# PRÁCTICA 3. DISTRIBUCIONES DISCRETAS

## Objetivo

El objeto de la presente sesión de práctica informática es complementar y afianzar los conceptos relativos a las distribuciones discretas vistos en clase (apartados 1, 2.1 y 2.2 de la UD4).

Para ello, en primer lugar (Apartado 1), se proponen varios ejercicios para completar o reforzar cuestiones que no se hayan podido tratar en las clases de teoría y seminario.

En los apartados 2 y 3, que el alumno deberá hacer en casa, se repasan los aspectos básicos de la distribución Binomial y de Poisson. Asimismo se explican las posibilidades que ofrece el *Statgraphics* respecto a este tipo de variables, mediante la resolución de dos ejemplos sencillos.

> **NOTA:** Se recomienda que, durante el trabajo no presencial del alumno, los resultados obtenidos a partir del Statgraphics en esta práctica se calculen "a mano" y se cotejen con los mismos.

## 1. Ejercicios

### 1.1. Ejercicio 1

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.

¿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%?

SOLUCIÓN:

*X="nº de componentes defectuosas en la muestra de N extraídos de partidas que no satisfacen el requisito"*

*X~B(N, p≥0,1) Tomamos el caso más difícil de detectar p=0,1.*

$P(aceptar) \leq 0{,}05 \Rightarrow P(X=0) \leq 0{,}05 \Rightarrow \binom{N}{0} 0{,}1^0 0{,}9^N \leq 0{,}05$

$\Rightarrow\ 0{,}9^N \leq 0{,}05 \Rightarrow N \log 0{,}9 \leq \log 0{,}05 \Rightarrow$

$\Rightarrow\ N \geq (\log 0{,}05)/(\log 0{,}9) \Rightarrow N \geq 28{,}43 \Rightarrow N_{mínimo}=29$

### 1.2. Ejercicio 2

Un fabricante de circuitos integrados desea garantizar que la proporción de chips defectuosos en los lotes que vende es inferior al 5 por mil. Para ello selecciona al azar N chips de un lote, rechazando el lote en caso de que 3 o más de los chips sean defectuosos.

Determinar el valor mínimo de N si se desea que la probabilidad de rechazar un lote que no satisfaga el requisito exigido sea al menos del 99%.

SOLUCIÓN:

*X="nº de chips defectuosos en los N extraídos de un lote que no satisface el requisito"*

*X~ Poisson (λ≥N 0,005) Caso más difícil de detectar λ=N 0,005.*

*P(rechazar) ≥ 0,99⇒ P(X≥3) ≥ 0,99 ⇒*

⇒ *P(X≤2)≤0,01⇒ ⇒P(Poisson(λ=N0,005)≤2)≤0,01*

⇒ *Mirando en el ábaco λ≥8,5 ⇒ N≥8,5/0,005*

⇒ *N≥1700 ⇒* $\boldsymbol{N_{mínimo}=1700}$

## 2. Distribuciones de probabilidad

En primer lugar debemos elegir el tipo de v.a. con la que vamos a trabajar, para ello seleccionamos **Plot > Probability Distributions...** (**Figura 1**).

Así llegamos a una ventana en la que nos aparecen las diferentes distribuciones de probabilidad que se pueden manejar con el Statgraphics (**Figura 2**).

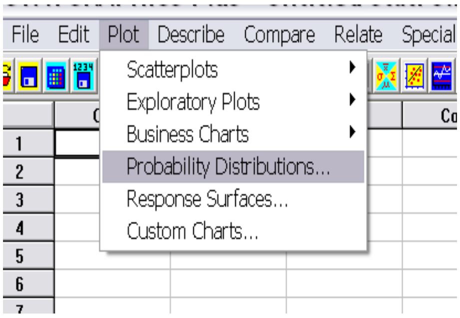



**Figura 1.** Opción del menú para trabajar con una distribución de probabilidad.

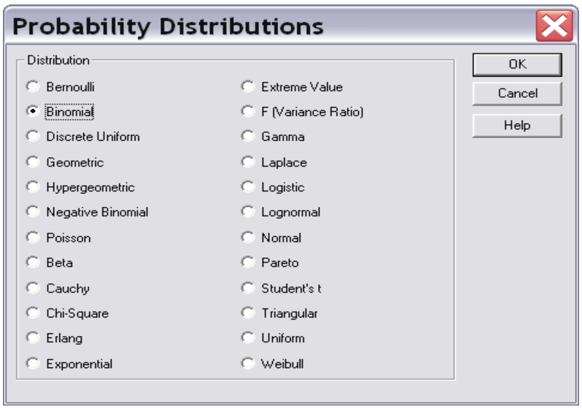



**Figura 2.** Cuadro de selección del modelo de distribución de probabilidad.

Si seleccionamos una de ellas, la distribución Binomial, por ejemplo, obtenemos por defecto los siguientes cuatro paneles (**Figura 3**):

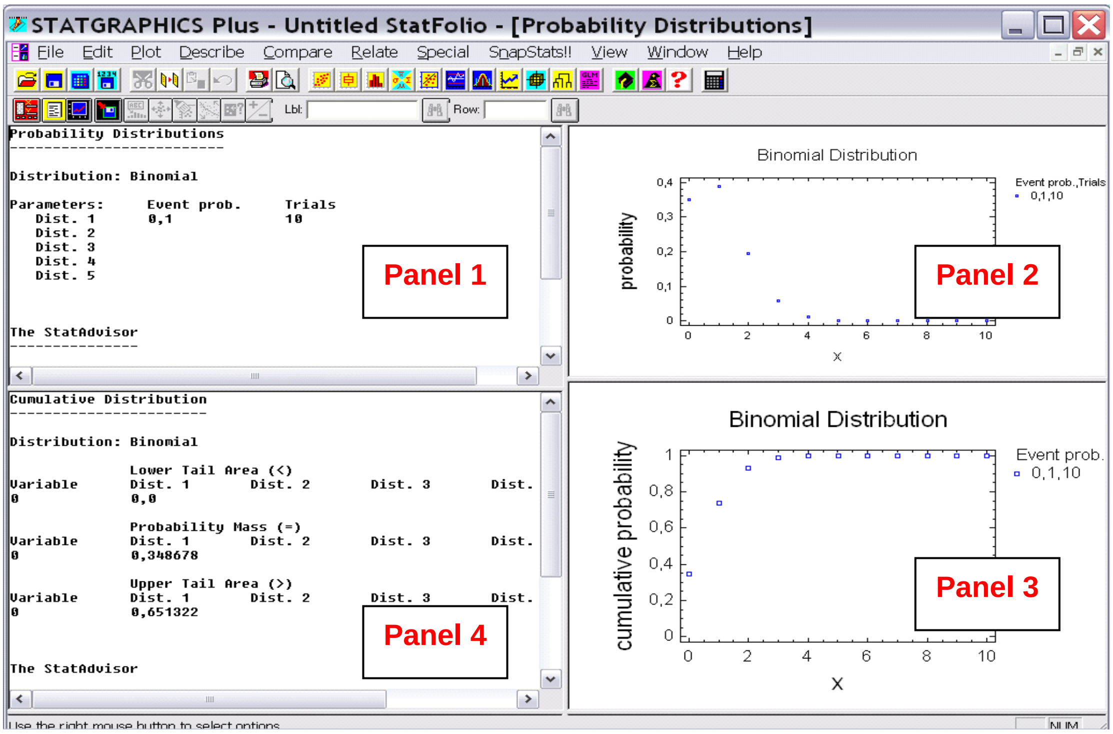



**Figura 3.** Ventanas (paneles) con los resultados de los distintos análisis sobre una v.a con una determinada distribución de probabilidad.

Mediante el botón derecho del ratón y picando "**Analysis Options**" sobre cualquiera de los paneles se especifican los **parámetros** que caracterizan a la distribución de probabilidad elegida (Poisson, Binomial,...).

Se puede especificar hasta 5 distribuciones del mismo tipo con diferentes parámetros simultáneamente. Esta información se recoge en el **panel 1.**

En el **panel 2** se muestra la f**unción de probabilidad P(x)** (***Mass function***) si la v.a. es discreta o la **función de densidad f(x)** (***Density function***) si la v.a. es continua.

Análogamente, en el **panel 3** se muestra **la función de probabilidades acumuladas** (***CDF Cumulative Distribution Function***) de dicha distribución.

En el **panel 4** se puede calcular la probabilidad de que una v.a. tome un valor mayor, igual o menor que uno introducido por nosotros. Inversamente, visualizando el panel *Inverse CDF* con el icono **Tabular Options**, se puede obtener el valor de una v.a. (*critical value*) para el cual existe una probabilidad predeterminada de que la v.a. tome como máximo ese valor.

## 3. Cálculo de probabilidades y representaciones gráficas

Una vez elegida la distribución de probabilidad que se adecua a la pauta de variabilidad de la v.a. en estudio, se pueden calcular las probabilidades asociadas a los valores de dicha v.a. de acuerdo al modelo elegido (Poisson, Binomial, Normal,…). También se pueden obtener las representaciones gráficas de la **Función de Densidad** o **Probabilidad**.

### 2.1. Ejemplo 1 (Ejercicio 2 del Boletín de Ejercicios de la UD4)

Una factoría fabrica disquetes de baja calidad a bajo precio. Sea la v.a. X={nº de disquetes defectuosos en una muestra de tamaño n}. Se sabe que dicha factoría produce un <u>20% de disquetes defectuosos de cada muestra de 11 disquetes</u> que se toma. ¿Cuál es la probabilidad de que 3 disquetes sean defectuosos? ¿Y la probabilidad de que encontremos más de 3?

En primer lugar hay que determinar cuál es la <u>distribución de probabilidad</u> de la v.a., así como sus <u>parámetros</u>, en este caso:

**X ~ Binomial (n=11, p=0,2)**

Variable aleatoria

Distribución

parámetros

**RECUERDA**: **n** y **p** son los **parámetros de la distribución**. **n** es el tamaño de la muestra y **p** la probabilidad de que en esa muestra un diskette sea defectuoso o lo que es lo mismo, proporción de diskettes defectuosos ("tanto por uno") en la muestra.

Para calcular dichas probabilidades seleccionamos la distribución pertinente, Binomial en este ejemplo (**Figura 1** y **Figura 2**) y seguidamente proporcionamos los parámetros de esta distribución Binomial. Para ello, seleccionamos "**Analysis Options**" con el botón derecho del ratón sobre cualquiera de los paneles, entonces aparece la correspondiente ventana de diálogo.

En esta ventana introducimos los parámetros **n** (*Trials*) y **p** (*Event Probability)* de la distribución, en este ejemplo 11 y 0,2 respectivamente (**Figura 4**).

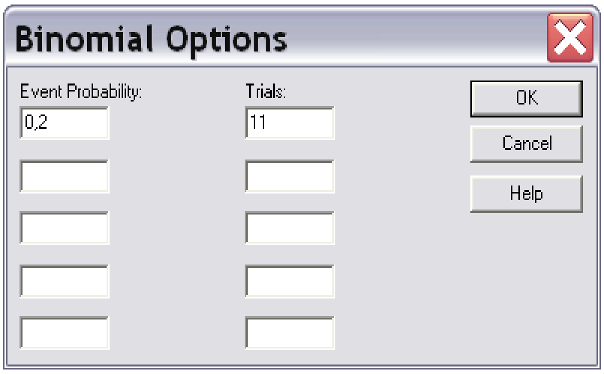



**Figura 4.** Cuadro de diálogo para la introducción de los valores de los parámetros de la distribución Binomial.

Automáticamente aparecen en los paneles de la derecha la función de probabilidad y la de probabilidades acumuladas.

Antes de responder a las preguntas planteadas, veamos que forma tienen las funciones de probabilidad y distribución para la v.a. **X**.

Observando el gráfico correspondiente a la función de probabilidad del **panel 2** (gráfica de arriba) se tiene que la probabilidad de que 3 disquetes sean defectuosos es aproximadamente del 22%, esto es, P(X=3)=0,22 (**Figura 5**) .

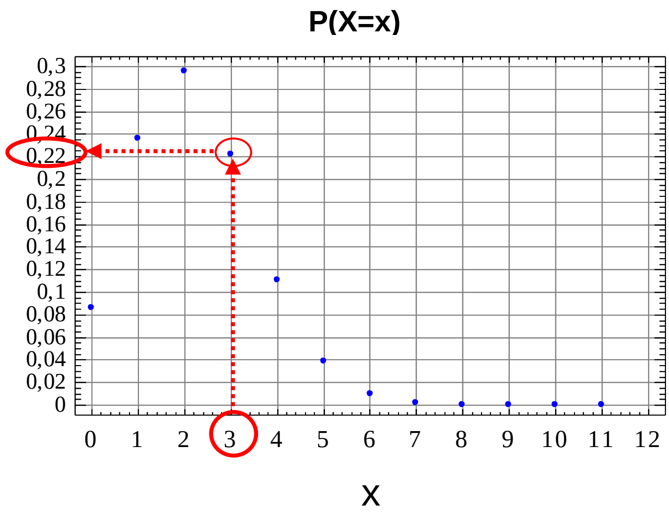



**Figura 5.** Representación gráfica de la **función de probabilidad P(x)** de la v. a X ~ B(n=11, p=0,2).

Igualmente, se obtiene que la probabilidad de que no se encuentre ningún disquete defectuoso (P(X=0)) es, con carácter aproximado, 0,086.

Con el objeto de poder calcular estas probabilidades con precisión, debemos utilizar el **panel 4**.

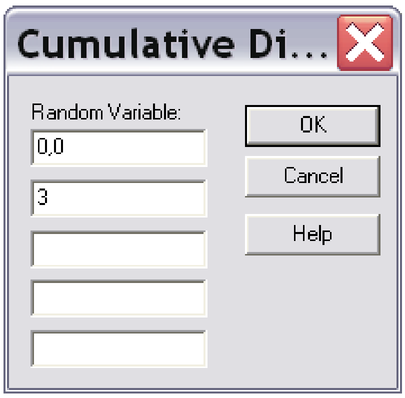



**Figura 6.** Cuadro de diálogo para la introducción de los valores de la v.a sobre los que deseamos calcular probabilidades.

Para indicar el valor de la v.a. para el que queremos conocer la probabilidad (*Random Variable*), seleccionamos "**Pane Options"** con el botón derecho del ratón sobre el **panel 4**, entonces aparece la ventana de diálogo de la **Figura 6** en la que introducimos los valores 0 y 3 de nuestro ejemplo (el valor 0 siempre aparece por defecto en cualquier distribución).

Los valores obtenidos en el cuadro `Cumulative Distribution` de la siguiente página indican que:

$P(X < 0) = 0$

$P(X < 3) = P(X \leq 2) = 0,617402$

$P(X = 0) = 0,0858993$

$P(X = 3) = 0,221459$

$P(X > 0) = P(X \geq 1) = 0,914101$

P(X > 3) = P(X ≥ 4) = 0,161139

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

Distribution: Binomial

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,0
3             0,617402

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,0858993
3             0,221459

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,914101
3             0,161139
```

Es obvio que a partir de los valores anteriores se puede calcular, por ejemplo:

P(X ≤ 3) = P(X < 3) + P(X = 3) = 0,617402 + 0,221459 = 0,838861

**RECORDAR.** Por definición P(X ≤ 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)

Asimismo, la última probabilidad calculada (P(X ≤ 3)) también se puede obtener de manera aproximada a partir del gráfico de la función de distribución (**Figura 7**):

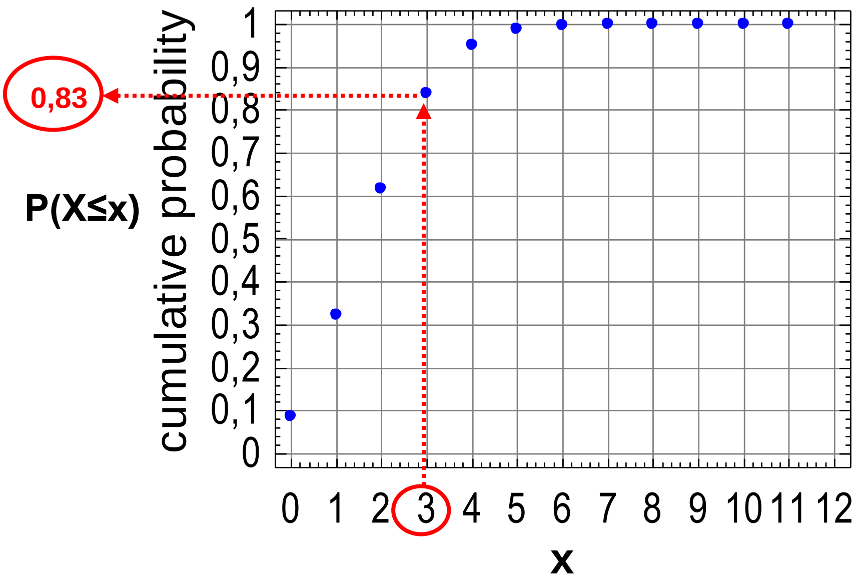



**Figura 7.** Representación gráfica de **P(X≤x)** de la v. a X ≈ B(n=11, p=0,2).

**Pregunta 1.** Calcular la probabilidad de encontrar 4 disquetes defectuosos.

**Pregunta 2.** Calcular la probabilidad de encontrar al menos 2 disquetes defectuosos.

**Pregunta 3.** Calcular la probabilidad de encontrar 5 disquetes defectuosos en una muestra de 20 disquetes. ¿Cuál sería ahora la distribución de la v.a.?

El Statgraphics también permite calcular los valores que toma la v.a. para determinadas probabilidades, para ello, abrimos el panel "**Inverse CDF**" con el icono **Tabular Options** (como ya se ha indicado con anterioridad) y aparece el cuadro que se muestra a continuación:

```
Inverse CDF
-----------

Distribution: Binomial

CDF            Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0,01           0
0,1            1
0,5            2
0,9            4
0,99           6
```

El cuadro indica:

Si P(X≤**x**)=0,01 → **x** = 0

Si P(X≤**x**)=0,5 → **x** = 2

Si P(X≤**x**)=0,99 → **x** = 6

Si P(X≤**x**)=0,1 → **x** = 1

Si P(X≤**x**)=0,9 → **x** = 4

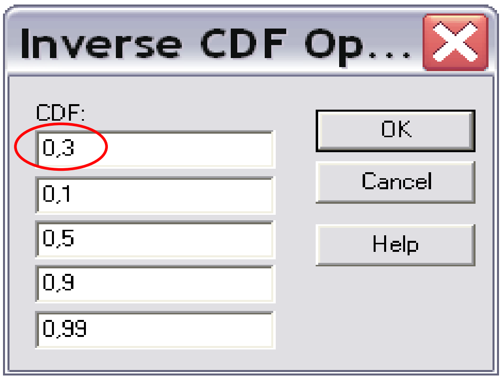



**Figura 8.** Cuadro de diálogo para la introducción de los valores de probabilidades sobre los que deseamos calcular el valor asociado de la v.a.

Para indicar una probabilidad distinta de la que aparece por defecto en el cuadro, seleccionamos "**Pane Options**" con el botón derecho del ratón para que se muestre la correspondiente ventana de diálogo.

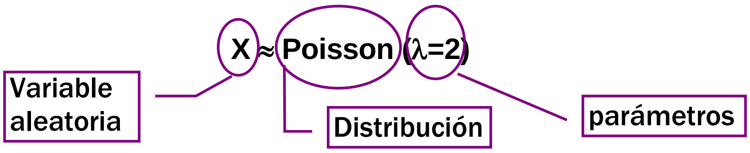
Por ejemplo, para obtener el valor de **x** asociado a una probabilidad de 0,3, introducimos este valor en el campo **CDF** (**Figura 8**).

**Pregunta 4.** Siguiendo con el primer ejemplo, si la probabilidad de encontrar una cantidad de disquetes defectuosos menor o igual que **x** es 0,95, ¿cuál es el valor de **x**?

**IMPORTANTE.** Todo lo visto hasta ahora se utiliza de igual forma para cualquiera que sea el tipo de la distribución elegida, sólo cambia, como es obvio, el tipo y valor de los **parámetros** que la determinan en cada caso ("**Analysis Options**").

## 2.2. Ejemplo 2

Los códigos CRC utilizados en el envío de paquetes a través de la red son capaces de corregir como máximo 5 errores por paquete. Se sabe que el número medio de errores por paquete en un determinado envío es de 2. Si queremos saber cuál es la probabilidad de que tan solo haya un error en un paquete tomado al azar, deberemos antes que nada determinar la v.a., su distribución y los parámetros de la misma:



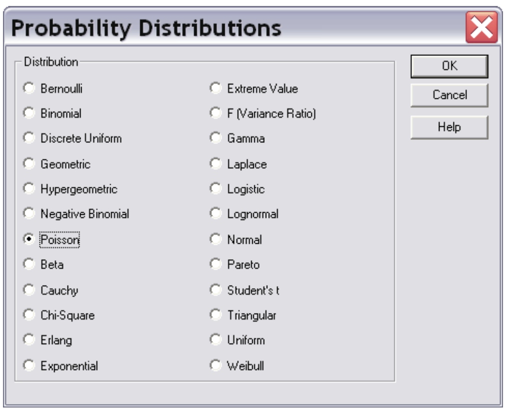

Siendo X = {nº errores en un paquete}.

**NOTA**: $\lambda$ es el **parámetro de la distribución**. $\lambda$ es el número de medio de errores en un paquete.

Para calcular la probabilidad que nos piden procedemos como se ha explicado en los apartados 1 y 2.

Primero elegimos la distribución correspondiente como se muestra en las **Figuras 1** y **2**, pero seleccionando **Poisson** en la ventana **Probability Distributions** (**Figura 9**), y a continuación introducimos los parámetros ("**Analysis Options**" con el botón derecho del ratón sobre cualquiera de los paneles).

Para este ejemplo, en la ventana introducimos el parámetro $\lambda$ (*Mean*) cuyo valor es 2 (**Figura 10**).



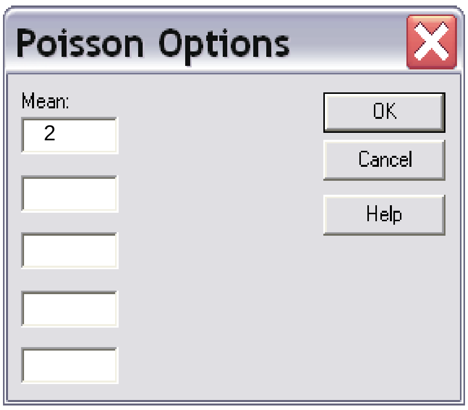



**Figura 9.** Cuadro de selección del modelo de distribución de probabilidad.

**Figura 10.** Cuadro de diálogo para la introducción de los valores de los parámetros.

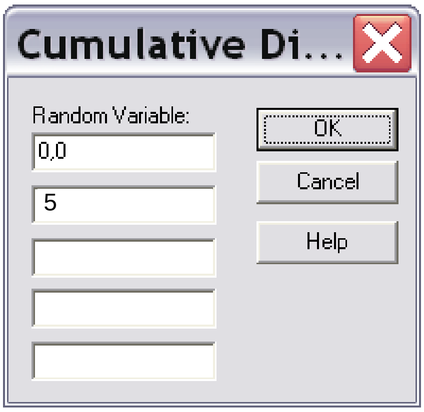



**Figura 11.** Cuadro de diálogo para la introducción de los valores de la v.a sobre los que deseamos calcular probabilidades.

Si además nos piden calcular la probabilidad de que un paquete haya de ser rechazado por no poder ser corregido, o sea, P(**X** > 5), entonces, introducimos el valor 5 (valor de la v.a. para el que queremos conocer la probabilidad o *Random Variable*) seleccionando "**Pane Options"** (botón derecho del ratón sobre el **panel 4**), tal y como se describió en el apartado 2 (**Figura 11**).

Las probabilidades obtenidas son:

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

Distribution: Poisson

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,0
5             0,947347

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,135335
5             0,0360894

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,864665
5             0,0165636
```

Estos resultados revelan que:

La probabilidad de que un paquete elegido al azar no contenga ningún error es

P(X = **0**) = **0,135335**

Y la probabilidad de que un paquete haya de ser rechazado por no poder ser corregido es P(X > **5**) = **0,0165636**

**Pregunta 5.** Si tomamos al azar 10 paquetes consecutivos, ¿cuál es la probabilidad de que el número total de errores sea mayor que 20?

## Respuestas a las preguntas propuestas del los apartados 2 y 3

```
Distribution: Binomial

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
2             0,322123
4             0,838861

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
2             0,295279
4             0,11073

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
2             0,382598
4             0,0504092
```

**Cuadro 1.** Probabilidades asociadas a una v.a. que sigue una distribución Binomial (n=11, p=0,2).

### Pregunta 1

Siendo **X** = {nº de disquetes defectuosos en una muestra de tamaño 11}

**X** ~ Binomial (n=**11**, p=0,2)

Como se observa en el **Cuadro 1**, P(**X**=**4**) = **0,11073**

La probabilidad de encontrar 4 disquetes defectuosos es **0,11073**.

---

### Pregunta 2

Siendo **X** = {nº de disquetes defectuosos en una muestra de tamaño 11}

**X** ~ Binomial (n=**11**, p=0,2) (misma v.a. que antes: misma distribución de probabilidad y mismos parámetros)

**$P(X \geq 2) = P(X > 2) + P(X = 2)$** y como se observa en el **Cuadro 1**:

**$P(X > 2) = 0,382598$**

**$P(X = 2) = 0,295279$**

**$P(X \geq 2) = P(X < 2) + P(X = 2) = 0,382598 + 0,295279 =$**

La probabilidad de encontrar al menos 2 disquetes defectuosos es de **0,677877**.

## Pregunta 3

La v.a. es ahora **X** = {nº de disquetes defectuosos en una muestra de tamaño **20**}

Por tanto, **X ~ Binomial (n=20, p=0,2)**

Cambiando **n** por **20** mediante "**Analysis Options**" en el **Panel 4** se obtiene:

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

Distribution: Binomial

               Lower Tail Area (<)
Variable       Dist. 1        Dist. 2        Dist. 3        Dist. 4        Dist. 5
5              0,629648

               Probability Mass (=)
Variable       Dist. 1        Dist. 2        Dist. 3        Dist. 4        Dist. 5
5              0,17456

               Upper Tail Area (>)
Variable       Dist. 1        Dist. 2        Dist. 3        Dist. 4        Dist. 5
5              0,195792
```

**Cuadro 2.** Probabilidades asociadas a una v.a. que sigue una distribución Binomial (n=20, p=0,2).

Observando el **Cuadro 2** se tiene que: **P(X=5) = 0,17456**

La probabilidad de encontrar 5 disquetes defectuosos en una muestra de 20 disquetes es **0,17456**.

---

## Pregunta 4

Volviendo a la v.a. **X** = {nº de disquetes defectuosos en una muestra de tamaño 11}

**X ~ Binomial (n=11, p=0,2)**

Para contestar a esta pregunta tendremos que abrir el panel "**Inverse CDF**" con el icono **Tabular Options** e introducir el valor de la probabilidad **0,95** (si no está por defecto) en el cuadro de diálogo que se muestra con **"Pane Options"**. Así aparecen los siguientes resultados:

```
Inverse CDF
-----------

Distribution: Binomial

CDF            Dist. 1        Dist. 2        Dist. 3        Dist. 4        Dist. 5
0,01           0
0,1            1
0,5            2
0,9            4
0,95           5
```

**Cuadro 3.** Valores de la v.a. asociados a las probabilidades de que tomen vales menores o iguales que estos en una distribución Binomial (n=11, p=0,2).

Si **P(X≤x)=0,95**, entonces **x** = **5** disquetes defectuosos.

**Pregunta 5**

Ahora tenemos una nueva v.a. que se construye a partir de otras v.a. de Poisson (**Ps** para abreviar):

v.a. **Z** = {nº errores en **10** paquetes consecutivos}

$\mathbf{Z = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10}}$, donde:

$\mathbf{X_i}$ = {nº de errores en el paquete i} ≈ **Ps** ($\boldsymbol{\lambda_i=2}$)

Como la suma de v.a. de Poisson también sigue una distribución de Poisson con el λ igual a la suma de los λ de las v.a. que la compone, entonces:

**Z ≈ Ps ($\boldsymbol{\lambda_Z}$) donde $\boldsymbol{\lambda_Z = 10x\lambda_i = 10x2 \rightarrow}$ Z ≈ Ps ($\boldsymbol{\lambda_Z=20}$)**

Los resultados que da el Statgraphics para esta v.a. se muestran en el **Cuadro 4** y a la vista del mismo, la probabilidad que nos piden es:

**P(z > 20) = 0,440907**

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

Distribution: Poisson

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             0,0
20            0,470257

              Probability Mass (=)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             2,06115E-9
20            0,0888353

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2       Dist. 3       Dist. 4       Dist. 5
0             1
20            0,440907
```

**Cuadro 4.** Probabilidades asociadas a una v.a. que sigue una distribución Ps ($\lambda_z$=20).

Por lo tanto, la probabilidad de que al tomar al azar 10 paquetes consecutivos el número total de errores sea mayor que 20 es **0,440907**.

## Fuentes

http://www.statgraphics.net/