EJERCICIOS U.D. 4- DISTRIBUCIONES DE PROBABILIDAD

1. Un laboratorio afirma que un medicamento causa efectos secundarios en 5 de cada 100 pacientes. Si hay un grupo de 10 personas tomando este medicamento:

a) ¿Cuál es la probabilidad de sufran efectos secundarios al menos dos de ellos?

X=nº de personas que sufren efectos secundarios en un grupo de 10
X seguirá una distribución Binomial de parámetros n=10 y p=0,05.

$P(X\geq 2)=1-P(X\leq 1)=1-P(B(10,0,05)\leq 1)=1-\binom{10}{0}0{,}05^0 0{,}95^{10}-\binom{10}{1}0{,}05^1 0{,}95^9 =$

$=1-0{,}9139=0{,}0861\Rightarrow 8{,}61\%$

b) ¿Y la probabilidad de que no sufran efectos secundarios ninguno?

$P(X=0)=P(B(10,0,05)=0)=\binom{10}{0}0{,}05^0 0{,}95^{10}=0{,}5987\Rightarrow 59{,}87\%$

2. Una factoría fabrica disquetes de baja calidad a bajo precio, y produce un 20% de disquetes defectuosos.

Si se toma una muestra de 11 disquetes,

a) ¿Cuál es la probabilidad de que 3 disquetes sean defectuosos?

b) ¿Y la probabilidad de que encontremos más de 3?

c) ¿Qué promedio de disquetes defectuosos hay en las muestras de tamaño n=11?

SOLUCIÓN:

v.a. **X**={nº de disketes defectuosos en la muestra de tamaño11} ~ B(n=11, p=0,2)

a) $P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}=P(X=3)=\binom{11}{3}0{,}2^3(1-0{,}2)^{11-3}=\binom{11}{3}0{,}2^3(0{,}8)^8=0{,}22$

b) $P(\mathbf{X}>3)=1-P(\mathbf{X}\leq 3)=1-0{,}84=0{,}16$

$P(\mathbf{X}\leq 3)=P(\mathbf{X}=0)+P(\mathbf{X}=1)+P(\mathbf{X}=2)+P(\mathbf{X}=3)=$

$=\sum_{x=0}^{3}P(X=x)=\sum_{x=0}^{3}\binom{n}{x}p^x(1-p)^{n-x}=\sum_{x=0}^{3}\binom{11}{x}0{,}2^x(1-0{,}2)^{11-x}=\binom{11}{0}0{,}2^0(1-0{,}2)^{11-0}+$

$\binom{11}{1}0{,}2^1(1-0{,}2)^{11-1}+\binom{11}{2}0{,}2^2(1-0{,}2)^{11-2}+\binom{11}{3}0{,}2^3(1-0{,}2)^{11-3}=0{,}84$

c) $m=E(X)=np=11x0{,}2=2{,}2$ disquetes

3. El número medio de libros relacionados con la asignatura Estadística que presta en un día una biblioteca de informática de una determinada universidad es constante e igual a 4.

a) ¿Cuál será la probabilidad de que en esta biblioteca hoy se presten exactamente 4 libros relacionados con Estadística?

X=nº de libros relacionados con la asignatura Estadística prestados diariamente
X sigue distribución de Poisson de parámetro λ=4

$P(X=4)=P(Poisson(\lambda=4)=4)=e^{-4}\frac{4^4}{4!}=0{,}1954 \Rightarrow 19{,}54\%$

b) ¿Y la probabilidad de que en este mes de septiembre se presten como máximo 120 libros relacionados con Estadística en la misma biblioteca?

El mes de septiembre tiene 30 días. Sea Y el número de libros prestados en dicho mes. Por ser la suma de variables de Poisson que se suponen independientes, Y seguirá también distribución de Poisson de parámetro λ_Y=4 x 30=120. Como la media es mucho mayor que 9, Y se puede aproximar a la distribución normal, de la forma:

$P(Y\leq 120)\approx P(N(0,1)<\frac{120.5-120}{\sqrt{120}})=P(N(0,1)<0{,}045)=1-P(N(0,1)>0{,}045)=$
$=1-0{,}482055=0{,}5179\Rightarrow 51{,}79\%$

4. La centralita telefónica de L'Alqueria Blanca recibe, durante las 16 horas que está operativa cada día, una tasa media constante de 5 llamadas por hora para hablar con algún vecino del pueblo (llamadas entrantes). Por otro lado, la tasa de llamadas salientes (del pueblo al exterior) que gestiona es constante e igual a 4 llamadas/hora. Se puede considerar que ambas variables son independientes y que siguen una distribución de Poisson.

A partir de estos datos, contesta las siguientes cuestiones:

a) ¿Cuál es la probabilidad de que en la próxima hora la centralita reciba exactamente 3 llamadas entrantes?

X = Llamadas entrantes en una hora ~ Po(λ_X=5)

Y = Llamadas salientes en una hora ~ Po(λ_Y=4)

$P(X=3) = e^{-\lambda}(\lambda^3 / 3!) = e^{-5}(5^3 / 3!) = 0{,}1404 = 14{,}04\%$

b) En la última hora la centralita ha gestionado 2 llamadas salientes. ¿Cuál es la probabilidad de que en la próxima hora gestione exactamente otras 2 llamadas salientes más?

La distribución de Poisson posee la propiedad de "falta de memoria": la probabilidad de lo que ocurra en la próxima hora es independiente de lo que haya ocurrido en las horas anteriores. Por tanto, la probabilidad que se pide es:

$P(Y = 2) = e^{-\lambda} (\lambda^2 / 2!) = e^{-4} (4^2 / 2!) =$ 0,1465 = 14,65%

c) ¿Cuál es la probabilidad de que en todo un día la centralita reciba más de 90 llamadas entrantes?

V = llamadas entrantes recibidas en un día (16 horas).

$X \sim \text{Po}(\lambda_X=5) \Rightarrow V \sim \text{Po}(\lambda_V=5\cdot16=80)$

Como $\lambda_V=80 > 9$, podemos calcular la probabilidad que nos piden aplicando la aproximación por la distribución Normal, con corrección por continuidad:

$P(V > 90) \approx P(Z > (90+0{,}5-80)/\sqrt{80}) = P(Z > 1{,}17) =$ [tablas] = 0,1210 = 12,10%

d) ¿Cuál es la probabilidad de que en la próxima hora la centralita gestione exactamente 3 llamadas entrantes y 2 o menos llamadas salientes?

$P(X=3 \cap Y\leq2) =$ [son v.a. independientes] $= P(X=3)\cdot P(Y\leq2) =$ $0{,}1404 \cdot (P(Y=0) + P(Y=1) + P(Y=2)) = \ldots = 0{,}1404\cdot0{,}2381 =$ 0,0334 = 3,34%

5. El número de descargas de un archivo de la asignatura Estadística que está colgado en la plataforma PoliformaT sigue una media de 25 descargas por semana.

a) ¿Cuál es la probabilidad de que en una semana haya más de 10 descargas de dicho archivo?

X = nº de descargas semanales del archivo
$X \approx$ Poisson ($\lambda = 25$)
$P(X > 10) = 1 - P(X \leq 10) = 1 - 0{,}0005 = 0{,}9995$ (mirando en el ábaco)

b) ¿Cuál es la probabilidad de que en una semana haya 20 descargas del archivo?

$P(X = 20) = e^{-25} \cdot 25^{20} / 20! = 0{,}0519$

c) Si consideramos 4 semanas, ¿cuál es la probabilidad de que en 2 de ellas hayan más de 10 descargas del archivo?

Y = nº de semanas de las 4 consideradas en las que hay más de 10 descargas del archivo
$Y \approx$ Binomial ($n = 4$, $p = 0{,}9995$)

$$P(Y = 2) = \binom{4}{2}\ 0{,}9995^2 \cdot 0{,}0005^2 = 0{,}0000014985$$

6. En un sistema informático de una empresa se reciben un promedio constante de 4 paquetes de datos por segundo en horas de máximo tráfico. Si el procesador del sistema tiene una capacidad tal que puede atender a lo sumo 8 paquetes de datos por segundo:

a) ¿Cuál es la variable aleatoria considerada? ¿Qué distribución sigue?

Sea la variable X={número de paquetes de datos recibidos por segundo en horas de máximo tráfico}. El problema ya nos dice que el promedio de esta variable es de 4 paquetes/segundo. Al distribuirse X como una variable de Poisson deducimos que el parámetro λ es de 4 paquetes/segundo (ya que, por definición, debe coincidir con el promedio).

b) ¿Cuál es la probabilidad de que en un segundo determinado no sea posible procesar todos los paquetes de datos que lleguen al sistema?

Debemos calcular la probabilidad de que X sea mayor de 8, que es la capacidad máxima del sistema:

$P(X > 8) = 1 - P(X \leq 8) = 1 - \sum_{j=0}^{8} \frac{e^{-\lambda}\lambda^j}{j!} = 0{,}0214$ Resultado que también se puede sacar mirando en el ábaco de Poisson con λ=4 y para el valor x=8.

c) ¿Y de que en un segundo no se reciba ningún paquete de datos?

Sería la P(X=0)=$e^{-4}\,\frac{4^0}{0!}$=0.018

7. En un departamento en el que se producen dispositivos MP4 se quiere realizar un plan de inspección para comprobar la calidad de dichos dispositivos. Para ello, de cada lote se extraen N unidades aleatoriamente y se rechaza el lote si se encuentra más de un MP4 defectuoso.

¿Cuál debe ser el mínimo valor de N para que la probabilidad de aceptar un lote con una proporción de dispositivos defectuosos mayor que 0,10 sea menor del 2%?

SOLUCIÓN:

Mirando en el ábaco de Poisson para un valor de la ordenada < 0,02, lambda toma el valor ≥6, aproximadamente.
Así, como mínimo N= 6/0,10 = 60.

8. Una industria que utiliza masivamente en sus productos cierto componente electrónico desea garantizar que el porcentaje de componentes defectuosos en cada partida que compra es inferior al 10%. Para ello prueba en cada partida N unidades seleccionadas al azar, aceptando ésta sólo si todas las unidades resultan correctas.

a) ¿Cuánto debe valer como mínimo N para que la probabilidad de admitir una partida con un 10% o más de unidades defectuosas no supere el 5%?

b) Una partida se considera muy satisfactoria si el porcentaje de unidades defectuosas en la misma es menor o igual que el 2%. ¿Qué probabilidad tiene el sistema de control establecido en a) de rechazar una partida con un 2% de unidades defectuosas?

SOLUCIÓN:

a) Sea X=nºcomponentes defectuosas en la muestra de tamaño N$\Rightarrow$
X seguirá distribución binomial (N, p) donde p es 0,1
Se desea P(aceptar)$\leq$0,05 $\Rightarrow$ P(X=0)$\leq$0,05$\Rightarrow$
(1-0,1)N$\leq$0,05 $\Rightarrow$ Nlog(0,9) $\leq$ log(0,05)$\Rightarrow$
N$\geq$log(0,05)/log(0,9)=28,4$\Rightarrow$
Nmínimo=29

b) Siendo N=29 y p=0,02
P(rechazar)=1-P(aceptar)=
=1-P(Binomial(29,0,02)=0)=
= 1-(1-0,02)29=0,44

9. Una industria que utiliza masivamente en sus productos cierto componente electrónico selecciona de cada partida que llega, N unidades al azar y rechaza la partida si hay mas de C componentes defectuosos en la muestra. Se desea diseñar el plan de muestreo (es decir tomar los valores de N y C) de forma que se verifiquen los siguientes requisitos:

1) la probabilidad de admitir una partida con un 10% o más de unidades defectuosas no debe superar el 5%,

2) se debe tener una probabilidad inferior al 10% de rechazar partidas buenas (o sea con un porcentaje de defectos $\leq$ 2%).

¿Cuánto debe valer como mínimo N y cuál debe ser el valor correspondiente de C para que el plan de inspección satisfaga los dos requisitos deseados?

SOLUCIÓN:

Las dos condiciones son:

1. P(aceptar/p=0,1)=P(Poisson(λ=Nx0,1) $\leq$C)$\leq$0,05

2. P(aceptar/p=0,02)=P(Poisson(λ=Nx0,02) $\leq$C)$\geq$0,9

Vamos a tantear en función de C, qué requisitos debe cumplir N para verificar estas dos condiciones, empezando desde C=0 y aumentándolo progresivamente hasta encontrar un plan factible.

Primer paso:Si C=0

Para cumplir cond. 1 con el ábaco se obtiene $\lambda \geq 2,9 \Rightarrow N \geq 2,9/0,1=29$

Para cumplir cond. 2 con el ábaco se obtiene $\lambda \leq 0,11 \Rightarrow N \leq 0,11/0,02=5,5$

Como las dos condiciones $N \geq 29$ y $N \leq 5,5$ son incompatibles no hay solución con C=0.

Segundo paso: Si C=1

Para cumplir cond. 1 con el ábaco se obtiene $\lambda \geq 4,65 \Rightarrow N \geq 4,65/0,1=46,5$

Para cumplir cond. 2 con el ábaco se obtiene $\lambda \leq 0,53 \Rightarrow N \leq 0,53/0,02=26,5$

Como las dos condiciones $N \geq 46,5$ y $N \leq 26,5$ son incompatibles no hay solución con C=1.

Tercer paso: Si C=2

Para cumplir cond. 1 con el ábaco se obtiene $\lambda \geq 6,2 \Rightarrow N \geq 6,2/0,1=62$

Para cumplir cond. 2 con el ábaco se obtiene $\lambda \leq 1,08 \Rightarrow N \leq 1,08/0,02=54$

Como las dos condiciones $N \geq 62$ y $N \leq 54$ son incompatibles no hay solución con C=2.

Cuarto paso: Si C=3

Para cumplir cond. 1 con el ábaco se obtiene $\lambda \geq 7,7 \Rightarrow N \geq 7,7/0,1=77$

Para cumplir cond. 2 con el ábaco se obtiene $\lambda \leq 1,7 \Rightarrow N \leq 1,7/0,02=85$

Las dos condiciones $N \geq 77$ y $N \leq 85$ son compatibles y N=77 es el menor valor buscado. Por tanto el plan de inspección tendrá N=77(como mínimo) y C=3.

10. Para controlar la calidad de las partidas que recibe, el servicio de control de calidad de una empresa toma en primer lugar una muestra de 50 piezas, aceptando la partida si todas son correctas y rechazándola si encuentra más de dos defectuosas.

Si esta primera muestra no le permite tomar una decisión, se toma una segunda muestra de otras 50 piezas, rechazándose la partida si en total (entre la 1ª y 2ª muestra) se han encontrado más de 3 defectuosas y aceptándose en caso contrario.

a) ¿Qué probabilidad tiene este plan de inspección de aceptar partidas que tengan un 1% de unidades defectuosas?

b) ¿Cuál será el número medio de piezas a inspeccionar si las partidas tienen un 1% de unidades defectuosas?

SOLUCIÓN:

Sean X_1=nº de defectuosas en la primera muestra

X_2=nº de defectuosas en la segunda muestra (en el caso de que se tome)

X_1 y X_2 serán cada una Binomial(n=50, p=0,01)

a) El suceso {aceptar la partida} es la suma de los siguientes sucesos, mutuamente excluyentes:

$\{aceptar\}=\{X_1=0\} +\{\{X_1=1\}\{X_2\leq 2\}\} +\{\{X_1=2\}\{X_2\leq 1\}\}$

Por tanto como además X_1 y X_2 son independientes

$P(aceptar)= P(X_1=0) +P(X_1=1)P(X_2\leq 2) +P(X_1=2)P(X_2\leq 1)$

Y sustituyendo en la fórmula de la función de probabilidad de la binomial, la P(aceptar) será:

$$\mathrm{P(aceptar)} = 0{,}99^{50} + \binom{50}{1} 0{,}01 \, 0{,}99^{49} \mathrm{x} \left[0{,}99^{50} + \binom{50}{1} 0{,}01 \, 0{,}99^{49} + \binom{50}{2} 0{,}01^2 0{,}99^{48} \right] +$$

$$+ \binom{50}{2} 0{,}01^2 0{,}99^{48} \left[0{,}99^{50} + \binom{50}{1} 0{,}01 \, 0{,}99^{49} \right] = 0{,}975$$

b) ¿Cuál será el número medio de piezas a inspeccionar si las partidas tienen un 1% de unidades defectuosas?

El tamaño de la muestra a tomar será N=50 si $X_1=0$ ó $X_1\geq 3$, ó N=100 en caso contrario. Como:

$$\mathrm{P}(X_1 \geq 3) = 1 - \mathrm{P}(X_1 \leq 2) = 1 - \left[0{,}99^{50} + \binom{50}{1} 0{,}01 \, 0{,}99^{49} + \binom{50}{2} 0{,}01^2 0{,}99^{48} \right] = 0{,}0138$$

El valor medio de N será:

E(N)=50x(0,9950+0,0138)+ 100x(1-0,9950-0,0138)=

= **<u>69,1 piezas</u>**

11. En un sistema, la vida de un determinado virus informático sigue una distribución exponencial con mediana 60 horas.

a) ¿Cuál es la probabilidad de que dicho virus permanezca en el sistema más de 70 horas?

X = tiempo de vida de un determinado virus informático
$X \approx \mathrm{Exp}(\alpha)$
Mediana = 60 h $\rightarrow$ $P(X > 60) = 0{,}5$ $\rightarrow$ $e^{-\alpha \cdot 60} = 0{,}5$ $\rightarrow$ $-\alpha \cdot 60 = \ln 0{,}5$

$\rightarrow \quad \alpha = -(\ln 0{,}5) / 60 \quad \rightarrow \quad \alpha = 0{,}01155$

Así:

$P(X > 70) = e^{-0{,}01155 \cdot 70} = 0{,}4455$

b) Si sabemos que el virus ya está residente en el sistema 50 horas, ¿cuál es la probabilidad de que supere las 60 horas?

Como la distribución exponencial no tiene memoria:

$P(X > 60 \mid X > 50) = P(X > 10) = e^{-0{,}01155 \cdot 10} = 0{,}8909$

12. El tiempo transcurrido entre dos fallos consecutivos de una determinada máquina sigue una distribución exponencial de media 100 horas.

a) Si acaba de ocurrir un fallo, calcular la probabilidad de que el próximo fallo tarde más de 100 horas en producirse

$X = \text{Tiempo entre fallos (horas)} \sim \text{Exp}(\lambda)$

$E[X] = 100 \text{ horas} \Rightarrow \lambda = 1/100$

$P(X > 100) = e^{-\lambda \cdot 100} = e^{-100/100} = e^{-1} = 0{,}3679 = 36{,}79\%$

b) Si llevamos 200 horas sin que se produzca un fallo, ¿cuál es la probabilidad de que pasen 100 horas más sin observar ninguno?

La probabilidad que nos piden es idéntica a la anterior, ya que, debido a la propiedad de “falta de memoria” de la distribución exponencial, el tiempo que llevamos sin observar ningún fallo no influye en la probabilidad del tiempo hasta el próximo fallo. Es decir:

$P(\text{que pasen 100 horas más sin observarse un fallo, sabiendo que ya han pasado 200}) = P(X > 300 \mid X \geq 200) = [\text{propiedad de “falta de memoria”}] = P(X > 100) = [\text{apartado anterior}] = 0{,}3679 = 36{,}79\%$

13. La duración de un determinado componente electrónico sigue una distribución exponencial de media 100 horas.

a) ¿Cuál es la probabilidad de que un determinado componente tenga una duración superior a 120 horas?

Variable X= duración de un componente

$m=100=1/\alpha$

$P(X>120)=e^{-120/100}=0{,}3012$

b) Tenemos un dispositivo que dispone de 10 componentes de este tipo. El dispositivo está montado de manera que comienza a funcionar un componente hasta que se estropea y en ese preciso instante se pone a funcionar automáticamente el siguiente componente, y así sucesivamente hasta que se estropean los diez, en este momento el dispositivo deja de funcionar.¿Cuál es aproximadamente la probabilidad de que el dispositivo dure más de 1700 horas?

v.a. X_i = {duración del componente i (horas)} ~ EXP(α)

$m = E(X_i) = \dfrac{1}{\alpha} = 100\ h$ $\qquad$ $\sigma^2(X_i) = \dfrac{1}{\alpha^2} = 100^2\ h^2$

v.a. Y = {duración de 10 componentes (horas)} ~ **?** *(*)*

(*) Por el Teorema Central del Límite: la suma de suficientes v.a. INDEPENDIENTES tiende a distribuirse siguiendo el modelo NORMAL.

$Y = \sum_{i=1}^{10} X_i \quad X_i$ independiente de $X_j \quad \forall i \neq j$

$m_Y = E(Y) = 10 \times 100 = 1000\ h$ La media de una suma de v.a. es la suma de sus medias

$\sigma^2(Y) = 10 \times 100^2 = 10^5\ h^2$ La varianza de una suma de v.a. INDEPENDIENTES es la suma de sus varianzas.

Por tanto:

v.a. Y = {duración de 10 componentes (horas)} ~ N(m=10^3, σ^2=10^5)

$$\underline{P(Y > 1700)} = P\left(N(0,1) > \frac{1700 - 10^3}{\sqrt{10^5}}\right) = P(N(0,1) > 2{,}21) = 0{,}0136 \cong \underline{0{,}014}$$

14. Se ha comprobado que el tiempo de vida de cierto marcapasos sigue una distribución exponencial con media de 16 años.

a) ¿Cuál es la probabilidad de que a una persona a la que se le ha implantado este marcapasos se le deba reimplantar otro antes de 20 años?

b) Si el marcapasos lleva funcionando correctamente 5 años en un paciente, ¿cuál es la probabilidad de que haya que cambiarlo antes de 25 años?

SOLUCIÓN:

Sea T la variable aleatoria que mide la duración de un marcapasos en una persona. Tenemos que

a) $T \longrightarrow Exp\left(\alpha = \dfrac{1}{16}\right)$ $\qquad$ $\Leftrightarrow f(t) = \alpha e^{-\alpha t} \quad \forall t \geq 0$

$\Leftrightarrow F(t) = 1 - e^{-\alpha t}$

Entonces: $P\left[T \leq 20\right] = F(20) = 1 - e^{-\frac{20}{16}} = 0{,}7135$

b) Siendo:

$$P\left[5 \leq T \leq 25\right] = F(25) - F(5) = 1 - e^{-\frac{25}{16}} - 1 + e^{-\frac{5}{16}=} = 0{,}522$$

$$P\left[T \geq 5\right] = e^{-\frac{5}{16}} = 0{,}7316$$

Resulta: $P\left[T \leq 25_{|T \geq 5}\right] = \dfrac{P[5 \leq T \leq 25]}{P[T \geq 5]} = \dfrac{0{,}522}{0{,}7316} = 0{,}7135$

Luego como era de esperar, por ser propio a un mecanismo exponencial:
$P\left[T \leq 25_{|T \geq 5}\right] = P[T \leq 20]$

15. A un procesador llegan lotes de código para su procesamiento. El tiempo de procesamiento de un lote se distribuye exponencialmente con media 3 unidades de tiempo (u.t.).

Calcular:
a) Probabilidad de que el procesamiento de un lote dure más de 5 u.t.?

b) Si un lote lleva procesándose 2 u.t. ¿cuál es la probabilidad de que el tiempo total de procesamiento sea inferior a 5 u.t.?

c) Si se extraen 10 lotes al azar ¿cuál es la probabilidad de que 8 o más tengan un tiempo de procesamiento mayor de 3 u.t.?

SOLUCIÓN:

a) v.a. T = {Tiempo de procesamiento de un lote} ~ Exp(α=1/3)

P(T>5)=$e^{-(1/3)x5}$ = 0,189

b)
$$P\left(T < 5 \big/ T > 2\right) = \frac{P(2 < T < 5)}{P(T > 2)} = \frac{P(T > 2) - P(T > 5)}{P(T > 2)} = \frac{e^{2/3} - e^{5/3}}{e^{2/3}} = 0{,}631$$

c) v.a. Y = {Nº de lotes, en un grupo de 10 cuyo tiempo de procesamiento es > 3 ut}

Y~ B(n=10, p=P(T>3)=0,368)

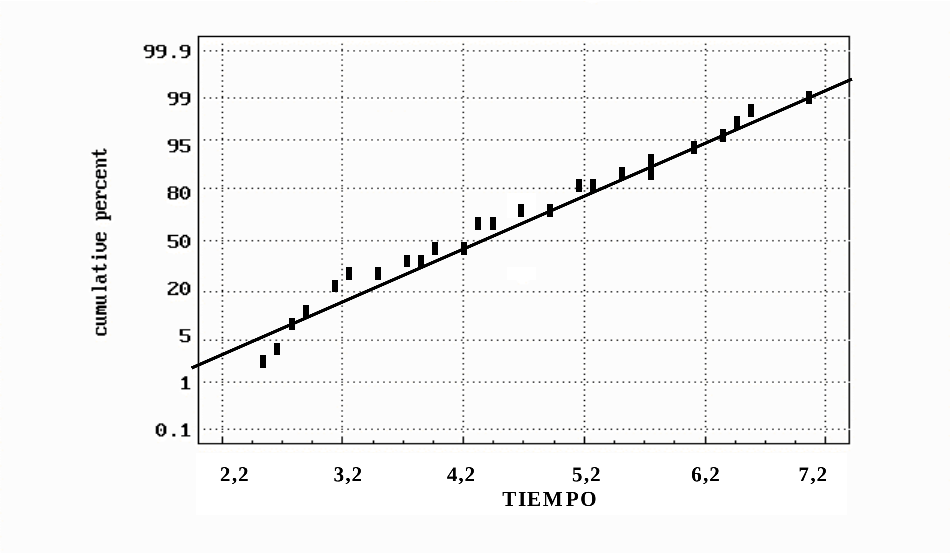
P(T>3)=$e^{-3/3}$ = 0,368

$$P(Y \geq \mathbf{8}) = P(T = \mathbf{8}) + P(T = \mathbf{9}) + P(T = \mathbf{10}) =$$

$$= \binom{\mathbf{10}}{\mathbf{8}} \mathbf{0{,}368^8} \times \mathbf{0{,}362^2} + \binom{\mathbf{10}}{\mathbf{9}} \mathbf{0{,}368^9} \times \mathbf{0{,}362^1} + \binom{\mathbf{10}}{\mathbf{10}} \mathbf{0{,}368^{10}} \times \mathbf{0{,}362^0} = \mathbf{0{,}007}$$

16. Con el fin de optimizar el funcionamiento de un centro de cálculo, se ha realizado un estudio en el que se ha medido el tiempo en ms de CPU requerido por 32 procedimientos. Los valores correspondientes se representan en la siguiente figura:

A la vista de la gráfica, contesta a los siguientes apartados:



a) Indicar quiénes son la población, la muestra, la variable aleatoria, y de qué tipo es ésta.

POBLACIÓN : Los procedimientos que realiza el centro de cálculo
MUESTRA : Los 32 procedimientos cuyos tiempos están representados
Variable aleatoria. : Tiempo en ms de CPU que requieren los procedimientos
TIPO : Cuantitativa continua

b) Indicar cómo se denomina la gráfica anterior.

Papel probabilístico normal

c) Estimar a partir de la figura el valor del intervalo intercuartílico, del coeficiente de asimetría y de curtosis.

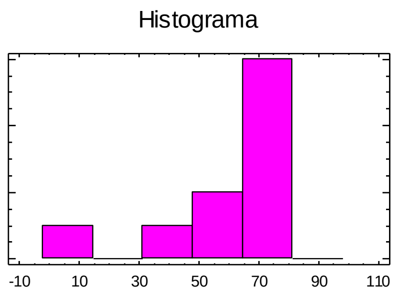
IIC: C3 – C1 = 5,2 – 3,5 = 1,7

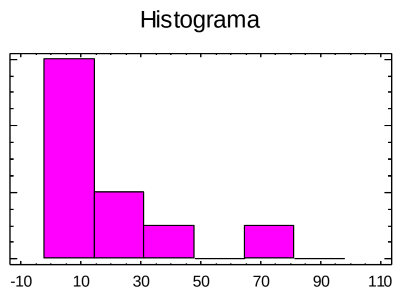
C.A. = 0 y C.C. = 3

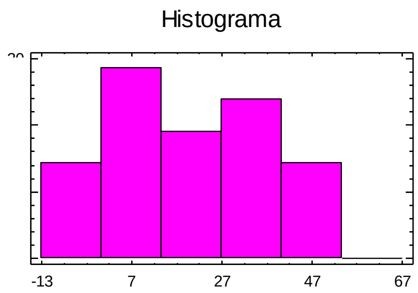
d) A la vista del gráfico, ¿cuáles crees que serán los mejores parámetros de posición y de dispersión? Justificar la respuesta.

Como, según el papel probabilístico normal, podemos ver que los datos siguen una distribución normal, los datos son simétricos, por lo que se tiene:
PARÁMETRO DE POSICIÓN: Media
PARÁMETRO DE DISPERSIÓN: Varianza o desviación típica

e) Seleccionar mediante un círculo el histograma que crees que se ajusta mejor a los datos y justificar la elección.







El tercer histograma, por ser aproximadamente simétrico

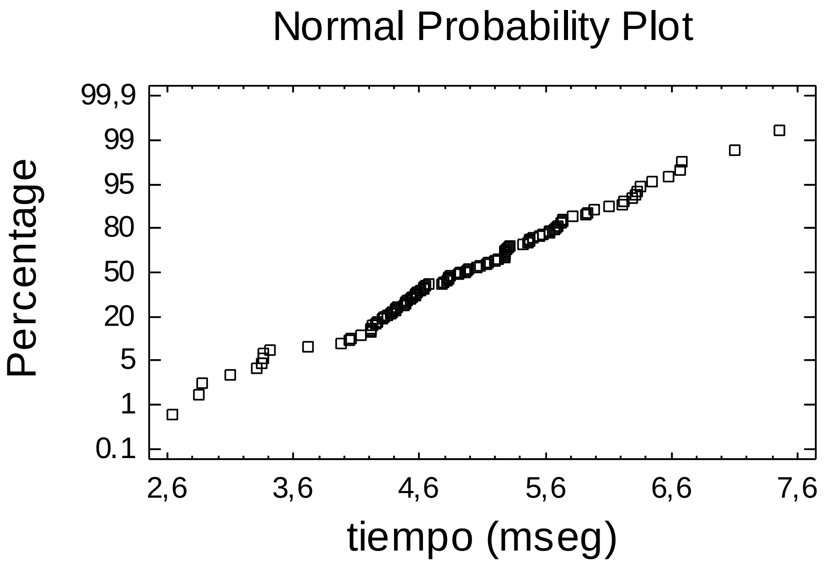
17. El tiempo de grabación de fichero fluctúa uniformemente entre 5 y 11 minutos. Disponemos de un tiempo máximo de grabación de 2 horas y media. Si consideramos la grabación consecutiva de 20 ficheros de este tipo, ¿qué porcentaje de las veces surgirán problemas por superar el tiempo máximo de grabación?

X = tiempo en minutos de grabación de un fichero
$X \approx$ Uniforme (5,11), por lo que $E(X) = 8$ min y $\sigma^2(X) = (11-5)^2 / 12 = 3$ min
Sea Y = tiempo en minutos de grabación consecutiva de 20 ficheros
$Y = X_1 + ... + X_{20}$
$E(Y) = 20 \cdot 8 = 160$ min y $\sigma^2(Y) = 20 \cdot 3 = 60$ min^2
$Y \approx N(m = 160, \sigma^2 = 60)$,
El tiempo máximo de grabación son 2 h y media, es decir, 150 minutos, por lo que nos piden $P(Y > 150)$
$P(Y > 150) = P(Z > (150 - 160) / \sqrt{60}) = P(Z > -1,29) = 1 – P(Z < -1,29) =$

= 1 – 0,09853 = 0,90147

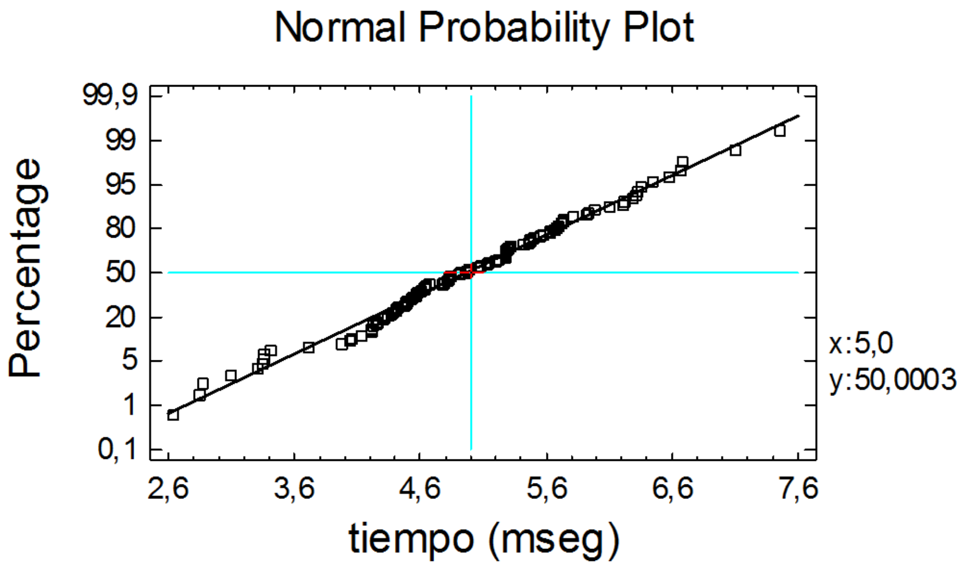
18. El tiempo de acceso a los datos almacenados en un sistema informático fluctúa aleatoriamente en función de su situación respecto a la cabeza de lectura.
El gráfico adjunto recoge la representación en papel probabilístico normal de 100 valores de dicho tiempo de acceso (medidos en milisegundos).

Si un programa tiene que acceder consecutivamente a 10 datos al azar, calcular aproximadamente la probabilidad de que el tiempo de lectura de los 10 datos sea superior a 58 msegundos.

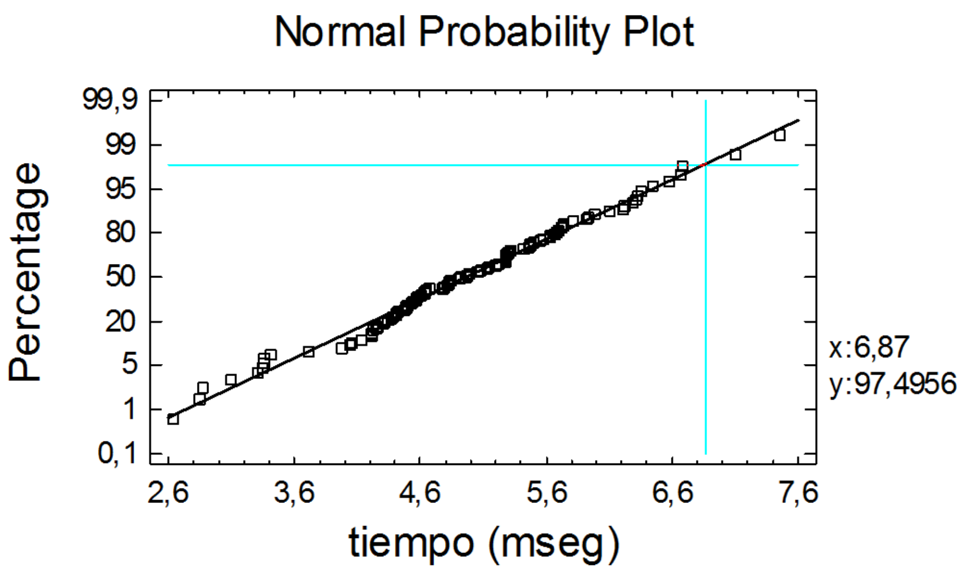


SOLUCIÓN:

De la representación se deduce que se puede admitir distribución normal para el tiempo de acceso. Para estimar la media m^* buscaremos la abcisa del punto sobre la recta de ordenada 50%, lo que resulta $m^*=5$, tal y como se observa en el gráfico adjunto:



Para estimar la $m^* + 2\,\sigma^*$ (siendo σ^* desviación típica), se buscará la abcisa del punto sobre la recta de ordenada 97,5%. Según se observa en el gráfico adjunto $m^*+2\sigma^*=6.87$. Por lo que $\sigma^*=(6.87-5)/2=0{,}935$



El tiempo de acceso consecutivamente a 10 datos al azar es normal de media= 5x10=50, y desviación típica= $\sqrt{10x0{,}935^2}$ =2.96.

La probabilidad de que se tarde mas de 58 milisegundos en leer consecutivamente 10 datos al azar es:

$$P(N(50,2.96)>58)=P(N(0,1)>\frac{58-50}{2{,}96})=P(N(0,1)>2.7)=0.00347$$

19. La velocidad del viento en una zona con molinos de viento para la producción de energía, se distribuye normalmente con media 50 km/h y desviación típica 15 km/h.

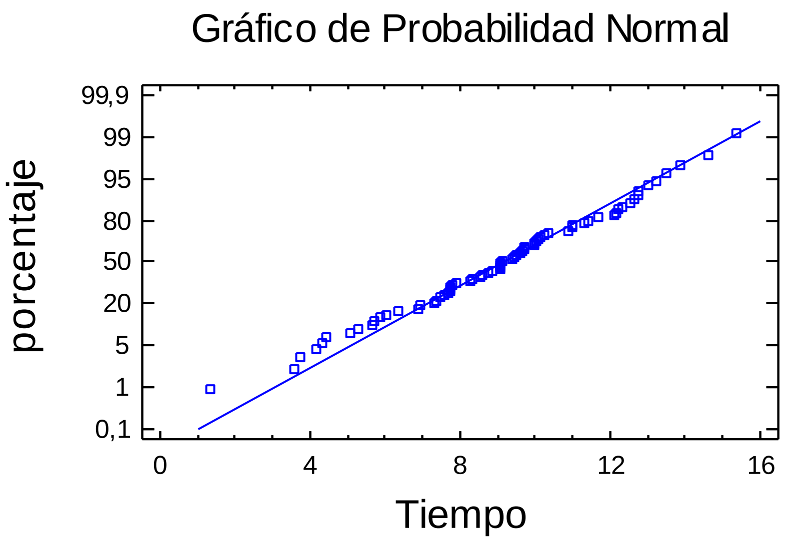
La velocidad máxima del viento, soportada por los motores generadores de electricidad, conectados a las aspas de los molinos, fluctúa normalmente con media 70 km/h y desviación típica de 5km/h.
A la vista de estos datos. ¿Cuál es la probabilidad de que el motor no soporte la velocidad del viento, y si no se desengancha previamente de las aspas, se queme?

SOLUCIÓN:

V = {Velocidad del viento}
S = {Velocidad máxima del viento soportada por el motor generador de electricidad}
V ~ N(50, 15)
S ~ N(70, 5)

El motor se quema si: V > S
Sea R = V – S por lo tanto se quemará cuando R> 0
R ~ N(50-70=-20, sqrt(15^2+5^2)=15,81)
P (R > 0) = P(Z > (0 – (-20))/15,81) = P(Z>1.27) = 0,1020

20. Tenemos el siguiente Gráfico Probabilístico Normal, que refleja los tiempos de respuesta en segundos de unos procesos distribuidos que se encargan de obtener cierta información de una base de datos.



A la vista del gráfico anterior, contesta a las siguientes cuestiones:
a) Obtener aproximadamente los siguientes parámetros:

- Media:≈9
- Desviación típica:≈2,5
- Primer cuartil:≈7,5

- Segundo cuartil:≈9
- Tercer cuartil:≈11
- Rango Intercuartílico:≈3,5
- Coeficiente de asimetría:≈0
- Coeficiente de curtosis:≈0

b) A la vista de los datos, ¿qué parámetros de posición y de dispersión representan bien los datos? ¿Por qué?

Puesto que la distribución es simétrica al seguir una distribución normal el mejor parámetro de posición sería la media y el mejor parámetro de dispersión la desviación típica.

21. Un PC tiene una CPU que soporta una temperatura operativa máxima[1] (TM) que fluctúa normalmente con media 70ºC y σ = 1ºC. Se sabe que la temperatura operativa óptima[2] (TO) también sigue una distribución normal con σ = 2ºC
Calcular la TO media de modo que, trabajando el PC 24 h al día durante largos periodos de tiempo, la probabilidad de superar la TM sea inferior al 1%.

NOTA:
[1] Temperatura máxima a la que puede trabajar la CPU sin producir daños en el sistema.
[2] Temperatura de trabajo de la CPU (incluidos o no disipadores, ventiladores u otros componentes).

SOLUCIÓN:

Variables aletorias.:
TM = {temperatura operativa máxima (ºC)} ~ Normal (m_{TM} = 70ºC, σ_{TM} = 1ºC)
TO = {temperatura operativa óptima (ºC)} ~ Normal (m_{TM} = **?** ºC, σ_{TM} = 2ºC)

P(TO > TM) < 0,01 → Si TO > TM → (TO – TM > 0) ó (TM – TO < 0)

Variable aleatoria D = {Diferencia temperaturas: TO - TM (ºC)} ~ **?** *(*)*
(*) La suma de v.a. normales es otra v.a. Normal

$m_D = m_{TO} - 70$ *(**)*

TO y TM se suponen independientes →

$$\sigma_D^2 = \sigma_{TM}^2 + \sigma_{TO}^2 = 1^2 + 2^2 = 5\ ^{\circ}C \quad (***)$$

(**) La media de una resta de v.a. es la resta de sus medias
(***)La varianza de una resta de v.a. INDEPENDIENTES es la suma de sus varianzas.

D = {Diferencia temperaturas: TO - TM (ºC)} ~Normal(media=m_{TO}-70, varianza=5ºC)

$$P(D>0)= P\left(N(0,1) > \frac{0-(m_{TO}-70)}{\sqrt{5}}\right) < 0{,}01 \Rightarrow \frac{70-m_{TO}}{\sqrt{5}} > 2{,}33 \Rightarrow m_{TO} < 64{,}79$$

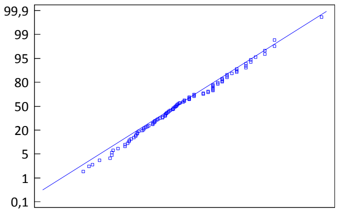
Tabla Normal

22. Se han estudiado cuatro variables correspondientes a diferentes experiencias aleatorias. A la izquierda se muestran algunos parámetros o gráficos descriptivos de cada una de las muestras, y a la derecha se muestran diversas conclusiones que se pueden extraer de cada muestra a partir de la su análisis descriptivo.

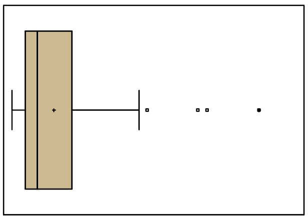
Empareja correctamente cada muestra con su conclusión, razonando de manera breve tu respuesta.

Muestra A

- Coef. de asimetría estandarizado = −0,01
- Coef. de curtosis estandarizado = 0,1



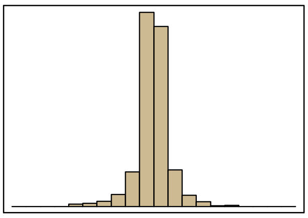
Muestra B



Muestra C

- Media = 34,5
- Mediana = 75

Muestra D



Conclusión 1
Los datos presentan asimetría negativa, o bien hay presencia de valores anómalamente pequeños.

Conclusión 2
Los datos presentan una fuerte asimetría positiva.

Conclusión 3
Los datos no muestran asimetrías ni una presencia relevante de valores anómalos.

Conclusión 4
Los datos se distribuyen aproximadamente según una campana de Gauss.

SOLUCIÓN:

Muestra A –Conclusión 4
La muestra tno tiene ni asimetría ni curtosis, y en el papel probabilístico normal se presenta alineada.

Muestra B – Conclusión 2
El diagrama Box-Whisker muestra que entre m´nimo y mediana hay mucha menos distancia que entre mediana y máximo.

Muestra C – Conclusión 1
La media es mucho más pequeña que la mediana, lo que puede ser debido a la presencia de valores extremos por la izquierda y/o asimetría por la izquierda.

Muestra D – Conclusión 3
El histograma es aproximadamente simétrico, sin presencia de barras separadas.

23. El tiempo de lectura de un dato en una unidad de disco se distribuye exponencialmente con media 3 milisegundos. En un determinado proceso deben leerse consecutivamente 500 datos situados aleatoriamente en el disco. Calcular la probabilidad de que el tiempo total de lectura de los datos sea mayor que 1 segundo.

SOLUCIÓN:

Sea X_i el tiempo de lectura de un dato. Se distribuye exponencial con media 0,003 segundos y varianza $0{,}003^2$ segundos^2.

Sea Y el tiempo total de lectura de 500 datos $\Rightarrow$ $Y = \sum_{i=1}^{500} X_i$ Como Y es la suma de 500 variables independientes, por el Teorema Central del Límite se aproximará a la distribución normal. Su media será m_Y=500x0,003= 1,5 segundos y su desviación típica $\sigma_Y = \sqrt{500x0{,}003^2} = 0{,}067$ segundos. Por tanto:

$P(Y>1) \approx P(N(1{,}5, 0{,}067)>1) = P(N(0,1) > \frac{1-1{,}5}{0{,}067}) = P(N(0,1)>-7{,}46) \approx 1$

24. Un montacargas se utiliza para transportar paquetes cuyo peso fluctúa normalmente con media 200 Kg. Se sabe que el 30,5 % de los paquetes supera los 210 Kg.
¿Qué probabilidad existe de que al tomar un paquete al azar su peso no supere los 195Kg?

SOLUCIÓN:

$P(\text{peso} > 210)=0{,}305 \Rightarrow P(N(0,1)> \frac{210-200}{\sigma})=0{,}305 \Rightarrow \frac{10}{\sigma}=0{,}51 \Rightarrow \sigma=19{,}6$

$P(\text{peso} < 195)=P(N(0,1)< \frac{195-200}{19{,}6}) = P(N(0,1)<-0{,}25)=0{,}4013$

25. El peso de las naranjas de un determinado calibre fluctúa normalmente con media m=150 grs y σ=30 grs. Una bolsa se llena con 15 naranjas seleccionadas al azar.
a) ¿Cuál es la probabilidad de que la mayor naranja de la bolsa pese más de 200 grs?
b)¿Cuál es la probabilidad de que la menor naranja de la bolsa pese menos de 120 grs?
c) ¿Cuál es la probabilidad de que el peso total de la bolsa sea inferior a 2 Kgs?

SOLUCIÓN:

a) Probabilidad de que una naranja seleccionada al azar pese más de 200 grs:

$$P(\text{peso} > 200) = P\left(\frac{\text{peso}-150}{30} > \frac{200-150}{30}\right) = P(N(0,1) > 1{,}67) = 0{,}0475$$

P(mayor>200)=1-P(mayor<200)=1-P(todas<200)=
$=1-(1-0{,}0475)^{15}=0{,}518$

b)
Probabilidad de que una naranja seleccionada al azar pese menos de 120 grs:

$$P(\text{peso} < 120) = P\left(\frac{\text{peso}-150}{30} < \frac{120-150}{30}\right) = P(N(0,1) < -1) = 0{,}1587$$

P(menor<120)=1-P(menor>120)=1-P(todas>120)=
$=1-(1-0{,}1587)^{15}=0{,}925$

c)
Sea X_i el peso de la naranja i-ésima (i=1,...,15) e Y el peso neto total de la bolsa:
$Y=X_1+...+X_{15} \Rightarrow Y\sim$Normal con
$m_Y=150+\ldots+150=2250$
$\sigma^2_Y=30^2+\ldots+30^2=13500$

$$P(Y < 2000) = P\left(\frac{Y-2250}{\sqrt{13500}} < \frac{2000-2250}{\sqrt{13500}}\right) = P(N(0,1) < -2{,}15) = 0{,}0158$$

26. Una partida defectuosa de tornillos tiene una resistencia a la torsión X que fluctúa normalmente con media 25 nwt y σ = 4 nwt. Los tornillos se utilizan para cerrar las cajas que contiene ciertos módulos electrónicos y se atornillan mediante una atornilladora que produce un par de apriete Y que fluctúa normalmente con una media de 12 nwt y una σ = 3 nwt.
¿Qué porcentaje de los tornillos se partirán al apretarse?

SOLUCIÓN:

P(se parta el tornillo) = P(Y > X) = P((Y-X) > 0)
Z = (Y-X) será Normal, por ser combinación lineal de dos variables normales independientes con $m_Z = m_Y - m_X = 12 - 25 = -13$ y $\sigma_Y = \sqrt{4^2+3^2} = 5$

P(parta tornillo) = P(Z>0) = $P\left(N(0,1) > \frac{0-(-13)}{5} = 2.6\right)$ = 0.0047

(se partirán el 4.7 por mil de los tornillos)

27. Calcular aproximadamente la probabilidad de sacar más de 80 puntos al lanzar 20 dados.

SOLUCIÓN:

Sea X_i el número de puntos al lanzar el dado i. Se tiene:

$E(X_i)=1(1/6)+2(1/6)+...+6(1/6)=3,5$

$\sigma^2(X_i)=(1-3,5)2(1/6)+(2-3,5)2(1/6)+... +(6-3,5)2(1/6)=2,92$

Total puntos $Y=X_1+...+X_{20} \Rightarrow Y \approx$Normal $m_y=20x3,5=70$ y $\sigma_Y=\sqrt{20x2,92}=7,64$

$$P(Y>80) \approx P(N(70,7,64)>80,5)=P(N(0,1)>\frac{80,5-70}{7,64})=P(N(0,1)>1,37)=0,0853$$