Dpto. Matemática Aplicada y Estadística

Ingeniero Técnico de Minas
Asignatura: **Estadística**
Soluciones de la hoja de problemas 1

Estadística Descriptiva

1. Por el histograma a) deducimos que la distribución de los datos es aproximadamente simétrica, unimodal, sin largas colas, y aparentemente sin datos atípicos. La media y la desviación típica son por lo tanto medidas de centralización y de dispersión adecuadas.

   La distribución de los datos representados en el histograma b) es algo asimétrica, unimodal, con una cola de la derecha relativamente larga y aparentemente sin datos atípicos. En este caso la mediana y el rango intercuartílico son medidas apropiadas de centralización y de dispersión aunque, al no ser la asimetría muy pronunciada, es probable que la media tampoco se aleje demasiado de la mediana, lo que nos deja la elección de la medida de centralización.

   La distribución correspondiente al histograma c) en cambio, es muy asimétrica, unimodal y muy concentrada en la primera clase. A priori, la mediana y el rango intercuartílico son medidas apropiadas de centralización y de dispersión.

   El histograma d) es bimodal, y hay presencia de dos subgrupos bastante claros, en este caso, no es razonable hablar del centro o de la dispersión de los datos sino que conviene, si es posible, separar los dos subgrupos y tratar los datos correspondientes por separado.

2. La respuesta es que la varianza aumenta y tenemos dos maneras de justificarla:

   1)Por razones intuitivas: empezamos por notar que la media corresponde al centro de gravedad del conjunto de datos, es decir que si representamos los datos en un segmento como puntos sólidos, la media es el punto de equilibrio del segmento. Por lo tanto si quito uno de los datos iguales a 7, quito un punto situado justo en el punto de equilibrio, lo que no modifica el centro de gravedad de los datos: la media no cambia, y para el conjunto de datos modificados sigue siendo igual a 7. Ahora en cuanto a la varianza, representa el cuadrado de la dispersión de los datos respeto de la media. Cuanto más datos están cerca de la media, menos dispersión presenta el conjunto, por lo tanto si quito un 7, aumentará la dispersión, y por consiguiente la varianza.

   2) Por razones matemáticas: Admitimos primero que la media de los datos modificados sigue siendo igual a 7. Denotemos por $x_1, x_2, \ldots, x_n$ los datos iniciales, y supongamos que el último dato es igual a 7 (siempre podemos cambiar nuestra manera de numerar los datos para que sea así). La varianza de los datos originales es

$$\begin{aligned} s^2_{originales} &= \frac{(x_1-7)^2+(x_2-7)^2+\cdots+(x_{n-1}-7)^2+(x_n-7)^2}{n-1} \\ &= \frac{(x_1-7)^2+(x_2-7)^2+\cdots+(x_{n-1}-7)^2+(7-7)^2}{n-1} \end{aligned}$$

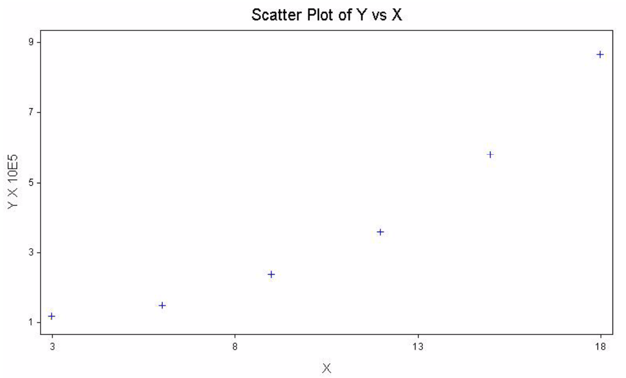
$$= \frac{(x_1-7)^2+(x_2-7)^2+\cdots+(x_{n-1}-7)^2}{n-1}$$

mientras que la varianza de los datos modificados es

$$s^2_{\mathrm{mod}\,ificados} = \frac{(x_1-7)^2+(x_2-7)^2+\cdots+(x_{n-1}-7)^2}{n-2}.$$

El numerador es el mismo mientras que el denominador es más pequeño: deducimos que $s^2_{\mathrm{mod}\,ificados}$ es mayor que $s^2_{originales}$.

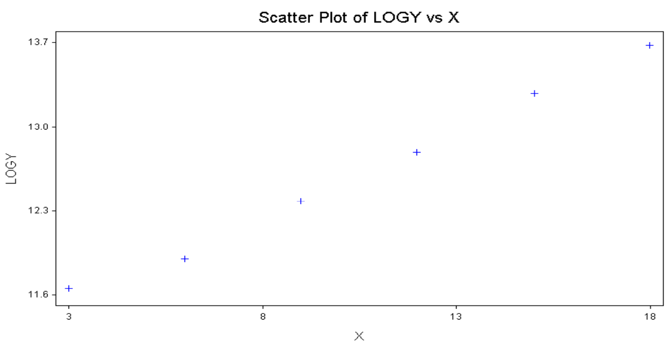
3. **a)**A partir de los datos obtenemos la siguiente representación gráfica en la que se observa claramente una tendencia exponencial:



b) Con el fin de poder realizar el ajuste exponencial, transformaremos los datos correspondientes a la característica y tomando logaritmos (neperianos por supuesto). La tabla resultante será:

x	3	6	9	12	15	18
y'=ln(y)	11.65	11.90	12.38	12.78	13.27	13.67

La representación gráfica que se obtiene ahora es:



en la que sí se observa una dependencia lineal entre ambas características. Con el fin de realizar el correspondiente ajuste lineal con estos nuevos datos, calcularemos en primer lugar:

$\overline{x}$ =10.50	$\overline{x^2}$ =136.50	s_x^2 =31.5	$s_{x;y'}$ =4.374
$\overline{y'}$ =12.61	$\overline{(y')}^2$ =159.51	$s_{y'}^2$ =0.612	$\overline{x \cdot y'}$ =136.05

(Nota.-En este ejercicio hemos calculado varianzas y covarianza **dividiendo por** $n-1$. La ecuación de la recta asustada es la misma que si se divide en varianzas y covarianza por n.)

Si realizamos el ajuste por mínimos cuadrados correspondiente obtenemos:

$$y' = 11.150 + 0.139x \Rightarrow y = 69564 \cdot e^{0.139x}$$

con $r^2_{x,y'} = 0.9942$ con lo cual podríamos preguntarnos ¿cuántas bacterias se inocularon? serían precisamente unas 70.000, es decir, el valor que estimamos para y cuando x=0, este valor debe de utilizarse con ciertas reservas puesto que el valor $x = 0$ no se encuentra dentro del rango de valores observados para la característica x.

c) Como el valor que nos piden no se encuentra dentro del rango de valores que observamos para la característica x debemos hacer las predicciones con cierta cautela, para $x = 20$ obtendríamos $y = 69564 * e^{0.139*20} = 1.1213 \times 10^6$, por tanto debemos tener nuestras reservas a la hora de utilizar dicho valor.

4.

Peso en gramos									
448	450	453	451	447	449	446	451	448	447

a) El peso medio de la muestra es $\bar{x} = 449$

b)La varianza muestral, tenemos $\overline{x^2} = 201605$; deducimos $s_X^2 = 4.89$ (se ha tomado la convención $s_X^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$.)

La desviación típica : $s_X = 2.21$

Para calcular la mediana y los cuartiles, empezamos por ordenar los datos:

$$446 \quad 447 \quad 447 \quad 448 \quad 448 \quad 449 \quad 450 \quad 451 \quad 451 \quad 453$$

Al disponer de un número par de datos, la mediana se calcula como el promedio entre el dato ordenado n^o $n/2 = 5$ y el dato ordenado $n/2 + 1 = 6$. Encontramos $Me = \frac{448+449}{2} = 448.5$. Para calcular el primer cuartil, nos quedamos con el grupo de datos a la izquierda de Me, y calculamos su mediana:

$$446 \quad 447 \quad 447 \quad 448 \quad 448$$

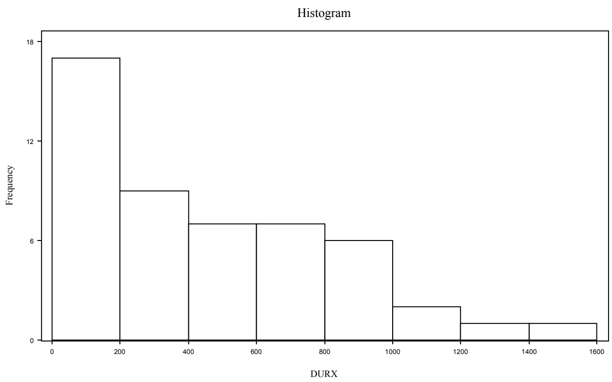
Encontramos $Q_1 = 447$. Para calcular Q_3, hacemos lo propio con el grupo a la izquierda de Me

$$449 \quad 450 \quad 451 \quad 451 \quad 453$$

y encontramos $Q_3 = 451$.

c) La conclusión en el caso anterior es que la empaquetadora funciona correctamente. El método de decisión adoptado por esa empresa sólo se basa sobre la media muestral. Podría por lo tanto ocurrir que se encuentren valores muy dispersos, pero que, al compensarse, tengan una media en el intervalo $448 \leq X \leq 452$, lo que llevaría a afirmar que la máquina funciona correctamente. Para mejorar ese método de decisión se debería tener en cuenta la desviación típica de los datos.

5. a) La variable estadística es la duración hasta el fallo del componente, sus unidades de medida son las horas.



Clases	n_i	f_i	F_i	Clases	n_i	f_i	F_i
$0 < X \leq 200$	17	0.34	0.34	$800 < X \leq 1000$	6	0.12	0.92
$200 < X \leq 400$	9	0.18	0.52	$1000 < X \leq 1200$	2	0.04	0.96
$400 < X \leq 600$	7	0.14	0.66	$1200 < X \leq 1400$	1	0.02	0.98
$600 < X \leq 800$	7	0.14	0.80	$1400 < X \leq 1600$	1	0.02	1.00

El histograma es muy asimétrico, con una cola larga a la derecha, lo que hace que la mediana es más representativa del centro de la distribución que la media.

b) Puesto que 800 es el límite superior de la clase n^{o}4, la proporción de dispositivos que tienen una duración menor o igual que 800h es la frecuencia absoluta acumulada de esta clase es decir 0.8. Por lo tanto, el 80% de los dispositivos tienen una duración menor o igual que 800h. Puesto que 200 es el límite superior n^{o}1, la proporción de dispositivos que tienen una duración mayor a 200 es $1-$ la frecuencia relativa acumulada de la clase 1 es decir 0.66. Por lo tanto, el 66% de los dispositivos tienen una duración mayor a 200h.

De la misma manera se encuentra que el 34% de los dispositivos tienen una duración mayor a 600h.

La proporción de dispositivos que tienen una duración comprendida en el intervalo $200 < X \leq 400$ es por definición la frecuencia relativa de esta clase Por lo tanto el 18% de los dispositivos analizados tienen una duración comprendida en ese intervalo.

Para calcular la proporción de dispositivos que tienen una duración comprendida en el intervalo $200 < X \leq 800$, podemos proceder de dos maneras : primero, sumar las frecuencias relativas de las clase $2, 3$ y 4, encontramos 0.46. Segundo, restar la frecuencia relativa acumulada de la segunda clase a la de la cuarta clase es $F_4 - F_1 = 0.8 - 0.34 = 0.46$.

Conclusión : el 46% de los dispositivos tienen una duración comprendida en el intervalo $200 < X \leq 800$.

c) Los datos se nos presentan agrupados por clases, el cálculo de la media se hace por lo tanto usando las marcas de clase.

$$\bar{x} = \frac{17\times 100 + 9\times 300 + \cdots + 1\times 1500}{50} = 464$$

$$s_X^2 = \frac{50}{49}\left[\frac{17\times(100)^2 + 9\times(300)^2 + \cdots + 1\times(1500)^2}{50} - (464)^2\right] = 135004$$

(hemos tomado la convención $s_X^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$)

$$s_X = \sqrt{s_X^2} = 367.4 \qquad cv = \frac{s_X}{\bar{x}} = 0.79$$

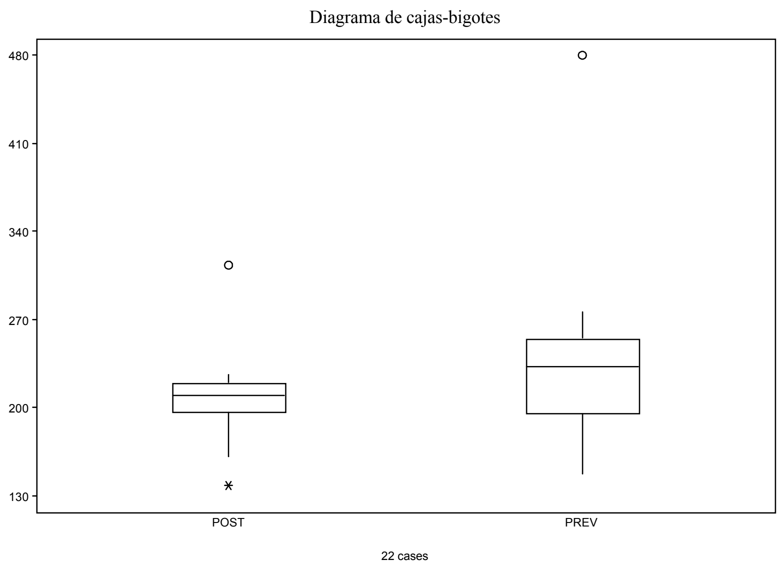
La desviación típica se expresa en las mismas unidades que la variable, en este caso son horas, mientras que el coeficiente de variación es adimensional.

d) Para determinar la clase de la mediana, observamos que tenemos un número par de datos, la mediana es por lo tanto el promedio entre el dato ordenado n^o $n/2 = 25$ y el dato ordenado n^o $n/2+1 = 26$. Después de examinar la tabla de frecuencias, deducimos que estos dos datos se encuentran en la segunda clase.

6. (a) $\overline{prev} = \frac{182+232+\cdots+262}{11} = 243.09$, $s_{prev} = \sqrt{\frac{11}{10}\left[\left(\frac{182^2+232^2+\cdots+262^2}{11}\right) - (243.09)^2\right]} = 87.17$ y $\overline{post} = 209.91$, $s_{post} = \sqrt{\frac{11}{10}\left[\left(\frac{198^2+210^2+\cdots+226^2}{11}\right) - (209.91)^2\right]} = 43.73$.

Datos correspondientes a la dureza previa: la mediana es igual a 232 mientras que $Q_1 = 191$, y $Q_3 = 262$.

Para la dureza posterior, la mediana es igual a 210, $Q_1 = 194$, y $Q_3 = 220$. Obtenemos los diagramas de cajas-bigotes:

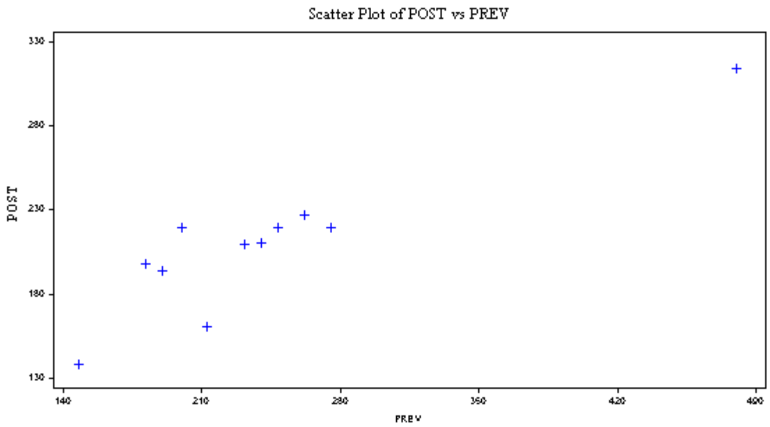


Por estos diagramas, deducimos que el centro de la distribución de la dureza no ha cambiado mucho por el proceso de templado (incluso ha bajado algo) por lo que no podemos afirmar que se haya mejorado la dureza. En cambio, la dispersión de los datos ha disminuido después del proceso de templado, lo que representa una mejora, porque corresponde a una producción más homogénea.

También es conveniente notar que detectamos gracias a esos diagramas un dato atípico para cada conjunto, concretamente, corresponden al par n^o 10, donde tenemos una dureza previa igual a 480 y una dureza posterior igual a 313.

(b) Puesto que hemos detectado un par correspondiente a datos atípicos, necesitamos decidir si los incluimos en el análisis posterior o si simplemente los eliminamos del conjunto de datos.

Queremos encontrar un modelo teórico que nos permita relacionar la dureza previa y la dureza posterior. Vamos a intentar ajustar la nube de puntos por una curva teórica, cuya ecuación conocemos. Tenemos que empezar por representar al nube de puntos:



Por el examen de la nube de puntos, parece que hay una tendencia lineal entre las dos variables, vamos por lo tanto a ajustar una recta de la forma : $y = ax + b$.

♦Supongamos primero que conservamos el par $(480, 313)$ que hemos identificado como atípico.

Necesitamos $s_{prev;post} = \frac{11}{10}\left[\frac{182 \cdot 198 + 232 \cdot 210 + \cdots + 262 \cdot 226}{11} - 243.09 \cdot 209.91\right] = 3451.8$

y la ecuación de la recta de regresión de *post* sobre *prev* es

$$post - \overline{post} = \frac{s_{prev,post}}{s^2_{prev}}(prev - \overline{prev})$$

$$post = 0.45 \times prev + 99.47$$

♦Si eliminamos de nuestro conjunto el par $(480, 313)$, obtenemos :

$\overline{prev} = 219.4 \quad s^2_{prev} = 1582.3 \quad s_{prev;post} = 850.07$
$\overline{post} = 199.6 \quad s^2_{post} = 826.71$

lo que nos lleva a la ecuación:

$$post = 81.73 + 0.54 prev$$

(c) Para estudiar la precisión del ajuste, calculamos r :

$$r = \frac{s_{prev,post}}{s_{prev}s_{post}}$$

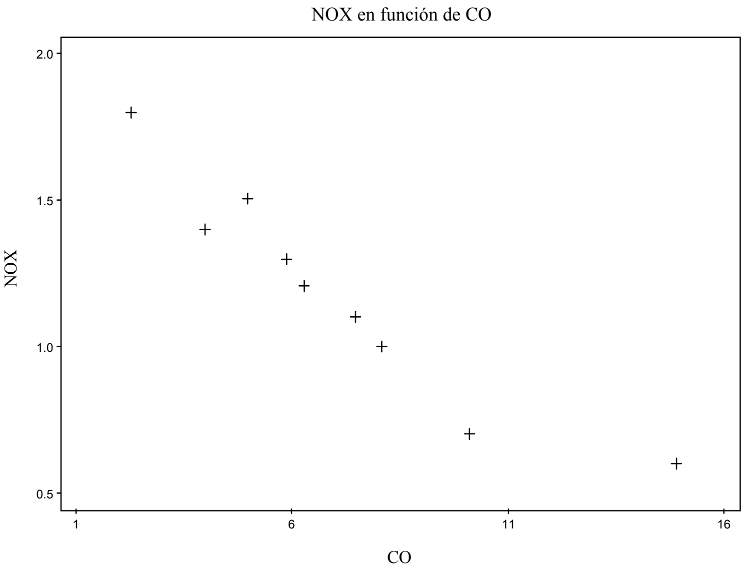
♦con el dato $(480, 313)$, encontramos $r = 0.91$.

♦Sin el dato $(480, 313)$, encontramos $r = 0.74$

lo que implica que el dato atípico mejora de manera artificial y engañosa el ajuste por una recta, tal como nos damos cuenta al observar la nube de puntos.

Para encontrar el valor esperado de la dureza posterior si la dureza previa es 195, basta con sustituir en la ecuación de la recta ajustada *prev* por 195. Por ejemplo, sin el dato atípico $(480, 313)$, encontramos $post = 81.73 + 0.54 * 195 = 187.03$. Notar sin embargo que la precisión de esta predicción es dudosa puesto que hemos visto que el ajuste lineal no es muy bueno.

7. a) Nube de puntos:



Si buscamos un modelo teórico para modelizar la relación entre las dos variables, observamos por la nube de puntos que parece existir una relación lineal entre las concentraciones de CO y de NO_x

Se trata de una asociación negativa.(cuando crece NO_x , decrece CO)

b) Para determinar la ecuación de la recta de regresión necesitamos las siguientes cantidades

$\overline{CO} = 7.12$; $s^2_{CO} = 13.77$ s_{CO,NO_x} =-1.35
$\overline{NO_x} = 1.18$; $s^2_{NO_x} = 0.14$

Deducimos la ecuación de la recta de regresión de CO sobre NO_x :

$$NO_x = -0.098CO + 1.88$$

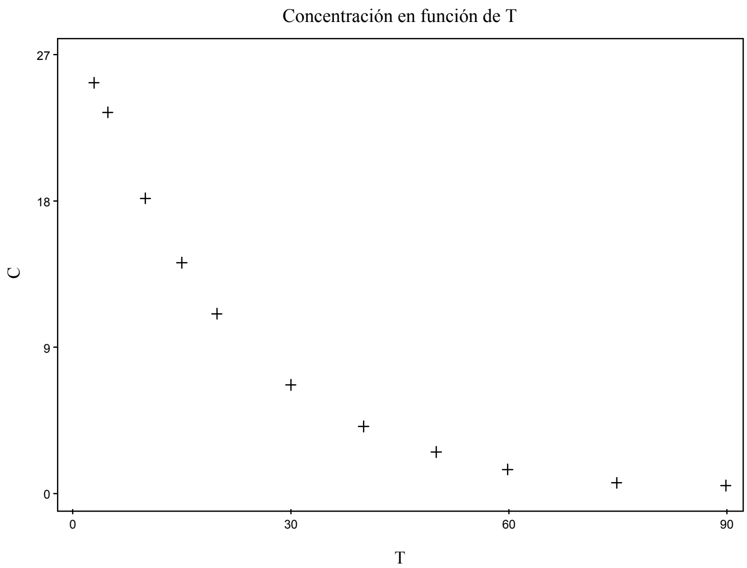
c)El coeficiente de determinación r^2 es igual a $\left(\frac{s_{CO,NO_x}}{s_{CO}s_{NO_x}}\right)^2$. Encontramos $r^2 = 0.92$

d) Hemos visto que existe una asociación de tipo negativa entre CO y NO_x. Por lo tanto, el periodista está equivocado, puesto que si encontramos poco NOx, el valor de CO es más alto. Debemos por consiguiente medir las dos concentraciones para controlar las emisiones del motor.

8. La hidrólisis de un cierto producto químico tiene lugar en medio ácido según un proceso cinético de primer orden. Partiendo de una concentración inicial desconocida, se han medido las concentraciones del mismo a diferentes tiempos obteniéndose los resultados siguientes.

t(min)	3	5	10	15	20	30	40	50	60	75	90
c.10^3(M)	25.5	23.4	18.2	14.2	11	6.7	4.1	2.5	1.5	0.7	0.4

a)Nube de puntos de las dos variables:



(a) El modelo teórico que nos interesa es del tipo exponencial

$$C_t = C_0 e^{-kt}.$$

Es por lo tanto adecuado aplicar el logaritmo sobre los dos términos para obtener

$$\ln C_t = \ln C_0 - kt,$$

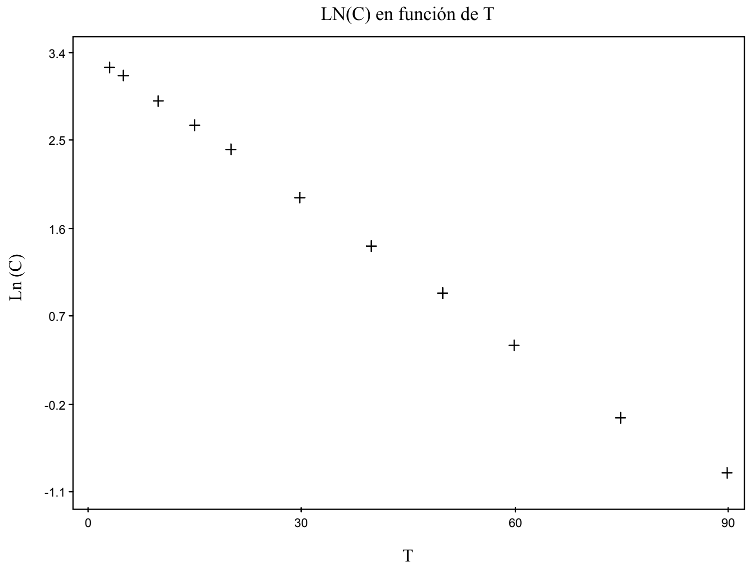
lo que implica que la relación entre $\ln C_t$, y t es del tipo lineal. Si denotamos por $C' = \ln(C_t)$, tenemos la relación

$$C' = \ln C_0 - kt,$$

obtenemos la nueva tabla

t	3	5	10	15	20	30	40	50	60	75	90
C'	3.23	3.15	2.90	2.65	2.40	1.90	1.41	0.92	0.41	-0.36	-0.92

cuya nube de puntos es



Notar que la nube de puntos de los datos transformados sí presenta una tendencia lineal. Podemos por lo tanto determinar la recta de regresión de $\ln C$ sobre el tiempo. La ordenada al origen será $\ln(C_0)$ mientras que la pendiente será $-k$.
Obtenemos

$$\begin{array}{ll} \bar{t} = 36.18 & s_t^2 = 868.36 \\ \bar{c'} = 1.61 & s_{c'}^2 = 2.07 \\ s_{c't} = -42.37 & \end{array}$$

La ecuación de la recta de regresión que obtenemos es

$$\ln C_t = 3.37 - 0.049t.$$

con un coeficiente de determinación $r^2 = 0.999$, lo que indica un ajuste buenísimo. Deducimos que

$$\begin{aligned} C_0 &= e^{3.37} = 29.1\ (10^{-3}M) \\ k &= 0.049. \end{aligned}$$

b) Tenemos ahora información adicional que podemos incluir en nuestro modelo teórico: nuestra curva teórica que ajuste la concentración en función del tiempo debe pasar por el punto $(0, 30)$. Para los datos transformados gracias al logaritmo, la recta deberá pasar por el punto $(0, \ln 30)$. En clase vimos como ajustar una recta que pase por el origen, es decir cuya ecuación sea de la forma $y = ax$. Para encontrarnos en esta situación, es conveniente realizar una nueva transformación de los datos C', y considerar $C'' = C' - \ln 30$.
Los datos que consideraremos son t y C'' :

t(min)	3	5	10	15	20	30	40	50
C''	-0.17	-0.25	-0.5	-0.75	-1.00	-1.50	-1.99	-2.48
t(min)	60	75	90					
C''	-2.99	-3.76	-4.32					

y realizamos un ajuste por una recta de ecuación $y = ax$. La fórmula para el valor de $\hat{\mathbf{a}}$ es

$$\hat{a} = \frac{\overline{xy}}{\overline{x^2}}$$

Necesitamos por lo tanto $\overline{c''t} = -103.34$ y $\overline{t^2} = 2095.54$. Encontramos : $\hat{a} \simeq -0.049$. Puesto que nuestro modelo es

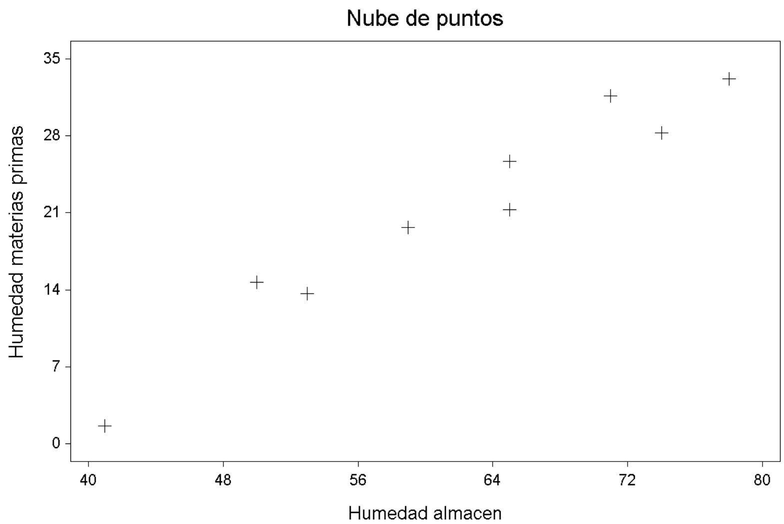
$$C'' = \ln C - \ln 30 = -kt,$$

deducimos que $k \simeq 0.049$.

9. Las materias primas empleadas en la producción de una fibra sintética son almacenadas en un lugar en donde no se tiene control de la humedad. La siguiente tabla refleja en porcentajes la humedad relativa del almacén (X) y la humedad observada en la materias primas (Y) durante un estudio que tuvo lugar durante 9 días.

X	41	53	59	65	71	78	50	65	74
Y	1.6	13.6	19.6	25.6	31.6	33.2	14.7	21.2	28.3

a) Empezamos por representar la nube de puntos,



Puesto que los datos parecen presentar una tendencia lineal, vamos a ajustar una recta a la curva experimental. Para ello necesitamos las siguientes cantidades:

$\bar{x} = 61.78 \quad s_x^2 = 146.69 \quad s_{xy} = 118.58$
$\bar{y} = 21.04 \quad s_y^2 = 101.06$

Obtenemos la recta de regresión

$$y = -28.8 + 0.8x$$

con un coeficiente de determinación $r^2 = 0.94$, lo que indica un buen ajuste.

b) Si utilizamos el modelo teórico del apartado anterior para predecir la humedad en las materias primas si se consigue una humedad relativa de 35,

$$y = -28.8 + 0.8 \times 35 = -0.8$$

lo que no tiene sentido físico. Esto quiere decir que la relación lineal entre la humedad en el almacén y la humedad en las materias primas que hemos descrito gracias a datos de x observados en el rango 41; 74 no sigue válida para $x_0 = 35$. Es mejor por lo tanto no proporcionar ningún valor para la pregunta.

Probabilidad

1. El espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$. La información que se nos da se traduce de la manera siguiente : $P(2) = 2 \times P(1)$; $P(3) = 3 \times P(1)$, $P(4) = 4 \times P(1)$; $P(5) = 5 \times P(1)$; $P(6) = 6 \times P(1)$.

   Por otra parte sabemos que $P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = P(S) = 1$. Por lo tanto $P(1)+2P(1)+3P(3)+4P(1)+5P(1)+6P(1) = 1$; i.e $21 \times P(1) = 1$ La probabilidad de sacar 1 es por lo tanto $1/21$, y la probabilidad de sacar un 4 es $4/21$.

2. Ayudándonos del diagrama de Venn deducimos que el espacio muestral Ω es la unión : $\Omega = (A \cap B) \cup (A^C \cap B) \cup (A \cap B^C) \cup (A^C \cap B^C)$ y que además los cuatro sucesos $(A \cap B), (A^C \cap B), (A \cap B^C)$ y $(A^C \cap B^C)$ son mutuamente excluyentes. Del tercer axioma de la probabilidad, obtenemos por consiguiente que $P(\Omega) = P(A \cap B) + P(A^C \cap B) + P(A \cap B^C) + P(A^C \cap B^C)$. Utilizando las hipótesis, esa igualdad se traduce de la manera siguiente : $1 = 10h + 15h + 20h + 5h$. Por lo tanto, $h = \frac{1}{50}$

   El conjunto de los elementos de A que no son de B es por definición $A \cap B^C$, y sabemos que $P(A \cap B^C) = 15h = 3/10$. Si todos los sucesos elementales son equiprobables, la definición de Laplace nos indica que

   $$P(A \cap B^C) = \frac{\text{num. elementos de } A \cap B^C}{\text{num. elementos de } \Omega}$$

   Por lo tanto (num. elementos de $A \cap B^C$) = (num. elementos de Ω) $\times 15h = 50 \times 3/10 = 15$.

3. Sea A el suceso "la suma de los puntos de las dos bolas es impar", queremos calcular $P(A)$. Empecemos por calcular el número de casos posibles : Sacamos dos bolas, sin reposición, donde el orden no importa y los números son distintos : se trata por lo tanto de combinaciones sin repeticiones, el número de casos posibles es $C_{5,2} = 10$. Podemos enumerar fácilmente todos los casos favorables para que la suma de los puntos sea impar : bolas 1 y 2, 1 y 4, 2 y 3, 2 y 5, 3 y 4, 4 y 5. Hay por lo tanto 6 casos favorables. Deducimos $P(A) = \frac{6}{10} = \frac{3}{5}$.

4. Sea A el suceso " todos los números son pares". Empezamos por numerar los dados del 1 al 5. Suponemos que los resultados de los dados son independientes. Introducimos los sucesos A_1="el resultado del primer dado es par", A_2="el resultado del segundo dado es par", y así hasta A_5.Nos piden $P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5)$. De la hipótesis de independencia de los dados deducimos

   $$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = P(A_1) \times P(A_2) \times P(A_3) \times P(A_4) \times P(A_5).$$

Por otra parte sabemos que $P(A_i) = 1/2,\ i = 1, \cdots, 5$. Obtenemos finalmente $P(A) = (1/2)^5 = 1/32$. Sea B el suceso " sale al menos un 6". Es más fácil calcular la probabilidad de B^C="ninguno de los cinco números sacados es igual a 6". Introducimos los sucesos C_1="el resultado del primer dado está comprendido entre 1 y 5.", C_2="el resultado del segundo dado está comprendido entre 1 y 5.", y así hasta C_5.Tenemos $P(B^C) = P(C_1 \cap C_2 \cap C_3 \cap C_4 \cap C_5)$. Usando la hipótesis de independencia, obtenemos $P(B^C) = P(C_1) \times P(C_2) \times P(C_3) \times P(C_4) \times P(C_5)$. Por otra parte sabemos que $P(Ci) = 5/6,\ i = 1, \cdots, 5$ y se comprueba que se obtiene $P(B^C) = 3125/7776 = 0.40$, y por lo tanto $P(B) = 4651/7776 \simeq 0.60$.

5. Denotemos por $S_i,\ i = 1, 2, 3$ el suceso "el atleta supera el intento número i", y por SU el suceso " el atleta supera la prueba en la competición" Nos piden $P(SU)$. El enunciado se traduce por $P(S_1) = 0.7, P(S_2|S_1^C) = 0.7$ y $P(S_3|(S_1^C \bigcap S_2^C)) = 0.7$. Para calcular $P(SU)$ es más fácil calcular la probabilidad del suceso complementario SU^C=" el atleta no supera la prueba en la competición". Tenemos que $SU^C = S_1^C \cap S_2^C \cap S_3^C$, utilizaremos por lo tanto la regla del producto $P(SU^C) = P(S_3^C|(S_1^C \bigcap S_2^C))P(S_1^C|S_2^C)P(S_1^C) = 0.3 \cdot 0.3 \cdot 0.3 = 0.027$ y $P(SU) = 1 - 0.027 = 0.973$.

6. Definimos los sucesos F_1, F_2 y F_3 como "Falla el cañón 1", "Falla el cañón 2", "Falla el cañón 3" respectivamente y los sucesos B_0, B_1, B_2, B_3 como "Ningún cañón acierta", "sólo un cañón acierta", "sólo dos cañones aciertan" y"los tres cañones aciertan". Del enunciado deducimos $P(F_1) = 1 - 0.1 = 0.9,\ P(F_2) = 0.8$, y $P(F_3) = 0.7$ y que nos piden $P(B_0)$, $P(B_1)$, $P(B_2)$ y $P(B_3)$.

   Tenemos que $B_0 = F_1 \cap F_2 \cap F_3$, y al ser los tres cañones independientes $P(B_0) = P(F_1) \times P(F_2) \times P(F_3)$. Por lo tanto $P(B_0) = 0.9 \times 0.8 \times 0.7 = 0.504$.

   Tenemos $B_1 = \big((F_1)^C \cap F_2 \cap F_3\big) \bigcup \big(F_1 \cap (F_2)^C \cap F_3\big) \bigcup \big(F_1 \cap F_2 \cap (F_3)^C\big)$,

   como los tres sucesos $\big((F_1)^C \cap F_2 \cap F_3\big)$, $\big(F_1 \cap (F_2)^C \cap F_3\big)$ y $\big(F_1 \cap F_2 \cap (F_3)^C\big)$ son mutuamente excluyentes, deducimos de la regla de adición y usando la independencia de los tres cañones,

$$
\begin{aligned}
P(B_1) &= P(F_1^C) \times P(F_2) \times P(F_3) + P(F_1) \times P(F_2^C) \times P(F_3) + P(F_1) \times P(F_2) \times P(F_3^C) \\
&= 0.1 \times 0.8 \times 0.7 + 0.9 \times 0.2 \times 0.7 + 0.9 \times 0.8 \times 0.3 = 0.398
\end{aligned}
$$

   Vamos a calcular ahora $P(B_3)$

   $B_3 = (F_1^C) \cap (F_2)^C \cap (F_3)^C$. y por lo tanto $P(B_3) = P(F_1^C) \times P(F_2^C) \times P(F_2^C) = 0.006$

   Para calcular $P(B_2)$, es suficiente observar que $P(B_2) = 1 - P(B_0) - P(B_1) - P(B_3) = 0.092$.

   Finalmente, si llamamos B el suceso "acierta al menos un cañón", tenemos que $B = B_0^C$ y $P(B) = 1 - P(B_0) = 0.496$

7. (a) Por la regla de adición, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, deducimos $P(A \cap B) = \frac{1}{3} + \frac{2}{5} - \frac{3}{5} = \frac{2}{15}$.

$P(A^C) = 1 - P(A) = \frac{2}{3}$. Por la ley de Morgan, $(A^C \bigcap B^C) = (A \cup B)^C$, por lo tanto $P(A^C \bigcap B^C) = 1 - P(A \cup B) = \frac{2}{5}$.
Finalmente $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2}{15} \cdot \frac{5}{2} = \frac{1}{3}$.

(b) Puesto que deducimos del apartado anterior que $P(A|B) = \frac{1}{3} = P(A)$, A y B son independientes. En cambio, puesto que $P(A \cap B) = \frac{2}{15} \neq 0$, es imposible que $A \cap B = \emptyset$, A y B no son incompatibles.

8. (a) Cálculo de $P(B)$ si A y B son independientes.
Por la regla de adición, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Al suponer que A y B son independientes, tenemos que $P(A \cap B) = P(A) \times P(B) = 0.5P(B)$. La regla de adición se escribe por consiguiente

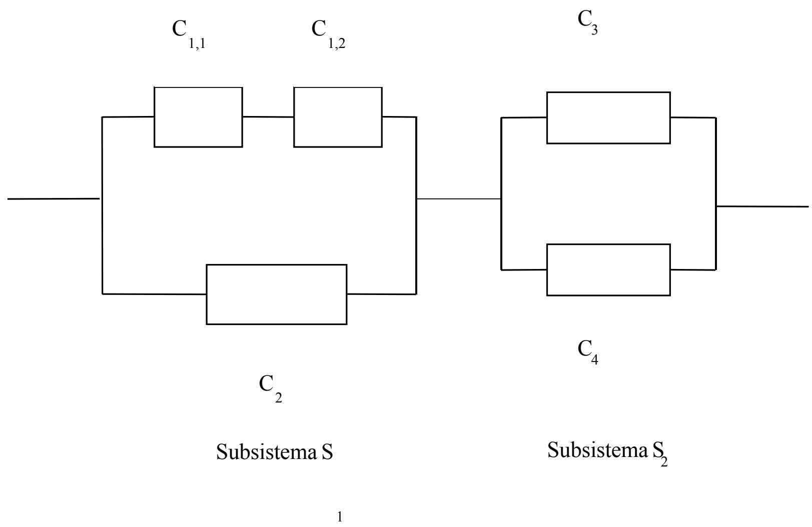
$$0.7 = 0.5 + P(B) - 0.5P(B)$$

de donde deducimos $P(B) = 0.4$

(b) Cálculo de $P(B)$ si A y B son mutuamente excluyentes.
En este caso el tercer axioma de la probabilidad implica $P(B) = P(A \cup B) - P(A)$, de donde deducimos $P(B) = 0.7 - 0.5 = 0.2$.

(c) Cálculo de $P(B)$ si $P(A|B) = 0.5$.
Si $P(A|B) = 0.5$, y como $P(A) = 0.5$ resulta que $P(A|B) = P(A)$, lo que significa que los sucesos A y B son independientes, por lo tanto esa pregunta ya se ha tratado en el apartado (a)

El circuito está constituido por dos subcircuitos en serie S_1 y S_2. S_1 está compuesto por dos subcircuitos en paralelo, el primero comporta dos componentes en serie $C_{1,1}$ y $C_{1,2}$ mientras que el segundo, comporta el componente C_2. El subcircuito S_2 comporta dos componentes en paralelo C_3 y C_4.



Nos piden $P((S_1 \text{ funciona }) \cap (S_2 \text{ funciona }))$. Al ser todos los componentes independientes tenemos

$$P((S_1 \text{ funciona }) \cap (S_2 \text{ funciona })) = P(S_1 \text{ funciona })P(S_2 \text{ funciona })$$

Para calcular $P(S_1 \text{ funciona })$ vamos a calcular la probabilidad del suceso complementario $P(S_1 \text{ no funciona })$

$$\begin{aligned} P(S_1 \text{ no funciona }) &= P((C_{1,1} \text{ no func. o } C_{1,2} \text{ no func. })) \cap (C_2 \text{ no func. })) \\ &= P(C_{1,1} \text{ no func. o } C_{1,2} \text{ no func. })P(C_2 \text{ no func. }) \end{aligned}$$

Por otra parte,

$$\begin{aligned} P(C_{1,1} \text{ no func. o } C_{1,2} \text{ no func. }) &= P(C_{1,1} \text{ no func. }) + P(C_{1,2} \text{ no func. }) \\ &\quad -P((C_{1,1} \text{ no func. }) \cap (C_{1,2} \text{ no func. })) \\ &= 0.01 + 0.01 - 0.01 \times 0.01 = 0.0199 \end{aligned}$$

y $P(C_2 \text{ no funciona }) = 0.1$. por lo tanto, $P(S_1 \text{ no funciona }) = 0.00199$.

Calculemos ahora $P(S_2 \text{ no funciona })$. Tenemos

$$\begin{aligned} P(S_2 \text{ no funciona }) &= P((C_3 \text{ no funciona }) \cap (C_4 \text{ no funciona })) \\ &= 0.1 \times 0.1 = 0.01 \end{aligned}$$

y

$$P((S_1 \text{ funciona }) \cap (S_2 \text{ funciona })) = (1 - 0.00199) \times (1 - 0.01) = 0.98803$$

9. Sea A el suceso "el número escogido es par", CA el suceso "sale cara", y "CR" el suceso "sale cruz". Se nos pide calcular $P(A)$.

   Por otra parte observamos que los dos sucesos CA y CR son mutuamente excluyentes y su unión es igual al espacio muestral. Estamos por lo tanto antes las condiciones de aplicación de la formula de probabilidad total,

$$P(A) = P(A|CA)P(CA) + P(A|CR)P(CR)$$

   Calculemos ahora $P(A|CA)$: Sabiendo que ha salido cara, los casos favorables para que se escoja un número par son 2,4,6,8 mientras que los casos posibles son todos los números del 1 al 9. Deducimos por lo tanto que $P(A|CA) = 4/9$. Calculemos $P(A|CR)$: Sabiendo que ha salido cruz, los casos favorables para que se escoja un número par son 2 y 4 mientras que los casos posibles son los números del 1 al 5 y por lo tanto $P(A|CR) = 2/5$. La formula de probabilidad total se escribe entonces :

$$P(A) = \frac{4}{9} \times \frac{2}{3} + \frac{2}{5} \times \frac{1}{3} = \frac{58}{135}$$

10. (a) Introducimos los sucesos D_1 ="el artículo presenta el defecto de tipo 1", y D_2 ="el artículo presenta el defecto de tipo 2". El suceso "el artículo tiene ambas clases de defectos" es $D_1 \cap D_2$. Al ser los dos sucesos independientes por hipótesis, $P(D_1 \cap D_2) = P(D_1)P(D_2) = 0.1 \times 0.05 = 0.005$

    (b) El suceso D="el artículo es defectuoso" es $D_1 \cup D_2$. (presenta el defecto 1 o el defectuo 2). Por la regla de adición $P(D) = P(D_1) + P(D_2) - P(D_1 \cap D_2)$.Usando el resultado del apartado a), obtenemos $P(D) = 0.1 + 0.05 - 0.005 = 0.145$.

    (c) El suceso "el artículo sólo tiene un tipo de defecto" se puede escribir $(D_1 \cap D_2^C) \cup (D_1^C \cap D_2)$, se nos pide la probabilidad de ese suceso condicionada al que el artículo es defectuoso, o sea $P((D_1 \cap D_2^C) \cup (D_1^C \cap D_2)|D)$.

$$P((D_1 \cap D_2^C) \cup (D_1^C \cap D_2)|D) = \frac{P(((D_1 \cap D_2^C) \cup (D_1^C \cap D_2)) \cap D)}{P(D)}$$

    Pero está claro que $((D_1 \cap D_2^C) \cup (D_1^C \cap D_2)) \cap D) = (D_1 \cap D_2^C) \cup (D_1^C \cap D_2)$, lo que implica

$$P((D_1 \cap D_2^C) \cup (D_1^C \cap D_2)|D) = \frac{P((D_1 \cap D_2^C) \cup (D_1^C \cap D_2))}{P(D)}$$

    Ahora, sabemos que $P(D_1 \cap D_2^C) = 0.1 \times 0.95 = 0.095$ y $P(D_1^C \cap D_2) = 0.9 \times 0.05 = 0.045$ y puesto que los sucesos $(D_1 \cap D_2^C)$ y $(D_1^C \cap D_2)$ son mutuamente excluyentes, $P((D_1 \cap D_2^C) \cup (D_1^C \cap D_2)) = 0.095 + 0.045 = 0.140$. Finalmente obtenemos que

$$P((D_1 \cap D_2^C) \cup (D_1^C \cap D_2)|D) = 0.140/0.145 = 0.9655$$

11. Empecemos por traducir los datos del enunciado : sean los sucesos $i = 1, 2, 3$, F_i =”la pieza ha sido producida por la fábrica número i”, y NS =” la pieza es non standard”, sabemos

$$P(F_1) = 0.2,\ P(F_2) = 0.46\ P(F_3) = 0.34$$

$$P(NS|F_1) = 0.03\ P(NS|F_2) = 0.02\ P(NS|F_3) = 0.01$$

Nos piden $P(F_1|NS)$. Puesto que los sucesos F_1, F_2, F_3 son mutuamente excluyentes y $F_1 \cup F_2 \cup F_3 = \Omega$. Estamos exactamente en las condiciones de aplicación del teorema de Bayes.

$$P(F_1|NS) = \frac{P(NS|F_1)P(F_1)}{P(NS|F_1)P(F_1) + P(NS|F_2)P(F_2) + P(NS|F_3)P(F_3)}$$

i.e

$$P(F_1|NS) = \frac{0.03 \times 0.2}{0.03 \times 0.2 + 0.02 \times 0.46 + 0.01 \times 0.34} \simeq 0.32$$

12. (a) Introducimos los sucesos DA =”La pieza tiene el defecto A”, DB =”La pieza tiene el defecto B”, C =”la pieza no tiene ningún defecto” y R =”La pieza se rompe durante el test”. $\begin{array}{lll} P(DA) = 0.08 & P(DB) = 0.05 & P(C) = 1 - 0.08 - 0.05 = 0.87 \\ P(R|DA) = 0.9 & P(R|DB) = 0.95 & P(R|C) = 0.01. \end{array}$

    (b) Nos piden $P(R)$. Como los sucesos DA, DB y C son incompatibles dos a dos, y su unión es igual al espacio muestral entero, estamos en las condiciones de aplicación de la fórmula de la probabilidad total.

$$\begin{aligned} P(R) &= P(R|DA)P(DA) + P(R|DB)P(DB) + P(R|C)P(C) \\ &= 0.9 \cdot 0.08 + 0.95 \cdot 0.05 + 0.01 \cdot 0.87 = 0.1282 \end{aligned}$$

    (c) Nos piden $P(C|R)$. Aplicamos el teorema de Bayes,

$$P(C|R) = \frac{P(R|C)P(C)}{P(R|DA)P(DA) + P(R|DB)P(DB) + P(R|C)P(C)} = \frac{0.0087}{0.1282} \simeq 0.0679$$

13. Como siempre, lo primero que tenemos que hacer es ir traduciendo los datos del problema. Sean los sucesos PE =” se produce un peligro”, A =” el alarma funciona”, nos indican que

$$P(PE) = 0.1,\ P(A|PE) = 0.95,\ P(A|PE^C) = 0.03$$

    (a) Nos piden $P(PE^C|A)$. Estamos en las condiciones de aplicación del teorema de Bayes,

$$P(PE^C|A) = \frac{P(A|PE^C)P(PE^C)}{P(A|PE^C)P(PE^C) + P(A|PE)P(PE)}$$

    y deducimos $P(PE^C|A) = \frac{0.03 \times 0.9}{0.03 \times 0.9 + 0.95 \times 0.1} = 0.22$.

(b) Nos piden $P(PE \cap A^C)$. $P(PE \cap A^C) = P(A^C|PE)P(PE)$, y por otra parte, $P(A^C|PE) = 1 - P(A|PE)$. Deducimos por lo tanto que $P(PE \cap A^C) = (1 - 0.95) \times 0.1 = 0.005$.

(c) Nos piden $P(PE|A^C)$. Estamos en las condiciones de aplicación del teorema de Bayes,

$$P(PE|A^C) = \frac{P(A^C|PE)P(PE)}{P(A^C|PE)P(PE) + P(A^C|PE^C)P(PE^C)}$$

lo que implica $P(PE|A^C) = \frac{0.005}{0.005+(1-0.03)0.9} \simeq 0.006$

Variables Aleatorias

1. Como sabemos, las dos condiciones que debe cumplir una función $f(x)$ para ser función puntual de probabilidad son:

   •$f(x) \geq 0$ para cualquier $x \in \mathbb{R}$

   • $\sum_i f(x_i) = 1$ *donde el índice i indica el número de posibles valores que toma* x_i.

   a) Es función puntual de probabilidad (f.p.p.) puesto que:

   $$\bullet f(x) = \tfrac{x}{15} \geq 0$$

   $$\bullet \sum_{i=0}^{5} f(x_i) = \tfrac{0}{15} + \tfrac{1}{15} + \tfrac{2}{15} + \tfrac{3}{15} + \tfrac{4}{15} + \tfrac{1}{15} = 1$$

   b) No es f.p.p. pues no se cumple la condición de ser positiva puesto que:

   $$f(3) = \frac{5-3^2}{6} = \frac{-4}{6} < 0$$

   c) Es f.p.p. puesto que:

   $$\bullet f(x) = 0.25 \geq 0$$

   $$\bullet \sum_{i=1}^{4} f(x_i) = 0.25 + 0.25 + 0.25 + 0.25 = 1$$

   d) No es f.p.p. pues no se cumple la condición de ser positiva puesto que:

   $$f(1) = \frac{1-2}{2} = -1 < 0$$

2. Como se debe verificar que $f(x) = \frac{k}{2^x} \geq 0$, tenemos que $\boxed{k \geq 0}$.
   Por otro lado, debe verificarse que $\sum f(x_i) = 1$, luego:

   $$1 = \frac{k}{2^0} + \frac{k}{2^1} + \frac{k}{2^2} + \frac{k}{2^3} + \frac{k}{2^4} = k \cdot \left[1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}\right] = k \cdot \frac{31}{16}$$

$$\boxed{k = \frac{16}{31}}$$

La función de distribución ($F(x) = \Pr(X \leq x) = \sum_{x_i \leq x} f(x_i)$) asociada a esta variable será:

$$\boxed{F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{16}{31} & 0 \leq x < 1 \\ \dfrac{16}{31}\left[1 + \frac{1}{2}\right] & 1 \leq x < 2 \\ \dfrac{16}{31}\left[1 + \frac{1}{2} + \frac{1}{4}\right] & 2 \leq x < 3 \\ \dfrac{16}{31}\left[1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8}\right] & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}}$$

3. **a)** Para que sea f.p.p. debe verificarse que sea siempre positiva y que

$$\sum_x \Pr(x) = \sum_{x=1}^{6} f(x) = 1$$

por tanto:

$$k(1+2+3+4+5+6) = 1 \Rightarrow k = \frac{1}{21}$$

b) Para que sea f.p.p. debe verificarse que sea siempre positiva y que

$$\sum_x \Pr(x) = \sum_{x=1}^{n} f(x) = 1$$

por tanto:

$$k(1+2+...+n) = 1 \Rightarrow k\left(\frac{(1+n)n}{2}\right) = 1 \Rightarrow k = \frac{2}{n^2+n}$$

ya que:

$$\sum_{x=1}^{n} x = \frac{(1+n)n}{2}$$

al tratarse de la suma de una progresión aritmética.

4. Del enunciado deducimos que los resultados del experimento no son equiprobables, **no** podremos usar la regla de Laplace para calcular probabilidades. Primero tenemos que calcular la probabilidad de obtener cara en esa moneda trucada. Si denotamos por CA el suceso = "al lanzar una vez, sale cara", el enunciado nos indica que $P(CA) = 2P(CA^C)$. De ello deducimos que $P(CA) = 2(1 - P(CA))$, y $P(CA) = 2/3$.

   Para describir un resultado del experimento usamos c y $+$ para indicar una cara y cruz respectivamente. El espacio muestral asociado al experimento será $S = \{ccc,\ cc+,\ c+c,\ +cc,\ ++c,\ +c+,\ c++,\ +++\}$

El rango de la v.a $X =$ "*Número de caras menos el de cruces*" es por lo tanto $R_X = \{-3, -1, 1, 3\}$ y la función puntual de probabilidad se obtiene como sigue:

Cálculo de $f_X(-3)$

Denotamos por C_1, C_2 y C_3, los sucesos "sale cara la primera vez que se lanza la moneda", "sale cara la segunda vez que se lanza la moneda", etc..

$f_X(-3) = P(X = -3) = P(C_1^C \bigcap C_2^C \bigcap C_3^C)$. Al suponer los tres lanzamientos de moneda independientes, obtenemos $f_X(-3) = P(C_1^C)P(C_2^C)P(C_3^C) = (\frac{1}{3})^3 = \dfrac{1}{27}$

Cálculo de $f_X(-1)$

$f_X(-1) = P(X = -1)$. El suceso $(X = -1)$ se puede descomponer

$(X = -1) = \left(C_1 \bigcap C_2^C \bigcap C_3^C\right) \bigcup \left(C_1^C \bigcap C_2 \bigcap C_3^C\right) \bigcup \left(C_1^C \bigcap C_2^C \bigcap C_3\right)$. Usando las reglas de adición y del producto deducimos $P(X = -1) = (\frac{2}{3})(\frac{1}{3})^2 + \frac{1}{3}(\frac{2}{3})(\frac{1}{3}) + (\frac{1}{3})^2(\frac{2}{3}) = 3(\frac{2}{3})(\frac{1}{3})^2 = \frac{2}{9}$

El cálculo de $f_X(1)$ y de $f_X(3)$ se realizan de manera muy parecida, y se obtiene $f_X(1) = P(X = 1) = 3 \cdot (\frac{2}{3})^2 \cdot (\frac{1}{3}) = \dfrac{12}{27} = \dfrac{4}{9}$

$f_X(3) = P(X = 3) = (\frac{2}{3})^3 = \dfrac{8}{27}$

La función de distribución dada por $F_X(x) = P(X \leq x) = \sum_{\substack{x_i \leq x \\ x_i \in R_X}} f_X(x_i)$ será:

$$F_X(x) = \begin{cases} 0, & x < -3 \\ \dfrac{1}{27}, & -3 \leq x < -1 \\ \frac{1}{27} + \frac{6}{27} = \dfrac{7}{27}, & -1 \leq x < 1 \\ \frac{1}{27} + \frac{6}{27} + \frac{12}{27} = \dfrac{19}{27}, & 1 \leq x < 3 \\ \frac{1}{27} + \frac{6}{27} + \frac{12}{27} + \frac{8}{27} = 1, & x \geq 3 \end{cases}$$

5. **(a)** Empezamos por describir el experimento aleatorio. Seleccionamos 4 bolas SIN REEMPLAZAMIENTO. Suponemos que podemos distinguir entre las bolas, el orden no importa, y no hay repeticiones posibles porque una misma bola no puede ser elegida dos veces. Los casos posibles asociados a este experimento son las $C_{10,4} = \binom{10}{4}$. Como no se permite reemplazamiento, el número máximo de bolas rojas que podemos extraer es cuatro y como mínimo extraeremos las tres bolas azules y una roja , por lo tanto el rango de X es $R_X = \{1, 2, 3, 4\}$ y la función puntual de probabilidad se obtiene como sigue:

   Cálculo de $f_X(1)$

   $f_X(1) = P(X = 1) = P($"sale una roja y 3 azules"$)$; tengo que escoger una roja entre 7 posibilidades, puedo hacerlo de $\binom{7}{1} = 7$ maneras; después tengo que escoger 3 azules entre

3 posibilidades, puedo hacer lo de $\binom{3}{3} = 1$ manera. Los casos favorables son por lo tanto $\binom{7}{1}\binom{3}{3} = 7$. y $P($"sale una roja y 3 azules"$) = \dfrac{\binom{7}{1}\binom{3}{3}}{\binom{10}{4}} = \dfrac{1}{30}$.

$f_X(2) = P(X = 2) = P($"sale 2 rojas y 2 azules"$) = \dfrac{\binom{7}{2}\binom{3}{2}}{\binom{10}{4}} = \dfrac{3}{10}$

$f_X(3) = P(X = 3) = P($"sale 3 rojas y 1 azul"$) = \dfrac{\binom{7}{3}\binom{3}{1}}{\binom{10}{4}} = \dfrac{15}{30} = \dfrac{1}{2}$

$f_X(4) = P(X = 4) = P($"salen 4 rojas"$) = \dfrac{\binom{7}{4}\binom{3}{0}}{\binom{10}{4}} = \dfrac{5}{30} = \dfrac{1}{6}$

La función de distribución dada por $F_X(x) = P(X \leq x) = \sum_{\substack{x_i \leq x \\ x_i \in R_X}} f_X(x_i)$ será:

$$F_X(x) = \begin{cases} 0, & x < 1 \\ \dfrac{1}{30}, & 1 \leq x < 2 \\ \frac{1}{30} + \frac{9}{30} = \dfrac{1}{3}, & 2 \leq x < 3 \\ \frac{1}{30} + \frac{9}{30} + \frac{15}{30} = \dfrac{5}{6}, & 3 \leq x < 4 \\ \frac{1}{30} + \frac{9}{30} + \frac{15}{30} + \frac{5}{30} = 1, & x \geq 4 \end{cases}$$

(b)

$P(1 \leq X \leq 3) = f_X(1) + f_X(2) + f_X(3) = \dfrac{25}{30} = \dfrac{5}{6}$

$P(1 < X \leq 3) = f_X(2) + f_X(3) = \dfrac{8}{10} = \dfrac{4}{5}$

$P(1 \leq X < 3) = f_X(1) + f_X(2) = \dfrac{10}{30} = \dfrac{1}{3}$

(c) Seleccionamos 4 bolas CON REEMPLAZAMIENTO

En este caso, estamos ante un experimento asociada a una situación dicotómica (la bola escogida es roja o azul) que repetimos 4 veces. La variable $X =$ "número de bolas rojas entre las cuatro escogidas" sigue una distribución Binomial de parámetros $n = 4$ y $p = P($" la bola escogida es roja"$) = \dfrac{7}{10}$.

La función puntual de probabilidad será

x	0	1	2	3	4
$f_X(x)$	0.0081	0.0756	0.2646	0.4116	0.2401

puesto que $P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$

La función de distribución $F_X(x) = P(X \leq x) = \sum_{\substack{x_i \leq x \\ x_i \in R_X}} f_X(x_i)$ será:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0.0081, & 0 \leq x < 1 \\ 0.0081+0.0756 = 0.0837, & 1 \leq x < 2 \\ 0.0081+0.0756+0.2646 = 0.3483, & 2 \leq x < 3 \\ 0.0081+0.0756+0.2646+0.4116 = 0.7599, & 3 \leq x < 4 \\ 0.0081+0.0756+0.2646+0.4116+0.2401 = 1, & x \geq 4 \end{cases}$$

$P(1 \leq X \leq 3) = f_X(1) + f_X(2) + f_X(3) = 0.7518$
$P(1 < X \leq 3) = f_X(2) + f_X(3) = 0.6762$
$P(1 \leq X < 3) = f_X(1) + f_X(2) = 0.3402$

6. Suponiendo que el dado está en perfecto estado, es decir, todos los valores tienen igual probabilidad, si apostamos la cantidad a, la variable aleatoria que mide la ganancia toma los valores $\{2a, a, -a\}$, puesto que si apostamos a pesetas y ganamos el triple de lo apostado, el beneficio será $2a$ pts. En este caso, el valor del juego (ganancia esperada media por jugada) es
$$E(X) = 2a\frac{1}{6} + a\frac{1}{6} - a\frac{4}{6} = -\frac{1}{6}a$$
que no es favorable al jugador.

7. Empecemos por la v.a X. Los valores posibles de X son $35p$ (si sale el número por el que se ha apostado) o $-p$ (si no sale). La probabilidad de que salga el número por el que se ha apostado se cálcula usando la regla de Laplace puesto que los resultados son equiprobables. $P(X = 35p) = P($ sale el n^o escogido$) = 1/37$ mientras que $P(X = -p) = 36/37$. Tenemos por lo tanto $E[X] = \sum_{i=1}^{2} x_i \cdot f_X(x_i) = \frac{1}{37} \cdot 35p + \frac{36}{37} \cdot (-p) = \frac{-p}{37}$
$Var(X) = E[X^2] - (E[X])^2 = \sum_{i=1}^{2} x_i^2 \cdot f_X(x_i) - \left(\frac{-p}{37}\right)^2$
$= \frac{1}{37} \cdot (35p)^2 + \frac{36}{37} \cdot (-p)^2 - \left(\frac{-p}{37}\right)^2 = \frac{46656}{1369}p^2 \simeq 34.1 \cdot p^2$
Sigamos por la v.a Y.

Los valores posibles de Y son p (si sale el color por el que se ha apostado) o $-p$ (si no sale). La probabilidad de que salga el color por el que se ha apostado se cálcula usando la regla de Laplace puesto que los resultados son equiprobables. $P(Y = p) = P($ sale el color escogido$) = 18/37$ mientras que $P(Y = -p) = 19/37$. Tenemos por lo tanto $E[Y] = \sum_{i=1}^{2} y_i \cdot f_Y(y_i) = \frac{18}{37} \cdot p + \frac{19}{37} \cdot (-p) = \frac{-p}{37}$
$Var(Y) = E[Y^2] - (E[Y])^2 = \sum_{i=1}^{2} y_i^2 \cdot f_Y(y_i) - \left(\frac{-p}{37}\right)^2$
$= \frac{18}{37} \cdot (p)^2 + \frac{19}{37} \cdot (-p)^2 - \left(\frac{-p}{37}\right)^2 = \frac{1368}{1369}p^2 \simeq 1.00p^2$
El primer comentario es que en los dos tipos de apuesta, la esperanza de la variable es negativa, es decir que el beneficio esperado es negativo (se pierde dinero, vamos). Además en los dos casos, el beneficio esperado es el mismo pero en cambio, la varianza del beneficio en el primer caso es mucho mayor, es decir que si se llega a ganar se puede ganar mucho más pero que también se puede perder mucho más. Por último, podemos notar que tanto

el beneficio esperado como la dispersión del beneficio alrededor de ese valor es proporcional a p, la cantidad apostada.

8. Sea X la .v.a "Beneficios para la compañía". X puede tomar dos valores $\begin{cases} R \text{ si no ocurre ningún sinie} \\ R - M \text{ si ocurre algún sini} \end{cases}$. Por lo tanto $P(X = R) = P(\text{no ocurre ningún siniestro}) = 1 - p$ y en cambio $P(X = R - M) = P(\text{ocurre algún siniestro}) = p$. Deducimos $E[X] = R \cdot (1-p) + (R-M) \cdot p = R - pM$. $Var(X) = R^2 \cdot (1-p) + (R-M)^2 \cdot p - (R-pM)^2 = p(1-p)M^2$ y $\sigma_X = M\sqrt{p(1-p)}$.

   Observamos que si $R > pM$, el beneficio esperado es positivo, por lo tanto la compañía asegurada debe evaluar el riesgo (p) de siniestro, y luego fijar la prima anual (R). Supongamos por ejemplo que la cantidad asegurada es 100.000pts, y la compañía estima a 0.001 la probabilidad de siniestro. Si impone una prima anual de 1000pts, el beneficio esperado será por lo tanto 1000-100=900pts mientras que la dispersión del beneficio alrededor de ese valor será $100.000\sqrt{0.001 \cdot 0.999} \simeq 3.1607$.

9. (Diciembre 99)

   (a) La suma de las probabilidades tiene que ser igual a 1. Deducimos que el valor que falta es $P(X = 560) = 0.1$.

   (b) Tenemos que $E[X] = \sum_x xP(X = x) = 540 \cdot 0.1 + 545 \cdot 0.25 + \cdots + 560 \cdot 0.1 = 550$. Para calcular la desviación típica de X, calculamos primero su varianza:

   $$\begin{aligned} Var(X) = E[X^2] - (E[X])^2 &= \sum_x x^2 P(X = x) - (E[X])^2 \\ &= \left(540^2 \cdot 0.1 + \cdots + 560^2 \cdot 0.1\right) - 550^2 \\ &= 30.0 \end{aligned}$$

   Deducimos que la desviación típica de X es $\sigma_X = \sqrt{30} \simeq 5.48$

   (c) Sea $D = X - 550$, nos piden $E[D]$ y σ_D. Tenemos que $E[D] = E[X] - 550 = 0$ y $\sigma_D = \sigma_X \simeq 5.48$.

   (d) Tenemos que $E[Y] = E[\left(\frac{9}{5}X + 32\right)] = \frac{9}{5}E[X] + 32 = 1022$, mientras que $\sigma_Y = \frac{9}{5}\sigma_X \simeq 9.86$.

10. **a)**Una función será una densidad si es positiva y su integral (el área debajo de ella) vale 1. Para calcular sus funciones de distribución calcularemos la integral $F(x) = \int_{-\infty}^{x} f(x)dx$.

    Así, puede comprobarse fácilmente que la función (1) no es una densidad ya que su integral vale $\int_0^2 x^2 dx = 8/3$.

    b)La segunda función, dada en (2), depende de la constante k, que deberá ser positiva y verificar

    $$k \int_1^3 (3x + 1)dx = 14k = 1$$

por lo que k debe ser $1/14$. Su función de distribución valdrá

$$F(x) = k \int_1^x (3x+1)dx = \frac{3}{28}x^2 + \frac{1}{14}x - \frac{5}{28}$$

si $1 < x < 3$, $F(x) = 0$ si $x < 1$ y $F(x) = 1$ si $x > 3$.

c) Se procede de forma similar, aunque hay que tener en cuenta que $(x^2 + 2x + 0.5) > 0$ si $0 < x < 2$, ya que sus raíces son -1.70 y -0.29. La ecuación que se obtiene es $7.6667k = 1$, por lo que $k = 0.13043$, y su distribución será $F(x) = k(x^3/3 + x/2 + x^2)$ si $0 < x < 3$.

d) Por otra parte, para $k > 0$, se tiene que la función dada es evidentemente positiva, por lo que debemos resolver

$$k \int_1^\infty x^{-3}dx = k[\lim_{x\to\infty} \frac{x^{-2}}{-2} - \frac{1^{-2}}{-2}] = \frac{1}{2}k = 1$$

lo que da $k = 2$. Su distribución es $F(x) = (x^2 - 1)/x^2$ si $x > 0$.

e) Para comprobar que la última función también es función de densidad, tendremos en cuenta que es positiva y que

$$\int_R f(x)dx = \int_0^1 xdx + \int_1^2 (2-x)dx = 1$$

Además, su distribución vale $F(x) = \int_0^x tdt = x^2/2$ si $0 < x < 1$ y

$$F(x) = \int_0^1 tdt + \int_1^x (2-t)dt = -1 + 2x - \frac{1}{2}x^2$$

si $1 < x < 2$, $F(x) = 0$ si $x < 0$ y $F(x) = 1$ si $x > 2$.

11. Como tenemos la función de distribución, resulta inmediato:

• $\Pr(X < 3) = \Pr(X \leq 3) = F(3) = 1 - \frac{4}{9} = \frac{5}{9}$

• $\Pr(4 < X < 5) = \Pr(4 \leq X \leq 5) = F(5) - F(4) = \left(1 - \frac{4}{25}\right) - \left(1 - \frac{4}{16}\right) = \frac{9}{100}$

Por último como $f(x) = F'(x)$, tenemos

$$f(x) = \begin{cases} 0 & x < 2 \\ \frac{8}{x^3} & x \geq 2 \end{cases}$$

12. Teniendo en cuenta que en los puntos x donde la función de distribución es derivable, se cumple que

$$\frac{dF(x)}{dx} = F'(x) = f(x)$$

resulta

$$f(x) = \begin{cases} nx^{n-1} & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}.$$

Además, utilizando la función de distribución, tenemos que

$$\Pr\left(X \geq \frac{3}{4}\right) = 1 - \Pr\left(X < \frac{3}{4}\right) = 1 - F\left(\frac{3}{4}\right) = 1 - \left(\frac{3}{4}\right)^n$$

Por último, observar que también podemos calcular esta probabilidad a través de su función de densidad:

$$\begin{aligned}\Pr\left(X \geq \frac{3}{4}\right) &= \int_{3/4}^{1} nx^{n-1}dx = n\int_{3/4}^{1} x^{n-1}dx \\ &= n \cdot \left[\frac{x^n}{n}\right]_{3/4}^{1} = 1 - \left(\frac{3}{4}\right)^n\end{aligned}$$

13. **a)** Dado que las dos condiciones que debe cumplir una función $f(x)$ para ser función de densidad son:

$$f(x) \geq 0, \text{ para cualquier } x \in \mathbb{R}$$

y

$$\int_{\mathbb{R}} f(x)\, dx = 1$$

se deduce de la primera condición que $k > 0$. Además, de la segunda condición, tenemos que

$$\begin{aligned}1 &= \int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{1} k(1-x^3)dx = k\int_{0}^{1}(1-x^3)dx \\ &= k\left[x - \frac{x^4}{4}\right]_0^1 = k\left(1 - \frac{1}{4}\right) = \frac{3}{4}k\end{aligned}$$

es decir,

$$k = \frac{4}{3}$$

b) Como la función de distribución asociada a una variable aleatoria está determinada por la probabilidad acumulada, y en el caso continuo, representada por el área encerrada por su función de densidad, tenemos que:

$$F(x) = \Pr(X \leq x) = \int_{0}^{x} f(t)dt, \forall x \in \mathbb{R}$$

Por tanto,

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \displaystyle\int_0^x \frac{4}{3}(1-t^3)dt = \frac{4x - x^4}{3} & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

c) A partir de la función de distribución obtenida en el apartado anterior, se obtiene que

$$\Pr\left(X \leq \frac{1}{2}\right) = F\left(\frac{1}{2}\right) = \frac{4 \cdot (\frac{1}{2}) - (\frac{1}{2})^4}{3} = 0.6458$$

$$\begin{aligned}
\Pr\left(\frac{1}{2} \leq X \leq \frac{3}{4}\right) &= F\left(\frac{3}{4}\right) - F\left(\frac{1}{2}\right) \\
&= \left(\frac{4 \cdot (\frac{3}{4}) - (\frac{3}{4})^4}{3}\right) - \left(\frac{4 \cdot (\frac{1}{2}) - (\frac{1}{2})^4}{3}\right) = 0.2487
\end{aligned}$$

y

$$\begin{aligned}
\Pr\left(X > \frac{3}{4}\right) &= 1 - \Pr\left(X \leq \frac{3}{4}\right) = 1 - F\left(\frac{3}{4}\right) \\
&= 1 - \left(\frac{4 \cdot (\frac{3}{4}) - (\frac{3}{4})^4}{3}\right) = 0.1055
\end{aligned}$$

Dpto. Matemática Aplicada y Estadística

Ingeniero Técnico de Minas
Asignatura: **Estadística**
Soluciones de la hoja de problemas 4

Modelos de Distribuciones Univariantes

1. Empezamos por reconocer que la densidad descrita corresponde a una variable exponencial con parámetro (media) $\lambda = 20$. Su función de distribución es $F(t) = 1 - e^{-\frac{1}{\lambda}t}$ si $t \geq 0$ y 0 en otro caso. Por lo tanto

   a)

$$\Pr(X \leq 10) = \left[1 - e^{-1/2}\right] = 1 - \frac{1}{\sqrt{e}} = 0.3934$$

   b) Para que la duración del neumático esté entre 16000 y 24000 kilómetros, obtenemos

$$\begin{aligned} \Pr(16 < X < 24) &= \Pr(16 < X \leq 24) \\ &= \left(1 - e^{-24/20}\right) - \left(1 - e^{-16/20}\right) = 0.1481 \end{aligned}$$

   c) Por último, la probabilidad de que dure más de 30000 kilómetros es

$$\begin{aligned} \Pr(X \geq 30) &= 1 - \Pr(X \leq 30) \\ &= 1 - \left(1 - e^{-30/20}\right) = 1 - 0.7768 = 0.\,2232 \end{aligned}$$

   d) Para una variable exponencial la media es igual a λ. Deducimos que la duración media es igual a 20000km.

2. **a)** Si X_1 es el número de gotas caídas en un cuarto de hora, sabemos que $X_1 \sim P(\lambda)$ con $\lambda = 4$, y nos piden $P(X_1 > 5)$. Pero $P(X_1 > 5) = 1 - P(X_1 \leq 5) =$ $1 - (P(X_1 = 0) + P(X_1 = 1) + \cdots P(X_1 = 5)) = 1 - (\sum_{i=0}^{5} \frac{\lambda^i e^{-\lambda}}{i!}) = 0.21$

   b) Si el número de gotas caídas en un cuarto de hora sigue una distribución de Poisson de media 4, el número de gotas caídas en tres cuartos de hora sigue una distribución de Poisson de media 12. Sea X_3 el número de gotas caídas en tres cuartos de hora, tenemos $X_3 \sim P(12)$ y nos piden $P(13 \leq X_3 \leq 18)$.
   $P(13 \leq X_3 \leq 18) = P(X_3 = 13) + P(X_3 = 14) + \cdots P(X_3 = 18)$
   $= e^{-12}(\frac{12^{13}}{13!} + \frac{12^{14}}{14!} + \cdots + \frac{12^{18}}{18!}) \simeq 0.39$

   c) Sea X_1 y X_2 el número de gotas caídas en el primer y segundo cuarto de hora respectivamente. Sabemos que $X_1 \sim P(4)$ y $X_2 \sim P(4)$. Nos piden $P(X_1 = 3 \bigcap X_2 = 3)$. Los dos sucesos $X_1 = 3$ y $X_2 = 3$ son independientes por lo tanto, por la regla del producto tenemos que $P(X_1 = 3 \bigcap X_2 = 3) = P(X_1 = 3)P(X_2 = 3) = (\frac{e^{-4}4^3}{3!})^2 \simeq 0.04$.

3. (a) Si X es el peso del paquete, el coste de producción de ese paquete es $X \times 0.5$ pts, mientras que el precio de venta del paquete es $500 \times 0.9 = 450$ pts, por lo tanto

$$B = 450 - 0.5X.$$

Deducimos que el beneficio promedio por paquete es $E[B] = 450 - 0.5E[X] = 201pts$.

(b) Puesto que $X \sim N(498; \sqrt{16})$, sabemos que $B = 450 - 0.5X$ también sigue una distribución normal. Por el apartado anterior, su media es 201 y su varianza es $Var(B) = Var(0.5X) = (0.5)^2 Var(X) = 4$. Nos piden $P(B > 200)$. Tipificando, $P(B > 200) = P(\frac{B-201}{\sqrt{4}} > \frac{200-201}{\sqrt{4}}) = 1 - \phi(-0.5) = \phi(0.5) = 0.69$. La empresa realiza un beneficio de más de 200 pts en aprox. 69% de la producción.

4. (a) Definamos la variable:

$$X = \text{"Número de piezas defectuosas en un lote de 100 unidades"}$$

Al escoger una pieza tenemos una situación dicotómica con dos sucesos posibles $A =$"la pieza es defectuosa" o $A^C =$"la pieza es correcta". Para completar un lote hay que repetir 100 veces la elección, como la probabilidad de que una pieza sea defectuosa es de 0.01 si tiene que:

$$X \sim B(100, 0.01)$$

por tanto:

$$\begin{aligned} P(X &\geq 2) = 1 - P(X \leq 1) = 1 - [P(X = 0) + P(X = 1)] \\ &= 1 - (0.99)^{100} - 100 \cdot 0.01 \cdot (0.99)^{99} \simeq 0.26 \end{aligned}$$

(b) La probabilidad de que un lote sea rechazado será: (utilizando los mismos cálculos que en el apartado anterior) $\boxed{P(rechazar) = P(X > 5) = 1 - P(X \leq 5) \simeq 0.0006}$
Si denotamos por Y a la variable:

$$Y = \text{"Número de lotes rechazados diariamente"}$$

Al escoger un lote, estamos antes una situación dicotómica con dos sucesos posibles : $A =$ "se rechaza el lote", y $A^C =$"se acepta el lote". Repetimos el experimento 100 veces, y deducimos que $Y \sim B(100, 0.0006)$, por consiguiente su esperanza será:

$$\boxed{E(Y) = n \cdot p = 100 \cdot 0.0006 = 0.06}$$

5. Sea:

$$X = \text{"Número de componentes que han fallado } (X = 0, 1, ..., 10)\text{"}$$

Al considerar un componente tenemos una situación dicotómica con $A =$"el componente ha fallado" con $p = P(A) = 0.3$.Se repite el exp. 10 veces y deducimos que $X \sim B(10, 0.3)$.

Nos piden $P(X \geq 2|X \geq 1)$.

$$\begin{aligned}
P(X &\geq 2|X \geq 1) = \frac{P(X \geq 2 \cap X \geq 1)}{P(X \geq 1)} = \frac{P(X \geq 2)}{P(X \geq 1)} \\
&= \frac{1 - P(X \leq 1)}{1 - P(X \leq 0)} = \frac{1 - [P(X = 0) + P(X = 1)]}{1 - P(X = 0)} \\
&= \frac{1 - \left[\binom{10}{0}0.3^0 0.7^{10} + \binom{10}{1}0.3^1 0.7^9\right]}{1 - \binom{10}{0}0.3^0 0.7^{10}} = \frac{0.8507}{0.9718} \simeq \boxed{0.87}
\end{aligned}$$

6. (a) Definamos la variable:

$$X = \text{"Número de bujias inservibles en un paquete } (X = 0, 1, 2, 3, 4)\text{"}$$

Al considerar una bujía, tenemos una situación dicotómica con dos sucesos $A =$"bujía inservible" y $A^C =$"bujía correcta" . Tenemos que $p = P(A) = 0.2$. X es el número de veces que ha ocurrido A. Deducimos $X \sim B(4, 0.2)$, por consiguiente:

$$P(X = i) = \binom{4}{i} 0.2^i \cdot (1 - 0.2)^{4-i}$$

y de ahí:

$$\begin{aligned}
P(X &\geq 2) = P(X = 2) + P(X = 3) + P(X = 4) \\
&= \binom{4}{2} 0.2^2 \cdot (1 - 0.2)^{4-2} + \binom{4}{3} 0.2^3 \cdot (1 - 0.2)^{4-3} + \binom{4}{4} 0.2^4 \cdot (1 - 0.2)^{4-4} \\
&= 0.1536 + 0.0265 + 0.0016 \simeq \boxed{0.18}
\end{aligned}$$

(b) Definamos la nueva variable:

$$Y = \text{"Número de defectuosos en una caja de 10 paquetes"}$$

El contexto es el mismo que en el apartado anterior pero ahora tenemos $10 \times 4 = 40$ bujías que considerar,deducimos por lo tanto:

$$Y \sim B(40, 0.2)$$

y por consiguiente:

$$P(Y > 10) = 1 - P(Y \leq 10) = 1 - \sum_{i=1}^{10} \binom{40}{i} 0.2^i \cdot (1 - 0.2)^{40-i}$$

Puesto que $np = 8$ y $n(1 - p) = 32$ estamos en las condiciones de aproximación satisfactoria de la binomial $B(40, 0.2)$ por una normal con parámetros $\mu = n \cdot p = 8$ y $\sigma^2 = n \cdot p \cdot (1 - p) = 6.4$. Usando la correción por continuidad, tenemos que si $W \sim N(8, \sqrt{6.4})$,
$P(Y > 10) \simeq P(W > 10.5)$. Nos falta tipificar y, usando la tabla, encontramos :
$P(Y > 10) \simeq P(\frac{W-8}{\sqrt{6.4}} > \frac{10.5-8}{\sqrt{6.4}}) = P(Z > 0.99) = 1 - \Phi(0.99) \simeq 0.161$

(c) Por último, si denotamos por:

$$Z = \text{"Número de paquetes sin bujias defectuosas } (Z = 0, 1, ..., 10)\text{"}$$

Al considerar un paquete, tenemos una situación dicotómica, con dos sucesos $A =$"el paquete no contiene bujías defectuosas" y $A^C =$"el paquete contiene al menos una bujía defectuosa". Para completar una caja, tenemos que repetir la elección de un paquete 10 veces. deducimos que $Z \sim B(10, p)$. Donde $p = P(A)$. Usando la notación del apartado a), es fácil ver que $p = P(A) = P(X = 0) = \binom{4}{0}0.2^0 0.8^4 \simeq 0.41$. Por lo tanto tenemos que:

$$Z \sim B(10, 0.4096)$$

de donde:

$$P(Z = 3) = \binom{10}{3} 0.4096^3 \cdot (1 - 0.4096)^{10-3} \simeq \boxed{0.21}$$

7. Definamos la variable:

$$X = \text{"Número de tornillos defectuosos en una caja de 170 } (X = 0, 1, ..., 170)\text{"}$$

Al igual que en los problemas anteriores, es fácil comprobar que, como la probabilidad de que un tornillo sea defectuoso es de $2/85$ ($\simeq 0.024$), se tiene que:

$$X \sim B(170, 0.024)$$

Definimos ahora la v.a

$$Y = \text{"Número de cajas con cero defectuosas en una muestra de tamaño 7"}$$

Al escoger una caja, tenemos una situación dicotómica, con dos sucesos posibles $A =$"la caja no contiene tornillos defectuosos" y $A^C =$"la caja contiene al menos un tornillo defectuoso". Repetimos el experimento 7 veces, y deducimos que

$$Y \sim B(7, p)$$

con $p = P(A) = P(X = 0) = \binom{170}{0}0.024^0(1 - 0.024)^{170} = 0.01\,6$
Por lo tanto

$$Y \sim B(7, 0.0183)$$

Entonces:

$$P(Y = 2) = \binom{7}{2} 0.01\,6^2 \cdot (1 - 0.016)^{7-2} = 0.005$$

8. (a) $P(X \leq 12) = P(\frac{X - 10}{2} \leq \frac{12 - 10}{2}) = P(Z \leq 1) = \Phi(1) = 0.841345$ donde Φ es la función de distribución de la variable normal $N(0, 1)$ que está tabulada.

(b) $P(\mid X \mid\leq 3) = P(-3 \leq X \leq 3) = P(\frac{-3-10}{2} \leq \frac{X-10}{2} \leq \frac{3-10}{2})$
$= P(-6.5 \leq Z \leq -3.5) = \Phi(-3.5) - \Phi(-6.5) = [1-\Phi(3.5)] - [1-\Phi(6.5)]$
$= \Phi(6.5) - \Phi(3.5) = 1 - 0.999676 = 0.000233$. (Notar que $\Phi(6.5)$ es aproximadamente 1)

(c) $P(\mid X-10 \mid\leq 1) = P(-1 \leq X-10 \leq 1) = P(\frac{-1}{2} \leq \frac{X-10}{2} \leq \frac{1}{2}) = P(\frac{-1}{2} \leq Z \leq \frac{1}{2}) = \Phi(\frac{1}{2}) - \Phi(-\frac{1}{2}) = \Phi(\frac{1}{2}) - \left[1-\Phi(\frac{1}{2})\right]$
$= 2\Phi(\frac{1}{2}) - 1 = 2 \times 0.691462 - 1 = 0.382924$

(d) Nos piden encontrar a tal que $P(\mid X-10 \mid\leq a) = 0.95$ $\underline{0.95}= P(\mid X-10 \mid\leq a) = P(-a \leq X-10 \leq a) = P(-\frac{a}{2} \leq \frac{X-10}{2} \leq \frac{a}{2}) = P(-\frac{a}{2} \leq Z \leq \frac{a}{2}) =$
$= \Phi(\frac{a}{2}) - \Phi(-\frac{a}{2}) = \Phi(\frac{a}{2}) - \left[1-\Phi(\frac{a}{2})\right] = \underline{2\Phi(\frac{a}{2}) - 1}$ De donde, $0.95 = 2\Phi(\frac{a}{2}) - 1 \Rightarrow$
$\frac{0.95+1}{2} = \Phi(\frac{a}{2}) \Rightarrow \quad \Phi(\frac{a}{2}) = 0.975$
Mirando en la tabla de la $N(0,1)$ tenemos que para el punto de abscisa 1.96, el área acumulada bajo la curva de la función de densidad es 0.975. De donde, $\frac{a}{2} = 1.96 \Rightarrow \boxed{a = 3.92}$

9. (a) Estamos interesados en la v.a $Y =$*Número de galletas incorrectas en un paquete de 15 galletas*. Estamos antes una situación dicotómica : al escoger una galleta, tenemos dos sucesos posibles $A =$"la galleta es incorrecta", o $A^C =$"la galleta es correcta". Repetimos la elección 15 veces y nos interesamos por $Y =$*Número de veces que ha ocurrido A*"
Deducimos que Y es una variable Binomial con parámetros $n = 15$ y $p = P($*Obtener una galleta incorrecta*$)$ *i, e* $\boxed{Y \sim B(15, p)}$. Debemos calcular p para seguir. Introducimos la variable $X=$ "peso de una galleta producida". Tenemos $p = P($*galleta incorrecta*$) = 1 - P($ *galleta correcta*$) = P(1.9 \leq X \leq 2.1)$. Tipificamos para encontrar

$$P(1.9 \leq X \leq 2.1) = P(-1.78 \leq Z \leq 1.78) = 2\phi(1.78) - 1 \simeq 0.92$$

Por lo tanto $p \simeq 1 - 0.92 \simeq 0.08$. Sabemos ahora que $\boxed{Y \sim B(15, 0.08)}$. Nos piden la probabilidad $P(Y \geq 2) = 1 - P(Y < 2) =$
$= 1 - [P(Y = 0) + P(Y = 1)] =$

$= 1 - \left[\binom{15}{0} (0.08)^0 (0.92)^{15} + \binom{15}{1} (0.08)^1 (0.92)^{14}\right]$
$\simeq 1 - 0.660 \simeq 0.340$

(b) Ahora tenemos $X \sim N(2, \sigma)$ y nos piden cual debe de ser la desviación típica para que la probabilidad de obtener una galleta incorrecta sea $P($*Obtener una galleta incorrecta*$) = P(\mid X-2 \mid> 0.1) = 0.006$.

$P(\mid X-2 \mid > 0.1) = 1 - P(\mid X-2 \mid \leq 0.1) = 1 - P(-0.1 \leq X-2 \leq 0.1) =$
$= 1 - P(\frac{-0.1}{\sigma} \leq \frac{X-2}{\sigma} \leq \frac{0.1}{\sigma}) = 1 - P(\frac{-0.1}{\sigma} \leq Z \leq \frac{0.1}{\sigma})$
$= 1 - \left(\Phi(\frac{0.1}{\sigma}) - \Phi(-\frac{0.1}{\sigma})\right) = 1 - \Phi(\frac{0.1}{\sigma}) + \left(1 - \Phi(\frac{0.1}{\sigma})\right) =$
$= 2 - 2\Phi(\frac{0.1}{\sigma})$
De donde nos queda que, $\quad 0.006 = 2\Phi(\frac{0.1}{\sigma}) - 1 \Rightarrow \Phi(\frac{0.1}{\sigma}) = \frac{0.994+1}{2} = \frac{1.994}{2} =$ 0.997
En la tabla de la $N(0,1)$ vemos que para el punto de abscisa de 2.75, el área acumulada bajo de la curva de la función de densidad es de 0.997. Entonces, $\frac{0.1}{\sigma} = 2.75 \Rightarrow$
$\boxed{\sigma = 0.036}$

(c) Ahora tenemos que la variable X se distribuye como en el apartado (b), esto es, $X \backsim N(2, \sigma)$ con $\sigma = 0.036$
$(0.036)^2 = 1.296 \times 10^{-3}$
Por lo que las probabilidad de obtener una galleta incorrecta es
$P(\textit{Obtener una galleta incorrecta}) = 0.006$
Nos piden la probabilidad que en 100 paquetes haya por lo menos 20 galletas defectuosas, es decir la probabilidad de que entre 1500 galletas, haya por lo menos 20 defectuosas. Si definimos la variable $W =$ *"Número de galletas incorrectas entre 1500"*, de la misma manera que en el apartado a), deducimos que $\boxed{W \sim B(1500, 0.006)}$
Entonces la probabilidad que nos piden sería
$P(W \geq 20) = 1 - P(W < 20) = 1 - P(W \leq 19)$
$= 1 - \sum_{i=1}^{19} \left[\binom{1500}{i} (0.006)^i (0.994)^{1500-i}\right]$
para facilitar el cálculo anterior,observamos que, puesto que $np = 9$ y $n(1-p) = 1491$, estamos en las condiciones de aproximación satisfactoria de una binomial por una normal con parámetros $\mu = np = 9$ y $\sigma^2 = np(1-p) = 8.946$. Si $V \sim N(9, \sqrt{8.946})$, usando la corrección de continuidad $P(W \geq 20) \simeq P(V \geq 19.5)$. Nos falta tipificar para encontrar $P(W \geq 20) \simeq P(\frac{V-9}{\sqrt{8.946}} \geq \frac{19.5-9}{\sqrt{8.946}}) = P(Z \geq 3.510)$
$= 1 - \Phi(3.510) = 0.0002$

10. (a) Introducimos la variable V, volumen envasado en una botella. La distribución de V es $N(150, \sqrt{4})$. La probabilidad de que una botella envasada por la fábrica no se pueda vender es $P(V \leq 147)$. Para calcular esta probabilidad tipificamos la variable V :

$$P(V \leq 147) = P(\frac{V-150}{2} \leq \frac{147-150}{2}) = P(Z \leq -1.5) \simeq 0.067$$

La proporción de invendibles es por lo tanto 6.7%.

(b) Nos encontramos ante la repetición 6 veces de un experimento dicotómico. Si consideramos la variable X número de botellas con menos de 147cl en un paquete de 6 botellas, deducimos que sigue una distribución Binomial con parámetros $n = 6$, y $p = P(V \leq 147) = 0.067$. Nos piden $P(X \geq 1)$ que será más fácil calcular a través del suceso complementario: $P(X \geq 1) = 1 - P(X = 0) = 1 - (1 - 0.067)^6 = 0.34$

(c) Llamemos Y la variable "número de botellas invendibles producidas en un día".Esta vez, repetimos 10000 veces el experimento dicotómico con los resultados posibles "la botella contiene menos de 147cl", y "la botella contiene más de 147cl". Al igual que en el apartado anterior, deducimos que la distribución de Y es Binomial, con parámetros, $n = 10000$, y $p = 0.067$. Nos piden $P(Y > 600)$. Para calcular esta probabilidad, es conveniente utilizar la aproximación normal a la Binomial, o utilizar una calculadora estadística como NCSSCALC. Con la aproximación normal: Y es aproximadamente normal con media $np = 670$ y varianza $np(1-p) = 625.11$. Tenemos por lo tanto que

$$\begin{aligned} P(Y &> 600) \simeq P\left(Z > \frac{600.5 - 670}{\sqrt{625.11}}\right) \simeq P\left(Z > -2.77\right) \\ &= 1 - \Phi(-2.77) = \Phi(2.77) \simeq 0,997 \end{aligned}$$

donde hemos utilizado la corrección de continuidad.

Si utilizamos la calculadora estadística NCSScalc, encontramos, con tres decimales, para una v.a. Binomial $B(10000, 0.067)$, $P(Y > 600) \simeq$0,997 , lo que coincide con la aproximación anterior.

(d) En este apartado estamos interesados en la variable $W =$ "número de días en el mes en los que se producen más de 600 botellas invendibles". Otra vez estamos ante la repetición de un experimento dicotómico: en un día, se pueden producir más de 600 botellas invendibles o no. W sigue por lo tanto una distribución Binomial con parámetros 30 y $p = P($ en un día se producen más de 600 botellas invendibles$) \simeq 0.997$. La media de W es $np \simeq 29.91$

11. Realizamos el test para detectar el síntoma de irritabilidad ocular, se trata de una situación dicotómica, puesto que tenemos dos sucesos posibles : $A =$" el trabajador padece irritabilidad en los ojos" y $A^C =$ " el trabajador no padece irritabilidad en los ojos". Del enunciado deducimos que $p = P(A) = 0.6$. Repetimos el test 1100 veces de manera independiente, y denotamos por X el número de trabajadores que padecen irritabilidad. Deducimos que X es una v.a Binomial con $n = 1100$ y $p = 0.6$, que se abrevia por $B(1100, 0.6)$

    (a) Nos piden $P(X \geq 680)$, puesto que $np = 660 > 5$, y $n(1-p) = 440 > 5$, estamos en las condiciones de aproximación satisfactoria de la distribución $B(1100, 0.6)$ por una distribución normal $N(\ np, \sqrt{np(1-p)}) = N(\ 660, \sqrt{264})$. Sea Y una v.a $N(660, \sqrt{264})$, usando la correción por continuidad, tenemos que

    $$P(X \geq 680) \simeq P(Y \geq 679.5)$$

    Pero $P(Y \geq 679.5) = P(\frac{Y-660}{\sqrt{264}} \geq \frac{679.5-660}{\sqrt{264}}) = P(Z \geq 1.2)$, donde Z es una v.a $N(0,1)$ y $P(Z \geq 1.2) = 1 - \phi(1.2) \simeq 1 - 0.885 = 0.115$, lo que implica $\boxed{P(X \geq 680) \simeq 0.115}$

(b) Nos piden $P(670 < X < 675)$ que podemos aproximar usando la corrección de continuidad por $P(670.5 < Y < 674.5) \simeq 0.0729$, o podemos calcular de manera exacta como $P(X = 671) + \cdots P(X = 674) \simeq 0.0732$

(c) Nos piden el valor esperado de X, i.e $E[X]$.Sabemos que para une v.a $B(n, p)$, $E[X] = np$, por lo tanto $\boxed{E[X] = 660}$.

12. Podemos describir la situación de la manera siguiente, realizamos un test dicotómico con dos sucesos posibles $A =$ "el artículo es defectuoso" y $A^C =$ "el artículo no es defectuoso" . Deducimos del enunciado que $p = P(A) = 0.02$.

(a) En ese apartado, repetimos el test $n =$1000 veces. El número de veces que ocurre A, es una v.a X que sigue una distribución Binomial de parámetros $n = 1000$ y $p = 0.02$, que denotamos por $B(1000, 0.02)$. Nos piden primero $P(20 \leq X \leq 30)$ y $P(X = 20)$. Puesto que $np = 20 > 5$, y $n(1-p) = 980 > 5$, estamos en las condiciones de aproximación satisfactoria de la distribución $B(1100, 0.6)$ por una distribución normal $N(\ np, \sqrt{np(1-p)}) = N(\ 20, \sqrt{19.6})$. Sea Y una v.a $N(20, \sqrt{19.6})$, usando la correción por continuidad, tenemos que

$$P(20 \leq X \leq 30) \simeq P(19.5 \leq Y \leq 30.5)$$

Pero $P(19.5 \leq Y \leq 30.5) = P(\frac{19.5-20}{\sqrt{19.6}} \leq \frac{Y-20}{\sqrt{19.6}} \leq \frac{30.5-20}{\sqrt{19.6}}) = P(-0.11 \leq Z \leq 2.37)$
$= \phi(2.37) - \phi(-0.11) \simeq 0.535$
Calculemos ahora $P(X = 20)$. Es posible calcularla de manera exacta, puesto que $P(X = 20) = \binom{1000}{20}(0.02)^{20}(0.98)^{980} \simeq 0.0088$ pero tenemos que disponer de una calculadora que admita esos cálculos. Cómo en el caso anterior, podemos calcular esa probabilidad usando la aproximación normal :

$$P(X = 20) \simeq P(19.5 \leq Y \leq 20.5)$$

De manera similar al cálculo anterior, obtenemos que $P(19.5 \leq Y \leq 20.5) = \phi(0.11) - \phi(-0.11) \simeq 0.088$
En conclusión,

$$\begin{aligned} P(20 &\leq X \leq 30) \simeq 0.535 \\ P(X &= 20) \simeq 0.088 \end{aligned}$$

(b) En este apartado se repite el test 8000 veces, la v.a $X =$n^o de artículos defectuosos sigue una distribución $B(8000, 0.02)$. Nos piden el valor esperado de X, i.e $E[X]$. Sabemos que $E[X] = np = 8000 \cdot 0.02 = 160$. Por lo tanto esperamos que a lo largo de la semana se produzcan 160 artículos defectuosos.

Dpto. Matemática Aplicada y Estadística

Muestreo y Distribuciones Muestrales

1. Si llamamos X a la variable contenido en ml de una botella de gaseosa, X sigue una distribución normal de media $\mu = 298$ y desviación típica $\sigma = 3$.

   a)Nos piden:

$$P(X < 295) = P(\frac{X-298}{3} < \frac{295-298}{3}) = P(Z < \frac{295-298}{3}) =$$
$$= \phi(-1) = 1 - \phi(1) = 0.16$$

   b)Si llamamos $\overline{X}_6 = (\sum_{i=1}^{6} X_i)/6$ a la variable contenido promedio de 6 botellas, por las propiedades de la distribución normal, $\overline{X}_6$ sigue un distribución normal de media $\mu_{\overline{X}_6} = 298$ y desviación típica $\sigma_{\overline{X}_6} = \sigma/\sqrt{6} = 3/\sqrt{6}$. Y la probabilidad requerida se calcula a partir de dicha distribución de la siguiente forma:

$$P(\overline{X}_6 < 295) = P(\frac{\overline{X}_6 - 298}{3/\sqrt{6}} < \frac{295-298}{3/\sqrt{6}}) = P(Z < \frac{295-298}{3/\sqrt{6}}) =$$
$$= \phi(-2.45) = 1 - \phi(2.45) = 0.007$$

2. a) Usando las propiedades de la distribución normal, el peso promedio de los tres pesos realizados en el laboratorio, $\overline{x}$, suponiendo que se realizan de forma independiente, tendrá una distribución normal con desviación típica:

$$\sigma_{\overline{x}} = \sigma/\sqrt{n} = 10/\sqrt{3} = 5.77$$

   b)Usando la anterior igualdad:

$$\sigma_{\overline{x}} = \sigma/\sqrt{n} = 10/\sqrt{n} = 5$$

   de donde se deduce $n = 4$. Hay que repetir la medición 4 veces.

3. a)Si el tamaño de la muestra n es grande y se verifica la condición $np(1-p) > 5$, usando el Teorema Central del Limite, la proporción muestral, $\hat{p}$, seguirá una distribución aproximadamente normal de media $p = 0.59$ y desviación típica $\sqrt{p(1-p)/n} = 0.4918/\sqrt{n} = 0.028$. Nos piden:

$$P(|\hat{p} - 0.59| < 0.03) = P(0.59 - 0.03 < \hat{p} < 0.59 + 0.03) =$$
$$P(-0.03/0.028 < (\hat{p} - 0.59)/0.028 < 0.03/0.028) = 2\phi(1.056) - 1 = 0.71$$

   b) Hemos de repetir los calculos del apartado a) sustituyendo el tamaño de la muestra n:

- Para $n = 600$ la desviación típica será $\sqrt{p(1-p)/n} = 0.4918/\sqrt{n} = 0.0201$, y

$$P(|\hat{p} - 0.59| < 0.03) = P(0.59 - 0.03 < \hat{p} < 0.59 + 0.03) =$$
$$P(-0.03/0.0201 < (\hat{p} - 0.59)/0.0201 < 0.03/0.0201) = 2\phi(1.49) - 1 = 0.86$$

- Para $n = 1200$ la desviación típica será $\sqrt{p(1-p)/n} = 0.4918/\sqrt{n} = 0.0142$, y

$$P(|\hat{p} - 0.59| < 0.03) = P(0.59 - 0.03 < \hat{p} < 0.59 + 0.03) =$$
$$P(-0.03/0.0142 < (\hat{p} - 0.59)/0.0142 < 0.03/0.0142) = 2\phi(2.11) - 1 = 0.97$$

Al aumentar el tamaño muestral la varianza de la distribución de la proporción muestral disminuye y los valores obtenidos se encuentran más concentrados en torno a la media de la población con lo que la extrapolación de los resultados a la población entera es más fiable (la probabilidad de que esten muy cerca ambos valores es mayor).

4. a) Si llamamos X a la variable resultado de la medición, X sigue una distribución normal con desviación típica $\sigma = 0.1$ y media μ desconocida. Nos piden:

$$P(|X - \mu| > 0.1) = 1 - P(|X - \mu| \leq 0.1) = 1 - P(-0.1 < X - \mu < 0.1) =$$
$$= 1 - P(-0.1 + \mu < X < 0.1 + \mu) = 1 - P(-0.1/0.1 < (X - \mu)/0.1 < 0.1/0.1) =$$
$$= 1 - P(-1 < Z < 1) = 1 - (2\phi(1) - 1) = 2(1 - \phi(1)) = 0.079$$

Si repetimos la medición $n = 5$ veces, la variable media muestral de las cinco mediciones, $\overline{X}_5$, por las propiedades de la distribución normal seguirá una distribución normal de media μ y desviación típica $\sigma_{\overline{X}_5} = 0.1/\sqrt{n} = 0.045$:

$$P(|\overline{X}_5 - \mu| > 0.1) = 1 - P(|\overline{X}_5 - \mu| \leq 0.1) = 1 - P(-0.1 < \overline{X}_5 - \mu < 0.1) =$$
$$= 1 - P(-0.1 + \mu < \overline{X}_5 < 0.1 + \mu) = 1 - P(-0.1/0.045 < (\overline{X}_5 - \mu)/0.045 < 0.1/0.045) =$$
$$= 1 - P(-2.24 < Z < 2.24) = 1 - (2\phi(2.24) - 1) = 2(1 - \phi(2.24)) = 0.025$$

5. Si suponemos que las condiciones son normales y la maquina esta bien ajustada, la proporción de piezas defectuosas producidas por la maquina será de $p = 0.01$. Sea $\hat{p}$ la proporción de piezas defectuosas en una muestra de tamaño $n = 100$, es decir, $\hat{p} = \frac{S_{100}}{100}$ con $S_{100} = \sum_{i=1}^{100} X_i$. Sabemos que la v.a. S_{100} mide el número de piezas de la muestra que son defectuosas y sigue una distribución Binomial $B(n = 100, p = 0.01)$. Nos piden:

$$P(\hat{p} > 0.02) = P(\frac{S_{100}}{100} > 0.02) = P(S_{100} > 2) =$$
$$1 - P(S_{100} = 0) - P(S_{100} = 1) - P(S_{100} = 2) = 0.08$$

Si hacemos uso del Teorema Central del Limite, la proporción muestral $\hat{p}$ seguirá una distribución aproximadamente normal de media $p = 0.01$ y desviación típica $\sigma_{\hat{p}} = \sqrt{p(1-p)/n} =$

0.0099. Sin embargo, para tener cierta seguridad en la eficacia de tal aproximación, debemos comprobar si se cumple la condición $np(1-p) > 5$, pero en este caso no se cumple. Si calculamos la probabilidad del enunciado haciendo uso de la aproximación Normal obtenemos:

$$P(\hat{p} > 0.02) = P(\frac{\hat{p} - 0.01}{0.0099} > \frac{0.02 - 0.01}{0.0099}) = 1 - \phi(\frac{0.02 - 0.01}{0.0099}) = 0.16$$

resultado que difiere bastante del obtenido inicialmente. Por tanto, en este caso no es aconsejable usar la aproximación Normal.

Si en un día tres piezas resultan defectuosas, podríamos concluir que la máquina no está bien ajustada porque en caso de estarlo la probabilidad de obtener más de un 2% de piezas defectuosas es muy baja y en ese caso se ha obtenido un 3%.

6. Teniendo en cuenta que le peso de cada individuo tiene una distribución normal $N(\mu = 71, \sigma = 7)$, si seleccionamos una muestra aleatoria de 4 personas, tenemos que

$$\begin{aligned}
\Pr\left(\sum_{i=1}^{4} X_i > 300\right) &= \Pr\left(\frac{\sum_{i=1}^{4} X_i}{4} > \frac{300}{4}\right) = \Pr\left(\overline{X} > 75\right) \\
&= \Pr\left(\frac{\overline{X} - 71}{\frac{7}{\sqrt{4}}} > \frac{75 - 71}{\frac{7}{\sqrt{4}}}\right) = \Pr\left(Z > 1.1429\right) \\
&= 1 - \Pr\left(Z \leq 1.1429\right)
\end{aligned}$$

donde Z tiene una distribución normal estándar, y por tanto,

$$\Pr\left(\sum_{i=1}^{4} X_i > 300\right) = 1 - 0.8735 = 0.1265$$

Dpto. Matemática Aplicada y Estadística

Ingeniero Técnico de Minas
Asignatura: **Estadística**
Soluciones de la hoja de problemas 6

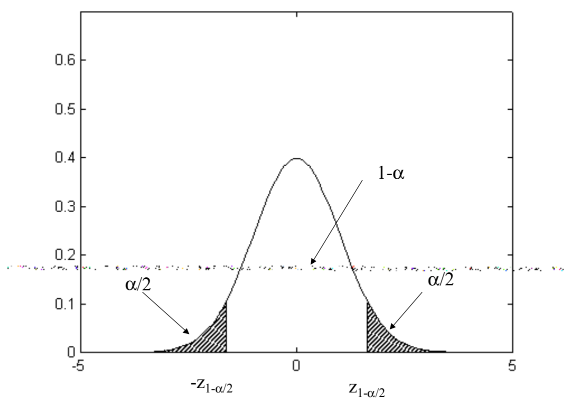
Estimación Paramétrica

1. El intervalo de confianza para la media poblacional es $\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$. Tenemos $\bar{x} = 166$, y $\alpha = 0.02$, luego necesitamos $z_{0.99}$ que encontramos en la tabla: $z_{0.99} \simeq 2.33$. Finalmente el intervalo pedido es $164.06 \leq \mu \leq 167.94$ o de forma equivalente : $\mu = 166 \pm 1.94$.

2. Su respuesta es incorrecta: la afirmación del artículo sólo concierne el centro, más concretamente la media, de la distribución de las alturas en la población española mayor de 18 años. El nivel de confianza de 95% se refiere a la confianza que tenemos en el método que hemos utilizado para proporcionar un intervalo para estimar la media. En este caso, el método que consiste en dos pasos (1) extraer una muestra al azar de n individuos, (2) calcular el intervalo para esa muestra $\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$, roporciona un resultado correcto (es decir un intervalo que efectivamente contiene la media poblacional) en el 95 % de las veces.

3. a) Introducimos la variable $X =$ "*valor* medido". La media poblacional de X representa el centro de los valores proporcionados por el aparato. En particular si el aparato es exacto, la media poblacional de X debe ser igual a 10 gramos. Si el aparato fuera perfecto, todos los valores medidos serían iguales a 10 y en particular la media de X valdría 10.

   b)Empezamos por la construcción detallada del intervalo para la media poblacional para un α dado. Sean $X_1, X_2, \ldots X_5$, las variables "valor obtenido en la 1^o medición", etc hasta "valor obtenido en la 5^o medición". Consideramos la media muestral $\bar{X} = \frac{X_1+X_2+\cdots+X_5}{5}$. Puesto que estamos en el caso en que la variable X sigue una distribución normal, tenemos que $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.



Tenemos $P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1-\alpha$. Lo que es equivalente a $P(-z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}) = 1-\alpha$. Despejando obtenemos $P(\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$. Deducimos que un intervalo de confianza al $100(1-\alpha)\%$ de confianza para la media poblacional es $\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$.

En nuestro caso particular, nos fijamos $\alpha = 0.02$, necesitamos $z_{0.99}$, que según la tabla es igual a 2.33. Sustituyendo obtenemos por lo tanto $10.\,0020 \leq \mu \leq 10.0025$.

c)El margen de error es $z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$, queremos que sea menor de 0.0001 :

$$z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq 0.0001$$

Despejamos para obtener $\sqrt{n} \geq z_{1-\alpha/2}\frac{\sigma}{0.0001}$, y elevando al cuadrado, $n \geq \left(z_{1-\alpha/2}\frac{\sigma}{0.0001}\right)^2$, sustituimos para obtener $n \geq 21.\,7$, por lo tanto, 22 mediciones son necesarias.

4. a) La variable X representa la cantidad de maiz de la nueva variedad cosechada en una parcela seleccionada al azar. La población de interés sería en este caso el conjunto de todas las cosechas recogidas con esa variedad de maíz. Por los resultados vistos en teoria, el intervalo de confianza a nivel $1-\alpha = 0.9$ pedido es:

$$[\overline{X} \pm z_{1-(\alpha/2)}\frac{\sigma}{\sqrt{5}}] = [\overline{X} \pm z_{0.95}\frac{10}{\sqrt{15}}] \tag{1}$$

donde $z_{0.95}$ es el percentil 0.95 de la distribución normal estandar:

$$\phi(z_{0.95}) = 0.95 \leftrightarrow z_{0.95} = 1.64$$

con lo que sustituyendo el anterior valor en la expresión 1 el intervalo que contiene la cosecha promedio con una confianza del 90% es:

$$[119.57, 128.03]$$

b)Razonando de la misma forma que en a) tenemos:

- A nivel $1-\alpha = 0.95$ pedido es:

$$[\overline{X} \pm z_{0.975}\frac{10}{\sqrt{15}}]$$

donde $z_{0.975}$ es el percentil 0.975 de la distribución normal estandar:

$$\phi(z_{0.975}) = 0.975 \leftrightarrow z_{0.975} = 1.96$$

y sustituyendo:

$$[118.74, 128.87]$$

- A nivel $1-\alpha = 0.99$ pedido es:

$$[\overline{X} \pm z_{0.995}\frac{10}{\sqrt{15}}]$$

donde $z_{0.995}$ es el percentil 0.995 de la distribución normal estandar:

$$\phi(z_{0.995}) = 0.995 \leftrightarrow z_{0.995} = 2.56$$

y sustituyendo:

$$[117.19, 130.41]$$

Podemos comprobar que a medida que exigimos un mayor nivel de confianza el intervalo obtenido es de mayor amplitud. El exigir un nivel de confianza mayor equivale a exigir una mayor seguridad de capturar el verdadero valor del parámetro mediante el intervalo construido a partir de la muestra (cuanto más ancho sea más valores posibles para μ considera y por tanto más fácil es que este valor sea 'capturado' por el intervalo).

c) Los intervalos anteriores se construyen basandose en la hipótesis de que la distribución de la variable (y por tanto también la de la media muestral) sigue una distribución normal. Si no conocemos la distribución de la variable no tenemos la hipótesis de normalidad, sin embargo el teorema Central del Limite nos permite afirmar que la distribución de la media muestral es aproximadamente normal cuando n es suficientemente grande con lo que podemos utilizar el mismo intervalo (siempre que asumamos que $n = 15$ es 'suficientemente grande').

d)El margen de error (E) cometido suponiendo un intervalo de confianza al 90% usando la expresión (1) será igual a:

$$\begin{aligned}
[\overline{X_n} \pm E] &= [\overline{X_n} \pm 1.64\frac{10}{\sqrt{n}}] \\
E &= 1.64\frac{10}{\sqrt{n}},
\end{aligned}$$

como queremos que sea inferior a 4 debemos exigir:

$$\begin{aligned}
4 &> E = z_{0.95}\frac{10}{\sqrt{n}} \\
&\leftrightarrow \\
n &> (1.64 * 10/4)^2 \\
&\leftrightarrow \\
n &> 16.81
\end{aligned}$$

con lo que debemos plantar al menos 17 parcelas.

5. a) Si llamamos R al verdadero valor de la resistencia, y X al valor obtenido al azar al realizar una medición,

$$X = R + e \tag{2}$$

donde e es el error que se comete en la medición. Como e sigue una distribución normal de media 0 y desviación típica $\sigma = 3$, y X está relacionada con la variable e mediante la expresión (2), por las propiedades de la distribución normal podemos concluir que X seguirá una distribución normal de media:

$$E(X) = E(R + e) = R + E(e) = R + 0 = R$$

y varianza:

$$Var(X) = Var(e) = 3^2$$

El aparato de medición es exacto en el sentido de que el valor esperado promedio de las mediciones es el valor correcto y desconocido R. En cuanto a la precisión sabemos que el 99.7% de las mediciones estarán comprendidas entre $(R \pm 3\sigma) = (R \pm 9)$. La desviación típica es de 3 unidades de medición lo cual es elevado y concluiriamos que no es muy preciso.

b) Como $R = \mu_X$, y dado que X sigue una distribución normal, por los resultados vistos en teoria, el intervalo de confianza a nivel $1 - \alpha = 0.9$ pedido es:

$$[\overline{X} \pm z_{1-(\alpha/2)} \frac{\sigma}{\sqrt{5}}] = [\overline{X} \pm z_{0.95} \frac{3}{\sqrt{5}}] \tag{3}$$

donde $z_{0.95}$ es el percentil 0.95 de la distribución normal estandar:

$$\phi(z_{0.95}) = 0.95 \leftrightarrow z_{0.95} = 1.64$$

con lo que sustituyendo el anterior valor en la expresión (3) y dado que $\overline{X_5} = 497.4$:

$$[497.4 \pm 1.64 \frac{3}{\sqrt{5}}] = [497.4 \pm E]$$

c) Si suponemos una confianza del 95% y dejamos n sin determinar el intervalo de confianza queda de la forma:

$$[497.4 \pm 1.96 \frac{3}{\sqrt{n}}] = [497.4 \pm E]$$

ya que $z_{0.975} = 1.96$ es el percentil $1 - \alpha/2$ para $1 - \alpha = 0.95$. El error E es en este caso:

$$E = 1.96 \frac{3}{\sqrt{n}}$$

de forma que como queremos que E no sea superior a 2 debemos exigir:

$$\begin{aligned} 2 &\geq E = z_{0.975}\frac{3}{\sqrt{n}} \\ &\leftrightarrow \\ n &\geq (1.96 * 3/2)^2 \\ &\leftrightarrow \\ n &\geq 8.64 \end{aligned}$$

con lo que debemos plantar al menos 4 mediciones más hasta llegar a $n = 9$.

6. Un estimador puntual de un parámetro es cualquier estadístico (función de las observaciones) diseñado para aproximar el parámetro que nos interesa. Se trata de una variable aleatoria y su valor depende de la muestra escogida. Dos propiedades deseables para un estimador puntual es a) Su media es igual al valor del parámetro poblacional, en este caso hablamos de un estimador insesgado. b) Si el estimador es insesgado, una buena propiedad adicional es que sea más preciso a medida que aumente el número de observaciones. Concretamente, es deseable que su varianza tienda a cero cuando n tiende hacia infinito. Un estimador insesgado con esta propiedad se llama consistente.

7. a) El intervalo pedido con un nivel de confianza $1 - \alpha = 0.95$ tiene la forma:

$$[\overline{X} \pm z_{1-(\alpha/2)}\frac{\sigma}{\sqrt{5}}] = [\overline{X} \pm z_{0.975}\frac{3}{\sqrt{5}}] \tag{4}$$

donde $z_{0.975}$ es el percentil 0.975 de la distribución normal estandar:

$$\phi(z_{0.975}) = 0.795 \leftrightarrow z_{0.975} = 1.96$$

con lo que sustituyendo el anterior valor en la expresión (4) y dado que $\overline{X_5} = 555.5$:

$$[550.179, 559.821] \tag{5}$$

b) Puesto que la relación entre las variables viene dada por:

$$Y = (9/5)X + 32$$

la relación que liga ambos valores esperados será:

$$\mu_Y = E(Y) = (9/2)E(X) + 32 = (9/2)\mu_X + 32$$

usando esta relación, y dado que tenemos un intervalo de confianza para μ_X (expresión 5), podemos obtener el intervalo para μ_Y transformando los extremos del intervalo:

$$\begin{aligned} LI &= (9/5)550.179 + 32 = 1022.322 \\ LS &= (9/5)559.821 + 32 = 1039.678 \end{aligned}$$

con lo que el intervalo queda:

$$[1022.322, 1039.678]$$

nota: alternativamente se podría haber calculado directamente el intervalo para μ_Y sabiendo que:

$$\begin{aligned} \overline{Y} &= (9/5)\overline{X} + 3 \\ \sigma_Y &= (9/5)\sigma_X \end{aligned}$$

8. a) El valor de 47% que se ha obtenido es relativo a la muestra escogida, no podemos por lo tanto afirmar sin más que, en la población entera de mujeres españolas, la proporción que nos interesa sigue siendo ésta.

   La afirmación se realiza con un nivel de confianza del 95% en el sentido de que si repitiesemos el proceso para todos los valores posibles que puede tomar la proporción muestral si extraemos al azar muestras de tamaño 1025, el 95% de los intervalos calculados captarían la verdadera proporción y el 5% restante no captaria el verdadero valor de la proporción de esta característica en la población de mujeres.

   b) El margen de error es mayor porque éste depende de n y para el caso de los hombres se han observado menos valores.

9. Empecemos por calcular la media y la varianza muestral. Tenemos

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{165 + 162 + \cdots + 168}{9} = 166.0 \\ s_X^2 &= \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\cdot(\overline{x})^2 \\ &= (\frac{165^2 + 162^2 + \cdots + 168^2}{8}) - \frac{9}{8}(166.0)^2 = 6.5 \end{aligned}$$

   Se trata de determinar el intervalo de confianza para la media poblacional de una distribución normal, estamos en el caso en que la varianza poblacional es desconocida. Necesitamos por lo tanto el resultado siguiente : la distribución del estadístico

$$T = \frac{\overline{X} - \mu}{S_X/\sqrt{n}} \sim t_{n-1}$$

   Al ser la distribución de Student simétrica, tenemos que

$$P(-t_{n-1,1-\alpha/2} \leq T \leq t_{n-1,1-\alpha/2}) = 1 - \alpha$$

   Es fácil deducir de este resultado que

$$P(\overline{X} - t_{n-1,1-\alpha/2}\cdot S_X/\sqrt{n} \leq \mu \leq \overline{X} + t_{n-1,1-\alpha/2}\cdot S_X/\sqrt{n}) = 1 - \alpha$$

y por lo tanto, para la media de una distribución normal, varianza desconocida, un intervalo de confianza del $100(1-\alpha)$ % está dado por

$$\overline{x} - t_{n-1,1-\alpha/2} \cdot s_X/\sqrt{n} \leq \mu \leq \overline{x} + t_{n-1,1-\alpha/2} \cdot s_X/\sqrt{n}$$

En nuestro caso, nos piden un nivel de confianza del 98%, lo que quiere decir que $\alpha = 0.02$. En la tabla, encontramos $t_{n-1,1-\alpha/2} = t_{8,0.99} \simeq 2.896$ y el intervalo de confianza pedido es

$$166 - 2.896 \cdot \sqrt{6.5}/\sqrt{9} \leq \mu \leq 166 + 2.896 \cdot \sqrt{6.5}/\sqrt{9}$$

finalmente, un intervalo de confianza al 98% para la media poblacional es

$$\boxed{163.54 \leq \mu \leq 168.46}$$

10. La población que nos interesa es el conjunto de los estudiantes que realizaron el examen de ingreso MIR y la variable X que nos interesa es la puntuación conseguida. Empecemos por traducir los datos del enunciado : X sigue una distribución normal, el tamaño muestral es $n = 20$, la media y la cuasi-desviación muestral son respectivamente $\overline{x} = 250$ y $s_X = 50$. Se trata de determinar el intervalo de confianza para la media poblacional de una distribución normal, estamos en el caso en que la varianza poblacional es desconocida. Usamos el resultado siguiente : la distribución del estadístico

$$T = \frac{\overline{X} - \mu}{S_X/\sqrt{n}}$$

es una distribución de Student con $n-1$ grados de libertad y, al igual que en el problema 1, deducimos que, para la media de una distribución normal, varianza desconocida, un intervalo de confianza del $100(1-\alpha)$ % está dado por

$$\overline{x} - t_{n-1,1-\alpha/2} \cdot s_X/\sqrt{n} \leq \mu \leq \overline{x} + t_{n-1,1-\alpha/2} \cdot s_X/\sqrt{n}$$

En nuestro caso, nos piden un nivel de confianza del 95%, lo que quiere decir que $\alpha = 0.05$. En la tabla, encontramos $t_{n-1,1-\alpha/2} = t_{19,0.975} \simeq 2.093$ y el intervalo de confianza pedido es

$$540 - 2.093 \cdot 50/\sqrt{20} \leq \mu \leq 540 + 2.093 \cdot 50/\sqrt{20}$$

finalmente, un intervalo de confianza al 95% para la media poblacional es

$$516.60 \leq \mu \leq 563.40$$

11. Estamos interesados en dos poblaciones distintas : la primera corresponda al cemento estándar y la segunda al cemento contaminado con plomo. Queremos estudiar la v.a. X_1="contenido en calcio" en la población del cemento estándar y la v.a. X_2="contenido en calcio" en la población del cemento contaminado. De manera más precisa, queremos investigar la diferencia entre el contenido medio en calcio del cemento estándar y el del

cemento contaminado. Vamos a establecer un intervalo de confianza al 95 % para la diferencia de medias entre las dos poblaciones. Estamos en el caso en el que las varianzas en las dos poblaciones son desconocidas pero iguales, X_1 y X_2 se asumen normales y independientes, para establecer el intervalo de confianza para la diferencia de medias usamos el resultado siguiente : el estadístico

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$, sigue una distribución t con $n_1 + n_2 - 2$ grados de libertad. Deducimos que un intervalo de confianza al $100(1-\alpha)\%$ para la diferencia entre medias de dos distribuciones normales, varianzas desconocidas pero iguales es

$$\begin{aligned} & (\overline{x}_1 - \overline{x}_2) - t_{n_1+n_2-2,1-\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ \leq\ & (\mu_1 - \mu_2) \leq \\ & (\overline{x}_1 - \overline{x}_2) + t_{n_1+n_2-2,1-\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \end{aligned}$$

En nuestro caso, los datos del enunciado se traducen por

$$\begin{array}{ll} \overline{x}_1 = 90.0 & \overline{x}_2 = 87.0 \\ s_1^2 = 25.0 & s_2^2 = 16.0 \\ n_1 = 10 & n_2 = 15 \end{array}$$

Deducimos $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{9\cdot25.0+14\cdot16.0}{23} = 19.5$. Nos piden un nivel de confianza del 95%, necesitamos $t_{n_1+n_2-2,1-\alpha/2} = t_{23,0.975} \simeq 2.069$. El intervalo pedido es

$$\begin{aligned} & (90.0 - 87.0) - 2.069\sqrt{19.5}\sqrt{\frac{1}{10} + \frac{1}{15}} \\ \leq\ & (\mu_1 - \mu_2) \leq \\ & (90.0 - 87.0) + 2.069\sqrt{19.5}\sqrt{\frac{1}{10} + \frac{1}{15}} \end{aligned}$$

Finalmente el intervalo de confianza al nivel del 95% para la diferencia de medias es

$$-0.73 \leq (\mu_1 - \mu_2) \leq 6.73$$

La conclusión es que, puesto que nuestro intervalo nos dice que es posible que $(\mu_1 - \mu_2)$ sea igual o muy próximo a cero, la muestra nos lleva a afirmar que no hay una diferencia significativa entre los dos tipos de cemento.

12. Estamos interesados en dos variables distintas, la primera, X_A, corresponde al número de unidades producidas por el operario A, y la segunda X_B, corresponde al número de unidades

producidas por el operario B. Queremos determinar si hay una diferencia entre el número medio de unidades producidas por el operario A y el operario B, es decir, queremos conocer la diferencia $\mu_A - \mu_B$. Las hipótesis del enunciado se traducen de la manera siguiente :X_A y X_B son dos variables normales independientes de media μ_A y μ_B, respectivamente, además se supone que sus varianzas son iguales. Disponemos de dos muestras, designamos con el subíndice A todas las cantidades relacionadas con la muestra del operario A, y con el subíndice B todas las cantidades relacionadas con la muestra del operario B. Calculamos

$$\begin{array}{ll} \overline{x}_A = \frac{48+\cdots 50}{5} = 52.4 & \overline{x}_B = \frac{56+\cdots 55}{5} = 60.6 \\ s_A^2 = \frac{48^2+\cdots 50^2}{4} - \frac{5}{4}(52.4)^2 = 54.8 & s_B^2 = \frac{56^2+\cdots 55^2}{4} - \frac{5}{4}(60.6)^2 = 27.8 \\ n_A = 5 & n_B = 5 \end{array}$$

Para establecer el intervalo de confianza para la diferencia de medias usamos el resultado siguiente : si $S_p^2 = \frac{(n_A-1)S_A^2+(n_B-1)S_2^2}{n_A+n_B-2}$, el estadístico

$$T = \frac{\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)}{S_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

sigue una distribución t con $n_A + n_B - 2$ grados de libertad. Deducimos que un intervalo de confianza al $100(1-\alpha)\%$ para la diferencia entre medias de dos distribuciones normales, varianzas desconocidas pero iguales es

$$\begin{array}{r} \overline{x}_A - \overline{x}_B - t_{n_A+n_B-2,1-\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \\ \leq (\mu_A - \mu_B) \leq \\ \overline{x}_A - \overline{x}_B + t_{n_A+n_B-2,1-\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}. \end{array}$$

En nuestro caso $s_p^2 = \frac{(n_A-1)s_A^2+(n_B-1)s_2^2}{n_A+n_B-2} = \frac{4\cdot 54.8+4\cdot 27.8}{8} = 41.3$. Nos piden un nivel de significación $\alpha = 0.1$, necesitamos $t_{n_1+n_2-2,1-\alpha/2} = t_{8,0.95} \simeq 1.86$. El intervalo pedido es

$$\begin{array}{r} 52.4 - 60.6 - 1.86\sqrt{41.3}\sqrt{\frac{1}{5} + \frac{1}{5}} \\ \leq (\mu_A - \mu_B) \leq \\ 52.4 - 60.6 + 1.86\sqrt{41.3}\sqrt{\frac{1}{5} + \frac{1}{5}} \end{array}$$

Finalmente el intervalo de confianza al nivel del 95% para la diferencia de medias es

$$-15.76 \leq (\mu_A - \mu_B) \leq -0.64$$

13. a) El valor medido X es igual al contenido exacto de carbonato de calcio de la caliza más el error ε que sigue una distribución normal con media cero y varianza desconocida σ^2. Por consiguiente, si denoto por $[CaC0_3]_{pob}$el contenido exacto de carbonato de calcio, tenemos

$$X = [CaC0_3]_{pob} + \varepsilon$$

y deduzco que X sigue una distribución normal de media $[CaC0_3]_{pob}$ y de varianza σ^2. Hay que destacar que la media del valor medido es igual al contenido exacto de carbonato de calcio de la caliza, es decir, $\mu = [CaC0_3]_{pob}$

b) Empecemos por calcular la media y la cuasi-varianza muestral

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{49.56 + 49.82 + 49.30 + 50.16 + 50.06}{5} \simeq 49.78$$

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\cdot(\overline{x})^2$$

$$= (\frac{49.56^2 + 49.82^2 + 49.30^2 + 50.16^2 + 50.06^2}{4}) - \frac{5}{4}(49.78)^2 \simeq 0.126$$

Tenemos que construir un intervalo de confianza para la media poblacional en el caso en que la varianza poblacional es desconocida. Usamos el resultado siguiente : la distribución del estadístico $T = \frac{\overline{X}-\mu}{S_X/\sqrt{n}}$,es una distribución de Student con $n-1$ grados de libertad y, al igual que en el problema 1, deducimos que, para la media de una distribución normal, varianza desconocida, un intervalo de confianza del $100(1-\alpha)$ % está dado por

$$\overline{x} - t_{n-1,1-\alpha/2}\cdot s_X/\sqrt{n} \leq \mu \leq \overline{x} + t_{n-1,1-\alpha/2}\cdot s_X/\sqrt{n}$$

En nuestro caso, nos piden un nivel de confianza del 90%, lo que quiere decir que $\alpha = 0.1$. En la tabla, encontramos $t_{n-1,1-\alpha/2} = t_{4,0.95} \simeq 2.13$ y el intervalo de confianza pedido es

$$49.78 - 2.13\cdot\sqrt{0.126}/\sqrt{5} \leq \mu \leq 49.78 + 2.13\cdot\sqrt{0.126}/\sqrt{5}$$

finalmente, un intervalo de confianza al 90% para la media poblacional es

$$49.\,44 \leq \mu \leq 50.\,12$$

que podemos presentar en la forma

$$\mu \simeq 49.78\% \pm 0.34$$

c) Estamos en el caso de una población infinita, el tamaño muestral necesario para cometer como máximo un error err sobre la media poblacional, con una confianza del $100(1-\alpha)$%, es dado por

$$n \geq \left(z_{1-\alpha/2}\right)^2 \cdot \frac{\sigma^2}{err^2}$$

Puesto que no conocemos la varianza poblacional σ^2, utilizamos el valor estimado con la muestra anterior, $s_X^2 = 0.126$. Tenemos $err = 0.2$ y nos piden $\alpha = 0.05$, encontramos en la tabla $z_{1-\alpha/2} = z_{0.975} = 1.96$ y por tanto:

$$n \geq (1.96)^2\,\frac{0.126}{0.2^2} \simeq 12.1$$

Luego debemos tomar como mínimo $n = 13$, por tanto, deberíamos hacer 8 mediciones más.

14. Estamos, al igual que en el problema anterior interesados en determinar la media poblacional de la variable X ="contenido medido de MnO_2". Esta media poblacional es el contenido real en bióxido de manganeso del mineral. De la misma manera que en el apartado b) del problema , construimos el intervalo de confianza al nivel de $100(1-\alpha)\%$ usando los percentiles de la distribución t de Student.

$$\overline{x} - t_{n-1,1-\alpha/2} \cdot s_X/\sqrt{n} \leq \mu \leq \overline{x} + t_{n-1,1-\alpha/2} \cdot s_X/\sqrt{n}$$

Calculemos la media y la cuasi varianza muestrales :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{37.62 + 38.23 + 38.44 + 37.62 + 38.71}{5} \simeq 38.12$$

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1} \cdot (\overline{x})^2$$

$$= (\frac{37.62^2 + 38.23^2 + 38.44^2 + 37.62^2 + 38.71^2}{4}) - \frac{5}{4}(38.12)^2 \simeq 0.24$$

y obtenemos el intervalo de confianza al nivel de 90% :

$$38.12 - 2.13 \cdot \sqrt{0.24}/\sqrt{5} \leq \mu \leq 38.12 + 2.13 \cdot \sqrt{0.24}/\sqrt{5}$$

es decir

$$37.66 \leq \mu \leq 38.59$$

que podemos presentar :

$$\mu \simeq 38.12 \pm 0.47$$

15. Estamos antes dos poblaciones : la primera corresponde a la producción de la primera semana mientras que la segunda corresponde a la producción de la segunda semana, introducimos las dos variables : X_1 ="puntuación de calidad de un artículo primera semana", y X_2 ="puntuación de calidad de un artículo, segunda semana". Estamos en el caso en el que las varianzas en las dos poblaciones son desconocidas pero iguales, X_1 y X_2 se asumen normales y independientes, para establecer el intervalo de confianza para la diferencia de medias usamos el resultado siguiente : el estadístico

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$, sigue una distribución t con $n_1 + n_2 - 2$ grados de libertad. Deducimos que un intervalo de confianza al $100(1-\alpha)\%$ para la diferencia entre medias de dos distribuciones normales, varianzas desconocidas pero iguales es

$$\begin{aligned} & \overline{x}_1 - \overline{x}_2 - t_{n_1+n_2-2,1-\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ \leq \ & (\mu_1 - \mu_2) \leq \\ & \overline{x}_1 - \overline{x}_2 + t_{n_1+n_2-2,1-\alpha/2} \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \end{aligned}$$

En nuestro caso, los datos del enunciado se traducen por

$$\begin{array}{ll} \overline{x}_1 = 91.5 & \overline{x}_2 = 89.9 \\ s_1^2 = 9.1 & s_2^2 = 17.8 \\ n_1 = 8 & n_2 = 8 \end{array}$$

Deducimos $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{7\cdot 9.1+7\cdot 17.8}{14} = 13.4$. Nos piden un nivel de confianza del 95%, necesitamos $t_{n_1+n_2-2,1-\alpha/2} = t_{14,0.975} \simeq 2.145$. El intervalo pedido es

$$\begin{aligned} & 91.5 - 89.9 - 2.145\sqrt{13.4}\sqrt{\frac{1}{8}+\frac{1}{8}} \\ \leq\ & (\mu_1 - \mu_2) \leq \\ & 91.5 - 89.9 + 2.145\sqrt{13.4}\sqrt{\frac{1}{8}+\frac{1}{8}}. \end{aligned}$$

Finalmente el intervalo de confianza al nivel del 95% para la diferencia de medias es

$$-2.31 \leq (\mu_1 - \mu_2) \leq 5.56$$

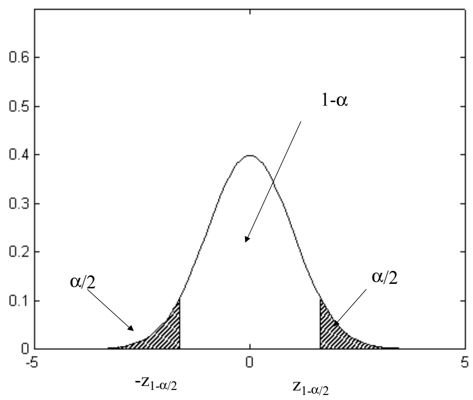
La conclusión es que, con los datos de la muestra, es posible que la diferencia de las medias poblacionales sea igual o muy próximo a cero, en consecuencia no podemos afirmar que ha habido un descenso significativo de la calidad entre las dos semanas.

Contrastes de Hipótesis

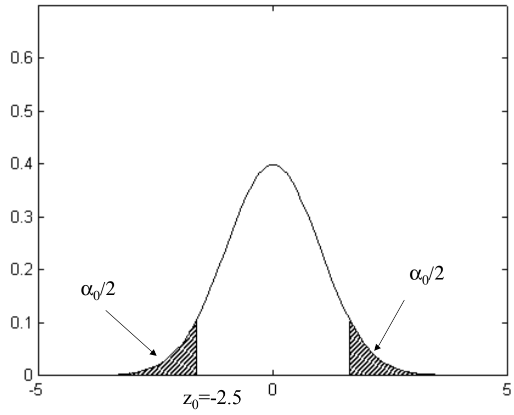
1. El primer contraste que realizamos es

$$\left.\begin{array}{l}\mathbf{H}_0 : \mu = 1600\ h. \\ \mathbf{H}_1 : \mu \neq 1600\ h.\end{array}\right\}$$

Escogeremos un nivel de confianza de 95% por ejemplo, lo que corresponde a $\alpha = 0.05$.
El estadístico de contraste es: $Z = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ que sigue, bajo H_0, aproximadamente una distribución normal estándar gracias al teorema central de límite.
Se trata de un contraste bilateral, y la región crítica o de rechazo vendrá dada por:



Para $\alpha = 0.05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$, . Para mi muestra el estadístico de prueba toma el valor :$z_0 = \dfrac{1570-1600}{120/\sqrt{100}} = -2.5$. z_0 cae en la región de rechazo, lo que implica que,al 95% de confianza, rechazamos H_0 y afirmamos que la vida útil promedio es significativamente diferente de 1600h. Para saber hasta qúe confianza rechazaríamos $H_{0,}$ calculamos el p-valor. Buscamos el valor α_0 de α más pequeño que nos permita rechazar $H_{0,}$ lo encontraremos haciendo coincidir un límite de nuestra región de rechazo con el valor del estadístico de prueba:

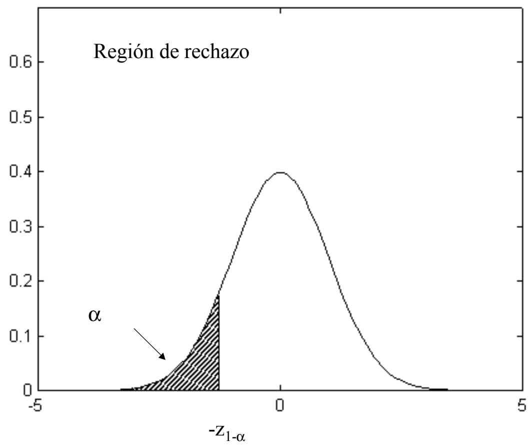


Deducimos que $\alpha_0/2 = P(Z_0 \leq -2.5) = 1 - \Phi(2.5) \simeq 0.006$. lo que implica un p-valor $\alpha_0 \simeq 0.012$.: podríamos por lo tanto rechazar H_0 hasta un máximo de 98.8% de confianza.

b) Ahora el contraste que nos piden es:

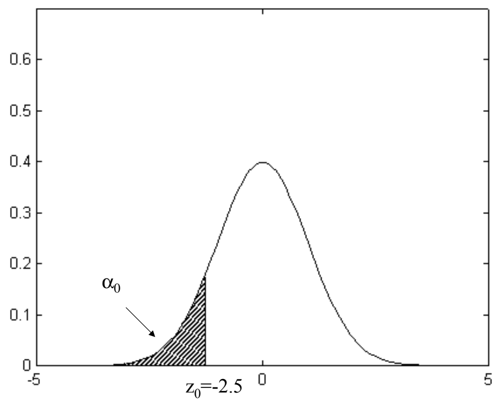
$$\left.\begin{array}{l} \mathbf{H}_0 : \mu = 1600\ h. \\ \mathbf{H}_1 : \mu < 1600\ h. \end{array}\right\}$$

Todo igual excepto la elección de la región de rechazo que es unilateral.



$\alpha = 0.05 \Rightarrow z_{1-\alpha} = z_{0.95} = 1.65$. Deducimos que, al 95% de confianza, también rechazamos H_0

Para calcular el p-valor, hacemos coincidir el límite de la región de rechazo con el valor de nuestro estadístico de prueba.



Deducimos $\alpha_0 = P(Z_0 \leq -2.5) = 1 - \Phi(2.5) \simeq 0.006$. Podríamos por lo tanto rechazar H_0 hasta un nivel de confianza de 99.4%

2. Planteamos el contraste
$$\mathbf{H}_0 : \mu = 0.5.$$
$$\mathbf{H}_1 : \mu \neq 0.5.$$
Seguimos los mismos pasos que en el ejercicio 1a) al que te puedes referir para los detalles: Encontramos un estadístico de prueba igual a $z_0 = \frac{0.4647-0.5}{0.2887/\sqrt{10}} \simeq -0.387$. El p-valor asociado es $\alpha_0 = 2P(Z_0 \leq -0.387) \simeq 0.700$. Sólo podríamos rechazar H_0 hasta una confianza de 30%: es muy insuficiente y admitimos H_0 : el centro de la distribución de los datos generados por Statistix parece coincidir con 0.5

3. Realizamos el contraste:
$$\begin{aligned} H_0 &: \mu \leq 45 \\ H_1 &: \mu > 45 \end{aligned}$$
puesto que lo que nos interesa es que como mínimo la producción media haya aumentado hasta al menos 45 unidades.

La variable poblacional X = "productividad diaria por persona de un sistema de producción" tiene desviación típica conocida, de manera que el estadístico del contraste anterior es:
$$Z_0 = \frac{\overline{X} - 45}{1.5/\sqrt{35}}$$
que sigue una distribución $N(0, 1)$ bajo la hipótesis nula. La región de rechazo de este contraste viene dada por:
$$\text{Si } \frac{\overline{X} - 45}{1.5/\sqrt{35}} > z_{1-\alpha} \Rightarrow \text{ rechazamos } H_0$$

y en este caso tenemos:

$$\frac{\overline{X}-45}{1.5/\sqrt{35}} = \frac{46.5-45}{1.5/\sqrt{35}} = 5.91 > z_{1-\alpha} = z_{0.95} = 1.64$$

luego rechazamos la hipótesis nula al 95%. Si calculamos el p-valor correspondiente:

$$\begin{aligned} p &= P(Z > \frac{46.5-45}{1.5/\sqrt{35}}) = \\ &= P(Z > 5.91) = 1 - \phi(5.91) \simeq 0 \end{aligned}$$

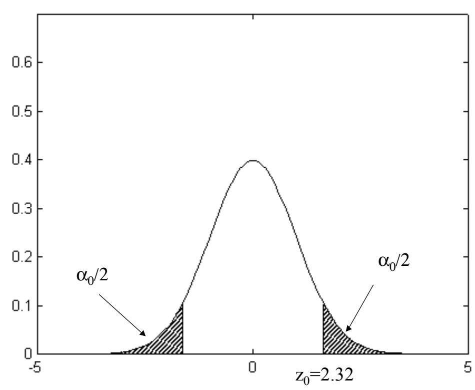
así que rechazaríamos la hipótesis nula practicamente a cualquier nivel de significación. Por tanto, podemos concluir que la productividad media es significativamente mayor de 45 unidades, así que resulta rentable aplicar la nueva tecnología.

4. Para un nivel de significación α se rechaza H_0 si $p-valor \leq \alpha$. Por tanto, llamando R a la decisión de rechazar H_0 y $\overline{R}$ a la decisión de no rechazar H_0, podemos resumir el ejercicio mediante la siguiente tabla:

p	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$
0.00012	R	R	R
0.54	$\overline{R}$	$\overline{R}$	$\overline{R}$
0.028	R	R	$\overline{R}$
0.17	$\overline{R}$	$\overline{R}$	$\overline{R}$

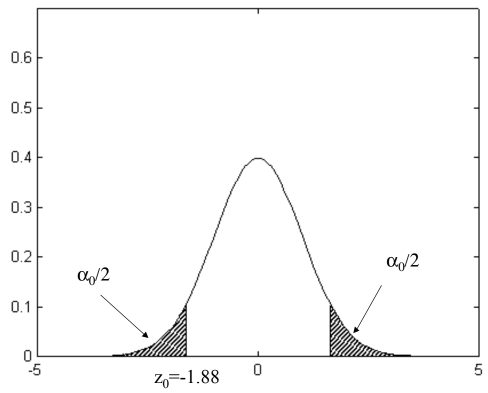
5. Para determinar el p-valor α_0, que es el valor de α más pequeño que nos permita rechazar H_0, hacemos coincidir el valor del estadístico de prueba con la frontera de la región de rechazo asociada a cada contraste.

a)



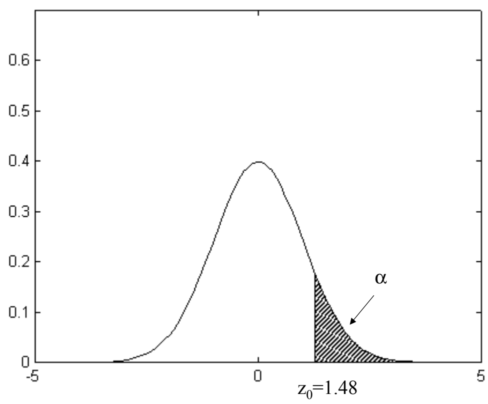
Deducimos $\alpha_0/2 = P(Z_0 \geq 2.32)$. Es decir $\alpha_0 = 2P(Z_0 \geq 2.32) = 2(1 - \Phi(2.32)) \simeq 0.02$. Rechazamos H_0 con gran confianza.

b)



Deducimos $\alpha_0/2 = P(Z_0 \leq -1.88)$. Por lo tanto, $\alpha_0 = 2P(Z_0 \leq -1.88) = 2(1-\Phi(1.88)) \simeq 0.06$. Rechazamos H_0 con gran confianza.

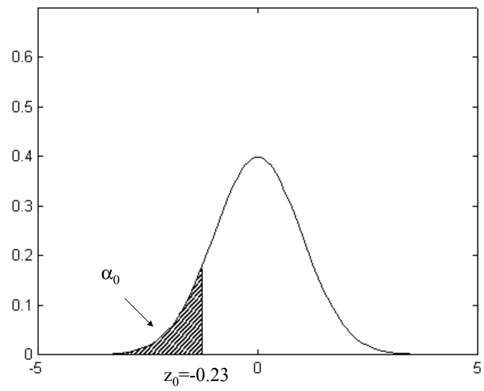
c)



Deducimos $\alpha_0 = P(Z_0 \geq 1.48) = 1 - \Phi(1.48) \simeq 0.07$. Rechazamos H_0 con gran confianza.

d)De la misma manera que en c): $\alpha_0 = P(Z_0 \geq 1.59) = 1 - \Phi(1.59) \simeq 0.06$. Rechazamos H_0 con gran confianza.

e)



Deducimos $\alpha_0 = P(Z_0 \leq -0.23) = 1 - \Phi(0.23) \simeq 0.40$. No podemos rechazar H_0 con una confianza suficiente: admitimos H_0.

6. Se quiere determinar si unos detectores de radón (un gas inodoro y incoloro ligeramente radioactivo) son fiables. Para ello, se colocan 12 de estos detectores en una cámara y se exponen durante 3 días a 105 picoCuries por litro de radón. Los datos obtenidos son los siguientes:

91.9	97.8	111.4	122.3	105.4	95.0
103.8	99.6	96.6	119.3	104.8	101.7

Supongamos que sabemos que la desviación típica de las mediciones para este tipo de detectores es $\sigma = 9$, y que podemos utilizar una distribución normal.

a) Construir un intervalo de confianza al 95% para el valor promedio de radon proporcionado por este tipo de detectores.

Queremos un intervalo de confianza para la media poblacional de una población normal de varianza conocida. El intervalo al $100(1-\alpha)\%$ de confianza está dado por

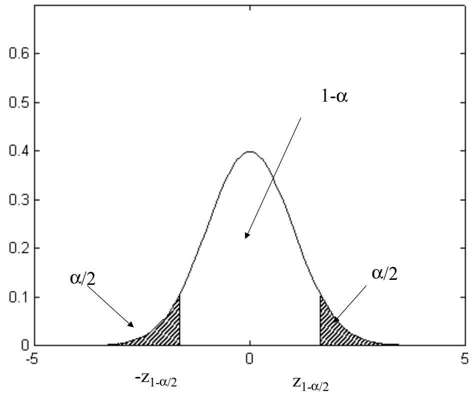
$$\left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Encontramos que, para nuestra muestra, $\bar{x} = 104.13$, y si trababjamos al 95 % de confianza necesitamos $z_{1-\alpha/2} = z_{0.975} = 1.96$, sustituyendo, obtenemos que el valor promedio de radón proporcionado por los detectores está comprendido entre $104.13 + 1.96 \cdot 9/\sqrt{12} = 99.04$ y 109.22 picoCuries por litro. Podemos formularlo también de la manera siguiente: el valor promedio de radón proporcionado por los detectores es 104.13 ± 5.09 picoCuries por litro.

b) Queremos comparar la media poblacional de los valores proporcionados por este tipo de detectores con 105. Planteamos el contraste

$$\mathbf{H}_0 : \mu = 105.$$
$$\mathbf{H}_1 : \mu \neq 105.$$

El estadístico de contraste es: $Z = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ que sigue, bajo H_0, una distribución normal. Se trata de un contraste bilateral, y la región crítica o de rechazo vendrá dada por:



Para $\alpha = 0.05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$, . Para mi muestra el estadístico de prueba toma el valor :$z_0 = \dfrac{104.13-105}{9/\sqrt{12}} \simeq -0.\,33$ z_0 no cae en la región de rechazo, lo que implica que,al 95% de confianza no podemos rechazar H_0: no podemos afirmar que el valor promedio difiera significativamente al 95% del valor real 105. Comprobamos que coincide con el resultado obtenido con el apartado anterior, puesto que el intervalo al 95% contenía el valor real 105.

7. El contraste que planteamos es:

$$\left.\begin{array}{l}\mathbf{H}_0: \ \mu = 28000 \\ \mathbf{H}_1: \ \mu < 28000\end{array}\right\}$$

Es un contraste unilateral, luego la región crítica se situará en la cola inferior de la distribución del estadístico de la prueba.

Tomamos $\alpha = 0.01$

Como el contraste se refiere a la media poblacional, y que la varianza poblacional es desconocida, el estadístico de la prueba será:

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Si suponemos que la hipótesis nula es cierta, el estadístico quedará

$$T = \frac{\overline{X}-28000}{S/\sqrt{n}} \sim t_{n-1}$$

La región de rechazo de este contraste viene dada por:

$$\text{Si } \frac{\overline{X}-28000}{S/\sqrt{n}} < -t_{n-1,1-\alpha} \Rightarrow \quad \textit{Se rechaza } H_0$$

En este caso, el valor particular del estadístico para la muestra que tenemos es

$$t_0 = \frac{27463-28000}{1348/\sqrt{40}} = -2.519 < -t_{39,0.99} = -2.42 \Rightarrow \quad \textit{Se rechaza } H_0.$$

Conclusión: La vida promedio de estos neumáticos es significativamente, al 99% de confianza, menor que 28.000 Km.

8. a) Para nuestra muestra, $n = 5$, $\bar{x} = 10.0023$ y $s = 0.0002$. El intervalo de confianza para la media poblacional es $\left[\bar{x} - t_{n-1,1-\alpha}\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha}\frac{s}{\sqrt{n}}\right]$. Al 98% de confianza, necesitamos $t_{n-1,1-\alpha/2} = t_{4,0.99} = 3.75$. Obtenemos que $\mu = 10.0023 \pm 0.00033$.

   b)El contraste de hipótesis que nos interesa es

$$\begin{aligned} H_0 &: \mu = 10 \\ H_1 &: \mu > 10. \end{aligned}$$

Si trabajamos al 95% de confianza, $\alpha = 0.05$. El estadístico de prueba es $T_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ que sigue una distribución t_{n-1} bajo H_0. La región de rechazo es unilateral $R = \{t : t \geq t_{n-1,1-\alpha} = t_{4,0.95} = 2.13\}$. Para nuestra muestra encontramos $t_0 = \frac{10.0023-10}{0.0002/\sqrt{5}} = 25.715$. En particular rechazamos H_0 al 95%, y afirmamos que μ es significativamente mayor de 10, al 95% de confianza. Notar que con este valor de t_0, rechazamos H_0 tambien al 99%. De hecho, el p-valor de la prueba es 7.10^{-6} : afirmamos con grandísima confianza que nuestro aparato de medición sobrevalora el peso real.