Tema 5. MODELOS DE DISTRIBUCIONES CONTINUOS.

Objetivos

Conceptos:

- Conocer los siguientes modelos continuos de probabilidad: uniforme, normal, exponencial, gamma y Pareto. De cada uno de ellos:

  * Tipo de experimentos que modelizan
  * Función de densidad
  * Esperanza y varianza
  * Propiedades gráficas y de asimetría
  * Propiedades de reproductividad (suma de v.a.)

- Conocer y comprender el Teorema Central del Límite

Saber hacer:

- Dada una variable **aleatoria continua**:
  * Reconocer si sigue algún modelo de probabilidad
  * Calcular probabilidades, utilizando las tablas o bien a partir de las fórmulas adecuadas
  * Hallar esperanza y varianza.

- Hallar probabilidades para una variable con distribución normal utilizando las tablas de N(0,1)

- Utilizar el TCL para poder calcular, mediante la normal, probabilidades de diversas v.a. (sumas de otras v.a independientes o medias aritméticas de dichas v.a.)

- Dada una variable **estadística continua**:
  * Estudiar si se ajusta a alguno de los modelos de probabilidad estudiados
  * Determinar los parámetros correspondientes a dicho modelo

Problemas de exámenes (web):

SIN: Diciembre 2005: Problema 2 (d)
Septiembre 2005. Problema 2
Septiembre 2004: Problema 2
Febrero 2003: Problema 2
Junio 2003: Problema 2 A

CON: Diciembre 2005: (b) (modelos A y B)
Junio 2006: (f)

5.1.- Introducción.

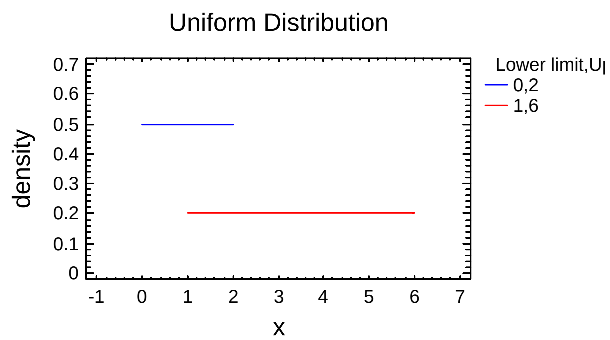
Al igual que en el caso discreto, hay situaciones que siguen modelos de distribución continua muy similares.

- Errores de redondeo, truncamiento
- Medidas de una característica de una población: estatura, peso, notas, ...
- Tiempo de vida de seres vivos, instrumentos, ...

Recordamos que los modelos son **simplificaciones** de la realidad, no se ajustan exactamente a ella, pero nos sirven para poder comprenderla mejor.

Además, veremos que uno de estos modelos, el modelo normal, se relaciona con todos los demás. Bajo determinadas condiciones, **la suma de v.a. de cualquier modelo tenderá a parecerse al modelo normal**.

5.2.- Distribución uniforme continua.

<table>
<tr><td>Modelo</td><td colspan="3">Uniforme continua U(a,b)</td></tr>
<tr><td>Tipo de experimento</td><td colspan="3">X: toma valores en un intervalo [a,b] de tal forma que subintervalos de igual amplitud tienen la misma probabilidad</td></tr>
<tr><td>Ejemplos</td><td colspan="3">1) X: Errores cometidos al redondear a un decimal un resultado X~U(-0.05,0.05)
2) Y: Tiempo de espera en una parada de autobús o metro que viene con una frecuencia de 10 minutos Y~U(0,10)</td></tr>
<tr><td>Función de densidad</td><td colspan="3">$f(x)=\frac{1}{b-a}, \; si \;\; x\in[a,b]$; 0 en el resto.</td></tr>
<tr><td colspan="2">E(X) = $\frac{a+b}{2}$</td><td>V(X)= $\frac{(b-a)^2}{12}$</td><td>CAF=0 (simétrica)</td></tr>
<tr><td colspan="4">
</td></tr>
</table>

Ejemplo:

Sabemos que un autobús pasa cada 10 minutos. Hallar la probabilidad de esperar menos de 2 minutos, el tiempo medio de espera y la desviación típica del tiempo de espera.

Sea Y: Tiempo de espera en la parada del autobús

Como el autobús pasa cada 10 minutos y no sabemos cuánto tiempo hace que ha pasado, nada nos indica que sea más probable esperar un tiempo u otro, por lo que Y sigue un modelo uniforme: **Y~U(0,10)** (0 es el tiempo mínimo de espera y 10 el tiempo máximo de espera)

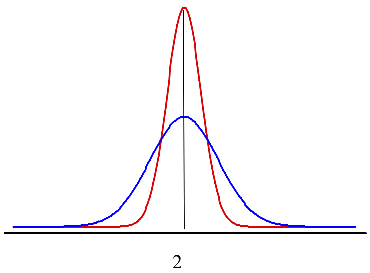
Su función de densidad será $f(x) = \frac{1}{10-0} = \frac{1}{10}$ $\textit{si } 0 < x < 10 \textbf{ ; } 0 \textit{ en el resto}$.

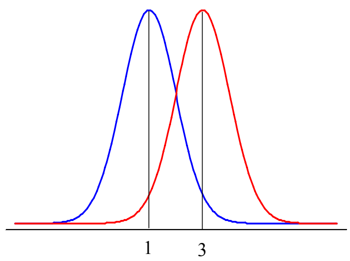
La probabilidad de esperar menos de 2 minutos, será $P(Y<2) = \int_{-\infty}^{2} f(x)dx = \int_{0}^{2} \frac{1}{10} dx = 0.2$

El tiempo medio de espera será $E(Y) = \frac{0+10}{2} = 5$ minutos.

La desviación típica: $dt = \sqrt{V(Y)} = \sqrt{\frac{(10-0)^2}{12}} = \frac{10}{\sqrt{12}} = 2.89$ minutos.

5.3.- Distribución normal.

<table>
<tr><td>Modelo</td><td colspan="3">Normal N(μ , σ)</td></tr>
<tr><td>Tipo de experimento</td><td colspan="3">X: medida de variables físicas: tamaños, temperaturas, errores, ...
X: suma de gran número de experimentos independientes ≈ Normal
(aunque es discreta, se puede aproximar por una normal)</td></tr>
<tr><td>Ejemplos</td><td colspan="3">1) Estatura de una población
2) Número de cafés vendidos en una semana en una cafetería ≈ Normal</td></tr>
<tr><td>Función de densidad</td><td colspan="3">$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \forall x \in \mathbb{R}$</td></tr>
<tr><td colspan="2">E(X) = μ</td><td>V(X)= σ^2 ; dt(X) = σ</td><td>$CAF = 0$: simétrica</td></tr>
<tr><td colspan="2">

Igual media, distintas desviaciones típicas</td><td colspan="2">

Distintas medias, igual desviación típica</td></tr>
</table>

Cálculo de probabilidades de la normal

Estandarización de la normal:

Si X ~ $N(\mu,\sigma)$, entonces $\dfrac{X-\mu}{\sigma} \sim N(0,1)$[1]. *Notación*: $Z \sim N(0,1)$.

Cálculo de probabilidades:

Si X ~ $N(\mu,\sigma)$, entonces:

$$P(a<X<b)=P\left(\frac{a-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{b-\mu}{\sigma}\right)=P\left(\frac{a-\mu}{\sigma}<Z<\frac{b-\mu}{\sigma}\right)$$, con $Z \sim N(0,1)$

Tablas de Z:

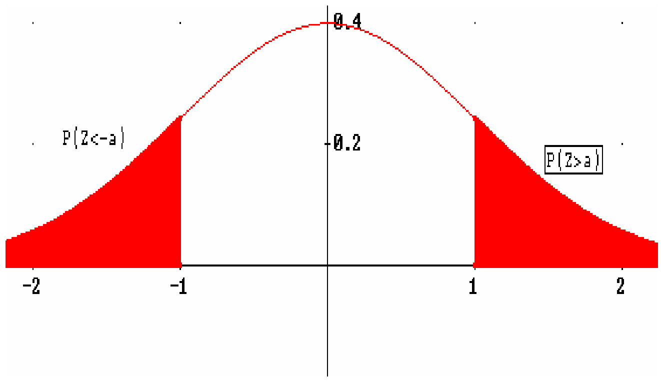
Da los valores de P(Z<**a**) con **a**>0.

Para el resto de los casos:

- P(Z>**a**)=1-P(Z<**a**)

Por la simetría de la normal:

- P(Z<**-a**)= P(Z>**a**)=1-P(Z<**a**)
- P(Z>**-a**)=P(Z<**a**)



Ejemplo:

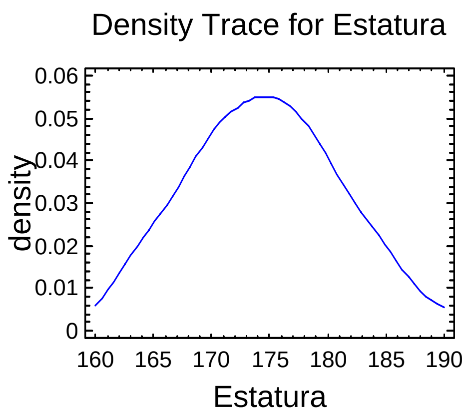
X: Estatura de los alumnos de Estadística, y suponemos X ≈ N(175,6)[2].

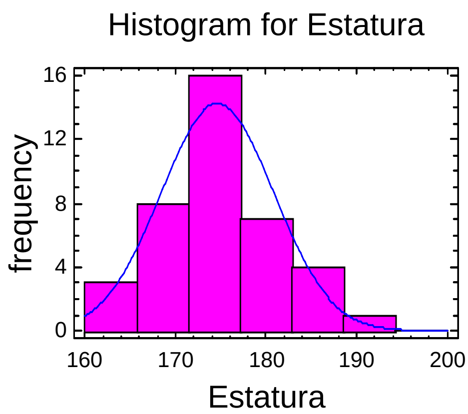
---

[1] Si X ~ $N(\mu,\sigma)$, E(X)=μ, V(X)=σ^2. Por tanto:

$$E\left(\frac{X-\mu}{\sigma}\right)=\frac{1}{\sigma}\big(E(X)-E(\mu)\big)=\frac{1}{\sigma}(\mu-\mu)=0, \text{ y } V\left(\frac{X-\mu}{\sigma}\right)=\frac{1}{\sigma^2}\big(V(X)+V(\mu)\big)=\frac{1}{\sigma^2}\left(\sigma^2+0\right)=1$$

[2] Ejemplo: Estatura de alumnos del GM23






```
Fitted normal distribution:
     mean = 174.615
     standard deviation = 6.24338
```

```
Approximate P-Value = 0.66543
```
> 0.3

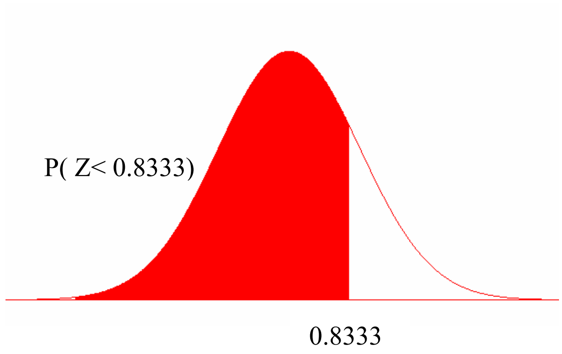
Aceptamos que sigue una distribución normal

Probabilidad de medir menos de 180:

$$P(X<180)=P\left(Z<\frac{180-175}{6}\right)=P\left(Z<0.8333\right)$$

Tablas:

$$P(X<180)=0.7967$$



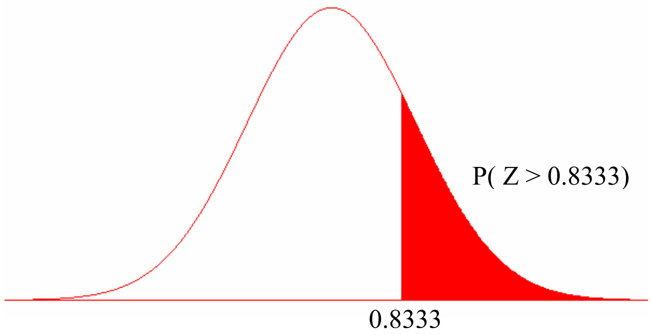
Probabilidad de medir más de 180:

$$P(X>180)=P\left(Z>\frac{180-175}{6}\right)=P\left(Z>0.8333\right)$$

Tablas:

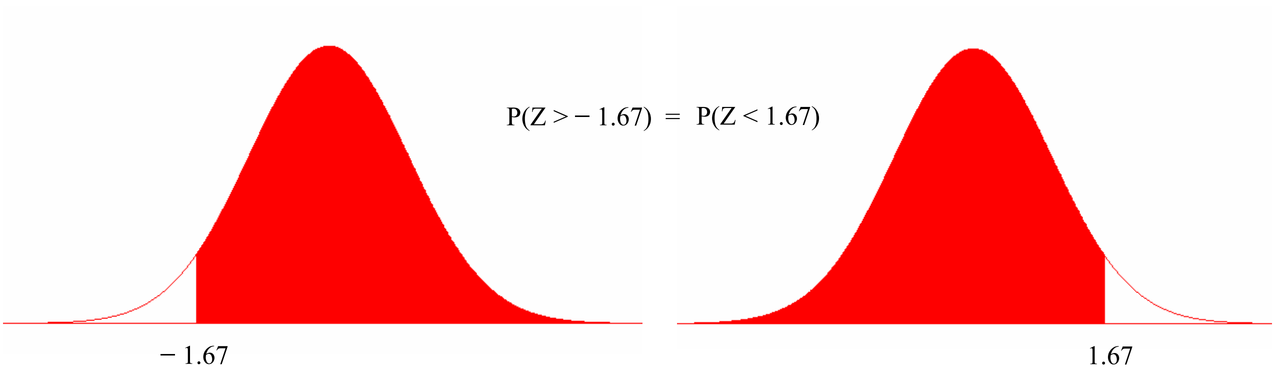
$P(X>180)=1$ – P(Z<0.83)= 1 - 0.7967=

= 0.2033



Probabilidad de medir más de 165:

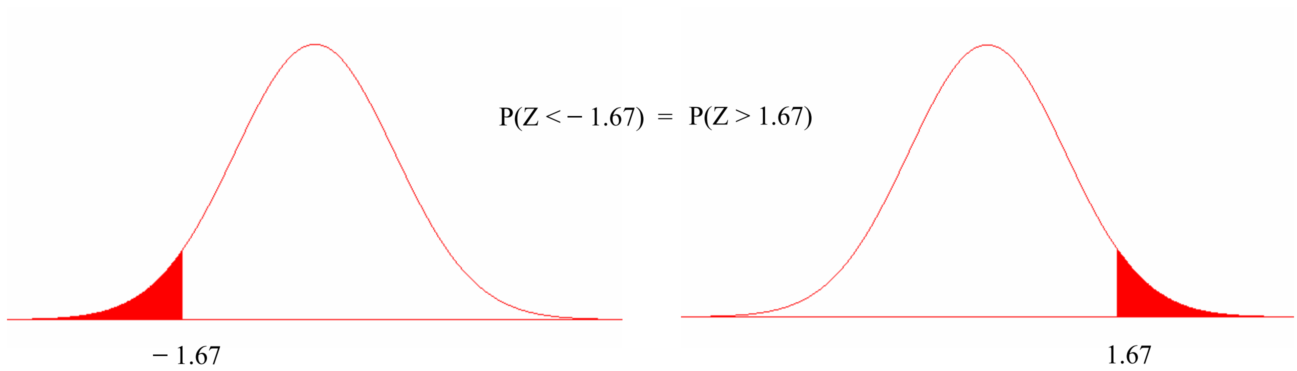
$$P(X>165)=P\left(Z>\frac{165-175}{6}\right)=P\left(Z>-1.67\right)$$; *Tablas:* $P\left(Z>-1.67\right)=P\left(Z<1.67\right)=$ 0.9525



Probabilidad de medir menos de 165:

$$P(X<165)=P\left(Z>\frac{165-175}{6}\right)=P\left(Z<-1.67\right)$$;

Tablas: $P\left(Z<-1.67\right)=P\left(Z>1.67\right)=1-P\left(Z<1.67\right)=1-0.9525=0.0475$

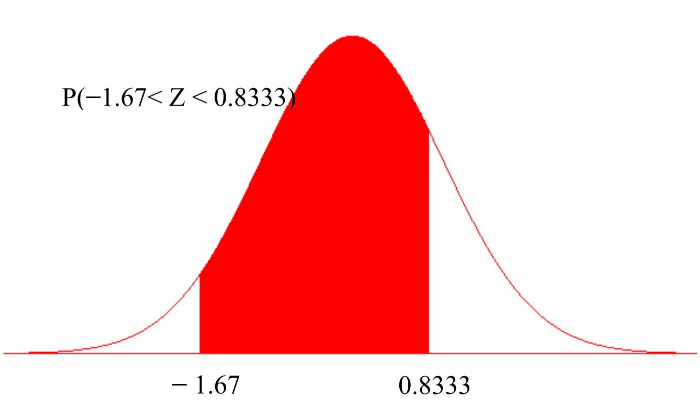


Probabilidad de medir entre 165 y 180:

$$P(165 < X < 180) = P\left(\frac{165-175}{6} < Z < \frac{180-175}{6}\right) = P\left(-1.67 < Z < 0.8333\right)$$

Tablas:

P(Z<0.83) – **P(Z<-1.67)** = P(Z<0.83) – **(1 - P(Z<1.67))** = 0.7967 – (1 – 0.9525) = 0.7492



Otras propiedades de la normal son:

Reproductividad de la normal:

Si X e Y son v.a. **independientes** tales que X ~ $N(\mu_X, \sigma_X)$, Y ~ $N(\mu_Y, \sigma_Y)$. Entonces:

i) **X+Y** ~ $N(\mu_X + \mu_Y, \sqrt{\sigma_X{}^2 + \sigma_Y{}^2})$.[3]

ii) *k* **X** ~ $N(k\mu_X, |k|\sigma_X)$.[4]

Ejemplos:

Si X e Y son v.a. **independientes** tales que X ~ $N(2,5)$, Y ~ $N(-3,1)$. Entonces:

i) **X+Y** ~ $N(2-3, \sqrt{5^2+1^2}) = N(-1, \sqrt{6})$.

ii) *(– 4)* **X** ~ $N(-8, 20)$.

[3] $E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$; por la independencia $V(X+Y) = V(X) + V(Y) = \sigma_X{}^2 + \sigma_Y{}^2$.

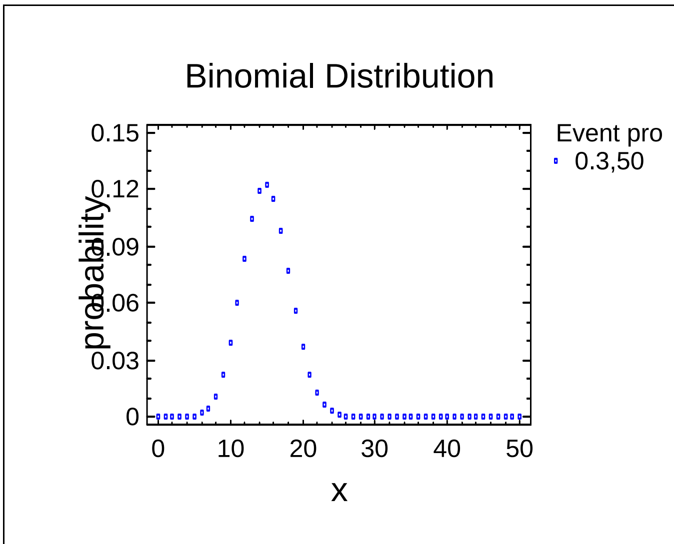
[4] $E(kX) = kE(X) = k\mu_X ; V(kX) = k^2 V(X) = k^2\sigma_X^2 \Rightarrow \sqrt{V(kX)} = \sqrt{k^2\sigma_X^2} = |k|\sigma_X$.

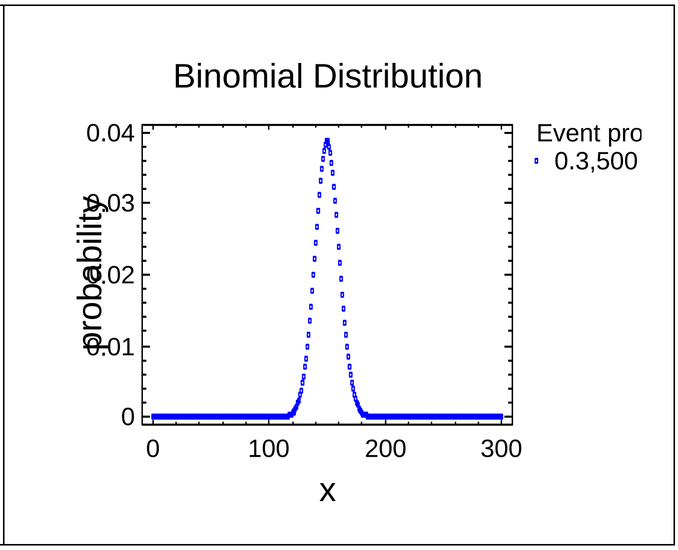
5.4.- Teorema Central del Límite.

Acabamos de ver que la suma de normales es otra distribución normal de media la media de la suma y varianza la varianza de la suma.

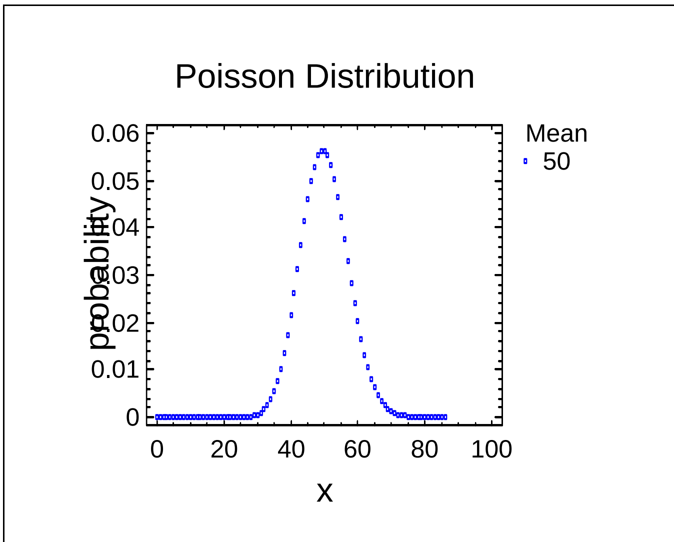
En determinadas condiciones, la **suma de v.a.** cualesquiera se podrá **aproximar por una distribución normal** de media la media de la suma y varianza la varianza de la suma.

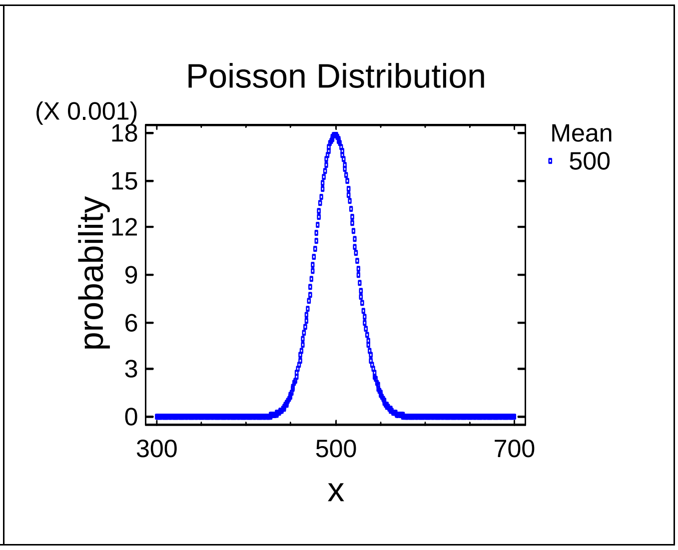
Hemos visto en el tema anterior que si X ~ B(n,p), se puede expresar como $X=X_1+...+X_n$ siendo X_i ~ B(1,p) v.a. **independientes**. Si **n** es **suficientemente grande** observamos que la distribución de probabilidad se asemeja a la de la normal:





Algo análogo ocurre con el modelo de Poisson. Si X ~ P(n), se puede expresar como $X=X_1+...+X_n$ siendo X_i ~ P(1) v.a. **independientes**. Si **n** es **suficientemente grande** observamos que la distribución de probabilidad también se asemeja a la de la normal:





¿Ocurre esto con otros modelos de distribución? ¿En qué condiciones?

Teorema Central del Límite. (TCL)

Sean X_1, ... , X_n v.a.i.i.d. (variables aleatorias **independientes con idéntica distribución**). Entonces, **cuando n → +∞**, la distribución de $\mathbf{Y = X_1 + \ldots + X_n}$ se puede aproximar por una distribución normal:

$$\mathbf{Y = X_1 + \ldots + X_n} \approx N(E(Y), \sqrt{V(Y)}), \quad \text{o bien} \quad \frac{Y - E(Y)}{\sqrt{V(Y)}} \approx N(0,1)$$

Si denotamos $E(X_i)=\mu$ y $V(X_i)=\sigma^2$, se tiene que:

- $\mathbf{X_1} + \ldots + \mathbf{X_n} \approx N(n\mu, \sigma\sqrt{n})$ [5]
- $\dfrac{\mathbf{X_1} + \ldots + \mathbf{X_n}}{n} \approx N(\mu, \dfrac{\sigma}{\sqrt{n}})$ [6]

NOTA: En general, se considera válido el TCL si **n≥100**.

Ejemplos:

Sea *X*: tiempo diario (horas) viendo la televisión, cuyo modelo de distribución no se conoce, pero sí se conoce E(X)=2 y V(X)=9.

a) Hallar la probabilidad de que al cabo de un año una persona haya visto la televisión durante más de 600 horas.

Si definimos *Y*: tiempo de televisión al cabo de un año, nos piden P(*Y*>600).

Como *Y* se puede expresar como $Y = X_1 + \ldots + X_{365}$. Suponiendo independencia entre los días, como n=365> 100, se tiene que:

$$Y \sim N(365 \cdot 2, 3\sqrt{365})$$

Por tanto, $P(Y > 600) = P\left(\dfrac{Y - 730}{3\sqrt{365}} > \dfrac{600 - 730}{3\sqrt{365}}\right) = P(Z > -2.27)$, siendo $Z \sim N(0,1)$.

Buscando en las tablas, $P(Y > 600) = P(Z > -2.27) = P(Z < 2.27) = 0.9884$.

b) Hallar la probabilidad de que, eligiendo una persona al azar, el tiempo medio diario que dedica a ver la televisión a lo largo de un año sea menor que hora y media.

Si definimos *M*: tiempo medio de televisión diaria al cabo de un año, nos piden P(*M*<1.5).

M se puede expresar como $M = \dfrac{\mathbf{X_1} + \ldots + \mathbf{X_{365}}}{365}$. Suponiendo independencia entre los días, como n=365> 100, se puede aplicar TCL y se tiene que:

---

[5] $E(Y) = E(X_1 + \ldots + X_n) = \mu + \ldots + \mu = n\mu$

$V(Y) = V(X_1 + \ldots + X_n) = \sigma^2 + \ldots + \sigma^2 = n\sigma^2 \Rightarrow \sigma_Y = \sqrt{n\sigma^2} = \sigma\sqrt{n}$ (es fundamental la independencia)

[6] $E\left(\dfrac{\mathbf{X_1} + \ldots + \mathbf{X_n}}{n}\right) = \dfrac{1}{n}\left(E(\mathbf{X_1}) + \ldots + E(\mathbf{X_n})\right) = \dfrac{1}{n}(\mu + \ldots + \mu) = \dfrac{1}{n}(n\mu) = \mu$

$V\left(\dfrac{\mathbf{X_1} + \ldots + \mathbf{X_n}}{n}\right) = \dfrac{1}{n^2}(V(\mathbf{X_1}) + \ldots + V(\mathbf{X_n})) = \dfrac{1}{n^2}(\sigma^2 + \ldots + \sigma^2) = \dfrac{1}{n^2}(n\sigma^2) = \dfrac{\sigma^2}{n} \Rightarrow \sigma_Y = \dfrac{\sigma}{\sqrt{n}}$

$$M \sim N(2, \frac{3}{\sqrt{365}})$$

Por tanto, $P(M < 1.5) = P\left(\dfrac{M-2}{3/\sqrt{365}} < \dfrac{1.5-2}{3/\sqrt{365}} \right) = P(Z < -9.55) \approx 0$.

Casos particulares:

- Si X ~ P(λ), consideraremos que $X \approx N(\lambda, \sqrt{\lambda})$ si λ > 10.
- Si X ~ B(n,p), consideraremos que $X \approx N(np, \sqrt{np(1-p)})$
    - si $n \geq 50$ y $0.1 < p < 0.9$.
    - si $n \geq 50$, $p \leq 0.1$ y $np > 10$ ([7]).
    - si $n \geq 50$, $p \geq 0.9$ y $n(1-p) > 10$ ([8]).

Ejemplos:

- Si X ~ B(75, 0.3), se verifica que $X \approx N\left(22.5, \sqrt{75 \cdot 0.3 \cdot 0.7}\right) = N\left(22.5, 3.97\right)$.
- Sea Y: Número de cafés vendidos en una semana en una cafetería

Por su definición, es una v.a. discreta ("número de"), que seguirá un modelo de Poisson: número de "exitos" (cafés vendidos) en un intervalo (semana). Si el número medio de cafés vendidos en una semana es de 10000, tendríamos Y ~ P(**10000**).

Se verifica que $X \approx N(10000, \sqrt{10000}) = N(10000, 100)$.

(Por la reproductividad de la Poisson, podríamos suponer que $Y = Y_1 + \ldots + Y_{100}$, siendo Y_i=número de cafés vendidos en media hora. El TCL, nos asegura que podemos aproximas Y por una normal.)

Para calcular probabilidades, conviene tener en cuenta que Y es discreta (probabilidad en puntos) y que la normal es continua (probabilidad en intervalos). En general, se asocia

$$P(Y=r) \approx P(r-0.5 < N < r+0.5)$$

Si queremos hallar la probabilidad de vender en una semana 10500, o más, cafés:

$$P(Y \geq 10500) \approx P\left(N > 10499.5\right) = P\left(\frac{N-10000}{100} > \frac{10499.5-10000}{100} \right) = P(Z > 4.995) \simeq 0$$

---

[7] Recordad que si $n \geq 50$, $p \leq 0.1$ y $np \leq 10$, X ~ B(n,p) se puede aproximar por $X \approx P(np)$ (tema 4, pg 7)

[8] Si si $n \geq 50$, $p \geq 0.9$ y $n(1-p) \leq 10$, se trabaja con Y ~ B(n,1-p), que se puede aproximar por $Y \approx P(n(1-p))$.

5.5.- Distribución gamma. Distribución exponencial.

Modelo	**Exponencial Exp(β)**
Tipo de experimento	X: **tiempo** transcurrido **hasta que ocurre un suceso de Poisson** (β: número medio de sucesos de Poisson por unidad de tiempo) X: tiempo de "funcionamiento" de "algo" (en el caso de que el "fallo" no depende del tiempo que lleva "funcionando")
Ejemplos	1) **Tiempo** (en minutos) transcurrido hasta recibir el primer e-mail del día. 2) Tiempo transcurrido entre la llegada de un cliente y la llegada del siguiente 3) Tiempo de servicio a un cliente
Función de densidad	$f(x) = \beta e^{-\beta x}$, si $x>0$, 0 en el resto.

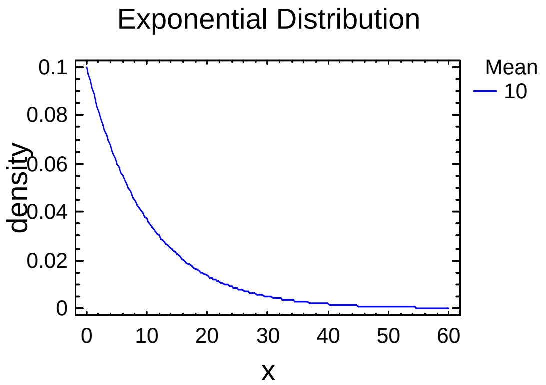
E(X) = $\dfrac{1}{\beta}$	**V(X) =** $\dfrac{1}{\beta^2}$

Simetría: $CAF = 2 > 0$ **:** simetría a la derecha.

Si el número de mensajes de correo-e que recibimos en un minuto sigue una distribución de Poisson de media β = 0.1, entonces

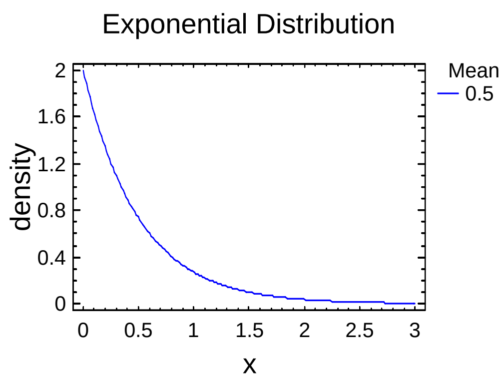
X: **Tiempo** (en minutos) transcurrido hasta recibir el primer e-mail del día, sigue una distribución exponencial con β = 0.1.

E(X)= $\dfrac{1}{\beta}$ = 10 (el **tiempo esperado** hasta recibir el primer e-mail es de 10 minutos)



Si nos dicen que X: **Tiempo** (en minutos) transcurrido hasta recibir el primer e-mail del día, sigue una distribución exponencial de media medio minuto, entonces:

E(X)= $\dfrac{1}{\beta} = \dfrac{1}{2} \Rightarrow \beta = 2$. Por tanto X ~ exp(2)



Otras propiedades:

La distribución exponencial no tiene memoria:

$$P\left(X \le t+h \big/ X > h\right) = P(X \le t)$$

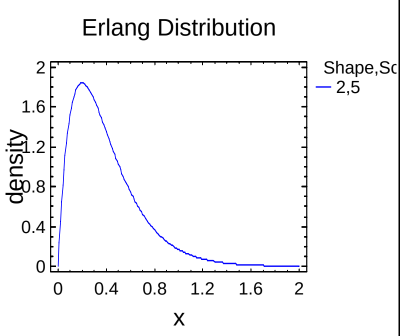
Ejemplo: Sea X es el tiempo empleado en encontrar una plaza de aparcamiento; si llevamos media hora de búsqueda infructuosa, la probabilidad de encontrar aparcamiento en los próximos 10 minutos es la misma que haberlo encontrado en menos de 10 minutos desde que empezamos a buscar:

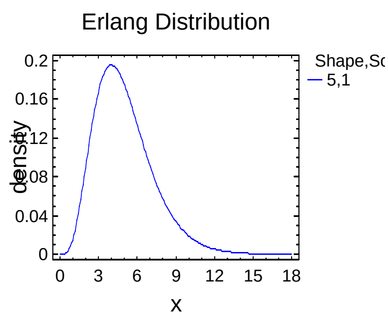
$$P\left(X \le 40 \big/ X > 30\right) = \frac{\int_{30}^{40} \beta e^{-\beta x} dx}{\int_{30}^{+\infty} \beta e^{-\beta x} dx} = \frac{-e^{-40\beta} + e^{-30\beta}}{e^{-30\beta}} = -e^{-10\beta} + 1$$

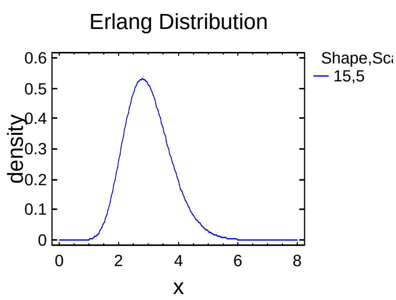
$$P(X \le 10) = \int_{0}^{10} \beta e^{-\beta x} dx = 1 - e^{-10\beta}$$

Reproductividad de la exponencial:

Si $X_1, \ldots, X_n$ son ***n*** v.a. independientes con $X_i \sim \exp(\beta)$, entonces $X_1 + \ldots + X_n \sim E(\boldsymbol{n}, \beta)$ (modelo que estudiaremos a continuación)

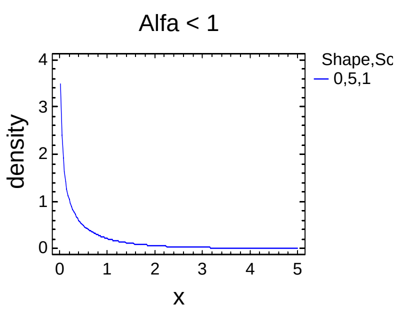
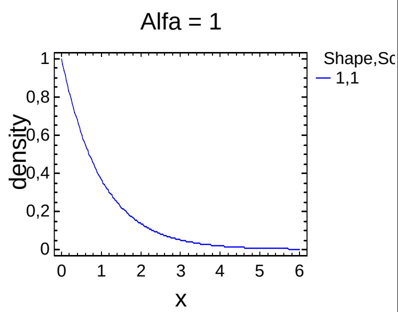
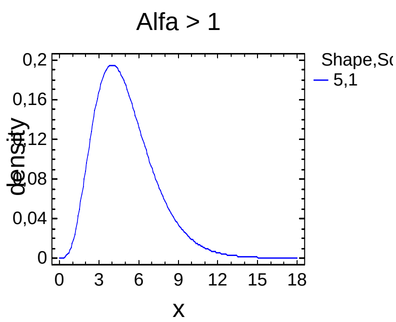
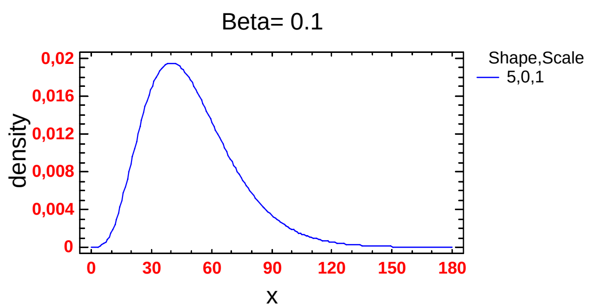
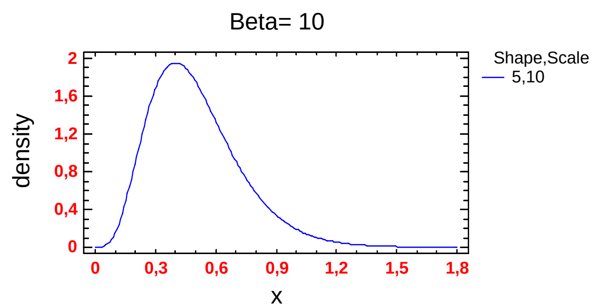
<table>
<tr><td>Modelo</td><td colspan="2">Erlang E(n , β)</td></tr>
<tr><td>Tipo de experimento</td><td colspan="2">X: tiempo transcurrido hasta que suceden n sucesos de Poisson
X: tiempo de vida de “algo”
(Si n=1, es el modelo exponencial.)</td></tr>
<tr><td>Ejemplos</td><td colspan="2">1) Tiempo transcurrido hasta que falla un sistema que depende del funcionamiento de un determinado número de dispositivos,</td></tr>
<tr><td>Función de densidad</td><td colspan="2">$f(x) = \frac{\beta^n}{(n-1)!} e^{-\beta x} x^{n-1}$, si $x > 0$, 0 en el resto.</td></tr>
<tr><td colspan="2">E(X) = $\frac{n}{\beta}$</td><td>V(X) = $\frac{n}{\beta^2}$</td></tr>
<tr><td>Simetría</td><td colspan="2">$CAF = \frac{2}{\sqrt{n}} > 0$: asimetría a la derecha.
Cuanto mayor es n, la distribución es más simétrica.</td></tr>
<tr><td>

E(2,5)</td><td>

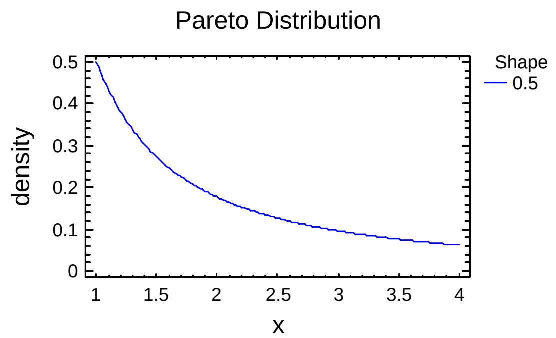
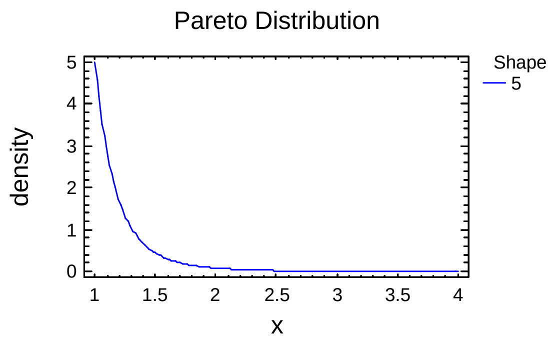
E(5,1)</td><td>

E(15,5)</td></tr>
<tr><td>Otras propiedades</td><td colspan="2">Reproductividad de la Erlang:
Si X e Y son v.a. independientes tales que X ~ E(n, β) , Y ~ E(m, β) entonces X+Y ~ E(n+m, β).
Si n es suficientemente grande $E(n,\beta) \approx N\left(\frac{n}{\beta}, \frac{\sqrt{n}}{\beta}\right)$ (TCL)</td></tr>
</table>

El modelo que generaliza al modelo exponencial y al modelo Erlang es el modelo gamma:

<table>
<tr><td>Modelo</td><td colspan="2">Gamma $\gamma\ (\alpha\ ,\ \beta\ )$</td></tr>
<tr><td>Tipo de experimento</td><td colspan="2">X: duración de “algo”</td></tr>
<tr><td>Ejemplos</td><td colspan="2">Duración de una llamada telefónica</td></tr>
<tr><td>Función de densidad</td><td colspan="2">$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1}$, si $x > 0$, 0 en el resto.
(Definición: $\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx$; si $\alpha = n \in$ N, $\Gamma(n) = (n-1)!$)
(Si α = 1, es el caso exponencial exp(β); si $\alpha = n \in$ N , es el caso Erlang E(n, β))</td></tr>
<tr><td colspan="2">E(X) = $\frac{\alpha}{\beta}$</td><td>V(X) = $\frac{\alpha}{\beta^2}$</td></tr>
<tr><td>Simetría</td><td colspan="2">$CAF = \frac{2}{\sqrt{\alpha}} > 0$: asimetría a la derecha.</td></tr>
<tr><td colspan="3">α es el parámetro que determina la forma (shape) (Cuanto mayor es α, mayor simetría):</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td colspan="3">β es el parámetro que determina la escala (scale):</td></tr>
<tr><td colspan="2">
</td><td>
</td></tr>
<tr><td>Otras propiedades</td><td colspan="2">Reproductividad de la Gamma:
Si X e Y son v.a. independientes tales que X ~ γ (α_X , β), Y ~ γ (α_Y , β) entonces X+Y ~ γ ($\alpha_X + \alpha_Y$, β).
Si α es suficientemente grande $\gamma(\alpha,\beta) \approx N(\frac{\alpha}{\beta}, \frac{\sqrt{\alpha}}{\beta})$ (TCL)</td></tr>
</table>

5.6.- Otras distribuciones continuas.

Modelo	**Pareto P(α , k)**
Tipo de experimento	X: **tiempo** de duración de "algo" , sabiendo que dicho tiempo es **superior a** k X: **cantidad** de "algo" , sabiendo que dicha cantidad es **superior a** k
Ejemplos	1) Tiempo en ser procesado un mensaje en un servidor web 2) Coste de una llamada telefónica, sabiendo que el primer minuto se paga siempre completo y es un precio fijo
Función de densidad	$f(x) = \alpha k^{\alpha} x^{-(\alpha+1)}$, **si** $x \geq k$, 0 en el resto. (α , k > 0)
E(X) = $\dfrac{\alpha k}{\alpha - 1}$, si α>1 ; **E(X) =** ∞ si $\alpha \leq 1$	**V(X)=** $\dfrac{\alpha k^2}{(\alpha-2)(\alpha-1)^2}$ si α>2 ; **V(X)=**∞ si $\alpha \leq 2$
Simetría	$CAF > 0$: simetría a la derecha ("cola pesada": más "densa" que en otros modelos; exponencial, por ejemplo)
	

Ejemplos:

1) Sea X: Tiempo (en segundos) en ser procesado un mensaje en un servidor web, X ~ P(3,2) (α=3 , k=2)

La función de densidad será: $f(x) = 3 \cdot 2^3 x^{-4}$, **si** $x \geq 2$.
La probabilidad de que un mensaje tarde más de 5 segundos en ser procesado será:

$$P(X \geq 5) = \int_5^{+\infty} 3 \cdot 2^3 x^{-4} dx = 0.064$$

El tiempo medio en ser procesado un mensaje es de 3 segundos:

$$E(X) = \int_2^{+\infty} x(3 \cdot 2^3 x^{-4}) dx = \int_2^{+\infty} \frac{24}{x^3} dx = 3 \qquad \text{o bien} \qquad \mathbf{E(X)} = \frac{3 \cdot 2}{3-1} = 3.$$

Ajuste de una variable estadística a un modelo teórico

Objetivo:

- elegir un modelo
- encontrar los parámetros del modelo

Medios:

- definición de la variable: ¿qué mide? ¿en qué condiciones?
- medidas descriptivas (media, varianza, simetrías, frecuencias, ...)
- representaciones gráficas

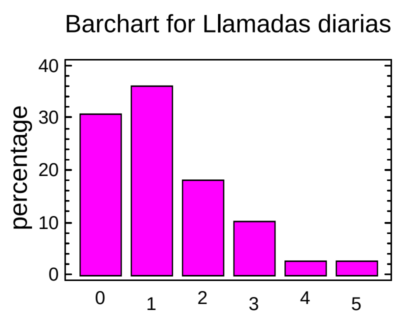
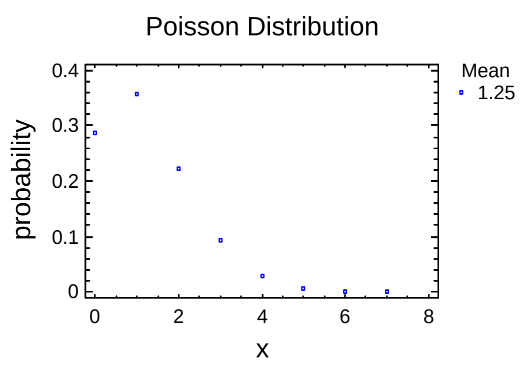
Verificación:

Contrastes no paramétricos con Statgraphics (Distribution Fitting):
p-valor > 0.3 para aceptar la hipótesis
(cuanto mayor sea, con más confianza se acepta el modelo propuesto)

Ejemplo:

X4: Número de llamadas diarias que se hacen por teléfono móvil
(Datos tomados en el curso 2004/05)

- Como la variable mide número de "exitos" (llamadas) en un intervalo de tiempo (un día), eso nos hace pensar en un modelo de Poisson.
- Como E(X)=1.25 y V(X)=1.51, que son valores más o menos similares, sería posible un modelo de Poisson de parámetro λ=1.25 (la media). (Con esos datos, sería menos posible un modelo binomial, pues en dicho modelo la varianza siempre es menor que la media)
- Observamos la representación gráfica del diagrama de barras (asimetría a la derecha) y lo comparamos con la distribución de probabilidad de la P(1.25), y vemos que son similares.

Diagrama de barras de frecuencias relativas	**Función de masa P(1.25)**
	

Por tanto, la hipótesis que tendríamos que verificar es si X4~P(1.25).

Utilizando la opción de Statgraphics Describe/Distributios/Distribution Fitting, y seleccionando la opción del modelo Poisson, obtenemos:

```
Goodness-of-Fit Tests for Llamadas diarias          :

        Fitted Poisson distribution:
             mean = 1.25641
```

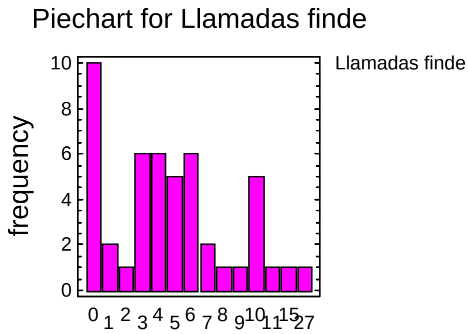
```
Chi-Square = 0.555238 with 2 d.f.   P-Value = 0.757586
```

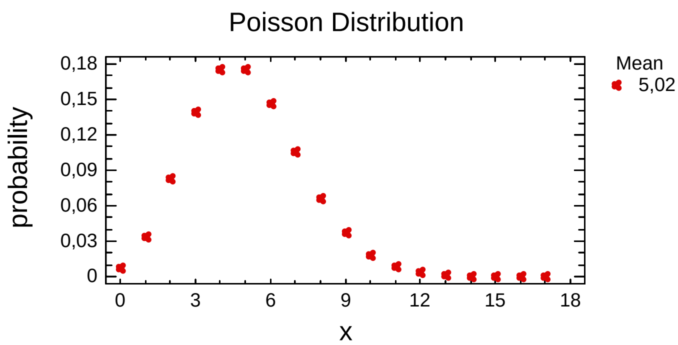
Como p-valor=0.75>0.3, aceptamos que la variable *número de llamadas diarias*, puede tener una distribución de Poisson de parámetro λ=1.25.

Si consideramos la variable Y: Número de llamadas que se hacen por teléfono móvil en un fin de semana, considerando los datos del curso 2006/07, tenemos lo siguiente:

- Como la variable mide número de "exitos" (llamadas) en un intervalo de tiempo (fin de semana), eso nos hace pensar en un modelo de Poisson.
- Como E(X)=5.02 y V(X)=23.5, que son valores muy dispares, esto nos indica que difícilmente sigue un modelo de Poisson.
- La representación gráfica del diagrama de barras y la de la distribución de probabilidad de la P(5), no se parecen.

Por tanto, no es muy probable que siga un modelo de Poisson. Esto se corrobora utilizando la opción de Statgraphics Describe/Distributios/Distribution Fitting, y seleccionando la opción del modelo Poisson, obtenemos P-Value = 0,000567705 < 0.3, por lo que rechazamos la idea de que la variable Y siga un modelo de Poisson.





Ejercicios:

Con los datos recogidos en este curso,estudiar si siguen algún modelo de distribución discreta las siguientes variables (http://www.eui.upm.es/~rafami/Estadistica/Material/material07.html):

- Nº de asignaturas matriculadas
- Nº de asignaturas aprobadas
- Nº de asignaturas aprobadas habiéndose matriculado de 11 asignaturas
- Nº de asignaturas aprobadas habiéndose presentado a 10 asignaturas