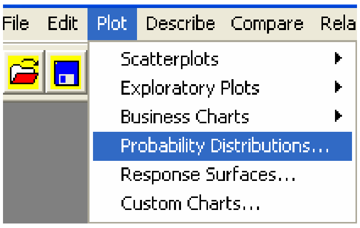
Modelos de Probabilidad con Statgraphics

1. Objetivos

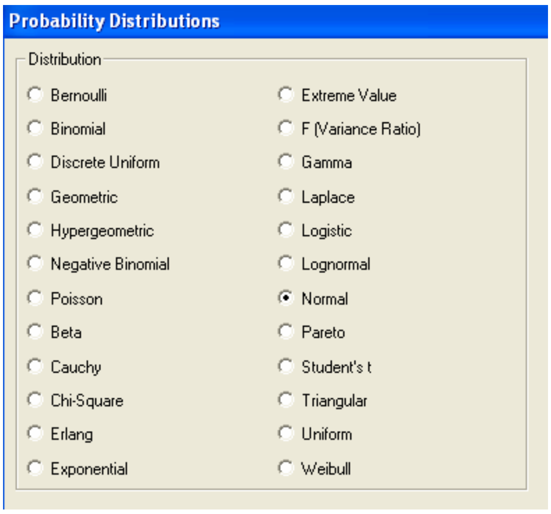
- Representar funciones de probabilidad/densidad y de distribución de diferentes modelos de variables aleatorias discretas/continuas
- Calcular probabilidades en distintas distribuciones
- Interpretar y comparar gráficos de distribuciones
- Modelizar situaciones reales mediante distribuciones de probabilidad

2. Modelos de distribuciones

Para acceder a los modelos de probabilidad que proporciona el Statgraphics seleccionamos *Plot / Probability distributions*



Observamos entonces que el Statgraphics Plus nos permite trabajar con veinticuatro distribuciones distintas de probabilidad. La ventana siguiente nos permite elegir el modelo de probabilidad.



De todos estos modelos, vamos a estudiar en detalle aquellos modelos de probabilidad que más frecuentemente surgen al analizar fenómenos de Ingeniería.

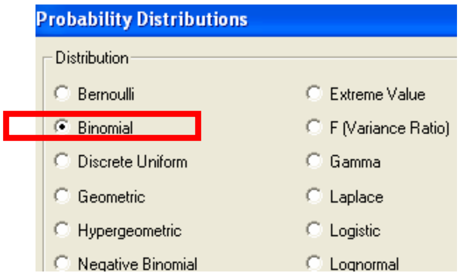
2.1. Distribuciones discretas: Binomial, Geométrica y Poisson.

2.1.1. Distribución Binomial, X~B(n,p)

Recordamos que una v.a. binomial con parámetros n y p representa una variable donde n es el número de repeticiones <u>independientes</u> del experimento (*number of trials*) , y p la probabilidad de éxito en cada repetición (event probability).

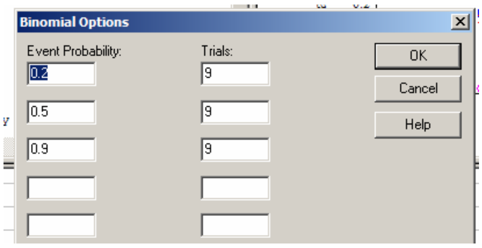
Representación gráfica de las funciones de probabilidad y distribución:

1. *Plot \ Probability Distributions \ Binomial \OK*

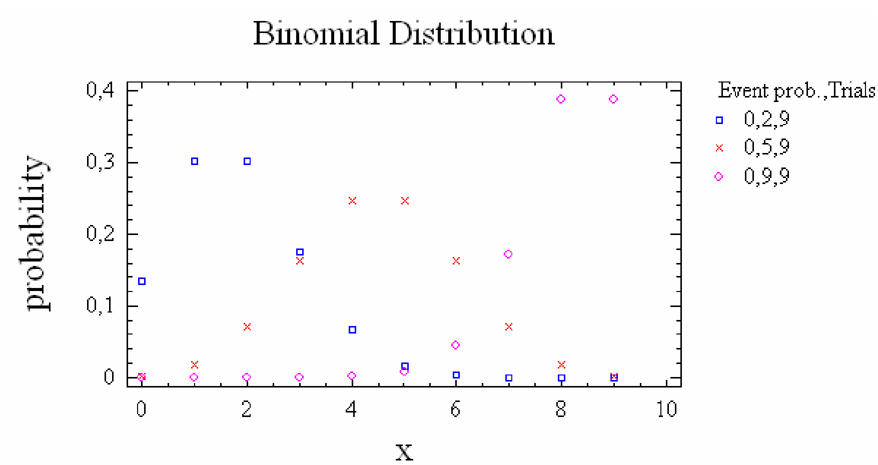


2. Ahora, se sitúa el cursor en cualquier punto de la pantalla, se pulsa el botón derecho del ratón y se selecciona *Análisis Options.* Aparecerá una pantalla que permite el estudio simultáneo de hasta cinco distribuciones del tipo seleccionado, con diferentes valores de los parámetros.

3. Introducimos los datos como en la figura adjunta y damos a *OK*. En el siguiente cuadro se han introducido los parámetros de las binomiales B(9,0.2), B(9,0.5) y B(9,0.9).

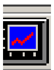


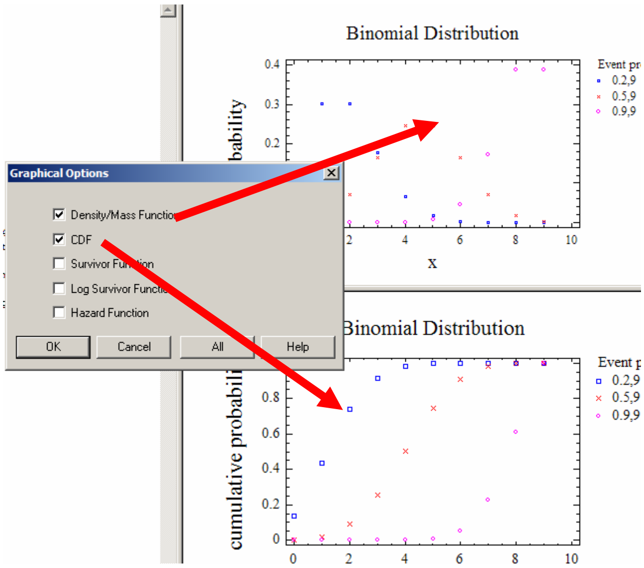
4. Maximizando el panel de gráficos se puede observar en detalle el gráfico adjunto, que representa la función de probabilidad de cada una de las tres binomiales anteriores.



Se puede observar que:

- para p=0.5, (en nuestro panel, los puntos marcados con cruces) la gráfica es **simétrica.**
- para p<0.5 (en nuestro panel, los puntos marcados con cuadrados, que corresponden a p=0.2), la gráfica es asimétrica a la derecha, es decir, la variable aleatoria presenta **asimetría positiva**
- para p>0.5 (en nuestro panel los puntos marcados con círculos, que corresponden a p=0.9), la gráfica es asimétrica por la izquierda, es decir, la variable aleatoria presenta **asimetría negativa**

5. Si también se quisiese mostrar la función de distribución, se hace clic sobre el botón *Graphical Options*  de la barra de herramientas y en el cuadro de diálogo que aparece se elige la opción correspondiente, en este caso *CDF* (CDF= cumulative distribution function) y *density/mass function*



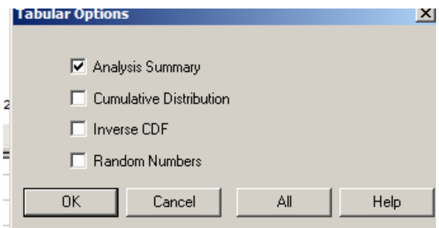
Cálculo de probabilidades

Supongamos una variable aleatoria X~B(12,0.4). Queremos calcular las probabilidades:

- P(X=7)
- P(X>3)
- P(X≤8)
- P(X<5)
-

Seguimos los siguientes pasos:

1. *Plot/Probability Distributions/ Binomial/OK*
2. Hacer clic sobre el botón  de *Tabular Options*. Aparecerá el siguiente cuadro de diálogo:

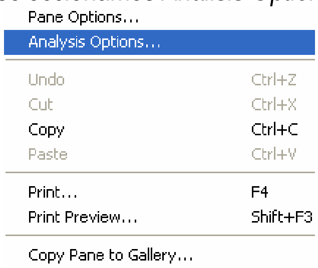


Este cuadro nos ofrece las siguientes opciones:

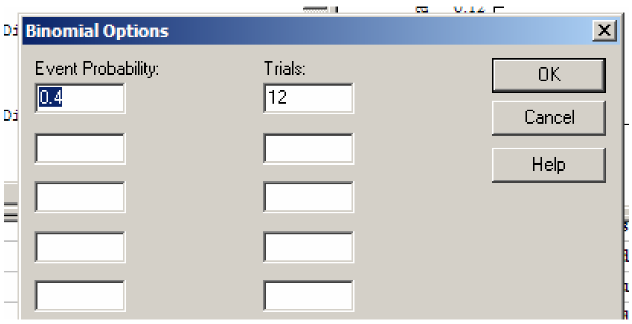
- *Análisis Summary*, opción que el programa activa por defecto, muestra los valores elegidos de los parámetros de la distribución. Por defecto el programa nos representa la B (10, 0.1)
- *Cumulative Distribution*, opción que nos proporciona las probabilidades de que la v.a. tome valores menores, iguales (únicamente en el caso de discretas) o mayores que un valor dado
- *Inverse CDF*, que nos permite obtener el valor de la variable X que deja a la izquierda una probabilidad concreta p, tal que P(X<=a)=p. Esta opción se utilizará en el siguiente apartado

- *Random Numbers* (= números aleatorios), es la opción que nos permite obtener una muestra de realizaciones al azar que siguen la distribución seleccionada

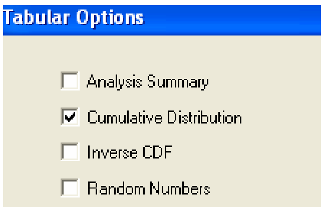
3. Introducir los parámetros de la binomial. Para ello situar el cursor en cualquier punto de la pantalla y, pulsar el botón derecho del ratón y seleccionamos *Análisis Options*



4. Se obtiene un cuadro de diálogo. Introducir los valores de los parámetros de la distribución (en este caso 0.4 y 12) en el cuadro de diálogo y hacer clic en OK

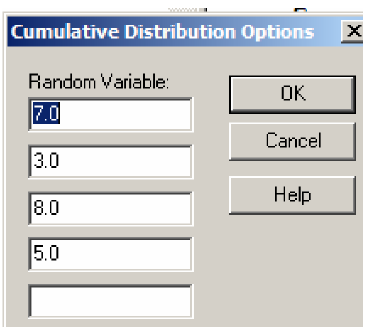


5. Activar la casilla de *Cumulative distribution* y hacer clic en OK



6. Situar el cursor en cualquier punto del panel de texto titulado *Cumulative Distribution* y pulsar el botón derecho del ratón seleccionando *Pane Options*

7. Introducir los valores deseados en el cuadro de diálogo como en la pantalla adjunta. Es decir, introducimos los valores de la v.a. (=random variable) para los cuales queremos obtener probabilidades; es decir, 7, 3, 8 y 5



8. Leyendo el panel de texto titulado *Cumulative Distribution, sabiendo que Upper Tail* significa cola superior (área a la derecha –por encima- del valor seleccionado) y Lower Tail significa cola inferior (área a la izquierda –por debajo- del valor seleccionado) observamos que :

- P (X=7) = 0.10092
- P (X>3) = 0.774663
- P(X≤8) = P (X<8) + P (X=8)= 0.94269 + 0.0420427 = 0.9847327 (Atención, estamos en v.a. discretas. Se puede comprobar que es también P(<9))
- P(X<5) = 0.438178

```
Cumulative Distribution
-----------------------

Distribution: Binomial

               Lower Tail Area (<)
Variable       Dist. 1        Dist. 2        Dist. 3        Dist. 4
7              0.841788
3              0.0834434
8              0.94269
5              0.438178

               Probability Mass (=)
Variable       Dist. 1        Dist. 2        Dist. 3        Dist. 4
7              0.100902
3              0.141894
8              0.0420427
5              0.22703

               Upper Tail Area (>)
Variable       Dist. 1        Dist. 2        Dist. 3        Dist. 4
7              0.0573099
3              0.774663
8              0.0152672
5              0.334791
```

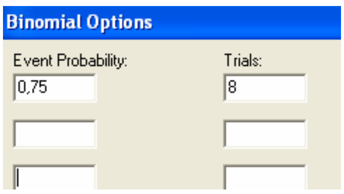
Cálculo de percentiles de la distribución

Nuestro interés ahora es calcular los valores de X que dejan por debajo cierto porcentaje de la población. Es decir, suministramos un valor de p y nos devuelve el valor a tal que P(X<=a)=p. Statgraphics denota por *Inverse CDF* a la función que asocia a cada percentil, el valor correspondiente de la variable aleatoria. La razón es que esta función hace precisamente la inversa de la función de distribución (CDF para el Statgraphics). En la CDF se proporciona el valor de a y nos devuelve el valor de p.

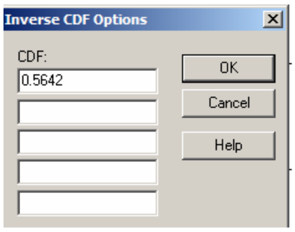
Supongamos X~B(8,0.75). Queremos hallar el valor de a, de forma que P(X<a)=0.5642. Procederemos como sigue:

1. Situar el cursor en cualquier punto de la pantalla, pulsar el botón derecho del ratón y seleccionar *Análisis Options.*

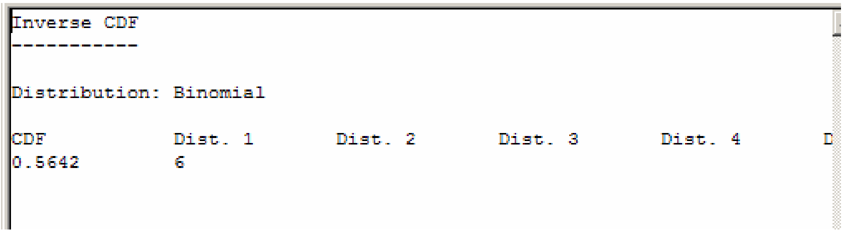
2. Introducir los parámetros de la distribución, y hacer clic en OK



3. Hacer clic sobre el botón  de *Tabular Options,* activar *Inverse CDF*, OK
4. Situar el cursor en cualquier punto del panel de texto titulado Inverse CDF, pulsar el botón derecho del ratón y seleccionar *Pane Options.*
5. Introducir el valor de la probabilidad como en la pantalla adjunta, y hacer clic sobre OK:. Por defecto el programa te ofrece una serie de probabilidades (0.01, 0.1, 0.5, 0.9 y 0.99)



6. El valor buscado es 6 , es decir, para X~B(8, 0.75), P(X<=a)=0.5642 ⇔a=6, y se obtiene leyendo el renglón del panel de texto que se adjunta



EJEMPLO APLICADO:

Un viajero de metro llega todas las mañanas a la misma hora a un andén. El 18% de las veces que llega al andén, el tren se encuentra en él, mientras que el resto de las veces ha de esperar.

a) En siete días consecutivos, ¿Cuál es la probabilidad de que se encuentre el tren estacionado uno sólo de esos días?
b) En quince días consecutivos, ¿Cuál es la probabilidad de que se encuentre el tren estacionado tres días como máximo?
c) En dieciocho días consecutivos, ¿Cuál es la probabilidad de que se encuentre el tren estacionado más de cinco días?

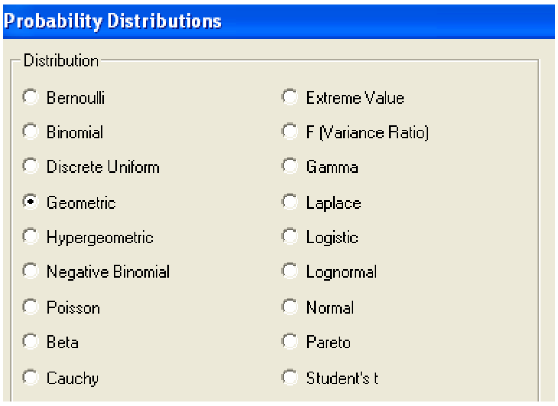
Definiremos la X como la v.a. X= nº de días en los que el viajero encuentra el tren en el andén de entre n días. X~B(n,p), Binomial con una probabilidad de éxito p= 0.18. Calculando con Statgraphic Plus, obtenemos:

a) X~B (7,0.18); P (X=1) = 0.383048
b) X~B (15,0.18); P (X≤3) = P (X<3) + P (X=3) = 0.721805
c) X~B (18,0.18); P (X>5) = 0.0889352

2.1.2. Distribución Geométrica, X ~ G (p)

Recordamos que, en un experimento consistente en la realización de pruebas de Bernouilli independientes, con una probabilidad de éxito p (*event probability*) , constante para todas ellas, la variable aleatoria X ~ G (p) representa el " nº de repeticiones del experimento para tener un éxito". Esta definición de distribución Geométrica no concuerda con la que utiliza el Statgraphics, así como otros libros de texto, donde se define variable geométrica a la variable Y=número de fallos ANTES de obtener un éxito; es decir Y=X-1. Puesto que en este guión nos apoyamos en el Statgraphics, usaremos esta segunda definición como variable geométrica. Esta ambigüedad en la definición de la distribución geométrica está bastante extendida, por lo que cuando utilicemos algún software para calcular distribuciones debemos averiguar qué definición de variable geométrica se está utilizando.

Con la distribución geométrica se trabaja en Statgraphic de manera similar a los procesos descritos para la Binomial. Para trabajar con la distribución Geométrica debemos seleccionar la distribución *Geometric*



EJEMPLO APLICADO:

El control de calidad de una fábrica de puertos USB inspecciona las unidades una vez montadas. Se sabe que la proporción de unidades defectuosas es de 0.02. ¿Cuál es la probabilidad de que la primera unidad defectuosa sea la que se inspeccione en décimo lugar?

Si X= número de puertos USB inspeccionados hasta encontrar uno defectuoso =>

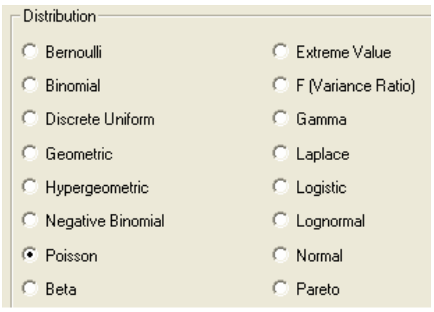
$X \sim$Geométrica ($p=0.02$)

$P(X=9)=0.016675$

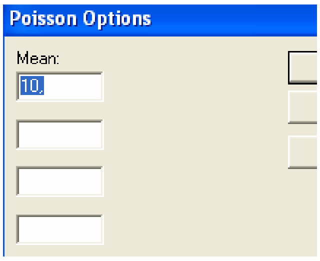
2.1. 3. Distribución de Poisson,

Recordamos que la v.a. $X\sim$Poisson (λ) representa el nº de sucesos aleatorios independientes que se observan en una unidad de medida. El parámetro de la distribución, λ, representa el nº medio de sucesos que ocurren por unidad de tiempo, longitud, superficie, volumen o cualquiera que sea la unidad de medida.

Para usar la Poisson seleccionamos *Poisson* en la ventana de selección de distribuciones



A continuación debemos introducir la media de la distribución, que es el parámetro que determina la función de probabilidad. Seleccionando *Analysis Options* (botón derecho del ratón) aparece la ventana de diálogo para introducir la media



EJEMPLO APLICADO

El número de usuarios que acceden a un ordenador, que hace de servidor de una red, es, por término medio, de 3000 cada hora. La red puede atender de manera óptima 100 accesos por minuto. Suponiendo que los accesos se producen de forma independiente y con media constante, se desea calcular la probabilidad de que en un minuto determinado accedan a dicha red ...

a)exactamente 40 usuarios
b)entre 40 y 50 usuarios
c)más de 100 accesos, y por tanto, se produzcan retrasos en el tráfico de la red.

Sea X=número de accesos en un minuto. Entonces $X \sim Poi(\lambda=50)$. Las probabilidades pedidas son

a) $P(X=40)=0.021$
b) $P(40 \leq X \leq 50)= P(39<X<51)=P(X<51)-P(X<39)=0.5375-0.0473 = 0.4902$
c) $P(X>100) =1.56$ e-10 (es decir 1.56×10^{-10})

Estos valores los obtenemos seleccionando CDF e introduciendo los valores que nos interesan. La salida de Statgraphics que se ha usado para calcular estas probabilidades es:

```
Cumulative Distribution
-----------------------

Distribution: Poisson

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2
51            0,537517
39            0,0473707
100           1,0
40            0,0645704

              Probability Mass (=)
Variable      Dist. 1       Dist. 2
51            0,0552206
39            0,0171997
100           1,63032E-10
40            0,0214996

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2
51            0,407263
39            0,93543
100           1,56973E-10
40            0,91393
```

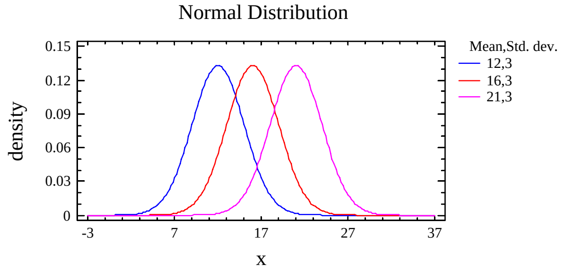
2.2. Distribuciones Continuas.

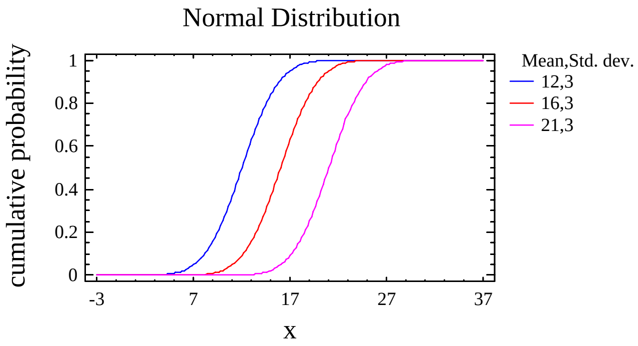
2.2.1. La distribución Normal.

La distribución Normal es simétrica. La media coincide con la moda y la mediana, y tiene forma de campana (se le denomina “campana de Gauss”).

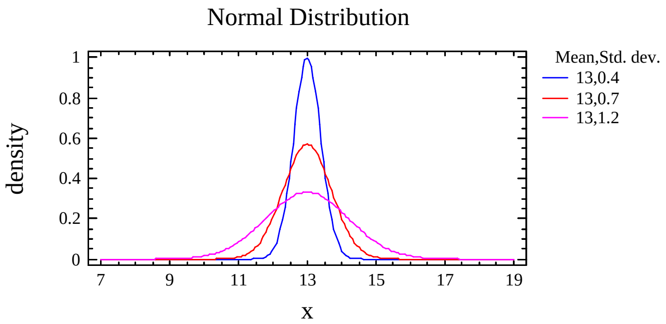
Comparación de funciones de densidad y de distribución

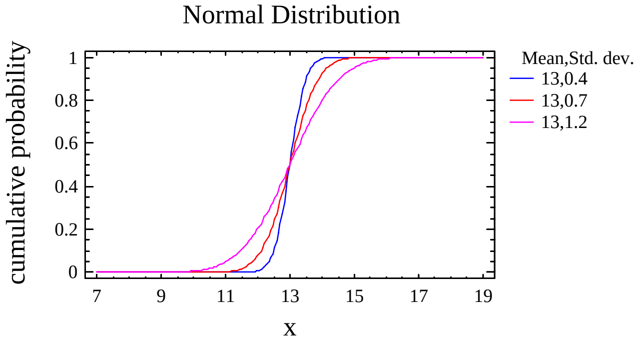
Vamos a dibujar las gráficas de las funciones de densidad de tres Normales con igual σ pero distinta μ: N(12,3), N(16,3) y N(21,3), donde se aprecia el desplazamiento de la campana según el cambio en la media.





Dibujaremos ahora las funciones de densidad de tres Normales con igual μ pero distinta σ: N(13, 0.4), N(13,0.7) y N(13,1.2), apreciándose cómo varía la dispersión de las distribuciones





Cálculo de probabilidades

De acuerdo con la definición de función de densidad, el cálculo de cualquier probabilidad se convierte en el cálculo de una integral. La probabilidad entre dos puntos es el área bajo la función de densidad entre esos dos puntos.

Ejemplo: Para la variable $X \sim N(8,2.6)$ calcular $P(X>11.3)$, $P(X<7.9)$, $P(-1<X<4)$ y $P(X \geqslant 18)$.

- $P(X>11.3) = 0.102179$
- $P(X<7.9) = 0.484657$
- $P(-1<X<4) = P(X<4)-P(X<-1) = 0.0619677-0.000268595=0.061699105$
- $P(X \geqslant 18)=0.000060015$

```
Cumulative Distribution
-----------------------

Distribution: Normal

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
11.3          0.897821
7.9           0.484657
-1.0          0.000268595
4             0.0619677
18            0.99994

              Probability Density
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
11.3          0.0685684
7.9           0.153326
-1.0          0.000383729
4             0.0469871
18            0.0000941238

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
11.3          0.102179
7.9           0.515343
-1.0          0.999731
4             0.938032
18            0.000060015
```

<u>Cálculo de percentiles</u>

Se procedería de forma similar a las distribuciones anteriores. Además, en el caso de la distribución Normal a veces interesa conocer cuál es la probabilidad de que el valor de la variable $X \sim N(\mu, \sigma)$ se encuentre a una, dos o tres desviaciones típicas de la media, es decir, cuál es la probabilidad de que X se encuentre en el intervalo $(\mu-k\sigma, \mu+k\sigma)$, generalmente para k=1,2 ó 3

2.2.2. Distribución Exponencial

La distribución exponencial es seguida por varios fenómenos:

- El tiempo de espera hasta la primera llegada (de llamadas, cartas, personas a la estación, etc) cuando estas llegadas se producen de forma que el nº de llegadas en el intervalo de tiempo [0,t] sigue una distribución de Poisson (λt) para cada t.

- El tiempo transcurrido desde un instante dado hasta que aparece una avería por causa fortuita

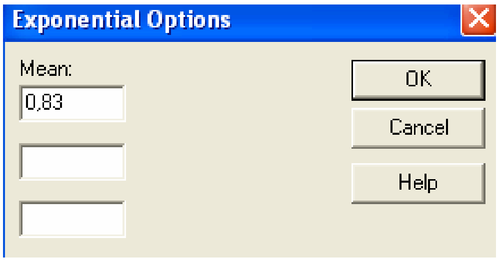
Es importante recordar que la distribución exponencial presenta ausencia de memoria. Para usar la distribución exponencial, seleccionamos *Exponential* en la ventana de modelos de distribución

EJEMPLO APLICADO

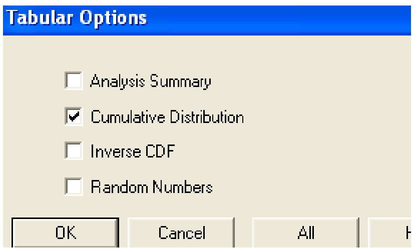
El número de usuarios que acceden a un ordenador, que hace de servidor de una red, es, por término medio, de 3000 cada hora. Suponiendo que los accesos se producen de forma independiente y con media constante, se desea calcular la probabilidad de que entre dos accesos consecutivos haya un intervalo de 5 segundos sin accesos

Solución:
Si los accesos por unidad de tiempo son independientes y de media constante, el número de accesos por unidad de tiempo es una variable de Poisson, y el intervalo entre accesos es Exponencial. Entonces, el tiempo en segundos entre 2 accesos consecutivos es T~Exp(λ=0.83 accesos/segundo). La probabilidad que se pide es
P(T>5)=0.0024

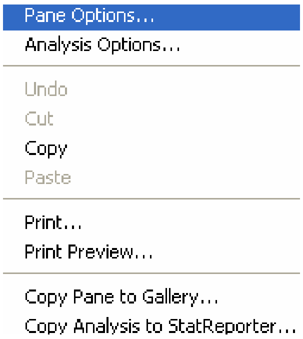
Para conseguir estos resultados, hemos introducido la media de la exponencial en la ventana que aparece tras pulsar Analysis Options (botón derecho del ratón).



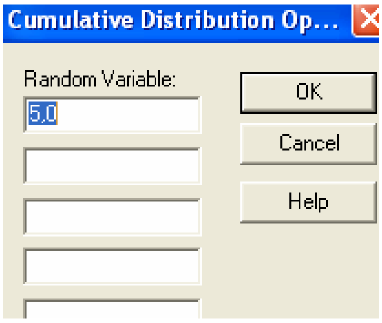
Seleccionando ahora la Opciones Tabulares (Tabular Options, icono), y allí elegimos Cumulative distribution



Nos colocamos en la ventana de resultados y con el botón derecho del ratón seleccionamos Pane Options



Aparece entoncs la ventana que nos pregunta por el valor de la distribución exponencial del que queremos calcular probabilidades. Introducimos el valor 5.



y obtenemos el siguiente resultado.

```
Cumulative Distribution
-----------------------

Distribution: Exponential

              Lower Tail Area (<)
Variable      Dist. 1        Dist. 2
5             0,99758

              Probability Density
Variable      Dist. 1        Dist. 2
5             0,00291535

              Upper Tail Area (>)
Variable      Dist. 1        Dist. 2
5             0,00241974
```