM®* dispone de una prensa de laminación de planchas de cobre. En el almacén encontramos un pedido de cuarenta láminas de cobre. Hemos medido la longitud de las láminas, obteniendo la siguiente tabla de frecuencias:

| **Clase (en metros)** | **Frecuencia** |
|---|---|
| $< 14.5$ | 5 |
| [ 14.5 y 15.0 ) | 14 |
| [ 15.0 y 15.5 ) | 14 |
| [ 15.5 y 16.0 ) | 5 |
| 16.0 | 2 |

a) Compruebe, mediante una bondad de ajuste, si es razonable pensar que la longitud de las láminas se distribuye según una distribución normal de media 15 m y varianza de 0.5 $m^2$. (Emplear $\alpha$=0.05)

b) Cuando la máquina está calibrada, produce unas láminas de cobre cuyo grosor puede modelarse mediante una variable aleatoria normal con media 5 mm y una desviación típica de 0.1 mm. Sin embargo, si no está calibrada, genera unas láminas cuyo grosor se distribuye según una normal cuya varianza y/o media pueden ser diferentes.

Tenemos un pedido de cinco láminas con los siguientes grosores:

5.16, 4.99, 4.92, 4.96, 5.09

Realice el contraste de hipótesis necesario para determinar si la **media** de la anterior muestra cumple con las especificaciones de la máquina calibrada. (Emplear $\alpha$=0.05)

c) Realice el contraste de hipótesis necesario para determinar si la **varianza** de la anterior muestra cumple con las especificaciones de la máquina calibrada. (Emplear $\alpha$=0.05)

d) Debido a la antigüedad de la máquina, un alto porcentaje de láminas son defectuosas. El año pasado, el 5% de láminas fabricadas resultaron defectuosas. Este año, de las 100 láminas fabricadas, 9 eran defectuosas.

Indique si se puede concluir que la proporción de defectuosas ha aumentado.

Calcule también la probabilidad de error de tipo II en el caso de que el porcentaje de defectuosas sea realmente un 10%. (Emplear aproximación a la normal y $\alpha$=0.05).

**Problema 3** (60 minutos – 10 puntos)

La fábrica manchega *IDEM*® dispone de una prensa de laminación de planchas de cobre. En el almacén encontramos un pedido de cuarenta láminas de cobre. Hemos medido la longitud de las láminas, obteniendo la siguiente tabla de frecuencias:

| **Clase (en metros)** | **Frecuencia** |
|---|---|
| $< 14.5$ | 5 |
| [ 14.5 y 15.0 ) | 14 |
| [ 15.0 y 15.5 ) | 14 |
| [ 15.5 y 16.0 ) | 5 |
| $\geq 16.0$ | 2 |

a) Compruebe, mediante una bondad de ajuste, si es razonable pensar que la longitud de las láminas se distribuye según una distribución normal de media 15 m y varianza de 0.5 $m^2$. (Emplear $\alpha$=0.05)

b) Cuando la máquina está calibrada, produce unas láminas de cobre cuyo grosor puede modelarse mediante una variable aleatoria normal con media 5 mm y una desviación típica de 0.1 mm. Sin embargo, si no está calibrada, genera unas láminas cuyo grosor se distribuye según una normal cuya varianza y/o media pueden ser diferentes.

   Tenemos un pedido de cinco láminas con los siguientes grosores:

   5.16, 4.99, 4.92, 4.96, 5.09

   Realice el contraste de hipótesis necesario para determinar si la **media** de la anterior muestra cumple con las especificaciones de la máquina calibrada. (Emplear $\alpha$=0.05)

c) Realice el contraste de hipótesis necesario para determinar si la **varianza** de la anterior muestra cumple con las especificaciones de la máquina calibrada. (Emplear $\alpha$=0.05)

d) Debido a la antigüedad de la máquina, un alto porcentaje de láminas son defectuosas. El año pasado, el 5% de láminas fabricadas resultaron defectuosas. Este año, de las 100 láminas fabricadas, 9 eran defectuosas.
   Indique si se puede concluir que la proporción de defectuosas ha aumentado.
   Calcule también la probabilidad de error de tipo II en el caso de que el porcentaje de defectuosas sea realmente un 10%. (Emplear aproximación a la normal y $\alpha$=0.05).

# SOLUCIÓN

## APARTADO A

Hemos de hacer el contraste:

$$H_0 : X_i \to \text{Normal}$$
$$H_1 : X_i \nrightarrow \text{Normal}$$

Primero calculamos la probabilidad correspondiente a cada tramo:

$X \sim N(15, 0.71)$

$p1 = P(X \leq 14.5) = 0.24$

$p2 = P(14.5 < X \leq 15.0) = 0.26$

$p3 = P(15.0 < X \leq 15.5) = 0.26$

$p4 = 0.16$ $p5 = 0.09$

Después calculamos las frecuencias esperadas:

| Clase (en metros) | Frec. Observada ($O_k$) | Frec. Esperada ($E_k$) | $\frac{(O_k - E_k)^2}{E_k}$ |
|---|---|---|---|
| $< 14.5$ | 5 | $p1 \cdot 40 = 9.6$ | 2.20 |
| 14.5 – 15.0 | 14 | $p2 \cdot 40 = 10.4$ | 1.25 |
| 15.0 – 15.5 | 14 | $p3 \cdot 40 = 10.4$ | 1.25 |
| 15.5 – 16.0 | 5 | 6.4 | 0.31 |
| $16.0 <$ | 2 | 3.2 | 0.45 |

Sabemos que:

$$\sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} \to \chi^2_{K-r-1}$$

donde r=0 (ya que no hemos estimado ningún parámetro), y K = 5.

$$\chi^2_0 = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = 5.45$$

De las tablas: $P(\chi^2_4 \leq 9.48) = 0.95$

Por tanto, como 5.45 < 9.48, aceptamos la hipótesis nula. No hay evidencia suficiente para rechazar el modelo propuesto.

## APARTADO B

Hemos de realizar el siguiente contraste bilateral:

$$\begin{cases} H_0:\ \mu = 5 \\ H_1:\ \mu \neq 5 \end{cases}$$

**Observación**: Como no sabemos si la máquina está calibrada o no, desconocemos la varianza poblacional. Por tanto, hemos de hacer un contraste de media con varianza desconocida.

Sabemos que

$$\frac{\bar{x}-\mu}{\frac{\hat{s}}{\sqrt{5}}} \to t_4$$

$$t_0 = \frac{\bar{x}-\mu}{\frac{\hat{s}}{\sqrt{5}}} = \frac{5.024-5}{\frac{0.09}{\sqrt{5}}} = 0.59$$

De las tablas obtenemos que P(–2.77 < $t_4$ < 2.77) = 0.95.

Como se cumple que '–2.77 < 0.59 < 2.77', se deduce que no hay evidencia estadística para rechazar $H_0$ con un nivel de confianza del 95%. **Aceptamos $H_0$.**

## APARTADO C

Hemos de realizar el siguiente contraste bilateral:

$$\begin{cases} H_0\text{: } \sigma^2 = 0.1 \\ H_1\text{: } \sigma^2 \neq 0.1 \end{cases}$$

Sabemos que

$$\frac{(n-1)\hat{s}^2}{\sigma^2} \to \chi^2_{n-1}$$

$$\chi^2_0 = \frac{(5-1)\cdot 0.0986^2}{0.01} = 3.89$$

De las tablas obtenemos que P(0.48 < $\chi^2_4$ < 11.14) = 0.95.

Como se cumple que '0.48 < 3.89 < 11.14', se deduce que no hay evidencia estadística para rechazar $H_0$ con un nivel de confianza del 95%. **Aceptamos $H_0$.**

## APARTADO D

Primero hemos de realizar el siguiente contraste unilateral:

$$\begin{cases} H_0\text{: } p = 0.05 \\ H_1\text{: } p > 0.05 \end{cases}$$

Si $H_0$ es cierto:

$$\hat{p} \sim N\left(0.05, \sqrt{\frac{0.05\cdot 0.95}{100}}\right)$$

$$P(\hat{p} > 0.086) = 0.05$$

Como 9/100 = 0.09 > 0.086, se rechaza la hipótesis nula, y se concluye que ha aumentado el porcentaje de defectuosas.

La probabilidad del error del tipo II se calcula como la probabilidad de aceptar $H_0$ cuando en realidad $H_0$ no es cierta. Sería la probabilidad de que:

$$P(\hat{p}' < 0.086), \quad \text{donde } \hat{p}' \sim N\left(0.1, \sqrt{\frac{0.1\cdot 0.9}{100}}\right) = N(0.1, 0.03)$$

Resulta que $P(\hat{p}' < 0.086) = P(\text{error tipo II}) = 0.32$

**Problema 3** (60 minutos – 4 puntos)

La fábrica manchega *IDEM®* dispone de una prensa de laminación de planchas de cobre. En el almacén encontramos un pedido de cuarenta láminas de cobre. Hemos medido la longitud de las láminas, obteniendo la siguiente tabla de frecuencias:

| **Clase (en metros)** | **Frecuencia** |
|---|---|
| $< 14.5$ | 5 |
| [ 14.5 y 15.0 ) | 14 |
| [ 15.0 y 15.5 ) | 14 |
| [ 15.5 y 16.0 ) | 5 |
| 16.0 | 2 |

a) Compruebe, mediante una bondad de ajuste, si es razonable pensar que la longitud de las láminas se distribuye según una distribución normal de media 15 m y varianza de 0.5 $m^2$. (Emplear $\alpha$=0.05)

b) Cuando la máquina está calibrada, produce unas láminas de cobre cuyo grosor puede modelarse mediante una variable aleatoria normal con media 5 mm y una desviación típica de 0.1 mm. Sin embargo, si no está calibrada, genera unas láminas cuyo grosor se distribuye según una normal cuya varianza y/o media pueden ser diferentes.

Tenemos un pedido de cinco láminas con los siguientes grosores:

5.16, 4.99, 4.92, 4.96, 5.09

Realice el contraste de hipótesis necesario para determinar si la **media** de la anterior muestra cumple con las especificaciones de la máquina calibrada. (Emplear $\alpha$=0.05)

c) Realice el contraste de hipótesis necesario para determinar si la **varianza** de la anterior muestra cumple con las especificaciones de la máquina calibrada. (Emplear $\alpha$=0.05)

d) Debido a la antigüedad de la máquina, un alto porcentaje de láminas son defectuosas. El año pasado, el 5% de láminas fabricadas resultaron defectuosas. Este año, de las 100 láminas fabricadas, 9 eran defectuosas.
Indique si se puede concluir que la proporción de defectuosas ha aumentado.
Calcule también la probabilidad de error de tipo II en el caso de que el porcentaje de defectuosas sea realmente un 10%. (Emplear aproximación a la normal y $\alpha$=0.05).

# SOLUCIÓN

## APARTADO A

Hemos de hacer el contraste:

$$H_0 : X_i \to \text{Normal}$$
$$H_1 : X_i \nrightarrow \text{Normal}$$

Primero calculamos la probabilidad correspondiente a cada tramo:

X ~ N(15, 0.71)

p1 = P(X 14.5) = 0.24

p2 = P(14.5 < X 15.0) = 0.26

p3 = P(15.0 < X 15.5) = 0.26

p4 = 0.16 p5 = 0.09

Después calculamos las frecuencias esperadas:

| Clase (en metros) | Frec. Observada ($O_k$) | Frec. Esperada ($E_k$) | $\frac{(O_k - E_k)^2}{E_k}$ |
|---|---|---|---|
| < 14.5 | 5 | p1 · 40 = 9.6 | 2.20 |
| 14.5 – 15.0 | 14 | p2 · 40 = 10.4 | 1.25 |
| 15.0 – 15.5 | 14 | p3 · 40 = 10.4 | 1.25 |
| 15.5 – 16.0 | 5 | 6.4 | 0.31 |
| 16.0 < | 2 | 3.2 | 0.45 |

Sabemos que:

$$\sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} \to \chi^2_{K-r-1}$$

donde r=0 (ya que no hemos estimado ningún parámetro), y K = 5.

$$\chi^2_0 = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = 5.45$$

De las tablas: $P(\chi^2_4 \leq 9.48) = 0.95$

Por tanto, como 5.45 < 9.48, aceptamos la hipótesis nula. No hay evidencia suficiente para rechazar el modelo propuesto.

## APARTADO B

Hemos de realizar el siguiente contraste bilateral:

$$\begin{cases} H_0\colon \ \mu = 5 \\ H_1\colon \ \mu \neq 5 \end{cases}$$

**Observación**: Como no sabemos si la máquina está calibrada o no, desconocemos la varianza poblacional. Por tanto, hemos de hacer un contraste de media con varianza desconocida.

Sabemos que

$$\frac{\bar{x}}{\frac{\hat{s}}{\sqrt{5}}}$$

$$\frac{\bar{x}}{\frac{\hat{s}}{\sqrt{5}}} \qquad \frac{\qquad}{\sqrt{5}}$$

De las tablas obtenemos que $P(-2.77 < t_4 < 2.77) = 0.95$.
Como se cumple que '$-2.77 < 0.59 < 2.77$', se deduce que no hay evidencia estadística para rechazar $H_0$ con un nivel de confianza del 95%. **Aceptamos $H_0$.**

## APARTADO C

Hemos de realizar el siguiente contraste bilateral:

$$\begin{cases} \\ \end{cases}$$

Sabemos que

$$\frac{(n \quad 1)\hat{s}}{\qquad}$$

$$\frac{(5 \quad 1)}{\qquad}$$

De las tablas obtenemos que $P(0.48 < \quad < 11.14) = 0.95$.
Como se cumple que '$0.48 < 3.89 < 11.14$', se deduce que no hay evidencia estadística para rechazar $H_0$ con un nivel de confianza del 95%. **Aceptamos $H_0$.**

## APARTADO D

Primero hemos de realizar el siguiente contraste unilateral:

$$\begin{cases} \\ \end{cases}$$

Si $H_0$ es cierto:

$$\hat{p} \sim N\left(0. \quad , \sqrt{\frac{\qquad}{\qquad}}\right)$$

$$P(\hat{p} \qquad 086)$$

Como $9/100 = 0.09 > 0.086$, se rechaza la hipótesis nula, y se concluye que ha aumentado el porcentaje de defectuosas.

La probabilidad del error del tipo II se calcula como la probabilidad de aceptar $H_0$ cuando en realidad $H_0$ no es cierta. Sería la probabilidad de que:

$$P(\hat{p} \qquad 086), \qquad \hat{p} \quad N\left(0.1, \sqrt{\frac{\qquad}{\qquad}}\right) \quad N(0. \qquad 03)$$

Resulta que $P(\hat{p} \qquad 086) \quad P(\text{error} \qquad \text{II})$

# EXÁMENES

## Curso 2013/14

## **Cuestiones** (45 minutos – 5 puntos)

**Cuestión 1:**

La función de densidad de una variable aleatoria X es la siguiente:

$$f_X(x) = \theta \cdot x, \quad 0 \leq x \leq 1.$$

Calcula el valor de $\theta$, la función de distribución, la media y la varianza de X.

**Cuestión 2:**

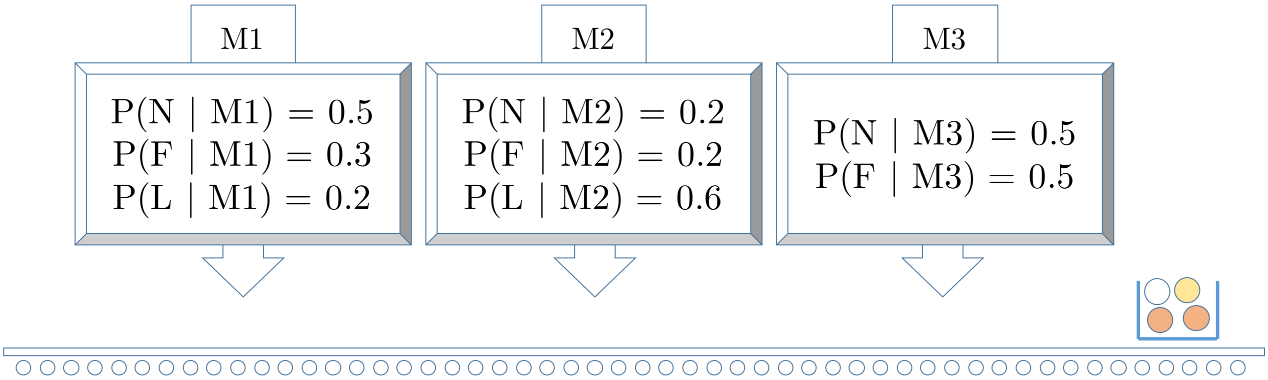
Para una determinada población se sabe que, de todos los jóvenes entre 18 y 25 años, un 60% no estudian y un 80% no trabajan.

Determina, justificando la respuesta, entre qué valor mínimo y máximo se encuentra el porcentaje de jóvenes que "ni estudia ni trabaja".

**Cuestión 3:**

En una industria se fabrican cajas con caramelos surtidos. Cada paquete contiene exactamente cuatro caramelos; y cada caramelo puede ser de Naranja (N), Fresa (F) o Limón (L). El sabor de cada caramelo es independiente del sabor del resto de caramelos del paquete.

Tenemos tres máquinas (M1, M2 y M3) que fabrican estas cajas, con las probabilidades que se muestran en la figura. Se sabe que la máquina M3 fabrica al doble de velocidad que el resto, es decir, que se cumple: P(M1) = P(M2) = P(M3)/**2**.




Cogemos del almacén un paquete al azar. ¿Cuál es la probabilidad de que haya sido fabricado por la máquina M1, si sabemos que la caja contiene únicamente caramelos de naranja?

**Problema** (45 minutos, 5 puntos)

La velocidad del viento en un parque eólico se suele modelar como una variable aleatoria con distribución de Rayleigh, cuya función de densidad viene determinada por (ver Figura 1):

$$f_X(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, x \geq 0, \sigma > 0.$$

El generador que se piensa colocar produce una potencia que depende de la velocidad del viento de acuerdo con la curva característica que se muestra en la Figura 2. En esta curva se observan tres zonas:

- Zona A: baja velocidad, por debajo de 3.5 m/s, donde la fuerza del viento es insuficiente para mover las palas.
- Zona B: con velocidades entre 3.5 m/s y 25 m/s, que es la zona de funcionamiento.
- Zona C: cuando la velocidad del viento está por encima de 25 m/s, la fuerza es excesiva y existe riesgo de daño en la estructura del generador, con lo que el molino se desconecta.

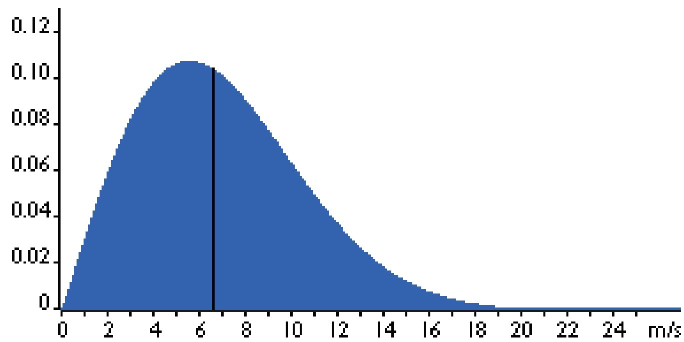



Figura 1.

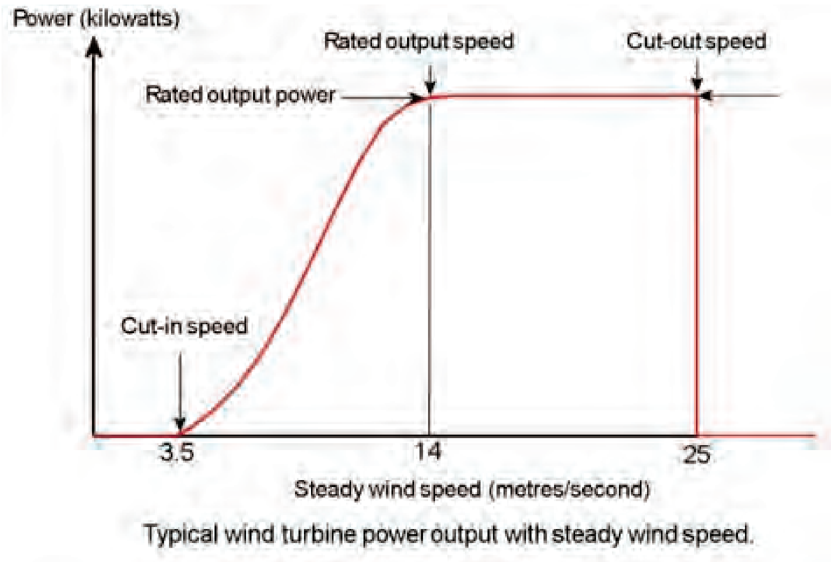



Figura 2.

1. Sabemos que, en el emplazamiendo donde se está diseñando la instalación, la velocidad del viento sigue la distribución anterior con una media de 7 m/s. Teniendo en cuenta que la media de la variable aleatoria de Rayleigh es
$$E[X] = \sigma\sqrt{\frac{\pi}{2}}$$
Calcula las probabilidades de que la velocidad se encuentre en la zona A, en la zona B y en la zona C.

2. Demuestra que si $U$ es una variable aleatoria con distribución uniforme en el intervalo (0,1), entonces
$$X = \sigma\sqrt{-2\log U}$$
tiene distribución de Rayleigh con parámetro $\sigma$.

   Un estudiante de ingeniería ha simulado 100 valores de la distribución anterior para la velocidad, resultando el siguiente diagrama de tallos y hojas:

```
 0 | 49
 1 | 0224
 2 | 5677
 3 | 01345677889
 4 | 1123345567788
 5 | 0056799
 6 | 0111356888
 7 | 11124477889
 8 | 0445555679
 9 | 0345679
10 | 24679
11 | 0012467
12 | 0045
13 | 3
14 | 9
15 | 138
```

   Compara la mediana, el cuartil 1 y el cuartil 3 de los datos con los correspondientes de la distribución de probabilidad.

   *Nota*: los tallos, valores a la izquierda de las barras son unidades y las hojas, cada cifra a la derecha de la barra, la primera cifra decimal. Así por ejemplo, la última fila corresponde a los valores 15.1, 15.3 y 15.8.

3. La "función de supervivencia" $S_X(x)$ se define como:
$$S_X(x) = P(X \geq x) \qquad , \qquad x \geq 0$$
Demuestra que, para la distribución de Rayleigh, se cumple:
$$\int_0^\infty S_X(x)dx = E[X]$$
*Nota*: Para evitarte el cálculo de la integral, ten en cuenta que:
$$f_T(t) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2\right\} \qquad , \qquad t \in \Re$$
es una función de densidad de probabilidad que es simétrica respecto al cero.

**Cuestión 1:**

Cálculo del valor de $\theta$:

$$\int_{-\infty}^{+\infty} f_X(x)dx \;=1 \qquad \rightarrow \qquad \int_{-\infty}^{+\infty} f_X(x)dx \;= \int_0^1 \theta\cdot x\,dx = \frac{\theta x^2}{2}\Bigg]_0^1 = \frac{\theta}{2} = 1 \qquad \rightarrow \qquad \theta = 2$$

Cálculo de la función de distribución:

$$F_X(x) = \int_0^x f_X(t)\,dt = \int_0^x 2\cdot t\,dt = x^2 \quad , \quad 0 \le x \le 1$$

Cálculo de la esperanza:

$$E[X] = \int_{-\infty}^{+\infty} x\cdot f_X(x)\,dx = \int_0^1 x\cdot 2\cdot x\,dx = \frac{2}{3}$$

Cálculo de la varianza:

$$Var[X] = \int_{-\infty}^{+\infty} x^2\cdot f_X(x)\;dx - E[X]^2 = \int_0^1 x^2\cdot 2\cdot x\,dx - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

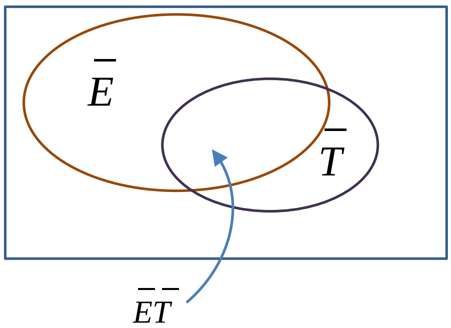
**Cuestión 2:**

La probabilidad de "ni estudia ni trabaja" se puede calcular de la siguiente manera:

$$P(\bar{E}\bar{T}) = P(\bar{E}) + P(\bar{T}) - P(\bar{E}\cup\bar{T})$$

$$P(\bar{E}\bar{T}) = 0.6 + 0.8 - P(\bar{E}\cup\bar{T})$$

Caso genérico:




Sabemos que la probabilidad de la unión ($P(\bar{E}\cup\bar{T})$) siempre cumple:

$$\max\big(P(\bar{E}), P(\bar{T})\big) \le P(\bar{E}\cup\bar{T}) \le 1$$

Substituyendo en la expresión anterior:

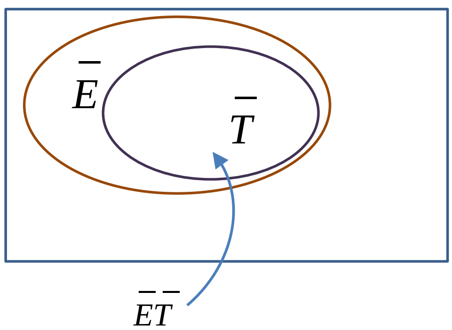
$$\max\big(P(\bar{E}), P(\bar{T})\big) \le 0.6 + 0.8 - P(\bar{E}\bar{T}) \le 1$$

$$\max(0.6{,}0.8) \le 0.6 + 0.8 - P(\bar{E}\bar{T}) \le 1$$

$$0.8 \le 0.6 + 0.8 - P(\bar{E}\bar{T}) \le 1$$

$$0.4 \le P(\bar{E}\bar{T}) \le 0.6$$

"Ni estudia ni trabaja" máximo:



**Cuestión 3:**

Nos piden calcular la probabilidad de que una caja esté fabricada por la máquina M1, sabiendo que los cuatro caramelos que contiene el paquete son de naranja. Es decir, hemos de calcular:

$$P(M1|NNNN)$$

Si la máquina M3 produce al doble de velocidad que el resto, sabemos que:

$$P(M1) = P(M2) = \frac{1}{4} \quad \text{y} \quad P(M3) = \frac{2}{4}$$

Empleando Bayes:

$$P(M1|NNNN) = \frac{P(M1 \cap NNNN)}{P(NNNN)}$$

$$= \frac{P(NNNN|M1) \cdot P(M1)}{P(NNNN|M1) \cdot P(M1) + P(NNNN|M2) \cdot P(M2) + P(NNNN|M3) \cdot P(M3)}$$

$P(NNNN|M1) = 0.5^4 = 0.063$

$P(NNNN|M2) = 0.2^4 = 0.002$

$P(NNNN|M3) = 0.5^4 = 0.063$

$$P(M1|NNNN) = \frac{0.063 \cdot \frac{1}{4}}{0.063 \cdot \frac{1}{4} + 0.002 \cdot \frac{1}{4} + 0.063 \cdot \frac{2}{4}} = 0.33$$

## Solución - Problema

### Apartado 1:

$$\sigma\sqrt{\frac{\pi}{2}} = 7 \quad \rightarrow \quad \sigma = 7 \cdot \sqrt{\frac{2}{\pi}} = 5.59$$

$$F_X(x) = \int_0^x \frac{t}{\sigma^2} e^{-\frac{t^2}{(2\cdot\sigma^2)}} dt = 1 - e^{-\frac{x^2}{(2\cdot\sigma^2)}} \quad x \geq 0$$

$$p_A = P(X \leq 3.5) = 1 - e^{-\frac{3.5^2}{(2\cdot\sigma^2)}} = 0.1780$$

$$p_B = P(3.5 \leq X \leq 25) = F_X(25) - F_X(3.5) = \left(1 - e^{-\frac{25^2}{(2\cdot\sigma^2)}}\right) - p_A = 0.8220$$

$$p_C = 1 - p_A - p_B \approx 0$$

### Apartado 2:

$$F_X(x) = P(X \leq x) = P\left(\sigma\sqrt{-2\log(U)} \leq x\right) = P\left(U \geq e^{-\frac{x^2}{(2\cdot\sigma^2)}}\right) = 1 - e^{-\frac{x^2}{(2\cdot\sigma^2)}} \quad , \ x \geq 0$$

Datos: $Q_1 = 4.3 \quad Q_2 = 6.8 \quad Q_3 = 9.5$

Distribución: $Fx(x_\alpha) = \alpha$

$$Q_1 \rightarrow x_{0.25} \quad \rightarrow \quad 1 - e^{-\frac{x_{0.25}^2}{(2\cdot\sigma^2)}} = 0.25$$

De igual forma:

$$x_{0.25} = \sigma\sqrt{-2\log(1-0.25)} = 4.24$$
$$x_{0.5} = \sigma\sqrt{-2\log(1-0.5)} = 6.58$$
$$x_{0.75} = \sigma\sqrt{-2\log(1-0.75)} = 9.30$$

### Apartado 3:

$$S_X(x) = e^{-\frac{x^2}{(2\cdot\sigma^2)}} \quad , \ x \geq 0$$

Teniendo en cuenta la nota:

$$\int_0^\infty e^{-\frac{x^2}{(2\cdot\sigma^2)}} dx = \frac{\sqrt{2\pi}\,\sigma}{2}$$ que cumple que es igual a E[X]

Otra forma:

$$\int_0^\infty S_X(x)\,dx = \int_0^\infty 1 - F_X(x)\,dx$$

Integrando por partes:

$$\begin{Bmatrix} v = 1 - F_X(x) \\ du = dx \end{Bmatrix} \rightarrow \begin{Bmatrix} dv = -f_X(x)dx \\ u = x \end{Bmatrix}$$

$$\int_0^\infty S_X(x)\,dx = x\left(1 - F_X(x)\right)\Big|_0^\infty + \int_0^\infty x \cdot f_X(x)\,dx = 0 + E[X] = \mathrm{E[X]}$$

**Cuestiones** (45 minutos, 5 puntos)

1) En un país, la distribución conjunta de las variables aleatorias *Oferta* y *Demanda* eléctrica en un día es normal bivariante con un coeficiente de correlación de $0,1$. La media y desviación típica marginales para la *Oferta* son 32000 $MW$ y 2200 $MW$ respectivamente, y para la *Demanda* 25000 $MW$ y 2000 $MW$ respectivamente ¿Cuál es la probabilidad de que la *Demanda* supere a la *Oferta* en un día determinado?

2) Se lanza una moneda equilibrada 6 veces. En este experimento se definen las variables aleatorias $X$: *número de caras obtenidas* e $Y$: *número de cruces obtenidas*. Calcule la covarianza entre X e Y ¿Son independientes?

3) Para controlar la calidad de unos lotes de 10000 piezas se toman muestras de 200 piezas y se rechaza el lote si resultan más de 8 defectuosas. El fabricante asegura que la proporción de piezas defectuosas de los lotes es del 2,5 % y el cliente considera totalmente inviable una proporción del 6 %. Calcule los riesgos del comprador y del vendedor. (Nota.- No utilice la corrección por continuidad).

**Problema** (45 minutos, 5 puntos)

A un puesto de atención personal llegan los clientes de forma que el tiempo que esperan en la cola es aleatorio. Una vez se les atiende, el tiempo que se tarda en hacerlo es también aleatorio. La función de densidad conjunta de ambos tiempos en horas ($X$: *tiempo de espera* e $Y$: *tiempo de atención*) es

$$f_{X,Y}(x,y) = ke^{-3x-4y}, x > 0, y > 0.$$

1) Calcule el valor de $k$ así como la media y la varianza de los dos tiempos $X$ e $Y$ ¿Son independientes ambas variables?

2) Sea $W$ la suma de los tiempos de espera $X$ y de atención $Y$. Calcule el coeficiente de correlación entre $W$ y $X$.

**Solución Cuestiones** (45 minutos, 5 puntos)

1) $S = O - D; S \rightarrow N(\mu = 32000 - 25000 = 7000MW; \sigma^2 = 2200^2 + 2000^2 - 2 \times 0,1 \times 2200 \times 2000 = 7960000MW^2)$;

$P(O < D) = P(O - D < 0) = P(S < 0) = P(Z < 0 - 7000)/\sqrt{7960000}) = P(Z < -2,48) = 1 - \phi(2,48) = 1 - 0{,}9934 = 0{,}0066$

2) $\sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)] = E[(X - 3)(6 - X - 3)] = E[-(X - 3)^2] = -\sigma_X^2 = -6 \times 0,5 \times 0,5 = -1,5;$

No son independientes porque la covarianza no es nula, pero además se ve que la dependencia es total, ya que la relación entre $X$ e $Y$ es determinista, para cada valor de $X$ existe un solo valor posible de $Y = 6 - X$

3) *Riesgo vendedor*$= P(\textit{Rechazar}|p = 0{,}025) = P(X > 8|X \rightsquigarrow N(\ 200 \times 0,025, \sqrt{200 \times 0,025 \times 0,975}) =$

$= P(Z > 1,36) = 1 - \phi(1,36) = 0{,}087$

*Riesgo comprador*$= P(\textit{Aceptar}|p = 0{,}06) = P(X \leq 8|X \rightsquigarrow N(200 \times 0,06, \sqrt{200 \times 0,06 \times 0,94}) =$

$= P(Z \leq -1{,}19) = 1 - \phi(1,19) = 0{,}117$

**Solución Problema** (45 minutos, 5 puntos)

1)

$k\int_0^\infty \int_0^\infty e^{-3x}e^{-4y}dxdy = k/12 = 1, k = 12;$

$g_X(x) = 12\int_0^\infty e^{-3x-4y}dy = 3e^{-3x}$

Se trata de una distribución exponencial, con parámetro $\lambda = 3; E[X] = 1/\lambda = 1/3;$

$var[X] = 1/\lambda^2 = 1/9;$

$h_Y(y) = 12\int_0^\infty e^{-3x-4y}dx = 4e^{-4y};$

Se trata también de una distribución exponencial, con parámetro $\lambda = 4; E[Y] = 1/\lambda = 1/4;$

$var[Y] = 1/\lambda^2 = 1/16;$

$f_{X,Y}(x,y) = g_X(x)h_Y(y)$ son independientes

2) $\rho(W,X) = \frac{\sigma_{W,X}}{\sigma_W\sigma_X};$

$\sigma_{W,X} =?$

$$V = \begin{bmatrix} W \\ X \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix};$$

$$var[V] = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1/9 & 0 \\ 0 & 1/16 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/9 & 1/16 \\ 1/9 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 25/144 & 1/9 \\ 1/9 & 1/9 \end{bmatrix};$$

$\rho(W,X) = \frac{1/9}{\sqrt{25/144}\sqrt{1/9}} = \sqrt{\frac{1/9}{25/144}};$

**Cuestiones** (45 minutos, 5 puntos)

1. Se plantea el siguiente contraste para el parámetro $p$ de una distribución binomial:

$$H_0 : p = 0.3 \quad ; \quad H_1 : p > 0.3$$

con $n = 200$ y $\alpha = 0.05$. Calcular la probabilidad de error tipo II cuando $p = 0.36$. Calcular $n$ si se desea reducir a la mitad la probabilidad del error tipo II.

2. Para estimar la varianza de una población normal, se tienen dos muestras de tamaños $n_1$ y $n_2$. Se toma como estimador de la varianza la siguiente expresión: $\hat{s}_R^2 = \alpha\hat{s}_1^2 + \beta\hat{s}_2^2$; siendo $\hat{s}_1^2$ y $\hat{s}_2^2$ las varianzas corregidas de cada muestra.

   Calcular $\alpha$ y $\beta$ (en función de $n_1$ y $n_2$) para que el estimador sea centrado y de mínima varianza.

   *Nota*: $Var(\hat{S}^2) = \frac{2\sigma^4}{n-1}$, siendo $\hat{S}^2$ la varianza corregida con datos de una normal con varianza $\sigma^2$.

3. Las puntuaciones obtenidas en un test de memoria por alumnos de Primero de Primaria, para chicos y chicas, son las siguientes:

| | Número de alumnos | Media puntuaciones | Desviación típica corregida |
|---|---|---|---|
| Chicos | $n_A = 64$ | $\overline{x}_A = 97.0$ | $\widehat{s}_A = 0.49$ |
| Chicas | $n_B = 95$ | $\overline{x}_B = 98.9$ | $\widehat{s}_B = 0.48$ |

   Asumiendo normalidad, calcular el intervalo de confianza para la diferencia de medias con un nivel de confianza del 95 %. Razonar, a partir del intervalo obtenido, si se acepta la igualdad de medias.

   *Observación*: suponer homocedasticidad (no hace falta comprobarla).

**Problema** (45 minutos, 5 puntos)

Se ha medido el tiempo de vida de 100 componentes fabricados por la Empresa Automovilística *Tord*$^e$. Siete de ellos han durado más de catorce días. La media del tiempo de vida del resto de componentes es 3.73 días. Asumiendo que el tiempo de vida tiene distribución exponencial de media $\mu$:

1) Estimar por máxima verosimilitud el parámetro $\mu$, justificando cada paso, empleando la información de las cien observaciones.

2) Los tiempos de vida de estos cien componentes se han agrupado en nueve intervalos, obteniendo:

| Intervalo | 0–1 | 1-2 | 2-3 | 3-5 | 5-7 | 7-9 | 9-11 | 11-14 | >14 |
|---|---|---|---|---|---|---|---|---|---|
| Frecuencia | 21 | 14 | 16 | 15 | 14 | 4 | 5 | 4 | 7 |

Contrastar la hipótesis de que los tiempos siguen una distribución exponencial estimada en el apartado anterior, con un nivel de significación del $\alpha = 0.05$.

*Nota*: Para el contraste, trabajar con los intervalos indicados en la tabla. (No agrupar intervalos.)

3) Se ha prolongado el experimento, hasta que se conocen los tiempos de vida exactos de los cien componentes.

Se sabe que $2n\overline{T}/\mu$ sigue una distribución $\chi^2_{2n}$, siendo $\mu$ la media de la distribución exponencial, y siendo $\overline{T}$ la media muestral de los tiempos de vida. Si $\overline{T} = 4.8$ días, contrastar que $\mu = 5$ días, frente a $\mu < 5$ días, con un nivel de confianza del 95 %.

*Nota*: $P(\chi^2_{200} \leq 163) = 0.025$; $P(\chi^2_{200} \leq 168) = 0.05$; $P(\chi^2_{200} \leq 234) = 0.95$; $P(\chi^2_{200} \leq 241) = 0.975$.

**Cuestiones** (45 minutos, 5 puntos)

1. **Se plantea el siguiente contraste para el parámetro $p$ de una distribución binomial:**

$$H_0 : p = 0.3 \quad ; \quad H_1 : p > 0.3$$

**con $n = 200$ y $\alpha = 0.05$. Calcular la probabilidad de error tipo II cuando $p = 0.36$. Calcular $n$ si se desea reducir a la mitad la probabilidad del error tipo II.**

$pr$(error tipo II) $= pr$(Aceptar $H_0 |\ H_0$ es falsa)

Utilizamos que $\frac{\widehat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \to Z$

Al ser un contraste unilateral, la región de rechazo para valores de $Z$ mayores que $Z_a = 1.645$, como en la Figura 1.

Luego en esa frontera entre la región de aceptación y rechazo $\widehat{p}$ vale:

$\widehat{p} = 1.645 \cdot \sqrt{\frac{0.3(1-0.3)}{200}} + 0.3 = 0.353.$

Entonces para calcular el error tipo II se ha de calcular la probabilidad:

$pr(\widehat{p} < 0.353|\ p = 0.36) =$

$= pr\left(\frac{\widehat{p}-0.36}{\sqrt{\frac{0.36(1-0.36)}{200}}} < \frac{0.353-0.36}{\sqrt{\frac{0.36(1-0.36)}{200}}}\right) = pr(Z < -0.2062) = pr(Z > 0.2062) =$

$= 1 - F_Z(0.2062) = 1 - 0.5832 = 0.4168$

Si se desea reducir ese error de tipo II a la mitad:

Llamando $c$ al valor que deseamos hallar para $\widehat{p}$ en la frontera, e imponiendo que error de tipo II valga la mitad que en el primer apartado planteamos el sistema de ecuaciones siguiente:

$\frac{c-0.3}{\sqrt{\frac{0.3(1-0.3)}{n}}} = 1.645$

$pr\left(Z < \frac{c-0.36}{\sqrt{\frac{0.36(1-0.36)}{n}}}\right) = pr\left(Z > \frac{0.36-c}{\sqrt{\frac{0.36(1-0.36)}{n}}}\right) =$

$= 1 - F_Z\left(\frac{0.36-c}{\sqrt{\frac{0.36(1-0.36)}{n}}}\right) = 0.4168/2 = 0.2084 \simeq 0.21 \to$

$\to F_Z\left(\frac{0.36-c}{\sqrt{\frac{0.36(1-0.36)}{n}}}\right) = 1 - 0.21 = 0.79 \to$ Buscando en tablas: $\frac{0.36-c}{\sqrt{\frac{0.36(1-0.36)}{n}}} = 0.805.$

Y despejando queda: $c \simeq 0.34$ y $n \simeq 355$.

2. **Para estimar la varianza de una población normal, se tienen dos muestras de tamaños $n_1$ y $n_2$. Se toma como estimador de la varianza la siguiente expresión: $\hat{s}_R^2 = \alpha\hat{s}_1^2 + \beta\hat{s}_2^2$; siendo $\hat{s}_1^2$ y $\hat{s}_2^2$ las varianzas corregidas de cada muestra.**

**Calcular $\alpha$ y $\beta$ (en función de $n_1$ y $n_2$) para que el estimador sea centrado y de mínima varianza.**

**Nota: $Var(\hat{S}^2) = \frac{2\sigma^4}{n-1}$, siendo $\hat{S}^2$ la varianza corregida con datos de una normal con varianza $\sigma^2$.**

$E[\hat{s}_R^2] = E[\alpha\hat{s}_1^2 + \beta\hat{s}_2^2] = \alpha E[\hat{s}_1^2] + \beta E[\hat{s}_2^2] = (\alpha + \beta)\sigma^2$

Para que sea insesgado $E[\hat{s}_R^2] = \sigma^2 \to (\alpha + \beta)\sigma^2 = \sigma^2 \to (\alpha + \beta) = 1 \to \beta = 1 - \alpha$.

La varianza del estimador: $var[\hat{s}_R^2] = var[\alpha\hat{s}_1^2 + \beta\hat{s}_2^2] = \alpha^2 var[\hat{s}_1^2] + \beta^2 var[\hat{s}_2^2] = \alpha^2 \frac{2\sigma^4}{n_1-1} + \beta^2 \frac{2\sigma^4}{n_2-1} = \alpha^2 \frac{2\sigma^4}{n_1-1} + (1-\alpha)^2 \frac{2\sigma^4}{n_2-1}$

Para minimizar la varianza del estimador, derivamos respecto de $\alpha$ :

$\frac{\partial[\alpha^2\frac{2\sigma^4}{n_1-1}+(1-\alpha)^2\frac{2\sigma^4}{n_2-1}]}{\partial\alpha} = \frac{2\sigma^4}{n_1-1}2\alpha - \frac{2\sigma^4}{n_2-1}2(1-\alpha) = \frac{4\alpha\sigma^4}{n_1-1} - \frac{4\sigma^4}{n_2-1} + \frac{4\alpha\sigma^4}{n_2-1} = \frac{(n_2-1)4\alpha\sigma^4+(n_1-1)[4\alpha\sigma^4-4\sigma^4]}{(n_1-1)(n_2-1)}$

$(n_2-1)4\alpha\sigma^4 + (n_1-1)[4\alpha\sigma^4 - 4\sigma^4] = 0$

$(n_2-1)\alpha + (n_1-1)[\alpha-1] = 0$

$\alpha(n_2-1+n_1-1) = (n_1-1)$

Entonces se tiene que:

$$\alpha = \frac{(n_1-1)}{(n_2-1+n_1-1)},$$

$$\beta = 1-\alpha = \frac{(n_2-1+n_1-1)}{(n_2-1+n_1-1)} - \frac{(n_1-1)}{(n_2-1+n_1-1)} = \frac{(n_2-1+n_1-1)-n_1+1}{(n_2-1+n_1-1)} = \frac{(n_2-1)}{(n_2-1+n_1-1)}$$

3. **Las puntuaciones obtenidas en un test de memoria por alumnos de Primero de Primaria, para chicos y chicas, son las siguientes:**

| | **N$^o$de alumnos** | **Media puntuaciones** | **Desv. Tip. corregida** |
|---|---|---|---|
| **Chicos** | **n**$_A$**= 64** | $\overline{x}_A$**= 97.0** | $\widehat{s}_A$**= 0.49** |
| **Chicas** | **n**$_B$**= 95** | $\overline{x}_B$**= 98.9** | $\widehat{s}_B$**= 0.48** |

**Asumiendo normalidad, calcular el intervalo de confianza para la diferencia de medias con un nivel de confianza del 95 %. Razonar, a partir del intervalo obtenido, si se acepta la igualdad de medias.**

**Observación: suponer homocedasticidad (no hace falta comprobarla).**

Utilizamos que $\frac{\overline{x}_1-\overline{x}_2-(\mu_1-\mu_2)}{\hat{s}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \to t_{n_1+n_2-2}$.

Para una $\alpha = 0.05$, es decir nivel de confianza 95 % se tiene que:

$(\mu_A - \mu_B) \in (\overline{x}_A - \overline{x}_B) \pm t_{n_A+n_B-2;\frac{0.05}{2}}\hat{s}_R\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}$

$(\mu_A - \mu_B) \in (97-98.9) \pm t_{64+95-2;\frac{0.05}{2}}\hat{s}_R\sqrt{\frac{1}{64}+\frac{1}{95}}$

$\hat{s}_R = \sqrt{\frac{(n_A-1)\hat{s}_A^2+(n_B-1)\hat{s}_B^2}{n_A+n_B-2}} = \sqrt{\frac{(95-1)\cdot 0.49^2+(64-1)\cdot 0.48^2}{64+95-2}} = \sqrt{0.2362} = 0.486.$

Entonces:

$(\mu_A - \mu_B) \in (97-98.9) \pm 1.96\cdot 0.486\sqrt{\frac{1}{64}+\frac{1}{95}}$

$(\mu_A - \mu_B) \in (97-98.9) \pm 1.96\cdot 0.486\sqrt{\frac{1}{64}+\frac{1}{95}}$

$(\mu_A - \mu_B) \in -1.9 \pm 0.154$

El intervalo de confianza 95 % para la diferencia de medias es:

$$(\mu_A - \mu_B) \in [-2.054, -1.746]$$

Pues el valor en tablas para la $t_{64+95-2;\frac{0.05}{2}}$ es 1.96, como en la $Z$.

**Solución del problema**

1) Se puede resolver el problema directamente en términos de la esperanza $\mu$ o en términos de $\lambda = 1/\mu$.

Para la función de verosimilitud, la contribución de los datos no censurados es la función de densidad y, la de los censurados la probabilidad de duración mayor que 14 horas.

$l = (1/\mu)^{93} e^{-93\times 3,73/\mu} e^{-7\times 14/\mu}$

$L = \log(l) = -93\log(\mu) - 93 \times 3,73/\mu - 7 \times 14/\mu;$

$\frac{\partial L}{\partial \mu} = -\frac{93}{\mu} + 93 \times 3,73/\mu^2 + 7 \times 14//\mu^2 = 0; \widehat{\mu} = \frac{93\times 3,73+7\times 14}{93} = 4,78$

2) Las frecuencias esperadas se obtienen como $100(1 - e^{-1/\widehat{\mu}})$ para el primer intervalo, $100(e^{-EI/\widehat{\mu}} - e^{-ES/\widehat{\mu}})$ para los 7 siguientes y $e^{-14/\widehat{\mu}}$ para el último.

La tabla de frecuencias esperadas es

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 18,86 | 15,3 | 12,42 | 18,25 | 12,01 | 7,91 | 5,2 | 4,67 | 5,35 |

$d = \sum \frac{(O_i - E_i)^2}{E_i} = 4,68 < \chi^2_{0,95,(9-1-1)} = 14,02$, por tanto no se rechaza la hipótesis de que los datos correspondan a una distribución exponencial

Como se ha estimado un parámetro, los grados de libertad de la $\chi^2$ son $9 - 1 - 1 = 7$.

3) Como la hipótesis alternativa es $H_1 : \mu < 5$, la región de rechazo es unilateral por la izquierda.

$\frac{200\times 4,8}{5} = 192 > \chi^2_{0,05,7} = 168$ no se rechaza $H_0$

**Problema 1** (60 minutos – 10 puntos)

Tres amigos: Ana (A), Benito (B), Carlos (C) lanzan una moneda equilibrada en el siguiente orden A, B, C, A, B, C,... y gana el primero que obtenga una *CARA*.

a) ¿Cuál es la probabilidad de que gane A? ¿Cuál es la probabilidad de que gane B? ¿Cuál es la probabilidad de que gane C?

b) Sabiendo que ha ganado A, ¿cuál es la probabilidad de que lo haya hecho en el primer lanzamiento?

   Obtener la distribución condicionada del número de lanzamientos efectuados por A, sabiendo que A ha ganado.

c) Si una vez concluido el juego se sabe que no ha ganado A, ¿cuál es la probabilidad de que gane B? ¿Cuál es la probabilidad de que gane C?

d) Se repite el juego una segunda vez, en la misma secuencia pero empezando por aquél al que le tocaba jugar. Por ejemplo si gana B en la primera ronda, en el segundo juego lanza primero C, después A y finalmente B y se repite C, A, B, C, A, B,... hasta que finaliza.

   ¿Cuál es la probabilidad de que gane A el segundo juego? ¿Cuál es la probabilidad de que gane B el segundo juego? ¿Cuál es la probabilidad de que gane C el segundo juego?

**Problema 2** (60 minutos – 10 puntos)

El banco C&R de inversión tiene 200 sucursales en España. El número de descubiertos en cualquier sucursal del banco en un día determinado sigue una distribución de Poisson de media 9. El número de descubiertos diarios de sucursales diferentes son variables aleatorias independientes.

a) ¿Cuál es el número esperado de oficinas con más de doce descubiertos en un día?
b) ¿Cuál es la probabilidad de que haya más de veinte sucursales con más de doce descubiertos en un día concreto?
c) ¿Cuál es la probabilidad de que el número total de descubiertos en cuatro sucursales del banco sea superior a 48 en un día?
d) El Banco Central Europeo establece en sus normas que un banco debe avisar a la autoridad monetaria cuando "el número medio de descubiertos observados en el total de sus oficinas en un día sea superior a $L = 9.5$".
¿Cuál es la probabilidad de que el banco no tenga que avisar ningún día de un mes?
¿Cuánto debería ser $L$ para que la probabilidad anterior sea igual a 0.95?

Nota: Suponer que el mes tiene 30 días y que el número de descubiertos en días distintos son variables aleatorias independientes.

Observación: Un *descubierto* es la situación en la que una cuenta corriente tiene saldo negativo.

**Problema 3** (60 minutos – 10 puntos)

La proporción $p$ de huevos que proporcionan una cría de tortuga de la especie Boba tras el proceso de incubación depende de múltiples factores. En 1999, una bióloga recogió datos de 20 nidos en la Costa Esmeralda de Cabo Verde, observando 1695 nacimientos de un total de 1820 huevos.

a) Aceptando que la probabilidad de éxito de nacimiento $p$ es la misma para todos los huevos incubados, proporciona un intervalo de confianza al 95% para $p$.

b) Según la literatura científica, en condiciones óptimas de incubación la probabilidad de éxito de nacimiento es 0.88. ¿Se puede afirmar que la probabilidad de éxito de nacimiento en Cabo Verde es superior? ($\alpha = 0.05$)

c) Calcular la probabilidad del error tipo II del contraste del '*apartado b*' en función de la probabilidad real de éxito de nacimiento ($p$), y evaluarla para los valores de $p$ igual a 0.88, 0.89, 0.90, 0.91.

d) El resultado fue tan sorprendente que un equipo de la Universidad de Florida se está planteando repetir el estudio en el verano de 2014. ¿Cuántos huevos debe observar en el estudio para que se verifique la condición siguiente: "si la probabilidad real de éxito en Cabo Verde es 0.90, la probabilidad de aceptar la hipótesis nula del contraste del '*apartado b*' sea 0.01"?

**Solución Problema 1**

*Apartado A:*

Definimos los siguientes sucesos:

A : "gana A el juego"

$a_i$ : "el jugador A obtiene cara en el $i$-ésimo lanzamiento"

$$P(A) = P(a_1) + P\left(\overline{a_1}\overline{b_1}\overline{c_1}a_2\right) + P\left(\overline{a_1}\overline{b_1}\overline{c_1}\overline{a_2}\overline{b_2}\overline{c_2}a_3\right) + \cdots$$
$$= \frac{1}{2} + \frac{1}{8}\cdot\frac{1}{2} + \left(\frac{1}{8}\right)^2\cdot\frac{1}{2} + \left(\frac{1}{8}\right)^3\cdot\frac{1}{2} + \cdots = \frac{1}{2}\cdot\frac{1}{1-\frac{1}{8}} = \frac{4}{7}$$
$$P(B) = P(\overline{a_1}b1) + P\left(\overline{a_1}\overline{b_1}\overline{c_1}\overline{a_2}b_2\right) + \cdots$$
$$= \frac{1}{4} + \frac{1}{8}\cdot\frac{1}{4} + \left(\frac{1}{8}\right)^2\cdot\frac{1}{4} + \left(\frac{1}{8}\right)^3\cdot\frac{1}{4} + \cdots = \frac{1}{4}\cdot\frac{1}{1-\frac{1}{8}} = \frac{2}{7}$$
$$P(C) = \frac{1}{7}$$

Por tanto, la probabilidad de que gane el jugador A, B y C es 4/7, 2/7, y 1/7, respectivamente. Se observa que el jugador con más probabilidad de ganar el juego es el A.

*Apartado B:*

La probabilidad de que gane A en el primer lanzamiento, sabiendo que A ha ganado la partida, se calcula como la siguiente probabilidad condicionada:

$$P(a_1|A) = \frac{P(a_1 \wedge A)}{P(A)} = \frac{P(a_1)}{P(A)} = \frac{1/2}{4/7} = \frac{7}{8}$$

Por tanto, sabiendo que gana A, la probabilidad de que gane A en el 1er lanzamiento es 7/8.

La probabilidad de que gane A en el segundo lanzamiento, sabiendo que A ha ganado la partida, se calcula como la siguiente probabilidad condicionada:

$$P(a_2|A) = \frac{P(a_2 \wedge A)}{P(A)} = \frac{P\left(\overline{a_1}\overline{b_1}\overline{c_1}a_2\right)}{P(A)} = \frac{1/8\ \cdot 1/2}{4/7} = \frac{7}{64}$$

Por tanto, la distribución condicionada del número de lanzamientos efectuados por A, sabiendo que A ha ganado:

$$P(a_i|A) = \frac{P(a_i \wedge A)}{P(A)} = \frac{P\left(\overline{a_1}\overline{b_1}\overline{c_1}\dots a_i\right)}{P(A)} = \frac{(1/8)^{i-1}\ \cdot 1/2}{4/7} = \frac{7}{2^{3\cdot i}}\ ; \quad i = 1,2,3,\dots$$

*Apartado C:*

La probabilidad de que gane B, sabiendo que A no ha ganado, se calcula como la siguiente probabilidad condicionada:

$$P(B|\bar{A}) = \frac{P(B \wedge \bar{A})}{P(\bar{A})} = \frac{P(B)}{1-P(A)} = \frac{2/7}{1-4/7} = \frac{2}{3}$$

Por tanto, la probabilidad de que gane B, sabiendo que A no ha ganado, es de 2/3. Finalmente, la probabilidad de que gane C, sabiendo que A no ha ganado, es de 1/3.

*Apartado D:*

Definimos:

$A_1$ : "gana A el primer juego"

$A_2$ : "gana A el segundo juego"

$$\begin{aligned} P(A_2) &= P(A_2 \wedge A_1) + P(A_2 \wedge B_1) + P(A_2 \wedge C_1) \\ &= P(A_2|A_1)P(A_1) + P(A_2|B_1)P(B_1) + P(A_2|C_1)P(C_1) \\ &= \frac{1}{7}\cdot\frac{4}{7}+\frac{2}{7}\cdot\frac{2}{7}+\frac{4}{7}\cdot\frac{1}{7}=\frac{12}{49} \end{aligned}$$

Obsérvese que:

$$P(A_2|A_1) = P(A_2|\text{saca B en el segundo juego}) = P\left(\overline{b_1}\,\overline{c_1}a_1\right) + P\left(\overline{b_1}\,\overline{c_1}\,\overline{a_1}\,\overline{b_2}\,\overline{c_2}a_2\right) + \cdots = \frac{1}{7}$$

$$\begin{aligned} P(B_2) &= P(B_2|A_1)P(A_1) + P(B_2|B_1)P(B_1) + P(B_2|C_1)P(C_1) \\ &= \frac{4}{7}\cdot\frac{4}{7}+\frac{1}{7}\cdot\frac{2}{7}+\frac{2}{7}\cdot\frac{1}{7}=\frac{20}{49} \end{aligned}$$

$$\begin{aligned} P(C_2) &= P(C_2|A_1)P(A_1) + P(C_2|B_1)P(B_1) + P(C_2|C_1)P(C_1) \\ &= \frac{2}{7}\cdot\frac{4}{7}+\frac{4}{7}\cdot\frac{2}{7}+\frac{1}{7}\cdot\frac{1}{7}=\frac{17}{49} \end{aligned}$$

La probabilidad de que A gane el segundo juego es 12/49.

La probabilidad de que B gane el segundo juego es 20/49.

La probabilidad de que C gane el segundo juego es 17/49.

Se observa que el jugador con más probabilidad de ganar el segundo juego es el B.

**Solución Problema 2**

*Apartado A:*

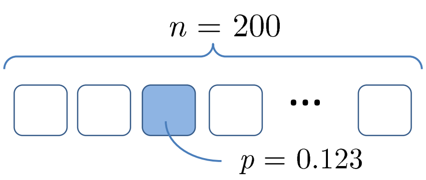
El número esperado de oficinas con más de doce descubiertos en un día se calcula como el producto del número de oficinas, por la probabilidad de que una oficina tenga más de doce descubiertos:

$$P(\text{una oficina con } X > 12 \text{ descubiertos } \mid X \sim \text{Poisson}(\lambda = 9))$$
$$\approx P(Y > 12.5 \mid Y \sim N(9,3)) = P\left(Z > \frac{12.5 - 9}{3}\right) = 0.123$$

Número esperado de oficinas: $0.123 \cdot 200 \approx 25$ oficinas

*Apartado B:*

Para calcular la probabilidad de que haya más de veinte sucursales con más de doce descubiertos en un día concreto, observamos que el "número de sucursales" sigue una distribución binomial, y empleamos el valor obtenido en el apartado anterior ($p = 0.123$).




Aplicamos la aproximación a la normal y la corrección por continuidad.

$$P(X > 20 \mid X \sim \mathit{Binomial}(n = 200, p = 0.123))$$
$$\approx P(Y > 20.5 \mid Y \sim N(200 \cdot 0.123, \sqrt{200 \cdot 0.123 \cdot 0.877}))$$
$$= P(Y > 20.5 \mid Y \sim N(24.6, 4.64))$$
$$= P\left(Z > \frac{20.5 - 24.6}{4.64}\right) = P(Z > -0.884) = 0.811$$

*Apartado C:*

El "número total de descubiertos en cuatro sucursales del banco" sigue una distribución de Poisson con el siguiente parámetro:

$$4 \text{ sucursales} \quad \rightarrow \quad \lambda' = 9 \cdot 4 = 36 \frac{\text{fallos}}{\text{día}}$$

Por tanto, hemos de calcular $P(\mathrm{X} > 48)$. Lo hacemos empleando la aproximación a la normal y la corrección por continuidad.

$$P(\mathrm{X} > 48 \mid X \sim \text{Poisson}(\lambda' = 36)) \approx P(Y > 48.5 \mid Y \sim N(36, 6)) = P\left(Z > \frac{48.5 - 36}{6}\right) = 0.0186$$

*Apartado D:*

Sabemos lo siguiente:

"nº de descubiertos en una oficina en un día" $\sim$ Poisson($\lambda$=9) $\sim N(9, 3)$

"nº medio de descubiertos en 200 oficinas en un día" $\sim N\left(9, \frac{3}{\sqrt{200}}\right) = N(9, 0.212)$

En primer lugar vamos a calcular la probabilidad de que el banco no avise en un día determinado; y después calcularemos la probabilidad de que el banco no avise en un mes.

$$P\big(X < 9.5 \mid X \sim N(9, 0.212)\big) = P\left(Z < \frac{9.5 - 9}{0.212}\right) = 0.991$$

La probabilidad de que el banco no avise en un día determinado es 0.991. Por tanto, la probabilidad de que no avise en un mes es: $0.991^{30} = 0.762$.

Para calcular $L$ que hace que la probabilidad anterior sea 0.95, hemos de calcular $L$ en la siguiente expresión:

$$\left(P\left(Z < \frac{L - 9}{0.212}\right)\right)^{30} = 0.95 \;\longrightarrow\; P\left(Z < \frac{L - 9}{0.212}\right) = \sqrt[30]{0.95} = 0.998$$

De las tablas obtenemos que $P(Z < 2.878) = 0.998$. Por tanto:

$$\frac{L - 9}{0.212} = 2.878 \longrightarrow L = 9 + 0.212 \cdot 2.878 = 9.61$$

## Solución Problema 3

*Apartado A:*

De los datos del enunciado, podemos calcular $\hat{p}$:

$$\hat{p} = \frac{1695}{1820} = 0.931$$

Ahora calculamos el intervalo de confianza para *p:*

$$p \in \hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$\in 0.931 \pm 1.96 \cdot \sqrt{\frac{0.931 \cdot (1 - 0.931)}{1820}} = [0.9193, 0.9426]$$

Por tanto, el intervalo de confianza al 95% para $p$ es $[0.9193, 0.9426]$.

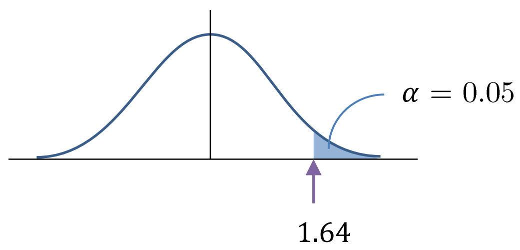
*Apartado B:*

En este apartado hemos de realizar el siguiente contraste unilateral:

$$\begin{cases} H_0: & p = 0.88 \\ H_1: & p > 0.88 \end{cases}$$

Sabiendo que:

$$\frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1 - p)}{n}}} \sim N(0,1)$$



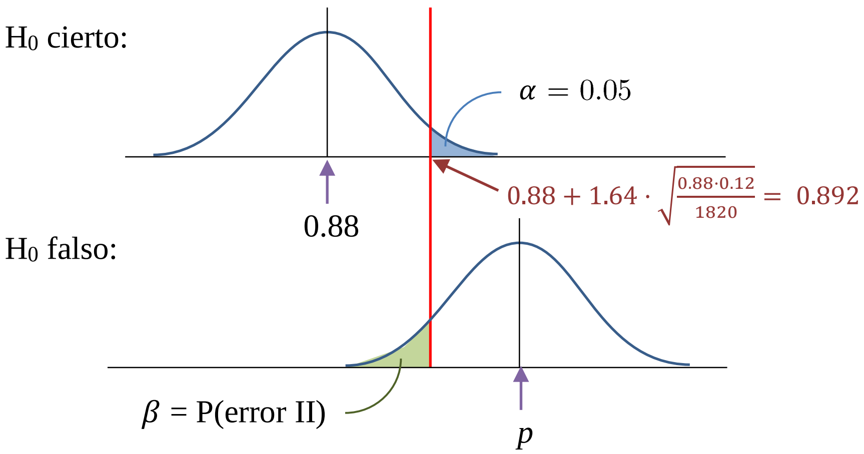

Si $H_0$ es cierto:

$$z_0 = \frac{0.931 - 0.88}{\sqrt{\frac{0.88 \cdot 0.12}{1820}}} = 6.69$$

Como se puede observar, el valor 6.69 > 1.64. Por tanto, rechazamos la hipótesis nula. Es decir, hay evidencia estadística para afirmar que la probabilidad de éxito de nacimiento en Cabo Verde es superior, con un nivel de confianza del 95%.

*Apartado C:*

El error del tipo II es la probabilidad de aceptar la hipótesis nula cuando, en realidad, es falsa.




$$P(error\ tipo\ II) = P(aceptar\ H_0 \mid\ H_0\ falsa)$$

$$= P\left(\hat{p} < 0.88 + 1.64\sqrt{\frac{0.88 \cdot 0.12}{1820}}\ \middle|\ \hat{p} \sim N\left(p, \sqrt{\frac{p \cdot (1-p)}{1820}}\right)\right)$$

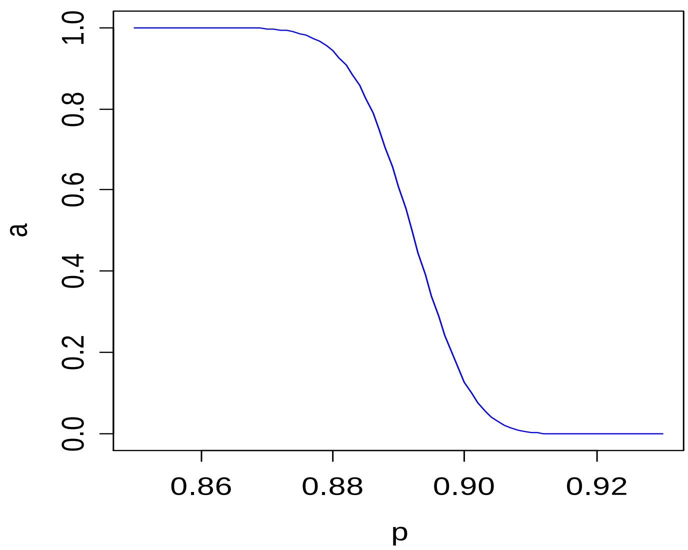
$$= P\left(Z < \frac{0.88 + 1.64\sqrt{\frac{0.88 \cdot 0.12}{1820}} - p}{\sqrt{\frac{p \cdot (1-p)}{1820}}}\right) = P\left(Z < \frac{0.892\ - p}{\sqrt{\frac{p \cdot (1-p)}{1820}}}\right)$$

$$= \Phi\left(\frac{0.892\ - p}{\sqrt{\frac{p \cdot (1-p)}{1820}}}\right)$$

Ahora evaluamos esta probabilidad para diversos valores de $p$:

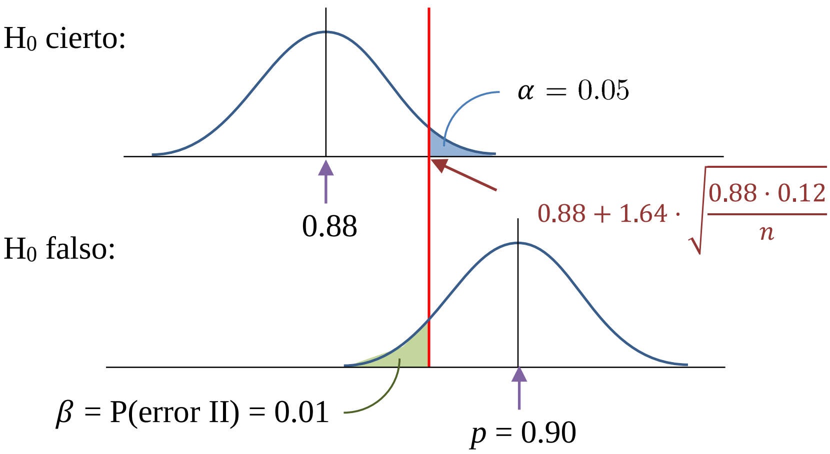
$$P(error\ tipo\ II \mid p = 0.88) = 0.95$$
$$P(error\ tipo\ II \mid p = 0.89) = 0.60$$
$$P(error\ tipo\ II \mid p = 0.90) = 0.13$$
$$P(error\ tipo\ II \mid p = 0.91) = 0.003$$

A continuación se muestra el gráfico de la probabilidad del error del tipo II, en función de la probabilidad real de éxito de nacimiento.



*Apartado D:*

Hemos de calcular $n$ para que la probabilidad de aceptar la hipótesis nula del contraste del '*apartado b*' sea 0.01, sabiendo que $p$ es igual a 0.90.




Por tanto, hemos de calcular $n$ para que la probabilidad del error del tipo II sea de 0.01:

$$\begin{aligned} P(error\ tipo\ II) &= P(aceptar\ H_0 \mid H_0\ falsa) \\ &= P\left(\hat{p} < 0.88 + 1.64\sqrt{\frac{0.88 \cdot 0.12}{n}} \;\middle|\; \hat{p} \sim N\left(0.90, \sqrt{\frac{0.90 \cdot 0.10}{n}}\right)\right) \\ &= P\left(Z < \frac{0.88 + 1.64\sqrt{\frac{0.88 \cdot 0.12}{n}} - 0.90}{\sqrt{\frac{0.90 \cdot 0.10}{n}}}\right) = 0.01 \end{aligned}$$

Sabemos que $P(Z < -2.32) = 0.01$. Por tanto:

$$\frac{0.88 + 1.64\sqrt{\frac{0.88 \cdot 0.12}{n}} - 0.90}{\sqrt{\frac{0.90 \cdot 0.10}{n}}} = -2.32$$

$$-0.02 + 1.64\sqrt{\frac{0.88 \cdot 0.12}{n}} = -2.32 \cdot \sqrt{\frac{0.90 \cdot 0.10}{n}}$$

$$-0.02 + 1.64 \cdot \sqrt{0.88 \cdot 0.12} \cdot \frac{1}{\sqrt{n}} = -2.32 \cdot \sqrt{0.90 \cdot 0.10} \cdot \frac{1}{\sqrt{n}}$$

$$-0.02 + 0.533 \cdot \frac{1}{\sqrt{n}} = -0.696 \cdot \frac{1}{\sqrt{n}}$$

De donde se obtiene que $n = 3775$ huevos.

**Cuestiones**. (45 minutos, 5 puntos)

**Cuestión 1**

A partir de una muestra aleatoria $X_1, ..., X_n$ de una distribución uniforme en el intervalo $(0, a)$, obtener el estimador de $a$ por el método de los momentos. Obtener el error cuadrático medio.

**Cuestión 2**

Se estima la media de una población normal con varianza conocida $\sigma^2 = 2$.

¿Cuál debe ser el tamaño muestral para que el error máximo de la estimación, con probabilidad 0,95, sea menor que 0,2?

**Cuestión 3**

Para dos poblaciones normales, se desea contrastar $H_0 : \sigma_1^2 = \sigma_2^2$ frente a $H_1 : \sigma_1^2 > \sigma_2^2$. Realizar el contraste sabiendo que $\widehat{s}_1^2 = 2{,}5$ y $\widehat{s}_2^2 = 2$, $n_1 = 30$ y $n_2 = 40$.

Calcular la probabilidad de error tipo II cuando $\sigma_1^2/\sigma_2^2 = 1{,}5$.

**Problema** (60 minutos, 5 puntos)

Diversos empresarios pretenden crear la empresa *LlevoTroncos,* que proporcionará sus servicios a la industria maderera mediante el transporte de troncos de árbol por carretera.

Apartado 1:

Se decide transportar madera a diez destinos diferentes: tres de ellos pertenecen a la zona Norte, dos a la zona Centro, y los cinco restantes a la zona Sur. Suponemos que el número de tráileres que se envía a cada destino es similar (es decir, los diez destinos son equiprobables).

La empresa ha comprado tráileres de dos capacidades: mediano (M) y grande (G): Para los destinos de zona Norte, emplean únicamente tráileres G; para los destinos de zona Centro, emplean indistintamente transportes de los tipos M y G; y para los destinos de la zona Sur, emplean el triple de tráileres M que de G.

La Oficina Central acaba de recibir una llamada de un transportista. Sabiendo que se trata de un conductor de un tráiler G, ¿cuál es la probabilidad de que este transportista se dirija a la zona Norte?

Apartado 2:

Actualmente está realizando un estudio para el dimensionamiento de sus tráileres, sabiendo que, para cada árbol talado, se transportará su tronco y su corteza.

Para cada árbol, se sabe que el peso de su tronco (variable $T$) y peso de su corteza (variable ) tienen una correlación de , y siguen la siguiente distribución:

$N(\qquad kg)$

$N(\qquad$ kg)

Si el tráiler ha de transportar el tronco y corteza de 50 árboles, calcular la probabilidad de que tenga que transportar más de 18.5 toneladas.

Apartado 3:

Se decide transportar madera a diez destinos diferentes. Tras realizar un estudio de mercado de una empresa similar que abastece esos diez destinos, se observa que, para un año determinado en el que se realizaron 200 pedidos, el número de pedidos/año observado para cada uno de los destinos es el siguiente:

| | Código del destino | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 |
| Pedidos/año | 18 | 22 | 20 | 21 | 19 | 26 | 20 | 14 | 16 | 24 |

Realizar el contraste correspondiente para comprobar si el número de pedidos/año de cada destino puede considerarse equidistribuido, con un nivel de confianza del 95%.

Apartado 4:

Finalmente la empresa decide comprar 200 tráileres. La duración de la batería del motor de cada tráiler es una variable aleatoria con distribución exponencial de media 5 años. ¿Cuál es el número esperado de tráileres que, durante el primer año, tendrán que sustituir la batería al menos una vez?

Nota: una batería se puede averiar varias veces en un año, y se supone que la sustitución es inmediata.

**Solución - Cuestiones**

Cuestión 1:
$\overline{x} = a/2; \widehat{a} = 2\overline{x};$
Esperanza del estimador: $E[\widehat{a}] = 2a/2 = a;$
Varianza del estimador: $var[\widehat{a}] = 4\sigma_x^2/n;$
$\sigma_x^2 = \int_0^a (x^2/a)dx - a^2/4 = a^2/12;$
$var[\widehat{a}] = 4a^2/12n = a^2/3n;$
Error cuadrático medio: $ECM = a^2/3n$

Cuestión 2:
Intervalo de confianza= $\overline{x} \pm 1,96\sqrt{2/n};$
El error máximo es la mitad del intervalo de confianza.
Error máximo $= 1,96\sqrt{2/n} = 0,2; n = (1,96^2 \times 2)/0,2^2 = 192.$

Cuestión 3:
$\widehat{s}_1^2/\widehat{s}_2^2 \sim F_{29,39}$ bajo $H_0$
$2,5^2/2^2 = 1,56 < F_{29,39}(0,95) = 1,76$; no se rechaza $H_0$
Cálculo del error tipo II:
$P(\widehat{s}_1^2/\widehat{s}_2^2 < 1,76|\sigma_1^2/\sigma_2^2 = 1,5) = P((\widehat{s}_1^2/\sigma_1^2)/(\widehat{s}_2^2/\sigma_2^2) < 1,76/1,5|\sigma_1^2/\sigma_2^2 = 1,5) =$
$= P(F_{29,39} < 1{,}17) = 0,68$

**SOLUCIÓN – PROBLEMA**

*Apartado 1:*

N: Zona Norte; C: Zona Centro; S: Zona Sur.

Información del enunciado:

$$P(N) \quad — \quad P(C) \quad — \quad P(S) \quad —$$

$$P(G|N) \qquad P(G|C) \qquad P(M|C) \qquad P(G|S) \qquad P(M|S)$$

Nos piden P(N|G):

$$P(N|G) \quad \frac{P(GN)}{P(G)} \quad \frac{P(G|N)\cdot P(N)}{P(G|N) \quad P(N) \quad P(G|C)\cdot P(C) \quad P(G|S)\cdot P(S)} \quad -$$

*Apartado 2:*

$$N(350\ 30) \qquad N(12\ 2)$$

Peso de un árbol: peso del tronco más peso de la corteza.

$$N\left(350+12, \sqrt{30 \qquad 6}\right) \quad N(362 \qquad 24)$$

$$\left(362\cdot 50, 31.24\cdot\sqrt{\ }\right) \quad N(18100\ 220)$$

$$P(X \qquad | X \sim N( \qquad 220))$$

*Apartado 3:*

| | Código del destino | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 |
| | 18 | 22 | 20 | 21 | 19 | 26 | 20 | 14 | 16 | 24 |
| | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| $(O \qquad)$ | 0.2 | 0.20 | 0 | 0.05 | 0.05 | 1.8 | 0 | 1.8 | 0.8 | 0.8 |

$$\sum \frac{(O \qquad)}{}$$

De las tablas: P(X | )

Como 5.7 < 16, aceptamos que son equiprobables.

*Apartado 4:*

$T$: duración de la batería de un tráiler. E[$T$] = 5 años.

$$f(T_i) \quad - \quad \mathrm{p}\left(--t\right)$$

Probabilidad en que se averíe antes del año: $p$

$$\int - \quad \mathrm{p}\left(--t\right) dt$$

$Y$: número de tráileres que se averían en un año

$$B(n \qquad 181)$$

tráileres

# EXÁMENES

## Curso 2014/15

**Cuestiones** (45 minutos, 5 puntos)

1. **En una clase de la École Centrale de París la proporción de estudiantes varones es del 60 %. De los varones, el 30 % son alumnos ERASMUS y en el caso de las mujeres un 40 % son ERASMUS. Entre los varones ERASMUS, un 50 % ha cursado Bachillerato en la enseñanza pública y entre los varones nacionales un 35 %. En las mujeres ERASMUS un 40 % proceden de la enseñanza pública en Bachillerato y entre las mujeres nacionales un 50 %.**

   **Se definen los sucesos:**

   **V = ser varón,**

   **E = ser alumno Erasmus,**

   **C = haber hecho el Bachillerato en colegio privado.**

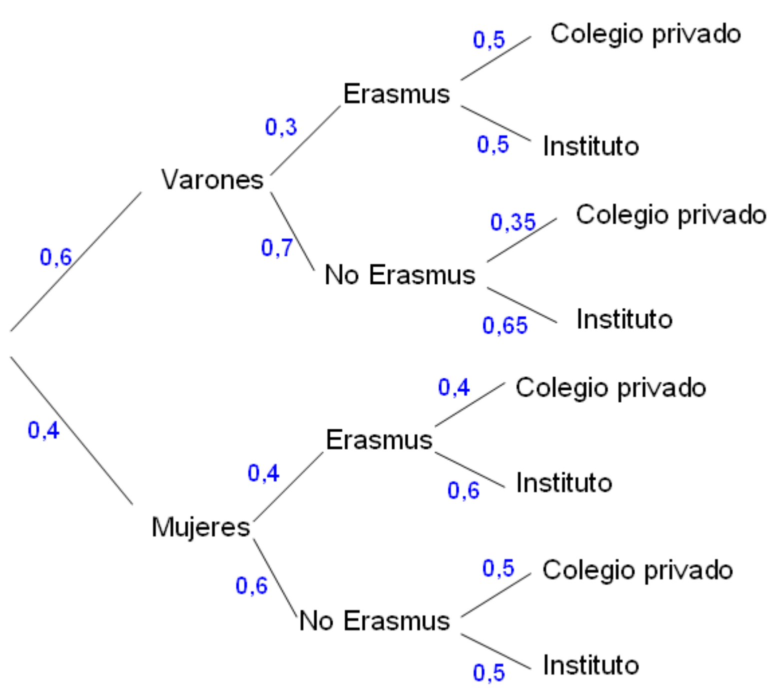
   **Se pide:**

   **a) Calcular la probabilidad de que ocurra alguno de los sucesos V, E ó C.**

   **b) Calcular la probabilidad de no ocurra ninguno de los sucesos V, E ó C.**

   **NOTA: Respecto al Bachillerato, sólo puede cursarse en colegios privados o institutos públicos.**

   Los datos se representan en la Figura siguiente:




   Y de acuerdo con ello resulta más sencillo resolver el apartado b) en primer lugar, ya que el a) será su complementario.

b) probabilidad de no ocurra ninguno de los sucesos V, E ó C = $pr(Mujer \cap Nacional \cap Instituto) =$ 0,4·0,6·0,5 = 0,12

a) $pr(V \cup E \cup C) = 1-$ probabilidad de no ocurra ninguno de los sucesos V, E ó C = 1 - 0.12 = 0.88

2. **Se tiene una caja que contiene dos monedas: una equilibrada y otra con dos caras.**

   **Se supone que se elige una de estas dos moneda al azar y que se lanza 3 veces, obteniéndose CARA las 3 veces. Calcular la probabilidad de que la moneda que se ha lanzado sea la moneda equilibrada.**

   Moneda equilibrada = $ME$

   Moneda no equilibrada = $\overline{ME}$

   $pr(CCC) = pr(3C)$

   $pr(ME) = pr(\overline{ME}) = \frac{1}{2}$

   $pr(3C|ME) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$

   $pr(3C|\overline{ME}) = 1$

   $pr(ME|3C) = \frac{pr(ME \cap 3C)}{pr(ME)} = \frac{pr(3C|ME) \cdot pr(ME)}{pr(3C|ME) \cdot pr(ME) + pr(3C|\overline{ME}) \cdot pr(\overline{ME})} = \frac{\frac{1}{8} \cdot \frac{1}{2}}{\frac{1}{8} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}} = \frac{\frac{1}{16}}{\frac{1}{16} + \frac{8}{16}} = \frac{1}{9}$

3. **Se tiene un conjunto de datos $\{x_1, x_2, ..., x_n\}$. Probar que para todo valor de $x_i$, $i = 1, ..., n$ se cumple que $s\sqrt{n} \geqslant |x_i - \overline{x}|$.**

   La varianza muestral se calcula como: $s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} \rightarrow s^2 n = \sum_{i=1}^{n} (x_i - \overline{x})^2$.

   Lo que se pide probar: $s\sqrt{n} \geqslant |x_i - \overline{x}|$ y elevando al cuadrado queda: $s^2 n \geqslant (x_i - \overline{x})^2$.

   Como $s^2 n = \sum_{i=1}^{n} (x_i - \overline{x})^2 = (x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2 \geqslant (x_i - \overline{x})^2$ c.q.d.

**Problema** (45 minutos, 5 puntos)

En una empresa $A$ se fabrican piezas que se reparten en cajas; en cada caja hay una proporción $X$ de defectuosas, siendo la función de densidad de $X$ la siguiente:

$$f_X(x) = kx(1-x) \qquad 0 < x < 1$$

1. Calcular la probabilidad de que, si se elige al azar una caja, su proporción de defectuosas sea menor que 0,2.
2. Una función de coste de procesado posterior de las piezas es $Y = X^2$, ¿cuál es la probabilidad de que $Y$ sea mayor que $0,5$?
3. Se elige una caja al azar y se extraen 5 piezas, resultando dos defectuosas. ¿Cuál es la probabilidad de que sean las dos primeras?

Solución al problema

1) $\int_0^1 kx(1-x)dx = 1; k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = k/6 = 1; k = 6;$

$P(X < 0,2) = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^{0,2} = 0,104$

2) $P(Y > 0,5) = P(X^2 > 0,5) = P(X > \sqrt{0,5}) = 6\int_{\sqrt{0,5}}^1 x(1-x)dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{\sqrt{0,5}}^1 = 0,207$

3) $P = \frac{1}{\binom{5}{2}} = 0,1$

**Cuestiones** (45 minutos)

1. Un pescador tiene comprobado que el número de peces que captura cuando va a pescar es una variable aleatoria de Poisson de media 0.6 peces a la hora. El próximo fin de semana decide ir dos horas de pesca. Si al finalizar las dos horas ha capturado algún pez se marchará a casa. En caso contrario, seguirá pescando hasta capturar un pez. Calcular la probabilidad de que esté pescando más de dos horas y el número esperado de peces que capturará.

2. Se lanzan dos monedas. A continuación se lanzan tantas monedas como número de caras han salido. Se definen las variables aleatorias X: "número de caras obtenido en la primera tirada", Y: "número de caras obtenido en la segunda tirada". Determinar la función de probabilidad conjunta de X e Y.

3. Sean 100 paquetes cuyos pesos son variables aleatorias independientes con distribución uniforme entre 5 y 50 kilos. ¿Cual es la probabilidad de que el peso total de los paquetes exceda 3000 kilos?

**Problema** (45 minutos)

Un empresario planea abrir 10 tiendas de alimentación en la Comunidad de Madrid. Un estudio previo de mercado indica que las recaudaciones de las tiendas son variables aleatorias independientes con función de densidad:

$$f_{X_i}(x) = \begin{cases} \frac{2x}{a^2} & x \in [0, a] \\ \\ 0 & \textit{en el resto} \end{cases},$$

donde $X_i$: "recaudación (en euros) de la tienda i-*ésima*" y $a$ es un parámetro conocido.

1. Calcule el número medio de tiendas cuya recaudación será superior a $\frac{2}{3}a$.
2. Calcule la función de distribución y de densidad de la variable aleatoria $M$: "máximo de la recaudación de las 10 tiendas".
3. Calcule la probabilidad de que al menos dos de las diez tiendas tengan una recaudación superior al valor esperado de $M$.
4. Dos de las tiendas están bastante cerca por lo que se sospecha que sus recaudaciones no son variables aleatorias independientes. Suponiendo que el coeficiente de correlación entre las recaudaciones de estas tiendas es -0.5, calcule el coeficiente de correlación entre los beneficios de ambas tiendas sabiendo que el beneficio de cada tienda se puede expresar como $B_i = 0{,}3X_i - 35000$.

**Solución de las cuestiones**

- Cuestion 1

  Se definen las siguientes variables aleatorias

  - X: número de peces capturados en las 2 primeras horas.
  - Y: número de peces capturados a partir de las 2 primeras horas.
  - Z: número de peces capturados en total.

  $$X \rightsquigarrow Poisson(1{,}2) \Rightarrow P(X=0) = e^{-1{,}2} = 0{,}30$$

  Además, $E[X] = \lambda = 1{,}2$

  Los valores que puede tomar Y son 0 y 1, con probabilidad

  $$P(Y=0) = P(X \geq 1) = 1 - P(X=0) = 0{,}70$$
  $$P(Y=1) = P(X=0) = 0{,}30$$

  Por lo tanto

  $$E[Y] = 0 \cdot P(Y=0) + 1 \cdot P(Y=1) = 0{,}30$$

  Finalmente

  $$E[Z] = E[X] + E[Y] = 1{,}2 + 0{,}3 = 1{,}5$$

- Cuestion 2

  Se definen las siguientes variables aleatorias

  - $X$: número de caras obtenido en la primera tirada.
  - $Y$: número de caras obtenido en la segunda tirada.

  Los valores que puede tomar $X$ son 0, 1, 2. Los valores que puede tomar $Y$ son 0, 1, 2. Las probabilidades son

  $$P(X=0,Y=0) = P(Y=0|X=0)P(X=0) = 1 \cdot \frac{1}{4} = \frac{1}{4}$$
  $$P(X=1,Y=0) = P(Y=0|X=1)P(X=1) = \frac{1}{2} \cdot \frac{2}{4} = \frac{1}{4}$$
  $$P(X=1,Y=1) = P(Y=1|X=1)P(X=1) = \frac{1}{2} \cdot \frac{2}{4} = \frac{1}{4}$$
  $$P(X=2,Y=0) = P(Y=0|X=2)P(X=2) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$$
  $$P(X=2,Y=1) = P(Y=1|X=2)P(X=2) = \frac{2}{4} \cdot \frac{1}{4} = \frac{2}{16}$$

$$P(X=2,Y=2)=P(Y=2|X=2)P(X=2)=\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}$$
$$P(X=k,Y=m)=0 \quad \text{si } m>k$$

Estos resultados los podemos organizar en la siguiente tabla

| | | X=k | | |
|---|---|---|---|---|
| | | 0 | 1 | 2 |
| | 0 | 1/4 | 1/4 | 1/16 |
| Y=m | 1 | 0 | 1/4 | 2/16 |
| | 2 | 0 | 0 | 1/16 |

Otra forma de resolverlo es utilizando la Binomial:

$$P(X=k)=\binom{2}{k}0{,}5^k 0{,}5^{2-k}=\binom{2}{k}0{,}5^2, \quad 0\leq k\leq 2$$

$$P(Y=m|X=k)=\binom{k}{m}0{,}5^m 0{,}5^{k-m}=\binom{k}{m}0{,}5^k, \quad 0\leq m\leq k$$

Por tanto

$$P(X=k,Y=m)=P(Y=m|X=k)P(X=k)=\binom{2}{k}\binom{k}{m}0{,}5^{k+2}, \quad 0\leq m\leq k\leq 2$$

- Cuestion 3

  Se definen las siguientes variables aleatorias

  - $P_i$: peso del paquete i-ésimo.

$$P_i \rightsquigarrow Uniforme(5,50) \Rightarrow$$
$$E[P_i]=\frac{5+50}{2}=27{,}5=\mu, \quad Var[P_i]=\frac{(50-5)^2}{12}=168{,}75=\sigma^2$$

  - $P_t$: peso de los 100 paquetes, $P_t=\sum_{i=1}^{n} P_i$

Por el Teorema Central del Límite

$$P_t=\sum_{i=1}^{100} P_i \rightsquigarrow N(100\mu,100\sigma^2)=N(2750,129{,}9^2)$$

$$P(P_t>3000)=1-P(P_t\leq 3000)=1-P\left(\frac{P_t-2750}{129{,}9}\leq\frac{3000-2750}{129{,}9}\right)$$
$$=1-P(Z\leq 1{,}92)=0{,}0274$$

**Solución del problema**

**a)** $Y$ =*v.a nº de tiendas que superan* $\dfrac{2}{3}a \rightarrow B(n=10,p)$

donde $p = P(X > \frac{2}{3}a) = \int\limits_{\frac{2}{3}a}^{a} \frac{2x}{a^2}dx = \left[\frac{x^2}{a^2}\right]_{\frac{2}{3}a}^{a} = \frac{1}{a^2}\left[a^2 - \frac{4}{9}a^2\right] = \frac{5}{9}$.

$E[Y] = n \times p = \frac{50}{9}$. Entre 5 y 6 tiendas superarán la recaudación mencionada.

**b)** $M = max\,\{X_1, X_2, X_3, ...., X_{10}\}$

$F_M(m) = P(M \leq m) = P(X_1 \leq m; X_2 \leq m; ....; X_{10} \leq m) = [P(X \leq m)]^{10} =$
$=[F_X(m)]^{10} = \frac{m^{20}}{a^{20}} \qquad m \in [0,a]$

$F_X(x) = \int\limits_0^x \frac{2x}{a^2}dx = \frac{x^2}{a^2} \qquad x \in [0,a]$

$f_M(m) = \frac{dF_M(m)}{dm} = \dfrac{20m^{19}}{a^{20}} \qquad m \in [0,a]$

**c)** $E[M] = \int\limits_0^a m\frac{20m^{19}}{a^{20}}dm = \frac{20}{a^{20}}\int\limits_0^a m^{20}dm = \frac{20}{21}a$

$T$ =*v.a nº de tiendas que superan este valor de un total de 10.* $B(n=10,p)$

$p = P(X > \frac{20}{21}a) = 1 - F_X(\frac{20}{21}a) = 1 - (\frac{20}{21}a)^2/a^2 = 1 - \left(\frac{20}{21}\right)^2 = 1 - 0{,}9070 = 0{,}093$

$$P(T \geq 2) = 1 - [\binom{10}{0}0{,}093^0(1-0{,}093)^{10} + \binom{10}{1}0{,}093^1(1-0{,}093)^9] = 1 - 0{,}763 = 0{,}237$$

**d)**

Hay dos posibles respuestas, ambas válidas.

**Respuesta 1.**:El coeficiente de correlación es adimensional, no se ve afectado por las transformaciones lineales y mantiene el signo cuando las constantes de proporcionalidad son del mismo signo.

En este caso $B_1$ es una transformación lineal (positiva) de $X_1$, y $B_2$ es una transformación lineal (positiva) de $X_2$, por lo tanto, $\rho(B_1, B_2) = \rho(X_1, X_2) = -0{,}5$.

**Respuesta 2:** Calcularlo para el caso concreto del problema

$B_1 = 0{,}3X_1 - 35000$
$B_2 = 0{,}3X_2 - 35000$

$$\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} 0{,}3 & 0 \\ 0 & 0{,}3 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} - \begin{pmatrix} 35000 \\ 35000 \end{pmatrix}$$

$M_B = AM_XA^T$

$$M_B = \begin{pmatrix} 0{,}3 & 0 \\ 0 & 0{,}3 \end{pmatrix}\begin{pmatrix} \sigma^2 & -0{,}5\sigma^2 \\ -0{,}5\sigma^2 & \sigma^2 \end{pmatrix}\begin{pmatrix} 0{,}3 & 0 \\ 0 & 0{,}3 \end{pmatrix} = \sigma^2\begin{pmatrix} 0{,}09 & -0{,}045 \\ -0{,}045 & 0{,}09 \end{pmatrix}$$

$\sigma^2_{X_1} = \sigma^2_{X_2} = \sigma^2;$
$\sigma_{X_1X2} = -0{,}5\sigma^2$

$$\rho(B_1, B_2) = \frac{\sigma_{B_1B_2}}{\sigma_{B_1}\sigma_{B_2}} = \frac{-0{,}045\sigma^2}{\sigma \times 0{,}3 \times \sigma \times 0{,}3} = \frac{-0{,}045}{0{,}09} = -0{,}5$$

**Cuestiones** - (45 minutos, 5 puntos)

1. Estimar por máxima verosimilitud $\sigma^2$ para la siguiente función de densidad

$$f_X(x) = \frac{x}{\sigma^2}\exp\{-\frac{x^2}{2\sigma^2}\}, \quad x > 0, \sigma > 0$$

y decir si es un estimador centrado, sabiendo que $E[X] = \sigma\sqrt{\frac{\pi}{2}}$ y $Var[X] = \frac{4-\pi}{2}\sigma^2$.

2. A un servidor de noticias internacional se puede acceder desde cualquier país del mundo. En la tabla se muestra la distribución de frecuencias de tiempos en minutos entre accesos consecutivos para los registros de los últimos días.

| Clases (en minutos) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | Más de 50 |
|---|---|---|---|---|---|---|
| Frecuencia Observada | 1680 | 554 | 218 | 67 | 23 | 18 |

Contrastar con $\alpha = 0.05$ si el tiempo entre accesos tiene distribución de probabilidad exponencial, teniendo en cuenta que el tiempo medio entre accesos calculado a partir de las observaciones anteriores es 9.6 minutos.

3. En la comparación de las medias de dos tratamientos (bajo las hipótesis de normalidad, homocedasticidad e independencia), el estimador de la varianza del modelo es

$$\widehat{s}_R^2 = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i} e_{ij}^2}{n_1+n_2-2} = \frac{(n_1-1)}{n_1+n_2-2}\widehat{s}_1^2 + \frac{(n_2-1)}{n_1+n_2-2}\widehat{s}_2^2$$

donde $e_{ij}$ son los residuos del modelo, $n_1$ es el número de datos del tratamiento 1, $n_2$ es el número de datos del tratamiento 2, $\widehat{s}_1^2$ la varianza muestral (corregida) del tratamiento 1 y $\widehat{s}_2^2$ la varianza muestral (corregida) del tratamiento 2. Esta estimación es una combinación lineal de las varianzas muestrales de los dos grupos. Sea $\widehat{\sigma}^2$ un estimador de $\sigma^2$ obtenido como

$$\widehat{\sigma}^2 = a\widehat{s}_1^2 + b\widehat{s}_2^2,$$

determinar los valores $a$ y $b$ (en función de $n_1, n_2$) que hacen que $\widehat{\sigma}^2$ sea un estimador centrado y de mínima varianza. Comparar el estimador obtenido con $\widehat{s}_R^2$. (Nota: la media y varianza de la distribución $\chi_n^2$ es $n$ y $2n$, respectivamente).

**Problema** - (45 minutos, 5 puntos)

Una línea fabrica rodamientos para maquinaria agraria, con una velocidad de procesamiento de 5000 piezas por minuto. Cuando el proceso funciona correctamente, debido a causas difíciles de evitar, la línea produce un 3% de rodamientos defectuosos. De vez en cuando el proceso se desajusta, fabricando un mayor número de piezas defectuosas. Para decidir si el proceso esta ajustado ($H_0 : p = 0.03$) o el proceso está desajustado ($H_1 : p > 0.03$), se toma una muestra de $n = 600$ piezas.

1. Determinar la región de aceptación y la región de rechazo del contraste para nivel de significación $\alpha = 0.01$. Indica cuál es la proporción máxima de piezas defectuosas que debe tener una muestra para aceptar que el proceso está bajo control. Obtener el p-valor del contraste si en una muestra se han observado 25 rodamientos defectuosos.

2. Si el proceso se desajusta y fabrica una proporción de defectuosas igual al 5%, ¿cuál es la probabilidad de detectarlo con una muestra de 600 piezas?

3. Determinar el tamaño muestral $n$ si se desea que la probabilidad de detectar el cambio del apartado 2 sea 0.90.

**Cuestiones** (45 minutos – 5 puntos)

**Cuestión 1:**

$$f_C(x_1,\dots,x_n) = \frac{x_1\cdot x_2\cdot\ldots\cdot x_n}{\sigma^{2n}}\exp\left(-\frac{1}{2\cdot\sigma^2}\sum x_i^2\right)$$

$$l(\sigma^2) = -n\cdot\log(\sigma^2) - \frac{1}{2\cdot\sigma^2}\sum x_i^2 + \sum(\log(x_i))$$

$$\frac{dl}{d\,\sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\sum x_i^2$$

$$-\frac{n}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum x_i^2 = 0 \quad\Rightarrow\quad \hat{\sigma}^2 = \frac{\sum x_i^2}{2n}$$

$$\frac{d^2l}{(d\,\sigma^2)^2} = +\frac{n}{\sigma^4} - \frac{1}{\sigma^6}\sum x_i^2 = \frac{n}{\sigma^4} - \frac{2n\hat{\sigma}^2}{\sigma^6}$$

$$\left.\frac{d^2l}{(d\,\sigma^2)^2}\right|_{\hat{\sigma}^2} = -\frac{n}{\hat{\sigma}^4} < 0 \;\Rightarrow\; \text{Es máximo}$$

$$E[\hat{\sigma}^2] = E\left[\frac{\sum x_i^2}{2n}\right] = \frac{1}{2n}\sum E[X_i^2]$$

$$E[X_i^2] = E[X_i]^2 + Var[X_i] = \frac{\sigma^2\pi}{2} + \frac{4-\pi}{2}\sigma^2 = 2\sigma^2$$

$$E[\hat{\sigma}^2] = \frac{1}{2n}(2n\sigma^2) = \sigma^2$$

Es un estimador centrado

**Cuestión 2:**

Para la distribución exponencial, tenemos: $F_T(t) = P(T < t) = 1 - e^{t/9.6}$

Con los datos hemos estimado un parámetro. Por tanto, $r = 1$.

| Clase | a | b | P(T<a) | P(T<b) | P(a<T<b) | Oi | Ei=n*p | (Oi-Ei)^2/Ei |
|---|---|---|---|---|---|---|---|---|
| [0-10) | 0 | 10 | 0,00 | 0,65 | 0,65 | 1680 | 1656,7 | 0,33 |
| [10-20) | 10 | 20 | 0,65 | 0,88 | 0,23 | 554 | 584,6 | 1,60 |
| [20-30) | 20 | 30 | 0,88 | 0,96 | 0,08 | 218 | 206,3 | 0,67 |
| [30-40) | 30 | 40 | 0,96 | 0,98 | 0,03 | 67 | 72,8 | 0,46 |
| [40-50) | 40 | 50 | 0,98 | 0,99 | 0,01 | 23 | 25,7 | 0,28 |
| [50-∞) | 50 | ∞ | 0,99 | 1,00 | 0,01 | 18 | 14,0 | 1,14 |
| | | | | | 1 | 2560 | 2560 | 4,47 |

$\chi^2_{6-1-1,\ 0.95} = 9.488$

Como $\sum\frac{(O_i-E_i)^\wedge 2}{E_i} = 4.47 < 9.488$, aceptamos la hipótesis nula.

**Cuestión 3:**

$$E[\hat{\sigma}^2] = a \cdot E[\hat{s}_1^2] + b \cdot E[\hat{s}_2^2] = a \cdot \sigma^2 + b \cdot \sigma^2 = (a+b) \cdot \sigma^2$$

Para que sea centrado, ha de cumplir: $a + b = 1$. Es decir: $b = a - 1$.

$$Var[\hat{\sigma}^2] = a^2 \cdot Var[\hat{s}_1^2] + b \cdot Var[\hat{s}_2^2]$$

$$\frac{(n_1 - 1) \cdot \hat{s}_1^2}{\sigma^2} \sim \chi^2_{n_1 - 1}$$

$$Var\left[\frac{(n_1 - 1) \cdot \hat{s}_1^2}{\sigma^2}\right] = 2(n_1 - 1)$$

$$Var[\hat{s}_1^2] = \frac{2\sigma^4}{(n_1 - 1)}$$

$$Var[\hat{\sigma}^2] = a^2 \cdot \frac{2\sigma^4}{(n_1 - 1)} + (1-a)^2 \cdot \frac{2\sigma^4}{(n_2 - 1)}$$

$$\frac{d\, Var[\hat{\sigma}^2]}{d\, a} = 4 \cdot a \cdot \frac{\sigma^4}{(n_1 - 1)} - 4 \cdot (1-a) \cdot \frac{\sigma^4}{(n_2 - 1)}$$

Igualando a cero:

$$\frac{a}{(n_1 - 1)} + \frac{a}{(n_2 - 1)} = \frac{1}{(n_2 - 1)}$$

$$a = \frac{\frac{1}{(n_2 - 1)}}{\frac{1}{(n_1 - 1)} + \frac{1}{(n_2 - 1)}} = \frac{n_1 - 1}{n_1 + n_2 - 2}$$

Por tanto: $a = \frac{n_1 - 1}{n_1 + n_2 - 2}$ y $b = \frac{n_2 - 1}{n_1 + n_2 - 2}$

**Problema** (45 minutos – 5 puntos)

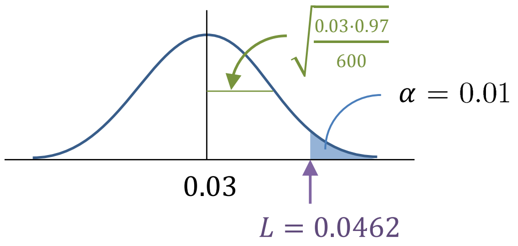
## Apartado 1:

Consideramos el siguiente contraste:

$$\begin{cases} H_0: & p = 0.03 \\ H_1: & p > 0.03 \end{cases}$$

Si $H_0$ es cierto: $\dfrac{\hat{p}-p}{\sqrt{\frac{p\cdot(1-p)}{n}}} \sim N(0,1)$

$$L = 0.03 + 2.33 \cdot \sqrt{\frac{0.03 \cdot 0.97}{600}} = 0.0462$$

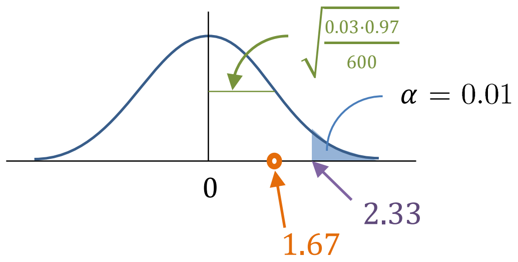



Obtenemos:

$$\begin{cases} \text{Región aceptación:} & \hat{p} \leq 0.0462 \\ \text{Región rechazo:} & \hat{p} > 0.0462 \end{cases}$$

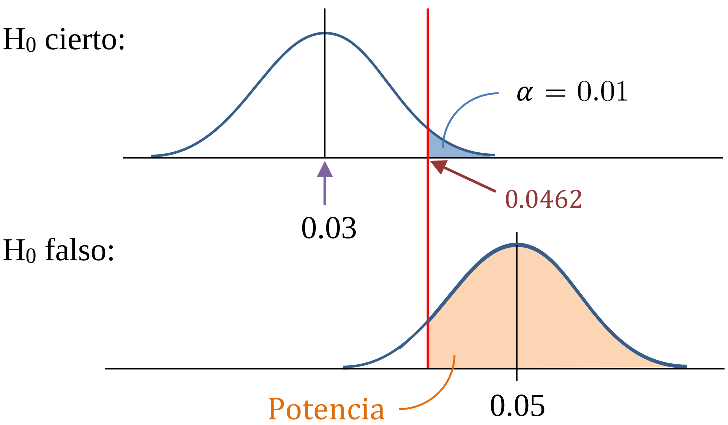
$$r = 25 \Rightarrow \hat{p} = \frac{25}{600} = 0.0417 \Rightarrow z_0 = \frac{0.0417 - 0.03}{\sqrt{\frac{0.03 \cdot 0.97}{600}}} = 1.67$$

$$P(Z > 1.67) = 1 - 0.9525 = 0.0475$$




El p-valor = 0.0475. Se observa que es mayor que $\alpha$, y por tanto aceptamos $H_0$.

## Apartado 2:




La probabilidad de detectar el cambio (potencia del contraste) se calcula como la probabilidad de que $\hat{p}$ sea mayor que $L$:

$P(\text{Detectar cambio}) = P(rechazar\ H_0 \mid H_0\ falsa)$

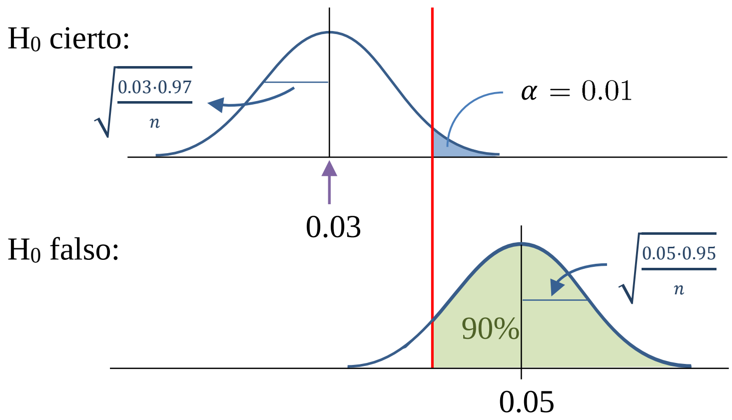
$$= P\left(\hat{p} > 0.03 + 2.33 \cdot \sqrt{\frac{0.03 \cdot 0.97}{600}} \;\middle|\; \hat{p} \sim N\left(0.05, \sqrt{\frac{0.05 \cdot 0.95}{600}}\right)\right)$$

$$= P\left(\hat{p} > 0.0462 \;\middle|\; \hat{p} \sim N\left(0.05, \sqrt{\frac{0.05 \cdot 0.95}{600}}\right)\right)$$

$$= P\left(Z > \frac{0.0462 - 0.05}{\sqrt{\frac{0.05 \cdot 0.95}{600}}}\right)$$

$$= P(Z > -0.4271) = 0.6664$$

**Apartado 3:**




$$L = 0.03 + 2.33\sqrt{\frac{0.03 \cdot 0.97}{n}}$$

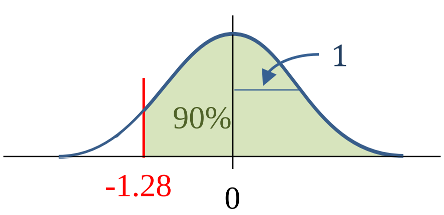
$$0.90 = P\left(\hat{p} > 0.03 + 2.33 \cdot \sqrt{\frac{0.03 \cdot 0.97}{n}} \;\middle|\; \hat{p} \sim N\left(0.05, \sqrt{\frac{0.05 \cdot 0.95}{n}}\right)\right)$$

$$0.90 = P\left(Z > \frac{0.03 + 2.33 \cdot \sqrt{\frac{0.03 \cdot 0.97}{n}} - 0.05}{\sqrt{\frac{0.05 \cdot 0.95}{n}}}\right)$$

$$\frac{0.03 + 2.33 \cdot \sqrt{\frac{0.03 \cdot 0.97}{n}} - 0.05}{\sqrt{\frac{0.05 \cdot 0.95}{n}}} = -1.28$$

$$n = \left(\frac{2.33 \cdot \sqrt{0.03 \cdot 0.97} + 1.28 \cdot \sqrt{0.05 \cdot 0.95}}{0.05 - 0.03}\right)^2 = 1143.9$$

$$n \approx 1144$$



**Problema 1** (50 minutos, 10 puntos)

Se dispone de seis bolas, cuatro de ellas Blancas y dos Negras, que se reparten aleatoriamente en tres urnas U1, U2 y U3 de la siguiente forma:

Se coloca al azar una bola en U1, luego dos bolas en U2 y las tres restantes en U3. Calcule:

1. La probabilidad de que la bola colocada en U1 sea de color Blanco.
2. La probabilidad de que en U2 haya:
   *a*) Dos bolas Blancas
   *b*) Dos bolas Negras
   *c*) Una bola de cada color
3. Al finalizar el reparto se saca una bola de U3 al azar y sale Blanca. ¿Cuál es la probabilidad de que en U1 haya una bola Blanca?

**Problema 2** (50 minutos, 10 puntos)

Las demandas anuales, $X_A$, $X_B$, $X_C$ de tres productos $A$, $B$ y $C$ siguen una distribución normal multivariante, con vector de medias y matriz de varianzas y covarianzas:

$$E[X] = \begin{pmatrix} 10 \\ 15 \\ 20 \end{pmatrix} \text{ y } Var[X] = \begin{pmatrix} 6 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 5 \end{pmatrix}.$$

1. Calcule la probabilidad de que $P(X_C > X_B - X_A)$

2. Si las organizaciones $O_1$ y $O_2$ fabrican esos tres productos, y los precios de venta de los mismos son:

| $\mathrm{O}_i$ | $\mathrm{p}_A$ | $\mathrm{p}_B$ | $\mathrm{p}_C$ |
|---|---|---|---|
| 1 | 10 | 12 | 15 |
| 2 | 8 | 14 | 14 |

   Obtenga la distribución conjunta de los ingresos anuales de estas dos organizaciones.

3. Probabilidad de que los ingresos anuales de la organización 1 sean superiores a los de la organización 2.

**Problema 1. Solución:**

**Solución:**

1. Se define el suceso:

   *Una bola Blanca en U1:* $B_1$

   $P(B_1) = \dfrac{\binom{4}{1}}{\binom{6}{1}} = \dfrac{2}{3};$

2. Se definen los sucesos:

   *Una bola Blanca en U1:* $B_1$; *Una bola Negra en U1:*$N_1$

   *Dos bolas Blancas en U2*: $BB_2$; *Dos bolas Negras en U2:* $NN_2$;

   *Una bola de cada color en U2:* $NB_2$

   *a)* $P(BB_2) = P(BB_2|B_1) \times P(B_1) + P(BB_2|N_1) \times P(N_1) = \dfrac{\binom{3}{2}\binom{2}{0}}{\binom{5}{2}} \times \dfrac{2}{3} + \dfrac{\binom{4}{2}\binom{1}{0}}{\binom{5}{2}} \times \dfrac{1}{3} =$

   $\dfrac{3}{10} \times \dfrac{2}{3} + \dfrac{3}{5} \times \dfrac{1}{3} = \dfrac{2}{5} = \dfrac{6}{15}.$

   *b)* $P(NN_2) = P(NN_2|B_1) \times P(B_1) + P(NN_2|N_1) \times P(N_1) = \dfrac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} \times \dfrac{2}{3} + 0 = \dfrac{1}{10} \times \dfrac{2}{3} = \dfrac{1}{15}.$

   *c)* $P(NB_2) = P(NB_2|B_1) \times P(B_1) + P(NB_2|N_1) \times P(N_1) = \dfrac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} \times \dfrac{2}{3} + \dfrac{\binom{1}{1}\binom{4}{1}}{\binom{5}{2}} \times \dfrac{1}{3} =$

   $\dfrac{3}{5} \times \dfrac{2}{3} + \dfrac{2}{5} \times \dfrac{1}{3} = \dfrac{8}{15}.$

3. Se define el suceso:

   *Extraer una bola Blanca de U3:* $EB_3$

   $P(B_1|EB_3) = \dfrac{P(EB_3|B_1) \times P(B_1)}{P(EB_3)} = \dfrac{\frac{3}{5} \times \frac{2}{3}}{\frac{3}{5} \times \frac{2}{3} + \frac{4}{5} \times \frac{1}{3}} = \dfrac{3}{5}.$

$P(EB_3) = P(EB_3|B_1) \times P(B_1) + P(EB_3|N_1) \times P(N_1).=\dfrac{3}{5} \times \dfrac{2}{3} + \dfrac{4}{5} \times \dfrac{1}{3}$

$P(EB_3|B_1) = P(EB_3|BB_2{\cdot}B_1)P(BB_2|B_1)+P(EB_3|NN_2{\cdot}B_1)P(NN_2|B_1)+P(EB_3|NB_2{\cdot}B_1)P(NB_2|B_1) = \dfrac{1}{3} \times \dfrac{3}{10} + \dfrac{3}{3} \times \dfrac{1}{10} + \dfrac{2}{3} \times \dfrac{6}{10} = \dfrac{3}{5}.$

$P(EB_3|N_1) = P(EB_3|BB_2{\cdot}N_1)P(BB_2|N_1)+P(EB_3|NN_2{\cdot}N_1)P(NN_2|N_1)+P(EB_3|NB_2{\cdot}N_1)P(NB_2|N_1) = \dfrac{2}{3} \times \dfrac{3}{5} + 0 + \dfrac{3}{3} \times \dfrac{2}{5} = \dfrac{4}{5}.$

**Problema 2. Solución:**

1. El enunciado pide $P(Y > 0)$,siendo $Y = X_A - X_B + X_C$.

   $\mu_Y = aE[X] = [1 - 11] \begin{bmatrix} 10 \\ 15 \\ 20 \end{bmatrix} = 15$

   $\sigma_Y^2 = aVar[X]a^T = [1 - 11] \begin{bmatrix} 6 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 9.$

   Así, $P(Y > 0|Y \sim N(15, 3)) = P(z > -5) \simeq 1$

2. Así,

   $O_1 = 10X_A + 12X_B + 15X_C$

   $O_2 = 8X_A + 14X_B + 14X_C$

   $O = AX,$

   $\mu_O = AE[X] = \begin{bmatrix} 10 & 12 & 15 \\ 8 & 14 & 14 \end{bmatrix} \begin{bmatrix} 10 \\ 15 \\ 20 \end{bmatrix} = \begin{bmatrix} 580 \\ 570 \end{bmatrix}$ y $Var[O] = AVar[X]A^T = \begin{bmatrix} 10 & 12 & 15 \\ 8 & 14 & 14 \end{bmatrix} \begin{bmatrix} 6 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 5 \end{bmatrix} \begin{bmatrix} 10 & 8 \\ 12 & 14 \\ 15 & 14 \end{bmatrix} =$ $\begin{bmatrix} 4041 & 3974 \\ 3974 & 3940 \end{bmatrix}.$

   Los ingresos anuales de estas dos organizaciones siguen una distribución normal bivariante con vector de medias $\begin{bmatrix} 580 \\ 570 \end{bmatrix}$ y matriz de varianzas y covarianzas $\begin{bmatrix} 4041 & 3974 \\ 3974 & 3940 \end{bmatrix}.$

3. La probabilidad de que los ingresos anuales de la organización 1 sean superiores a los de la organización 2 es $P(O_1 > O_2) = P[W > 0]$, siendo $W = O_1 - O_2$.

   $P(W > 0|W \sim N(10, \sqrt{4041 + 3940 - 2x3974})) = P(z > \frac{-10}{\sqrt{33}}) = P(z > -1{,}74) = 0{,}9591$

**Problema** (50 minutos, 5 puntos)

**En una empresa se fabrican piezas con una resistencia media de 100 y una desviación típica de 5 (distribucion normal). Se ha diseñado un nuevo procedimiento y en un ensayo con 20 probetas, su resistencia media ha sido 105.**

1. **Contrastar suponiendo que la varianza no ha cambiado, si la resistencia promedio con el nuevo procedimiento es mayor que la del habitual y calcular la probabilidad de error tipo II para un valor alternativo de 106.**

   Se pide contrastar:

   $H_0$: $\mu = 100$

   $H_1$: $\mu > 100$

   Como la varianza (y la desviación típica) es conocida, $\sigma = 5$, entonces utilizamos que $\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \rightarrow Z$.

   Particularizando $\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$ bajo la hipótesis nula quedaría: $\frac{105-100}{5/\sqrt{20}} = 4{,}4721 > 1{,}64$ que sería el valor que separa la región de aceptación (a la izquierda de este valor) y rechazo (a la derecha de este valor) en un contraste como el que se pide. Así pues se rechaza $H_0$ ($\mu = 100$) frente a la alternativa ($\mu > 100$).

   Ahora, para calcular la probabilidad de error de tipo II:

   $pr$(error de tipo II) $= pr$(aceptar $H_0|H_0$ falsa), y se pide esta probabilidad bajo el supuesto concreto de que $\mu = 105$.

   Justo en la frontera entre la región de aceptación y rechazo: $\overline{x} = 1{,}64 \cdot (5/\sqrt{20}) + 100 = 101{,}8336$ ya que en dicha frontera $\frac{\overline{x}-100}{5/\sqrt{20}} = 1{,}64$.

   $pr(\text{aceptar } H_0|H_0 \text{ falsa}) = pr(\overline{x} < 101{,}8336|\mu = 106) = pr(\frac{\overline{x}-106}{5/\sqrt{20}} < \frac{101{,}8336-106}{5/\sqrt{20}}) = pr(Z < \frac{101{,}8336-106}{5/\sqrt{20}}) = \Phi_Z(-3{,}7265) = 1 - 0{,}9999004$

2. **Los datos de un segundo ensayo (B) con el nuevo procedimiento se distribuyen en 6 intervalos tal y como se muestra en la tabla adjunta. Para los datos de dicha tabla la media muestral y desviación típica muestral corregida son, respectivamente: $\overline{x}_B = 102{,}8$ y $\widehat{s}_B = 4{,}7$.**

   | Intervalo | Frecuencia Obs. |
   |---|---|
   | **<97** | **5** |
   | **[97 - 99)** | **2** |
   | **[99-101)** | **3** |
   | **[101-103)** | **8** |
   | **[103-105)** | **4** |
   | **⩾105** | **13** |

   **Se pide contrastar la hipótesis de normalidad para la variable resistencia media"en este segundo ensayo (ensayo B).**

   Para las frecuencias esperadas, se utiliza que $\widehat{\mu}_B = \overline{x}_B = 102{,}8119$ y $\widehat{\sigma}_B = \widehat{s}_B = 4{,}7117$. Y el número total de datos (a partir de las frecuencias observadas) es 35.

$pr(Y < 97|Y \to N(102{,}8119, 4{,}7117)) = pr(Z < \frac{97-102{,}8119}{4{,}7117}) = pr(Z < -1{,}2335) = 1 - 0{,}8907 = 0{,}1093$.

$pr(97 < Y < 99|Y \to N(102{,}8119, 4{,}7117)) = pr(Z < \frac{99-102{,}8119}{4{,}7117}) - pr(Z < \frac{97-102{,}8119}{4{,}7117}) = pr(Z < -0{,}8090) - pr(Z < -1{,}2335) = (1 - 0{,}7910) - 0{,}1093 = 0{,}0997$.

$pr(99 < Y < 101|Y \to N(102{,}8119, 4{,}7117)) = pr(Z < \frac{101-102{,}8119}{4{,}7117}) - pr(Z < \frac{99-102{,}8119}{4{,}7117}) = pr(Z < -0{,}3846) - pr(Z < -0{,}8090) = (1 - 0{,}648) - (1 - 0{,}7910) = 0{,}1430$.

$pr(101 < Y < 103|Y \to N(102{,}8119, 4{,}7117)) = pr(Z < \frac{103-102{,}8119}{4{,}7117}) - pr(Z < \frac{101-102{,}8119}{4{,}7117}) = pr(Z < 0{,}0399) - pr(Z < -0{,}3846) = 0{,}5160 - (1 - 0{,}648) = 0{,}1640$

$pr(103 < Y < 105|Y \to N(102{,}8119, 4{,}7117)) = pr(Z < \frac{105-102{,}8119}{4{,}7117}) - pr(Z < \frac{103-102{,}8119}{4{,}7117}) = pr(Z < 0{,}4644) - pr(Z < 0{,}0399) = 0{,}6772 - 0{,}5160 = 0{,}1612$

$pr(Y > 105|Y \to N(102{,}8119, 4{,}7117)) = pr(Z > \frac{105-102{,}8119}{4{,}7117}) = pr(Z > 0{,}4644) = 1 - 0{,}6772 = 0{,}3228$

Con lo que la tabla para realizar el contraste de bondad de ajuste de la Chi-cuadrado quedaría:

| Intervalo | Frecuencia Obs. | Frec. Esperada | $(O_k - E_k)/\sqrt{E_k}$ |
|---|---|---|---|
| <97 | 5 | $35 \cdot 0{,}1093 = 3{,}8255$ | 0.6005 |
| [97 - 99) | 2 | $35 \cdot 0{,}0997 = 3{,}4895$ | -0.7974 |
| [99-101) | 3 | $35 \cdot 0{,}1430 = 5{,}0050$ | -0.8962 |
| [101-103) | 8 | $35 \cdot 0{,}1640 = 5{,}7400$ | 0.9433 |
| [103-105) | 4 | $35 \cdot 0{,}1612 = 5{,}6420$ | -0.6913 |
| >105 | 13 | $35 \cdot 0{,}3228 = 11{,}2980$ | 0.5064 |

$\sum_{k=1}^{6} \frac{(O_k - E_k)^2}{E_k} = 3{,}4238$ y este valor se compara con el de la $\chi^2_{K-r-1;0{,}05} \equiv \chi^2_{6-2-1;0{,}05} = 7{,}8$

3. **Se realiza un ensayo con $n_1 = 100$ probetas del procedimiento habitual y los resultados son $\overline{x}_1 = 99{,}4$ y $\widehat{s}_1^2 = 18{,}3$. Contrastar la igualdad de medias de los dos procedimientos y calcular al $p$-valor de dicho contraste para la muestra del segundo procedimiento que se detalla en el apartado 1 ($n_2 = 20$, $\overline{x}_2 = 105$) pero considerando en este caso que la varianza muestral corregida es $\widehat{s}_2^2 = 20{,}25$.**

En el apartado 1 teníamos para el segundo procedimiento: $n_2 = 20$, $\overline{x}_2 = 105$ y aquí nos piden usar como desconocida la sigma pues se proporciona la muestral. Queremos contrastar:

$H_0$: $\mu_1 = \mu_2$

$H_1$: $\mu_1 \neq \mu_2$

$\frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\widehat{s}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to t_{n_1+n_2-2}$

$\widehat{s}_R^2 = \frac{(n_1-1)\widehat{s}_1^2 + (n_2-1)\widehat{s}_2^2}{n_1+n_2-2} = \frac{(100-1)\cdot 18{,}3565 + (20-1)\cdot 20{,}25}{100+20-2} = 18{,}6614$.

Y particularizando $\frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\widehat{s}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ bajo la $H_0$ quedaría: $\frac{(99{,}4512 - 105) - (0)}{\sqrt{18{,}6614}\sqrt{\frac{1}{100} + \frac{1}{20}}} = -5{,}2436$.

La región de aceptación sería entre los valores $t_{n_1+n_2-2;0{,}025} \equiv t_{118;0{,}025}$ y $t_{n_1+n_2-2;075} \equiv t_{118;0{,}975}$, que se miran en las tablas de la $Z$ y son: -1.96 y 1.96. Con lo que se rechaza la hipótesis de igualdad de medias al estar $-5{,}2436$ fuera de este intervalo. El $p$-valor se calcula como $2 \cdot pr(Z < -5{,}2436)$ $=2 \cdot pr(Z > 5{,}2436)$ $=2 \cdot [1 - pr(Z < 5{,}2436)] = 0$

4. **Calcular un intervalo de confianza para el cociente de varianzas y realizar el contraste $H_0$: $\sigma_1^2 = \sigma_2^2$ frente a $H_1$: $\sigma_1^2 > \sigma_2^2$ .**

$\frac{\frac{\hat{s}_1^2}{\sigma_1^2}}{\frac{\hat{s}_2^2}{\sigma_2^2}} \to F_{n_1-1,n_2-1} \equiv F_{99,19}$

Para calcular el intervalo: $pr(F_a \leq \frac{\frac{\hat{s}_1^2}{\sigma_1^2}}{\frac{\hat{s}_2^2}{\sigma_2^2}} \leq F_b) = 0{,}95$, donde $F_a \equiv F_{99,19;\frac{\alpha}{2}=0{,}025} \simeq 1/1{,}85 = 0{,}5405$ y $F_b \equiv F_{99,19;0{,}975} = 2{,}22$.

$\frac{\hat{s}_2^2}{\hat{s}_1^2} = 20{,}25/18{,}3565$

$F_a \frac{\hat{s}_2^2}{\hat{s}_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_b \frac{\hat{s}_2^2}{\hat{s}_1^2} \to 0{,}5405 \cdot 1{,}1032 \leq \frac{\sigma_2^2}{\sigma_1^2} \leq 2{,}22 \cdot 1{,}1032$ y el intervalo quedaría: $0{,}5963 \leq \frac{\sigma_2^2}{\sigma_1^2} \leq 2{,}4491$.
No podemos rechazar la hipótesis nula de igualdad de varianzas, en el caso de un bilateral. Para el que nos piden hay que plantear el contraste unilateral, y buscar en tablas el valor de la $F_{99,19;\alpha=0{,}05} = 1{,}77$.
Y como $\frac{\hat{s}_1^2}{\hat{s}_2^2} = 18{,}3565/20{,}25 < 1{,}77$ cae en la región de aceptación de $H_0$ y no podemos rechazar la hipótesis nula de igualdad de varianzas.

# EXÁMENES

## Curso 2015/16

**Cuestiones** (45 minutos – 5 puntos)

**Cuestión 1:**

Un alumno tiene, en su escritorio, tres botes con bolígrafos de diferentes colores, con la siguiente distribución:

- Bote A: 5 negros y 2 rojos.
- Bote B: 10 negros y 3 verdes.
- Bote C: 2 negros, 2 rojos y 2 verdes.

Elige un bote al azar, y coge dos bolígrafos de ese bote. Observa que son negros los dos bolígrafos escogidos. Calcular la probabilidad de que no haya elegido el bote A.

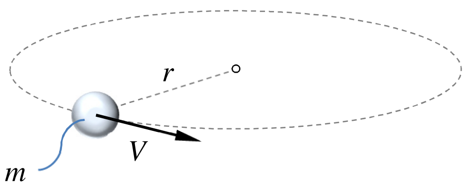
**Cuestión 2:**

Un grupo de 250 estudiantes universitarios están planificando el viaje fin de curso 2015/16 a Cracovia. La media de edad es de 20 años, con una desviación típica de 1.2 años.
¿Entre qué rango de edades podemos asegurar que se encuentra el 80% de los participantes?

**Cuestión 3:**

Una partícula de masa $m$ sometida a un movimiento circular uniforme experimenta una fuerza centrípeta $Y$, calculada como: $Y = m \cdot V^2/r$
donde $r$ es el radio, $m$ es la masa y $V$ es la velocidad.

El valor de la velocidad es una variable aleatoria de distribución uniforme comprendida en el intervalo: $-5 \leq v \leq 5$.
Calcular la función de densidad y la esperanza de la fuerza $Y$.



**Cuestión 1:**

Nos piden calcular la probabilidad de que no haya elegido el bote A, sabiendo que los dos bolígrafos escogidos son de color negro.
Definimos los siguientes sucesos:

A: "Elijo el bote A" B: "Elijo el bote B" C: "Elijo el bote C"
N: "Elijo un bolígrafo negro"

En base a las anteriores definiciones, nos piden calcular: $P(\bar{A}|NN)$

De los datos del enunciado, sabemos:

$$P(NN|A) = \frac{5}{7}\cdot\frac{4}{6} \qquad P(NN|B) = \frac{10}{13}\cdot\frac{9}{12} \qquad P(NN|C) = \frac{2}{6}\cdot\frac{1}{5}$$

$$P(A) = \frac{1}{3} \qquad P(B) = \frac{1}{3} \qquad P(C) = \frac{1}{3}$$

En primer lugar, vamos a calcular $P(A|NN)$, pues sabemos que: $P(\bar{A}|NN) = 1 - P(A|NN)$

Empleando Bayes:

$$P(A|NN) = \frac{P(NN|A)\cdot P(A)}{P(NN|A)\cdot P(A) + P(NN|B)\cdot P(B) + P(NN|C)\cdot P(C)}$$

$$= \frac{\left(\frac{5}{7}\cdot\frac{4}{6}\right)\cdot\left(\frac{1}{3}\right)}{\left(\frac{5}{7}\cdot\frac{4}{6}\right)\cdot\left(\frac{1}{3}\right) + \left(\frac{10}{13}\cdot\frac{9}{12}\right)\cdot\left(\frac{1}{3}\right) + \left(\frac{2}{6}\cdot\frac{1}{5}\right)\cdot\left(\frac{1}{3}\right)} = 0.423$$

Por tanto: $P(\bar{A}|NN) = 1 - P(A|NN) = 0.577$

**Cuestión 2:**

Empleando Chebyshev:

$$P(|x_i - \bar{x}| \le k\cdot s) > 1 - \frac{1}{k^2}$$

Del enunciado sabemos: $\bar{x} = 20$ y $s = 1.2$.
Si queremos calcular $k$ para que se encuentre más del 80% de observaciones dentro del intervalo $\bar{x} \pm k\cdot s$, tenemos que $k$=2.24. (Ya que: $1 - 1/k^2 = 0.80$)

Por tanto, el rango de edad es:

$$\bar{x} \pm k\cdot s = 20 \pm 1.2\cdot 2.24 = (17.3 \quad ; \quad 22.7) \text{ años}$$

**Cuestión 3:**

Cálculo del valor de $a$:

$$\int_{-\infty}^{+\infty} f_V(v)dv = 1 \quad \rightarrow \quad \int_{-5}^{+5} a\, dx = 10\, a = 1 \quad \rightarrow \quad a = \frac{1}{10}$$

Cálculo de la función de distribución de *V*:

$$F_V(v) = \int_{-\infty}^{v} f_V(v)\, dv = \int_{-5}^{v} \frac{1}{10} dv = \frac{v}{10} + \frac{1}{2} = \frac{1}{10}(v+5)\ , \qquad -5 \le v \le 5$$

Cálculo de la función de distribución de *F*:

$$F_Y(y) = P(Y \le y) = P\left(m \cdot \frac{V^2}{r} \le y\right) = P\left(V^2 \le y \cdot \frac{r}{m}\right)$$

$$= P\left(-\sqrt{y \cdot \frac{r}{m}} \le V \le \sqrt{y \cdot \frac{r}{m}}\right) = P\left(V \le \sqrt{y \cdot \frac{r}{m}}\right) - P\left(V \le -\sqrt{y \cdot \frac{r}{m}}\right)$$

$$= F_V\left(\sqrt{y \cdot \frac{r}{m}}\right) - F_V\left(-\sqrt{y \cdot \frac{r}{m}}\right) = \left(\frac{\sqrt{y \cdot \frac{r}{m}}}{10} + \frac{1}{2}\right) - \left(\frac{-\sqrt{y \cdot \frac{r}{m}}}{10} + \frac{1}{2}\right)$$

$$= \frac{1}{5}\sqrt{y \cdot \frac{r}{m}} \qquad 0 \le y \le \frac{5^2 \cdot m}{r}$$

Cálculo de la función de densidad de *F*:

$$f_Y(y) = \frac{d\, F_Y(y)}{d\, y} = \frac{1}{10}\sqrt{\frac{r}{m \cdot y}} \qquad 0 \le y \le \frac{5^2 \cdot m}{r}$$

Cálculo de la esperanza:

$$E[Y] = E\left[m \cdot \frac{V^2}{r}\right] = \frac{m}{r} E[V^2] = \frac{m}{r}\int_{-\infty}^{+\infty} v^2 \cdot f_V(v)\, dv = \frac{m}{r}\int_{-5}^{+5} v^2 \cdot \frac{1}{10}\, dv = \frac{m}{r}\frac{25}{3}$$

(Manera alternativa del cálculo de la esperanza:)

$$E[Y] = \int_{-\infty}^{+\infty} y \cdot f_Y(y)\, dy = \int_{0}^{\frac{5^2 \cdot m}{r}} y \cdot \frac{1}{10}\sqrt{\frac{r}{m \cdot y}} dy = \int_{0}^{\frac{5^2 \cdot m}{r}} \frac{1}{10}\sqrt{\frac{r \cdot y}{m}} dy$$

$$= \frac{1}{10}\sqrt{\frac{r}{m}} \int_{0}^{\frac{5^2 \cdot m}{r}} \sqrt{y} dy = \frac{25}{3}\frac{m}{r}$$

**Problema** (5 puntos - 45 minutos)

El tiempo (en miles de horas) X que dura una batería eléctrica es una variable aleatoria con la siguiente función de densidad

$$f_X(x) = \frac{x}{a^2} \exp\{-\frac{x^2}{2a^2}\}, \qquad x \geq 0$$

y tiene media

$$E[X] = a\sqrt{\frac{\pi}{2}}.$$

Según el fabricante, la duración media de la batería es de 20 mil horas.

1. (a) ¿Cuál es la probabilidad de que una batería dure más de dos años? (Un año tiene 8760 horas) (**1 punto**)

   (b) Se llama fiabilidad del sistema a la función $R(x) = P(X \geq x)$. Obtén la función $R(x)$ y dibújala, proporcionando los valores correspondientes a uno, dos y tres años. (**1 punto**)

2. Un equipo de comunicaciones tiene dos componentes, transmisor y receptor. Cada uno tiene una batería del tipo anterior cuyos tiempos de funcionamiento son variables aleatorias independientes. Para que el equipo funcione deben hacerlo las dos baterías.

   (a) ¿Cuál es la probabilidad de que el equipo dure más de dos años? (**0.5 puntos**)

   (b) Obtener la fiabilidad del equipo, es decir la probabilidad de que el equipo de comunicaciones dure más de $x$ miles de horas. (Ayuda: Sea $A$ el suceso "la batería del transmisor dura más de $x$ miles de horas" y $B$ el suceso "la batería del receptor dura más de $x$ miles de horas", obtén $P(A \cap B)$). (**1 punto**)

3. ¿Cuál será la duración media del equipo de comunicaciones? (**1.5 puntos**).

Nota: $\displaystyle\int_0^\infty \frac{x^2}{a^2} \exp\{-\frac{x^2}{2a^2}\} dx = a\sqrt{\frac{\pi}{2}}$

1. (a) Primero calculamos el valor de $a$ (medido en miles de horas):

$$E[X] = 20 = a\sqrt{\frac{\pi}{2}} \to a = 20\sqrt{\frac{2}{\pi}} \approx 16$$

Primero calculamos la función de distribución:

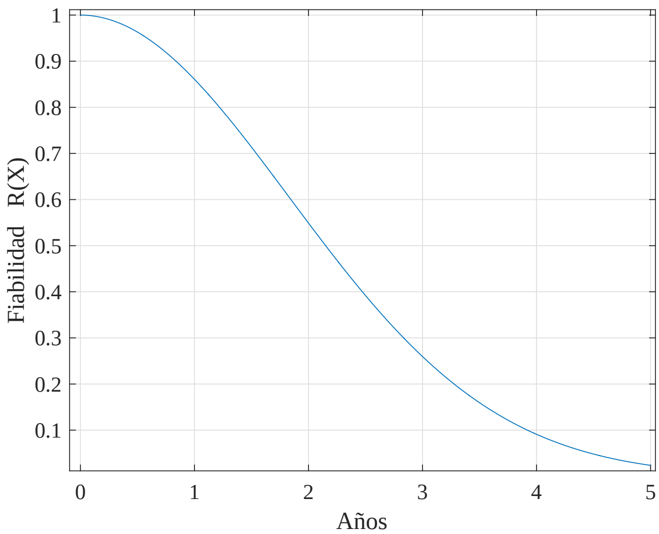
$$\begin{aligned} F_X(x) &= P(X \leq x) = \int_0^{\infty} f_X(x)dx \\ &= \int_0^{\infty} \frac{x}{a^2}\exp\{-\frac{x^2}{2a^2}\}dx = 1 - \exp\{\frac{-x^2}{2a^2}\} \quad x \geq 0 \end{aligned}$$

$$P(X \geq x) = \exp\{\frac{-x^2}{2a^2}\} \quad x \geq 0$$

$$P(X \geq 8{,}76) = \exp\{\frac{-8{,}76^2}{2 \cdot 16^2}\} = 0{,}86$$

(b) Calculamos la Fiabilidad:

$$R(x) = P(X \geq x) = 1 - F_X(x) = \exp\{\frac{-x^2}{2a^2}\} \quad x \geq 0$$




Evaluando:

$$R(1 \cdot 8{,}76) = 0{,}8608 \quad ; \quad R(2 \cdot 8{,}76) = 0{,}5491 \quad ; \quad R(3 \cdot 8{,}76) = 0{,}2595$$

2.

$$R_{equipo}(x) = P(X \geq x)P(X \geq x) = R(x)R(x) = \exp\{\frac{-x^2}{a^2}\} \quad x \geq 0$$

$$\mathrm{P}(\text{dos años}) = R_{equipo}(2 \cdot 8{,}76) = 0{,}549^2 = 0{,}3352$$

3.

$$f_{equipo}(x) = \frac{x}{2a^2} \exp\{-\frac{x^2}{a^2}\} = \frac{x}{(\frac{a}{\sqrt{2}})^2} \exp\{-\frac{x^2}{2(\frac{a}{\sqrt{2}})^2}\}, \qquad x \geq 0$$

$$E[X_{equipo}] = \frac{a}{\sqrt{2}}\sqrt{\frac{\pi}{2}} = \frac{20}{\sqrt{2}} = 14{,}14 \text{ mil horas}$$

**Cuestiones** (45 minutos)

1. El tiempo que una persona tarda en llegar desde su casa al trabajo es una variable aleatoria normal de media 26 minutos y desviación típica 4 minutos. Si todos los días sale de casa 30 minutos antes de la hora de entrada al trabajo, calcular la probabilidad de que llegue tarde más de un día a la semana (la semana laboral de esta persona tiene 5 días).

2. El número de goles que marca un equipo de fútbol es una variable aleatoria que sigue una distribución de Poisson de media 2 goles por partido. Si en un determinado partido el equipo marcó 3 goles, calcular la probabilidad de que se marcaran uno o dos goles en la segunda parte. Razonar la respuesta.

3. Sean $X$ e $Y$ dos variables aleatorias con función de densidad conjunta

$$f_{XY}(x, y) = x + y, \quad x, y \in [0, 1].$$

Sean $U$ y $V$ otras dos variables aleatorias que verifican $U = X+Y$, $V = 2X-Y$. Calcular el coeficiente de correlación entre $U$ y $V$.

Nota: para agilizar los cálculos, considerar que $\text{var}(X) = \text{var}(Y) = 11/144$.

**Problema** (45 minutos)

Los tiempos de vida (en horas) de dos componentes electrónicos A y B son variables aleatorias que denominaremos respectivamente $X$ e $Y$ y cuya función de densidad conjunta es:

$$f_{XY}(x, y) = \lambda_A \lambda_B \exp^{-(\lambda_A x + \lambda_B y)}, \quad \text{con } x \geq 0, y \geq 0, \lambda_A = \frac{1}{10}, \lambda_B = \frac{1}{15}.$$

1. Calcule las funciones de densidad marginales $f_X(x)$ y $f_Y(y)$, la función de distribución conjunta $F_{XY}(x, y)$ y las funciones de distribución marginales $F_X(x)$ y $F_Y(y)$ ¿Son variables aleatorias independientes? ¿Cuál es la probabilidad de que el componente A dure menos que el componente B?

2. Se definen $U_S$ y $U_I$ como el máximo y el mínimo de las duraciones de los componentes A y B respectivamente. Calcule $P(U_S > 30)$ y $P(U_I < 15)$.

3. Se diseña un sistema con $n$ componentes tipo A y $n$ componentes tipo B, de tal forma que el tiempo de vida del sistema se calcula como $T = \sum_{i=1}^{n} X_i + \sum_{j=1}^{n} Y_j$. Calcule el número de componentes de cada tipo (valor $n$) que debe tener el sistema para que la probabilidad de que el sistema dure menos de 10000 horas sea inferior a 0.01.

   Nota. Considere que los tiempos de vida de los componentes son variables aleatorias independientes.

**Solución de las Cuestiones**

- Cuestión 1

$X$: "tiempo en llegar al trabajo"

$Y$: "número de días que llega tarde en una semana"

$$P(\text{“llegar un día tarde”}) = P(X \geq 30 | X \to N(26,4)) =$$
$$= 1 - P\left(Z \leq \frac{30-26}{4}\Big| Z \to N(0,1)\right) = 1 - 0{,}8413 = 0{,}1587$$

$$Y \to B(5; 0{,}16) \Rightarrow P(Y \geq 2) = \binom{5}{2} 0{,}16^2 0{,}84^3 + \binom{5}{3} 0{,}16^3 0{,}84^2$$
$$+ \binom{5}{4} 0{,}16^4 0{,}84^1 + \binom{5}{5} 0{,}16^5 0{,}84^0 = 0{,}18$$

- Cuestión 2

$X$: "número de goles en la segunda parte"

$Y$: "número de goles en la primera parte"

$Z$: "número de goles en todo el partido"

donde

$Z \to Poisson(\lambda)$, $\quad X, Y \to Poisson(\lambda/2)$. Entonces

$$P(X = x|Z = 3) = \frac{P(X = x, Z = 3)}{P(Z = 3)} = \frac{P(X = x, Y = 3-x)}{P(Z = 3)} = \frac{P(X = x)P(Y = 3-x)}{P(Z = 3)}$$

$$= \frac{e^{-\lambda/2}\frac{(\lambda/2)^x}{x!} \cdot e^{-\lambda/2}\frac{(\lambda/2)^{3-x}}{(3-x)!}}{e^{-\lambda}\frac{\lambda^3}{3!}} = \frac{3!}{x!(3-x)!}\frac{1}{8} = \frac{1}{8}\binom{3}{x}, \quad x = 0, 1, 2, 3$$

Por tanto

$$P(X = 1|Z = 3) + P(X = 2|Z = 3) = \frac{1}{8}\binom{3}{1} + \frac{1}{8}\binom{3}{2} = \frac{3}{4}$$

- Cuestión 3

Primero se calcula la covarianza entre $X$ e $Y$:

$$E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy \cdot f_{XY}(x,y)dxdy = \int_0^1\int_0^1 (x^2y + xy^2)dxdy = \frac{1}{3}$$

$$cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{3} - \left(\frac{7}{12}\right)^2 = \frac{-1}{144}$$

La esperanza de $X$ y de $Y$ se calculan:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_0^1 (x+y)dy = x + \frac{1}{2}, \quad x \in [0,1]$$

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^1 (x^2 + \frac{x}{2}) dx = \frac{7}{12}$$

Aunque lo dan en el enunciado, vamos a calcular la varianza de $X$:

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^1 (x^3 + \frac{x^2}{2}) dx = \frac{5}{12}$$

$$var(X) = E(X^2) - E(X)^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$$

La esperanza y varianza de $Y$ se calculan de manera análoga.

Por último se calcula la varianza de $U$, la de $V$ y su covarianza:

$$\begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \Rightarrow var\left(\begin{bmatrix} U \\ V \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} var\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^T$$

$$var\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) = \begin{bmatrix} var(X) & cov(X,Y) \\ cov(X,Y) & var(Y) \end{bmatrix} = \frac{1}{144}\begin{bmatrix} 11 & -1 \\ -1 & 11 \end{bmatrix}$$

$$var\left(\begin{bmatrix} U \\ V \end{bmatrix}\right) = \begin{bmatrix} var(U) & cov(U,V) \\ cov(U,V) & var(V) \end{bmatrix} = \frac{1}{144}\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 11 & -1 \\ -1 & 11 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^T = \frac{1}{144}\begin{bmatrix} 20 & 10 \\ 10 & 59 \end{bmatrix}$$

Finalmente, el coeficiente de correlación es:

$$\rho_{UV} = \frac{cov(U,V)}{\sqrt{var(U)var(V)}} = \frac{10/144}{\sqrt{20/144 \cdot 59/144}} = \frac{10}{\sqrt{20 \cdot 59}} = 0{,}29$$

**Solución del Problema**

**1.**
a) Funciones de densidad marginales:
$f_X(x) = \int_0^\infty \lambda_A \lambda_B \exp^{-(\lambda_A x + \lambda_B y)}\, dy = \lambda_A \exp(-\lambda_A x), x \geq 0$
$f_Y(y) = \int_0^\infty \lambda_A \lambda_B \exp^{-(\lambda_A x + \lambda_B y)}\, dx = \lambda_B \exp(-\lambda_B y), y \geq 0$
Son variables aleatorias independientes
$\Longrightarrow f_{XY}(x, y) = f_X(x) \times f_Y(y). = \lambda_A \exp(-\lambda_A x) \lambda_B \exp(-\lambda_B y).$
b) Función de distribución conjunta:

$$F_{XY}(x, y) = \int_0^x \int_0^y \lambda_A \lambda_B \exp^{-(\lambda_A x + \lambda_B y)}\, dy dx = (1 - \exp^{-\lambda_A x})(1 - \exp^{-\lambda_B y}), \quad x \geq 0, \quad y \geq 0$$

c) Funciones de distrirbución marginales:

$F_X(x) = P(X \leq x, Y \leq \infty) = F_{XY}(x, \infty) = (1 - \exp^{-\lambda_A x}), \qquad x \geq 0$
O tambien: $F_X(x) = \int_0^x \lambda_A \exp(-\lambda_A x) dx = (1 - \exp^{-\lambda_A x}), \qquad x \geq 0$
$F_Y(y) = P(X \leq \infty, Y \leq y) = F_{XY}(\infty, y) = (1 - \exp^{-\lambda_B y}), \qquad y \geq 0.$
O tambien: $F_Y(y) = \int_0^y \lambda_B \exp(-\lambda_B y) dy = (1 - \exp^{-\lambda_B y}), \qquad y \geq 0.$

Se ve que $F_{XY}(x, y) = F_X(x) \times F_Y(y)$ al ser $X$ e $Y$ son independientes.
d) $P(X < Y) = \iint_{X<Y} \lambda_A \exp(-\lambda_A x) \lambda_B \exp(-\lambda_B y) dx dy =$

$$= \int_0^\infty \int_0^y \lambda_A \exp(-\lambda_A x) \lambda_B \exp(-\lambda_B y) dx dy =$$
$$= \int_0^\infty \lambda_B \exp(-\lambda_B y) \cdot [\int_0^y \lambda_A \exp(-\lambda_A x) dx] dy =$$
$$= \int_0^\infty \lambda_B \exp(-\lambda_B y)[1 - \exp(-\lambda_A y)] dy =$$
$$= \int_0^\infty \lambda_B \exp(-\lambda_B y) dy - \int_0^\infty \lambda_B \exp[-(\lambda_A + \lambda_B) y] dy =$$
$$= 1 - \frac{\lambda_B}{\lambda_A + \lambda_B} \int_0^\infty (\lambda_B + \lambda_A) \exp[-(\lambda_A + \lambda_B) y] dy =$$
$$= 1 - \frac{\lambda_B}{\lambda_A + \lambda_B} = \frac{\lambda_A}{\lambda_A + \lambda_B} = \frac{\frac{1}{10}}{\frac{1}{10} + \frac{1}{15}} = 0{,}6$$

**2.**
a) $U_S = Max(X, Y)$

$P(U_S > 30) = 1 - P(U_S \leq 30) = 1 - [P(X \leq 30 \cdot Y \leq 30)] = 1 - P(X \leq 30) \cdot P(Y \leq 30) =$
$= 1 - F_X(30) \cdot F_Y(30) = 1 - (1 - \exp^{-\lambda_A 30}) \cdot (1 - \exp^{-\lambda_B 30}) =$
$= 1 - (1 - \exp^{-3})(1 - \exp^{-2}) = 0{,}1783.$

b) $U_I = Min(X, Y)$

$P(U_I < 15) = 1 - P(U_I \geq 15) = 1 - [P(X \geq 15.Y \geq 15)] = 1 - (\exp^{-\lambda_A 15}) \cdot (\exp^{-\lambda_B 15}) =$
$=1 - \exp^{-1,5}\exp^{-1} = 0{,}9179.$

**3.**
$P(T < 10000) < 0{,}01$
$T = \sum_{i=1}^{n} X_i + \sum_{j=1}^{n} Y_j. \Longrightarrow$ La distribución de la variable aleatoria T puede considerarse asintóticamente Normal.
$E[X_i] = \frac{1}{\lambda_A} = 10 \qquad Var[X_i] = \frac{1}{\lambda_A^2} = 100$ ($X_i$ es exponencial)
$E[Y_i] = \frac{1}{\lambda_B} = 15 \qquad Var[Y_i] = \frac{1}{\lambda_B^2} = 225$ ($Y_i$ es exponencial)
Además son todas variables aleatorias independientes.
$E[T] = E[\sum_{i=1}^{n} X_i + \sum_{j=1}^{n} Y_j] = n \times E[X_i] + n \times E[Y_i] = n \times 10 + n \times 15 = 25n$
$Var[T] = Var[\sum_{i=1}^{n} X_i + \sum_{j=1}^{n} Y_j] = n \times Var[X_i] + n \times Var[Y_i] = n \times 100 + n \times 225 = 325n$
$P(T < 10000) < 0{,}01 \Longrightarrow P(T < 10000 | T \rightsquigarrow N(25n, 325n)) < 0{,}01$

$P(\frac{T-25n}{\sqrt{325n}} < \frac{10000-25n}{\sqrt{325n}}) < 0{,}01 \Longrightarrow P(Z < z_n) < 0{,}01 \Longrightarrow$Mirando en las tablas$\Longrightarrow$ $z_n = -2{,}33$

$$\frac{10000-25n}{\sqrt{325n}} = -2{,}33 \Longrightarrow n = 435{,}14 \Longrightarrow n \approx 436.$$

# Estadística

18 de diciembre de 2015

**Cuestiones**

(45 minutos, 5 puntos)

1. Se toma una muestra aleatoria simple $X_1, X_2, ..., X_n$ de la variable aleatoria "capacidad de un condensador" que sigue una distribución $N(\mu, \sigma)$. Se consideran los siguientes estimadores para $\sigma^2$ :

$$\hat{s}_1^2 = \frac{\sum(X_i - \bar{X})^2}{n}, \hat{s}_2^2 = \frac{\sum(X_i - \bar{X})^2}{n+1} \text{ y } \hat{s}_3^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

   Obtenga el estimador para $\sigma^2$ con menor Error Cuadrático Medio (ECM) entre los tres descritos anteriormente.(Nota $E[\chi_n^2] = n$ y $Var[\chi_n^2] = 2n$)

2. Se toma una muestra del peso en kg de 30 toros Holstein (muestra A) y una muestra del peso en kg de 28 vacas Holstein (muestra B). Las medias muestrales y las varianzas muestrales corregidas para cada una de las dos muestras son respectivamente $\bar{x}_A = 1100$ $kg$, $\hat{s}_A^2 = 15\ 10^3\ kg^2$ y $\bar{x}_B = 720\ kg$, $\hat{s}_B^2 = 10^4\ kg^2$ .¿Es el peso medio de los toros significativamente superior al de las vacas? Calcule el nivel crítico (p-valor) para el contraste anterior. Realice el contraste para la igualdad de las varianzas de ambas muestras (toros y vacas). (Nota: $\alpha = 0.05$)

3. Una empresa eléctrica dispone de un parque eólico con 500 aerogeneradores. Históricamente se han observado de media un 5% de averías al año. Durante el último año se han registrado 40 aerogeneradores averiados, ¿se ha producido un aumento de la proporción de aerogeneradores que se averían anualmente? Calcule la probabilidad de error tipo II del contraste anterior si la proporción de averias al año es 8%.(Nota: Aceptamos independencia y que un generador no se avería más de una vez al año). ($\alpha = 0.05$)

Estadística 18 de diciembre de 2015

## Cuestiones

1. Se toma una muestra aleatoria simple $X_1, X_{2,...}, X_n$ de la variable aleatoria "capacidad de un condensador" que sigue una distribución $N(\mu, \sigma)$.Se consideran los siguientes estimadores para $\sigma^2$ :

$$\hat{s}_1^2 = \frac{\sum(X_i - \bar{X})^2}{n}, \hat{s}_2^2 = \frac{\sum(X_i - \bar{X})^2}{n+1} \text{ y } \hat{s}_3^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

Obtenga el estimador con menos Error Cuadrático Medio (ECM) para $\sigma^2$entre los tres descritos anteriormente.(Nota $E[\chi_n^2] = n$ y $Var[\chi_n^2] = 2n$)

**Solución:**

El error cuadrático medio de un estimador $\theta$ ($\hat{\theta}$) es: $ECM(\hat{\theta}) = Var[\hat{\theta}] + (sesgo(\hat{\theta}))^2$.

Por lo que para cada uno de los tres estimadores hay que calcular la esperanza (y obtener su sesgo) y obtener su varianza. Posteriormente seleccionar el que tenga menor ECM.

Sabiendo que:

$$\frac{\sum(X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2 \longrightarrow \frac{n\hat{s}_1^2}{\sigma^2} = \frac{(n+1)\hat{s}_2^2}{\sigma^2} = \frac{(n-1)\hat{s}_3^2}{\sigma^2} \sim \chi_{n-1}^2.$$

**Esperanzas**

Por lo que la esperanzas de cada uno de esos tres estimadores es respectivamente:

$E[\hat{s}_1^2] = E[\frac{\sigma^2}{n}\chi_{n-1}^2] = \frac{n-1}{n}\ \sigma^2$, que es segado, con sesgo $-\frac{\sigma^2}{n}$

$E[\hat{s}_2^2] = E[\frac{\sigma^2}{n+1}\chi_{n-1}^2] = \frac{n-1}{n+1}\ \sigma^2$, que es segado, con sesgo $-\frac{2\sigma^2}{n+1}$

$E[\hat{s}_3^2] = E[\frac{\sigma^2}{n-1}\chi_{n-1}^2] = \sigma^2$, que es centrado para $\sigma^2$, por lo que su sesgo vale cero.

Por lo tanto, $\hat{s}_3^2$ es centrado y $\hat{s}_1^2$ y $\hat{s}_2^2$ son sesgados.

**Varianzas**

$Var[\hat{s}_1^2] = Var[\frac{\sigma^2}{n}\chi_{n-1}^2] = \frac{2(n-1)}{n^2}\ \sigma^4$

$Var[\hat{s}_2^2] = Var[\frac{\sigma^2}{n+1}\chi_{n-1}^2] = \frac{2(n-1)}{(n+1)^2}\ \sigma^4$

$Var[\hat{s}_3^2] = Var[\frac{\sigma^2}{n-1}\chi_{n-1}^2] = \frac{2(n-1)}{(n-1)^2}\ \sigma^4 = \frac{2\sigma^4}{n-1}$

Calculado el sesgo y la varianza para cada uno de los tres estimadores, ontenemos el error cuadrático medio según $ECM(\hat{\theta}) = Var[\hat{\theta}] + (sesgo(\hat{\theta}))^2$, y se obtiene:

$ECM(\hat{s}_1^2) = Var[\hat{s}_1^2] + (sesgo(\hat{s}_1^2))^2 = \frac{2(n-1)}{n^2}\ \sigma^4 + \frac{\sigma^4}{n^2}$

$ECM(\hat{s}_2^2) = Var[\hat{s}_2^2] + (sesgo(\hat{s}_2^2))^2 = \frac{2(n-1)}{(n+1)^2}\ \sigma^4 + \frac{4\sigma^4}{(n+1)^2}$

$ECM(\hat{s}_3^2) = Var[\hat{s}_3^2] + (sesgo(\hat{s}_3^2))^2 = \frac{2\sigma^4}{n-1} + 0$

El estimador que tiene menor $ECM$ es el $\hat{s}_2^2$.

2. Se toma una muestra del peso en kg de 30 toros Holstein (muestra A) y una muestra del peso en kg de 28 vacas Holstein (muestra B), las medias muestrales y las varianzas muestrales corregidas para cada una de las dos muestras son respectivamente $\bar{x}_A = 1100$ $kg$ $\hat{s}_A^2 = 15\ 10^3\ kg^2$ y $\bar{x}_B = 720\ kg$ $\hat{s}_B^2 = 10^4\ kg^2$ .¿Es el peso medio de los toros significativamente superior al de las vacas? Calcule el nivel crítico (p-valor) para el contraste realizado.Realice el contraste para la igualdad de las varianzas de ambas muestras (toros y vacas). (Nota: $\alpha = 0.05$)

**Solución:**

(a) Para concluir que el peso medio de los toros es significativamente superior al de las vacas se realiza el siguiente contraste unilateral:

$$H_0 : \mu_T = \mu_V$$

$$H_1 : \mu_T > \mu_V$$

Así $\bar{x}_T - \bar{x}_{V_T} \sim N(\mu_T - \mu_V, \sigma^2(\frac{1}{n_T} + \frac{1}{n_V})$, se obtiene:

$$\frac{(\bar{x}_T - \bar{x}_V) - 0}{\hat{s}_R\sqrt{\frac{1}{n_T} + \frac{1}{n_V}}} \sim t_{n_T+n_V-2},$$

siendo $\hat{s}_R^2 = \dfrac{(n_T-1)\hat{s}_T^2 + (n_V-1)\hat{s}_V^2}{n_T+n_V-2}$, y teniendo en cuenta los datos de los enunciados $\bar{x}_T = 1100\ kg$ $\hat{s}_T^2 = 15\ 10^3\ kg^2$ y $\bar{x}_V = 720\ kg$ $\hat{s}_V^2 = 10^4\ kg^2$, se obtiene $\hat{s}_R^2 = 1,2589x10^4$ y $\hat{s}_R = 112,202$. El valor de $t_0$ es:

$$t_0 = \frac{(\bar{x}_T - \bar{x}_V) - 0}{\hat{s}_R\sqrt{\frac{1}{n_T} + \frac{1}{n_V}}} = \frac{1100 - 720}{112.2\sqrt{\frac{1}{30} + \frac{1}{28}}} = 12.89,$$

que se compara con la $t$ con 56 grados de libertad ($\alpha = 0.05$), $t_{56;0.05} = 1.671$.
Como $12.89 > 1.671$, se rechaza la $H_0$ y se concluye que el peso medio de los toros es significativamente superior al de las vacas.
Para obtener el p-valor del contraste se calcula:
$p - valor = P(t_{56} > 12.89) \simeq 0$.

(b) A continuación se realiza el contraste para la igualdad de las varianzas de ambas muestras (toros y vacas), así:

$$H_0 : \sigma_T^2 = \sigma_V^2$$

$$H_1 : \sigma_T^2 \neq \sigma_V^2$$

Para ello se tiene en cuenta:

$\frac{(n_T-1)\hat{s}_T^2}{\sigma_T^2} \sim \chi^2_{n_T-1}$ y $\frac{(n_V-1)\hat{s}_v^2}{\sigma_v^2} \sim \chi^2_{n_V-1}$ y se obtiene:

$$F_0 = \frac{\hat{s}_T^2}{\hat{s}_V^2} \sim F_{n_T-1,n_V-1}$$

Utilizando los datos se obtiene un valor $F_0 = 1.5$,que se compara con los dos valores de las tablas de la $F$, $F_a$ y $F_b$. Siendo $F_b = F_{29,27;0.025} = 2.13$, y $F_a = F_{29,27;0.975} = \frac{1}{F_{27,29;0.025}} = \frac{1}{2.13} \simeq 0.47$.

Como $1.5 \in (0.47, 2.13)$, no se rechaza la $H_0 : \sigma_T^2 = \sigma_V^2$.

3. Una empresa eléctrica dispone de un parque eólico con 500 aerogeneradores. Históricamente se han observado de media un 5% de averías al año. Durante el último año se han registrado 40 aerogeneradores averiados, ¿se ha producido un aumento de la proporción de aerogeneradores que se averían anualmente? Calcule la probabilidad de error tipo II del contraste anterior si la proporción de averias al año es 8%.(Nota: Aceptamos independencia y que un generador no se avería más de una vez al año). ($\alpha = 0.05$)

   **Solución:**

   (a) Para decidir si se ha producido un aumento de la proporción de aerogeneradores que se averían anualmente, se realiza el siguiente contraste:

   $$H_0 : p = 0.05$$

   $$H_1 : p > 0.05$$

   sabiendo que $\hat{p} \sim N(p, \sqrt{\frac{pq}{n}})$,y con los datos obtenemos $\hat{p} = \frac{40}{500} = 0.08.Así:$

   $$z_0 = \frac{0.08 - 0.05}{\sqrt{\frac{0.05\text{x}0.95}{500}}} = 3.0779,$$

   Y comparando $z_0$ con el valor de $z$ de las tablas con $\alpha = 0.05$, se obtiene $z_\alpha = 1.64$.Como $3.08 > 1.64$, se rechaza la $H_0 : p = 0.05$ y se concluye que se ha producido un aumento en la proporción de aerogeneradores defectuoos.

   (b) Para calcular la probabilidad de error tipo II del contraste anterior si la proporción de averias al año es 8%,se calcula el valor de $\hat{p}$ que nos marca la región de aceptación y de rechazo. La región de aceptación sería:

   $$\hat{p} > 0.05 + 1.64\sqrt{\frac{0.05\text{x}0.95}{500}} = 0.066.$$

   Por lo tanto la probabilidad de errot tipo II es:

$\beta = P(\text{No rechazar } H_0|p = 0.08) = P(\ \hat{p} < 0.06|\ \hat{p} \sim N(0.08, \sqrt{\frac{0.08\text{x}0.92}{500}}) =$
$P(z < \frac{0.066 - 0.08}{\sqrt{\frac{0.08x0.92}{500}}}) = P(z < -1.1511) \simeq 0.12.$

# Estadística

18 de diciembre de 2015

**Problema**

(45 minutos, 5 puntos)

Un localizador de posición de drones comete un error de distancia en $cm$ que sigue una distribución exponencial parametrizada como sigue:

$$f_X(x) = \frac{1}{\lambda} exp(-\frac{x}{\lambda}), \quad x \geq 0,\ \lambda > 0$$

Para una muestra de 100 observaciones, el promedio de los 84 primeros errores ha sido $57,81$ cm, para el resto sólo se sabe que han sido mayores que 170 cm.

1. Estimar por máxima verosimilitud el parámetro $\lambda$ de la distribución exponencial.

   **Solución**

   La función de verosimilitud es:

   $l = (1/\lambda)^{84} \exp(-(1/\lambda) \sum_{i=1}^{84} x_i) \exp(-16 \times 170/\lambda);$

   $l = \log(l) = -84 \log(\lambda) - (1/\lambda) \sum_{i=1}^{84} x_i - 16 \times 170/\lambda;$

   $\frac{\partial L}{\partial \lambda} = -\frac{84}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^{84} x_i + \frac{16 \times 170}{\lambda^2} = 0;$

   $\widehat{\lambda} = \frac{\sum_{i=1}^{84} x_i + 16 \times 170}{84} = \frac{84 \times 57{,}81 + 16x170}{84} = 90,19.$

2. Las frecuencias de las observaciones se reparten de acuerdo con la tabla siguiente. Contrastar que los errores de distancia en cm siguen efectivamente una distribución exponencial.($\alpha = 0.05$)

| Intervalo | $< 15$ | $15-30$ | $30-45$ | $45-65$ | $65-90$ | $90-130$ | $130-170$ | $> 170$ |
|---|---|---|---|---|---|---|---|---|
| Fr. Observada | 15 | 14 | 12 | 12 | 9 | 12 | 10 | 16 |

**Solución**

La tabla que incopora las frecuencias esperadas es la siguient:

| Intervalo | $< 15$ | $15-30$ | $30-45$ | $45-65$ | $65-90$ | $90-130$ | $130-170$ | $> 170$ |
|---|---|---|---|---|---|---|---|---|
| Fr. Observada | 15 | 14 | 12 | 12 | 9 | 12 | 10 | 16 |
| Fr. Esperada | $15,31$ | $12,97$ | $10,98$ | $12,08$ | $11,78$ | $13,21$ | $8,48$ | $15,19$ |

Las frecuencias esperadas se calculan, para un intervalo genérico $(a, b)$, como $100 \times \int_a^b (1/90,19) \exp(-t/90,19) dt$

$D = \sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} = 1,26 < \chi^2_{0,95;8-1-1} = 12,59$. No se rechaza la hipótesis nula.

3. Se ha conseguido obtener los valores exactos de los 16 datos censurados y a partir de ellos saber que la media de la muestra completa es $86,41$.

(a) Obtener la distribución de la media muestral utilizando el teorema central del límite, indicando la media y la varianza.

(b) Utilizando el resultado anterior realice el contraste con $\alpha = 0.05$

$$H_0 : \lambda = 100$$

$$H_1 : \lambda < 100$$

(c) Calcule el p-valor.

(d) Si $\lambda = 90$, ¿cuál es la probabilidad de rechazar $H_0$?.($\alpha = 0.05$).

**Solución**

a. $\overline{x}$ es aproximadamente normal con media $\lambda$ y varianza $\lambda^2/n$.
b. $(86,41 - 100)/(100/\sqrt{100}) = -1,36 < z_{0.05} = -1,645$. No se rechaza $H_0$.
c. El $p - valor$ es $\phi(-1,36) = 0,09$.
d. El valor $c$ que separa las regiones de aceptación y rechazo es $c = 100 - 1,645 \times 100/\sqrt{100} = 83,55$.
$(83,55 - 90)/(90/\sqrt{100}) = -0,716$;
$P(\text{rechazar } H_0 \,|\lambda = 90) = \phi(-0,716) = 0,24$.

**Problema 1** (60 minutos, 10 puntos)

El tiempo que un coche de policía tarda en llegar a un robo desde que le avisan desde la central es una variable aleatoria con función de densidad:

$$f_X(x) = kx \text{ para } 0 < x < 10, \text{ y } 0 \text{ para el resto.}$$

Se sabe que el tiempo que tarda un camión de bomberos a un aviso desde su sede es otra variable aleatoria definida a partir de la anterior como $Y = X^{3/2}$.

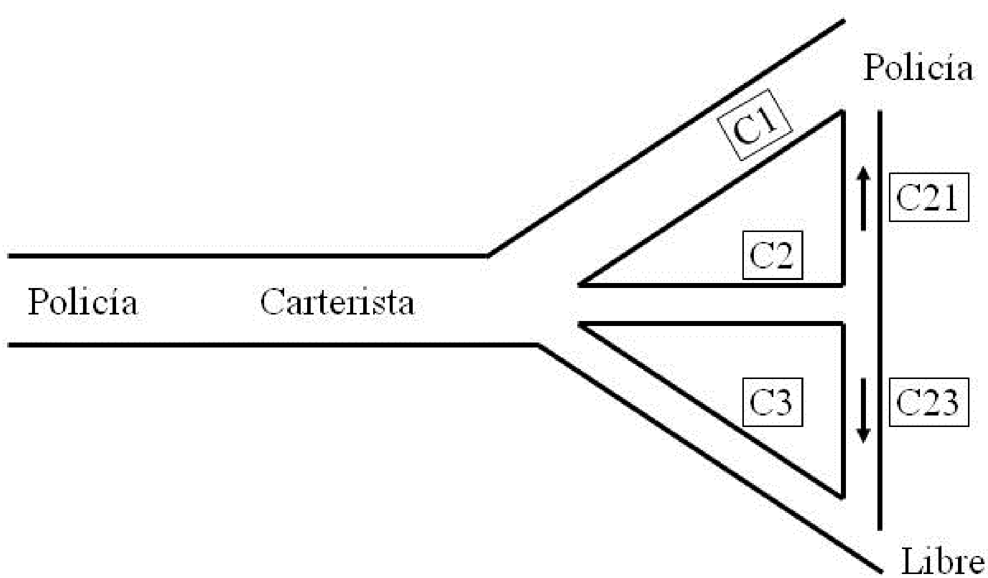
1. Obtenga la esperanza y la varianza de la variable tiempo que tarda en llegar un coche de policía desde el aviso de la sede hasta el atracador (**1 punto).**

2. Obtenga la esperanza y la varianza de la variable tiempo que tarda en llegar un coche de bomberos desde el aviso de la central hasta el accidente. **(3 puntos)**

Por otra parte, un carterista es perseguido por un coche de policía. Al llegar a una bifurcación puede huir por tres calles (C1, C2 y C3). Si huye por la calle C1 le atrapan seguro ya que al final de esa calle hay otro coche de policía cortando la calle. Las calles C2 y C3 son muy estrechas por lo que un coche de policía no cabe por ellas pero el carterista no es consciente de ello. Si huye por la calle C3 escapa seguro ya que no está vigilada. Si huye por la calle C2 se encuentra que está cortada y que se bifurca en dos calles: la C21, que llega a la calle C1 y la C23 que conduce a la calle C3.

3. El carterista elige los caminos al azar. Calcule la probabilidad de que el carterista sea atrapado. **(3 puntos)**

4. Sabiendo que escapó, ¿cuál es la probabilidad de que huyera por la calle C3 entrando por la C2 y llegando a la C3 por la callecita que las comunica, C2-C3? **(3 puntos)**



## Problema

El tiempo que un coche de policía tarda en llegar a un robo desde que le avisan desde la central es una variable aleatoria con función de densidad:

$$f_X(x) = kx \text{ para } 0 \ < x < 10, \text{ y } 0 \text{ para el resto.}$$

Se sabe que el tiempo que tarda un camión de bomberos a un aviso desde su sede es otra variable aleatoria definida a partir de la anterior como $Y = X^{3/2}$.

1. Obtenga la esperanza y la varianza de la variable tiempo que tarda en llegar un coche de policía desde el aviso de la sede hasta el atracador (**1 punto).**

   **Solución:**

   Obtenemos en primer lugar la constante $k$,

   $$\int_0^{10} kxdx = 1, \qquad k = \frac{1}{50}.$$

   Por lo que $f_X(x) = \dfrac{1}{50}x$ para $0 < x < 10$, y 0 para el resto.

   La esperanza y varianza de X son:

   $E[X] = \int_0^{10} \dfrac{1}{50}x^2dx = \dfrac{20}{3}$,

   $Var[X] = E[X^2] - [E(X)]^2 = \int_0^{10} \dfrac{1}{50}x^3dx - (\dfrac{20}{3})^2 = 50 - (\dfrac{20}{3})^2 = \dfrac{450-400}{9} = 5{,}56$

2. Obtenga la esperanza y la varianza de la variable tiempo que tarda en llegar un coche de bomberos desde el aviso de la central hasta el accidente. **(3 puntos)**

   **Solución:**

   Para obtener la esperanza y varianza de $Y$ (tiempo que tarda en llegar un coche de bomberos) se calcula la función de densidad $f_Y(y)$, utilizando la tranformación no lineal $Y = g(X) = X^{3/2}$. Calculada $f_Y(y)$, se obtiene la $E[Y]$ y la $Var[Y]$.

   $$f_Y(y) = f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}|$$

   $$f_Y(y) = \frac{1}{50}y^{2/3}(\frac{2}{3}y^{-1/3}) = \frac{1}{75}y^{1/3}\text{con } 0 < y < 10^{3/2}$$

   $$E[Y] = \int_0^{10^{3/2}} yf_Y(y)dy = \int_0^{10^{3/2}} y\frac{1}{75}y^{1/3}dy = \left[\frac{y^{7/3}}{\frac{7}{3}75}\right]_0^{10^{3/2}} = 18{,}07$$

   $$Var[Y] = E[Y^2] - [E(Y)]^2 = \int_0^{10^{3/2}} y^2\frac{1}{75}y^{1/3}dy - (18{,}07)^2 = 73{,}48$$

Por otra parte, un carterista es perseguido por un coche de policía. Al llegar a una bifurcación puede huir por tres calles (C1, C2 y C3). Si huye por la calle C1 le atrapan seguro ya que al final de esa calle hay otro coche de policía cortando la calle. Las calles C2 y C3 son muy estrechas por lo que un coche de policía no cabe por ellas pero el carterista no es consciente de ello. Si huye por la calle C3 escapa seguro ya que no está vigilada. Si huye por la calle C2 se encuentra que está cortada y que se bifurca en dos calles: la C21, que llega a la calle C1 y la C23 que conduce a la calle C3.

3. El carterista elige los caminos al azar. Calcule la probabilidad de que el carterista sea atrapado. **(3 puntos)**

   **Solución:**

   Sean A, C1, C2 y C3 los siguientes sucesos:

   $A$: atrapar al carterista

   $C1$ : carterista huye por la calle 1

   $C2$ : carterista huye por la calle 2

   $C3$ : carterista huye por la calle 3

   La probabilidad de atrapar al carterista es:

   $$P(A) = P(C1)P(A|C1) + P(C2)P(A|C2) + P(C3)P(A|C3),$$

   pero:

   $$P(A|C2) = P(C21)P(A|C21) + P(C23)P(A|C23),$$

   por lo que:

   $$\begin{aligned} P(A) &= P(C1)P(A|C1) + P(C2)[P(C21)P(A|C21) + P(C23)P(A|C23)] + P(C3)P(A|C3) \\ &= \frac{1}{3}1 + \frac{1}{3}(\frac{1}{2}1 + \frac{1}{2}0) + \frac{1}{3}0 = \frac{1}{2} \end{aligned}$$

4. Sabiendo que escapó, ¿cuál es la probabilidad de que huyera por la calle C3 entrando por la C2 y llegando a la C3 por la callecita que las comunica, C23? **(3 puntos)**

   **Solución:**

   Sean E y C23 los siguientes sucesos:

   $E$: el carterista escapó

   $C23$ : carterista huye por la calle C23

   La probabilidad pedida es:

   $$P(C23|E) = \frac{P(E|C23)P(C23)}{P(E)} = \frac{\frac{1}{6},1}{\frac{1}{2}} = \frac{1}{3}.$$

   ya que la probabilidad de escapar es: $P(E) = 1 - P(A) = \frac{1}{2}$

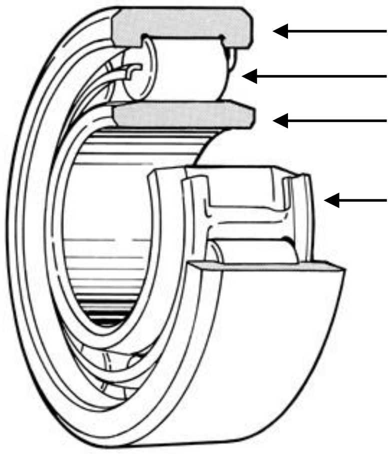
**Problema 2** (60 minutos – 10 puntos)

La fábrica de rodamientos RodaMax dispone de un servicio de atención al cliente compuesto por una centralita telefónica. El número de llamadas recibidas en una hora se distribuye según una distribución de Poisson, con $\lambda = 40$ llamadas por hora.

a) Si el servicio telefónico comienza a las 8:00h, ¿cuál es la probabilidad de que la primera llamada del día sea posterior a las 8:05?

b) Calcular el número esperado de días al año con más de 350 llamadas. (Asumir que la jornada laboral es de 8 horas al día, 260 días al año.)

Los rodamientos fabricados por RodaMax del modelo AC2014MC, se componen de un anillo exterior, **diez** rodillos, un anillo interior y una jaula. El peso de cada elemento se distribuye según una normal, con los parámetros siguientes:




| Elemento | Media del peso [gramos] | Desv. típica [gramos] |
|---|---|---|
| Anillo exterior | 50 | 0.03 |
| Anillo interior | 30 | 0.02 |
| Rodillo | 7 | 0.01 |
| Jaula | 10 | 0.01 |

c) Calcular cuál es el valor que nos asegura que el 95% de los rodamientos tienen un peso inferior a dicho valor.

Nota: Se asume independencia entre los pesos de cada elemento.

Esta empresa desea realizar un pedido a sus proveedores. Para verificar la calidad del suministro, desea realizar un control de recepción a un lote compuesto por 10.000 unidades. Se emplea una muestra aleatoria de 100 observaciones, y se rechaza el lote si hay dos o más observaciones defectuosas.

d) Calcular la probabilidad de aceptar el lote para el caso en el que la tasa de piezas defectuosas es de 3%.

Solución:

a) Tenemos una distribución de Poisson, con $\lambda = 40 \cdot 5\,/60$ llamadas cada cinco minutos. La probabilidad de que la primera llamada del día sea posterior a las 8:05, es la misma probabilidad que durante esos primeros cinco minutos no se reciba ninguna llamada:

$$P(X = 0) = \frac{\lambda^x}{x!} e^{-\lambda} = 0.036$$

b) Calculamos la probabilidad de que un día se reciban más de 350 llamadas, empleando con $\lambda = 40 \cdot 8$ llamadas cada día:

$$P\big(X > 350 \mid X \sim Pois(\lambda = 320)\big) \approx P\left(X > 350.5 \mid X \sim N\big(320, \sqrt{320}\big)\right)$$
$$= 1 - P\left(Z \le \frac{350.5 - 320}{\sqrt{320}}\right) = 1 - P(Z \le 1.70) = 0.955$$

Por tanto, el número esperado de días en los que se reciban más de 350 llamadas es $(1 - 0.955) \cdot 260 =$ 11.7 días

c) El peso del rodamiento es una variable aleatoria que sigue una distribución normal, ya que corresponde con la suma de variables aleatorias normales. A continuación, calculamos sus parámetros:

$$\begin{cases} E[Peso] = 50 + 30 + 7 \cdot 10 + 10 = 160 \\ Var[Peso] = 0.03^2 + 0.02^2 + 10 \cdot 0.01^2 + 0.01^2 = 0.0024 \end{cases}$$

Ahora calculamos el percentil 95:

$$P\left(Peso \le q_{0.95} \mid Peso \sim N\big(160, \sqrt{0.024}\big)\right) = 0.95$$
$$P\left(\frac{Peso - 160}{0.049} \le \frac{q_{0.95} - 160}{0.049}\right) = P\left(Z \le \frac{q_{0.95} - 160}{0.049}\right) = 0.95$$

De las tablas, observamos que $P(Z \le 1.64) = 0.95$, por tanto:

$$q_{0.95} = 160 + 0.049 \cdot 1.64 = 160.0804$$

d) La probabilidad de aceptar el lote se calcula de la siguiente manera:

$$P(aceptar\ lote \mid p = 0.03) = P(X \le 1 \mid X \sim Binom(n = 100, p = 0.03))$$
$$= 0.97^{100} + 100 \cdot 0.97^{99} \cdot 0.03 = 0.195$$

**Problema 3** (60 minutos)

El error cometido por una balanza puede modelarse como una variable aleatoria $N(0, \sigma^2)$. Un supermercado quiere comprobar que la varianza del error de su balanza es la que asegura el fabricante (25 gr$^2$). Para ello se pesa repetidas veces un mismo producto obteniendo los siguientes valores:

$$\{246.87,\ 250.92,\ 245.82,\ 257.98,\ 251.65,\ 245.90,\ 252.44,\ 253.69\}$$

1. Explicar paso a paso como estimar $\sigma^2$ por el método de máxima verosimilitud. Calcular el sesgo y el error cuadrático medio de dicho estimador.

2. Calcular un intervalo de confianza para $\sigma^2$. Utilizando la información proporcionada por el intervalo, contrastar si la varianza del error cometido por la balanza es 25 gr$^2$.

3. Contrastar si el peso del producto que se ha pesado es igual a 250 gr o es superior. Calcular el p-valor del contraste.

4. El supermercado ha comprado otra balanza de peor calidad y se sospecha que la varianza del error cometido por esta nueva balanza ($\sigma_2^2$) es superior a la del cometido por la primera balanza ($\sigma_1^2$). Para ello se ha pesado el mismo producto con la nueva balanza obteniendo los siguientes valores:

$$\{254.61,\ 247.56,\ 262.09,\ 253.12,\ 245.03,\ 232.28,\ 259.00,\ 249.64\}$$

   Contrastar si la varianza del error cometido por la nueva balanza es superior y calcular un intervalo de confianza para $\sigma_2^2/\sigma_1^2$.

**NOTA 1**: $\alpha = 0.05$ para todo el ejercicio.
**NOTA 2**: Si se pesa un producto cuyo peso es $P$ en una balanza que comete un error $N(0, \sigma^2)$, entonces los valores proporcionados por la balanza tienen distribución $N(P, \sigma^2)$.

**SOLUCION**

1. Sea la variable aleatoria X: “error cometido por la balanza”, $X \to N(0, \sigma^2)$.

   Muestra aleatoria simple de X:

$$\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\}$$

   Función de densidad conjunta de la muestra:

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)$$

   Logaritmo de la función verosimilitud:

$$L(\sigma^2) = \log f(x_1, x_2, \cdots, x_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{\sum x_i^2}{2\sigma^2}$$

   Estimador máximo verosimil de $\sigma^2$:

$$\frac{dL(\sigma^2)}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum x_i^2}{2\sigma^4} = 0 \Rightarrow \hat{\sigma}^2 = S^2.$$

   Esperanza y varianza del estimador. Sabemos que $nS^2/\sigma^2 \to \chi^2_{n-1}$, $\mathrm{E}(\chi^2_{n-1}) = n-1$, $\mathrm{Var}(\chi^2_{n-1}) = 2(n-1)$. Por tanto:

$$\mathrm{E}\left(\frac{nS^2}{\sigma^2}\right) = n - 1 \Rightarrow \mathrm{E}(S^2) = \frac{n-1}{n}\sigma^2$$

$$\mathrm{Var}\left(\frac{nS^2}{\sigma^2}\right) = 2(n-1) \Rightarrow \mathrm{Var}(S^2) = \frac{2(n-1)}{n^2}\sigma^4$$

   Sesgo y error cuadrático medio:
   
   $\mathrm{Sesgo}(\hat{\sigma}^2) = \mathrm{E}\ (\hat{\sigma}^2) - \sigma^2 = \frac{-\sigma^2}{n}$
   
   $\mathrm{ECM}(\hat{\sigma}^2) = \mathrm{Sesgo}^2(\hat{\sigma}^2) + \mathrm{Var}(\hat{\sigma}^2) = \frac{2n-1}{n^2}\sigma^4$

2. $\mathrm{IC} = \left[\frac{(n-1)\hat{S}^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)\hat{S}^2}{\chi^2_{\alpha/2}}\right] = \left[\frac{7\cdot 18.20}{16.01}, \frac{7\cdot 18.20}{1.69}\right] = [9.06, 58.79]$

   Queremos resolver el contraste $H_0 : \sigma^2 = 25$; $H_1 : \sigma^2 \neq 25$. Como $25 \in$ IC, $\mathrm{H}_0$ es cierta.

3. 

$$\begin{aligned} H_0 &: \ P = 250 \\ H_1 &: \ P > 250 \end{aligned}$$

   p-valor $= P\left(t \geq \frac{\bar{x}-250}{\sqrt{\hat{S}^2/n}} | t \to t_{n-1}\right) = P\left(t \geq \frac{250.66-250}{\sqrt{18.20/8}} | t \to t_7\right) = P\left(t \geq 0.44 | t \to t_7\right) =$ 0.34. Como p-valor $> \alpha$, se acepta la hipótesis nula.

4.

$$H_0: \ \sigma_1^2 = \sigma_2^2$$
$$H_1: \ \sigma_1^2 < \sigma_2^2$$

$F_0 = \frac{\hat{S}_2^2}{\hat{S}_2^1} = \frac{85.94}{18.20} = 4.72$

Como $F_0 > F_{(7,7);0.05} = 3.79$, se rechaza la hipótesis nula $\Rightarrow \sigma_1^2 < \sigma_2^2$ con un 95 % de confianza.

$\text{IC} = \left[\frac{\hat{S}_2^2}{\hat{S}_1^2} F_{1-\alpha/2}, \frac{\hat{S}_2^2}{\hat{S}_1^2} F_{\alpha/2}\right] = \left[\frac{85.94}{18.20} 0.20, \frac{85.94}{18.20} 4.99\right] = [0.95, 23.58]$

**Cuestiones** (45 minutos, 5 puntos)

1. **Las puntuaciones en un test que mide la variable creatividad siguen, en la población general de adolescentes, una distribución Normal de media 11.5.**

   **Un colegio ha implantado un programa de estimulación de la creatividad. Se dispone de los resultados obtenidos en este test de una muestra de 30 alumnos de dicho colegio:**

   **11, 9, 12, 17, 8, 11, 9, 4, 5, 9, 14, 9, 17, 24, 19, 10, 17, 17, 8, 23, 8, 6, 14, 16, 6, 7, 15, 20, 14, 15.**

   **¿Es efectivo este programa de estimulación de la creatividad implantado en este colegio? Calcule el p-valor del contraste.**

   **NOTA: Tomar como nivel de significación** $\alpha = 0{,}05$.

   Se trata de contrastar:

$$\begin{aligned} H_0 &: \mu = 11{,}5 \\ H_1 &: \mu > 11{,}5, \end{aligned}$$

   y si se rechaza la hipótesis nula en favor de la alternativa, entonces diremos que es efectivo el programa de estimulación de la creatividad

   Para ello se utiliza que $\frac{\overline{x}-\mu}{\widehat{s}/\sqrt{n}} \to t_{n-1}$, donde $n = 30$, $\overline{x} = 12{,}4667$ y $\widehat{s} = 5{,}3093$.

   Como es un contraste unilateral buscamos en las tablas de la $t_{30-1} \equiv t_{29}$ el valor que deja a su derecha un área $\alpha = 0{,}05$ es 1.699.

   Como $\frac{\overline{x}-\mu}{\widehat{s}/\sqrt{n}} = \frac{12{,}4667-11{,}5}{5{,}3093/\sqrt{30}} = 0{,}9973 < 1{,}699$ entonces no se rechaza la hipótesis nula de que $\mu = 11{,}5$.

2. **En una empresa el 20 % de los trabajadores habla dos idiomas (además de español), el 8 % tiene un MBA y el 6 % habla dos idiomas y tiene un MBA.**

   **a) ¿Qué porcentaje de trabajadores habla dos idiomas y no tiene un MBA?**

   **b) Sabiendo que un empleado tiene un MBA, ¿cuál es la probabilidad de que hable dos idiomas?**

   **c) Si la empresa tiene 150 trabajadores, ¿cuántos tienen un MBA y no hablan dos idiomas?**

   A: hablar 2 idiomas

   B: tener un MBA

   $pr(A) = 0{,}2$

   $pr(B) = 0{,}08$

   $pr(A \cap B) = 0{,}06$

   a) $pr(A \cap \overline{B})$, se calcula a partir de:

$$\begin{aligned} pr(A) &= pr(A \cap B) + pr(A \cap \overline{B}) \\ pr(A \cap \overline{B}) &= pr(A) - pr(A \cap B) = 0{,}2 - 0{,}06 = 0{,}14. \end{aligned}$$

Un 14 % de los trabajadores habla 2 idiomas y no tienen un MBA.

b) $pr(A|B) = \frac{pr(A \cap B)}{pr(B)} = \frac{0{,}06}{0{,}08} = 0{,}75$

c) $pr(B \cap \overline{A})$, que se calcula a partir de:

$$\begin{aligned} pr(B) &= pr(B \cap A) + pr(B \cap \overline{A}) \\ pr(B \cap \overline{A}) &= pr(B) - pr(B \cap A) = 0{,}08 - 0{,}06 = 0{,}02, \end{aligned}$$

luego un 2 % tienen un MBA y no hablan 2 idiomas, de un total de 150 trabajadores un 2 % son $150 \cdot 0{,}02 = 3$ trabajadores.

3. **La proporción de estudiantes universitarios que utilizan el metro para desplazarse diariamente es una variable aleatoria con función de densidad $f(x) = kx^2(1-x)$ si $0 < x < 1$, y 0 para el resto de valores de $x$. Se pide:**

   **a) Obtener el valor de $k$.**

   **b) Calcule la varianza de $X$.**

   **c) ¿Cuál es la proporción esperada de estudiantes que utilizan el metro diariamente?**

   a) Como es una función de densidad debe integrar a 1, entonces:

   $\int_0^1 f(x)dx = \int_0^1 kx^2(1-x)dx = k\int_0^1 (x^2 - x^3)dx = k\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = k\left[\frac{1}{3} - \frac{1}{4}\right] =$

   Quedaría: $k\left[\frac{1}{3} - \frac{1}{4}\right] = k\frac{4-3}{12} = k\frac{1}{12} = 1 \to k = 12.$

   b) Para calcular la varianza: $var[X] = E[(X-\mu)^2] = E[X^2] - (E[X])^2 = \frac{6}{15} - \left(\frac{3}{5}\right)^2$

   $E[X^2] = \int_0^1 xf(x)dx = 12\int_0^1 (x^4 - x^5)dx = 12\left[\frac{x^5}{5} - \frac{x^6}{6}\right]_0^1 = 12\frac{1}{30} = \frac{6}{15}$

   $\mu = E[X] = \int_0^1 xf(x)dx = \int_0^1 12x^3(1-x)dx = 12\int_0^1 (x^3 - x^4)dx = 12\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 = 12[\frac{1}{4} - \frac{1}{5}] = 12\frac{1}{20} = \frac{3}{5}.$

   c) Del apartado anterior tenemos que $\mu = 3/5$, valor esperado de $X$.

**Problema** (45 minutos, 5 puntos)

Una compañía aérea, en su línea de bajo coste, utiliza aviones Boeing 737-800 con una capacidad de 190 plazas. En los últimos 10 diez vuelos realizados en una ruta determinada, ha habido 30, 24, 29, 19, 26, 31, 36, 27, 29, 30 pasajeros que han decidido no tomar el avión. Suponiendo que cada pasajero decide de manera independiente:

1. Estime por el método de los momentos la proporción de pasajeros que ha decidido no tomar el avión. Construya un intervalo para dicha proporción con una confianza del 95 %.

2. La compañía aérea, conocedora de este hecho, decide vender más billetes que plazas tiene el avión. Suponiendo que el 15 % de los pasajeros con billete anulan el viaje, calcule el número esperado de vuelos en un mes, en los que al menos un pasajero se queda sin poder volar, sabiendo que la compañía realiza 30 vuelos mensuales en dicha ruta y vende 215 billetes para cada vuelo.

3. Cada vez que en un vuelo al menos un pasajero, con el billete comprado, se queda sin poder volar por falta de asiento en el avión, aviación civil impone una sanción a la compañía. El importe de dicha sanción depende de la frecuencia con la que se produce la irregularidad: Si sucede una vez en un mes, la sanción es 10.000 Euros; si ocurre 2 veces, 25.000 Euros, y si ocurre 3 veces o más, la sanción es de 100.000 Euros y entra en un proceso de reevaluación de permisos de vuelo. Calcule cuál sería el importe medio de la sanción que podría tener la compañía en un mes, si sigue la política de ventas de billetes mencionada anteriormente (215 billetes por vuelo).

**Solución:**

**Apartado 1:**

$\widehat{p} = \dfrac{281}{1900} = 0{,}1479$

$p \in \widehat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\widehat{p}(1-\widehat{p})}{1900}} = 0{,}1479 \pm 1{,}96\sqrt{\dfrac{0{,}1479(1-0{,}1479)}{1900}} \Longrightarrow p \in [0{,}1319; 0{,}1638]$ con confianza 0.95

**Apartado 2:**

$X$ : *v.a $n^o$ de personas que deciden tomar del vuelo de un total de 215 que han comprado el billete*$\Longrightarrow$

$$X \to B(215, 0{,}85)$$

$Y$ : *v.a $n^o$ de vuelos en un mes en los que al menos una persona se queda sin volar de un total de 30*

$Y \to B(30, p)$. Piden $E[Y]$

$p = P$(al menos una persona no tenga asiento (se quede sin volar) cuando se venden 215 billetes)

$p = P(X > 190 | X \to B(215, 0{,}85))$. $\Longrightarrow X \rightsquigarrow N(215 \times 0{,}85, \sqrt{215 \times 0{,}85 \times 0{,}15})$

$p = P(\dfrac{X - 215 \times 0{,}85}{\sqrt{215 \times 0{,}85 \times 0{,}15}} > \dfrac{190 - 215 \times 0{,}85}{\sqrt{215 \times 0{,}85 \times 0{,}15}}) = P(Z > 1{,}38) = 0{,}0838.$

$$E[Y] = 30 \times 0{,}0838 = 2{,}514.$$

**Apartado 3:**

Sean *Y: v.a $n^o$ de vuelos en un mes en los que al menos una persona se queda sin volar de un total de 30* y *S: v.a sanción impuesta a la compañía en un mes.*

La distribución de probabilidad de ambas variables se calcula en la tabla siguiente:

| $Y$ | $S$ | $P(S = s)$ |
|---|---|---|
| 0 | 0 | $\binom{30}{0} 0{,}0694^0(1-0{,}0694)^{30} = 0{,}1159$ |
| 1 | 10000 | $\binom{30}{1} 0{,}0694^1(1-0{,}0694)^{29} = 0{,}2586$ |
| 2 | 25000 | $\binom{30}{2} 0{,}0694^2(1-0{,}0694)^{28} = 0{,}2796$ |
| $\geq 3$ | 100000 | $1 - [0{,}1159 - 0{,}2586 - 0{,}2796] = 0{,}3459$ |

$$E[S] = 0 \times 0 + 10000 \times 0{,}2586 + 25000 \times 0{,}2796 + 100000 \times 0{,}3459 = 44166.Euros$$