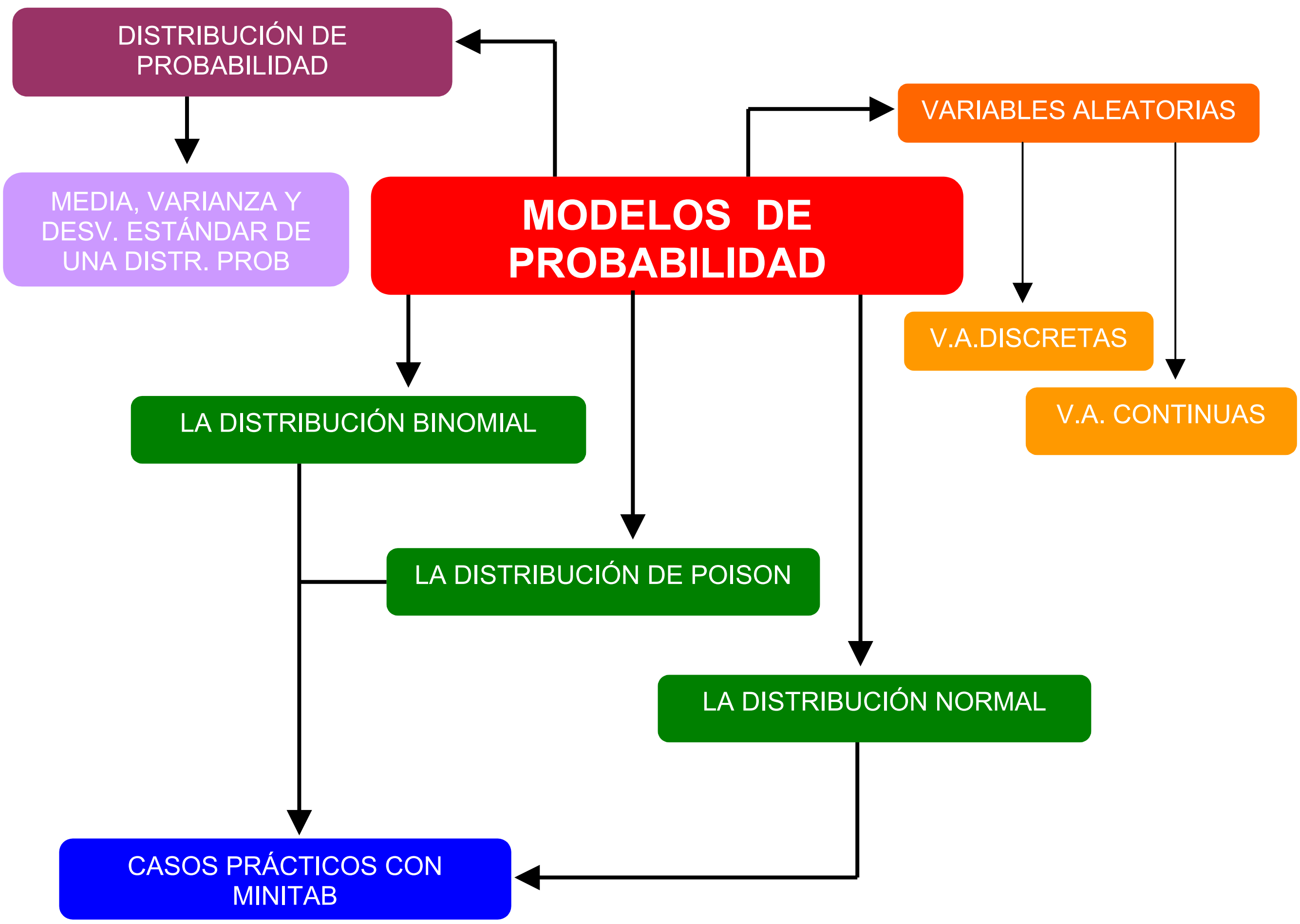
MODELOS DE PROBABILIDAD

Autores: Angel Juan (ajuanp@uoc.edu), Máximo Sedano (msedanoh@uoc.edu) , Alicia Vila (avilag@uoc.edu), José Francisco Martínez (jmartinezbos@uoc.edu), Anna López (alopezrat@uoc.edu)

MAPA CONCEPTUAL



Financiado por la Secretaría de Estado de Educación y Universidades (MECD)

INTRODUCCIÓN

Este *math-block* pretende introducir al concepto de distribución de probabilidad como el rango de sucesos susceptibles de ocurrir al realizar un determinado experimento (cuán probable es que ocurra un determinado suceso perteneciente a un experimento concreto).

Así, veremos cómo aplicar esta idea a los tipos de distribución más utilizadas como son la Distribución Binomial, la Distribución de Poisson y la Distribución Normal.

También veremos cómo utillizar estas distribuciones de probabilidad en casos prácticos resueltos con Minitab.

OBJETIVOS

- Definir los términos distribución de probabilidad y variable aleatoria
- Distinguir entre distribuciones de probabilidad discretas y continuas
- Calcular la media, varianza y desviación estándar de una distribución de probabilidad
- Describir las características de la distribución Binomial y entender su aplicación en casos prácticos
- Describir las características de la distribución de Poisson y entender su aplicación en casos prácticos
- Describir las características de la distribución normal y entender su aplicación en casos prácticos
- Utilizar la distribución normal para aproximar la distribución de probabilidad Binomial

CONOCIMIENTOS PREVIOS

Sería conveniente tener presente el *math-block* "Estadística Descriptiva con Minitab" para tener asimilados los conceptos básicos referentes a los parámetros estadísticos fundamentales, así como el documento asociado al uso del Minitab.

CONCEPTOS FUNDAMENTALES

- **Definición de variable aleatoria (v.a.):** Corresponde al valor resultante de un determinado experimento.

  Por ejemplo, si contamos el número de empleados ausentes en un determinado turno de trabajo, el resultado podría ser 0, 1, 2, ..., este número de ausencias es la variable aleatoria.

  Distinguiremos entre variables aleatorias discretas y continuas.

  Diremos que una variable aleatoria es discreta cuando sólo puede tomar un número contable de valores. Estos valores no necesariamente han de ser enteros, pero sí han de tener valores claramente definidos.
  Serían v.a. discretas, p.e., X1 = "nº de hermanos de cada uno de nuestros amigos", o X2 = "nota, con una cifra decimal, obtenida en un examen por cada alumno de un aula".

  Por el contrario, una v.a. **continua** es aquella que puede tomar cualquier valor dentro de un intervalo real.
  Serían v.a. continuas, p.e., X_3 = "altura, en cm., de los jugadores de un equipo de baloncesto" (1.9, 1.92, 1.923,...), o X_4 = "distancia entre dos ciudades".

- **Definición de distribución de probabilidad:** Es aquella que permite calcular todos los resultados probables de ocurrir de un experimento determinado, así como la probabilidad de ocurrencias de estos resultados. [2]

  Las características más importantes a tener en cuenta en una distribución de probabilidad son:

  - La probabilidad de un resultado específico está entre cero y uno.
  - La suma de las probabilidades de todos los resultados mutuamente excluyentes es 1.

- **Definición de función de distribución de probabilidad:** La función de probabilidad de una variable aleatoria es la probabilidad acumulada hasta un valor determinado de la variable. Dada una variable aleatoria X, diremos que *F(a)* es la **función de distribución** tal que:

  $$F(a) = P(X \leq a)$$

  La función de distribución de probabilidad cumple $0 \leq F(x) \leq 1$.

  En el caso de las variables discretas la función de probabilidad se asocia con la función de probabilidad, función que da la probabilidad de cada posible valor que toma la variable.

  En el caso de las continuas como estas pueden tomar infinitos valores en un intervalo su función de probabilidad viene definida como la probabilidad a intervalos de valores. De hecho, la probabilidad de que la variable tome un determinado valor es nula. Las variables aleatorias continuas se caracterizan por una función denominada **función de densidad**.

- **Definición de función de probabilidad para una variable aleatoria discreta:** Dada una variable aleatoria discreta X, diremos que $f(x_i)$ es la **función de probabilidad** que asocia a cada valor x_i de la variable su probabilidad, i.e., $f(x_i) = P(X=x_i)$.

De este modo: *F(a)* = *P(X≤a)* es igual a la suma de todos los *P(X=xi)* tales que *xi* son menores que *a*.

- **Definición de función de densidad para una variable aleatoria continua:** Dada una variable aleatoria continua X la **función de densidad** f(x) asociada a una variable aleatoria continua X caracteriza la función de distribución de probabilidad de X donde:

$$F(a) = P(X \le a) = \int_{-\infty}^{a} f(x)\, dx$$

- **La media, la varianza y la desviación estándar.**
Como sabemos, la media nos da información acerca de la tendencia central de los datos y la varianza describe la dispersión de éstos.

A la media de la distribución la denotaremos por μ , y a la desviación estándar por σ.

La **media** es el valor promedio ponderado en el que los valores posibles de la variable aleatoria se ponderan según las probabilidades correspondientes de ocurrencia, también se denomina *valor esperado E(X).*
Para una variable aleatoria discreta:

$$\mu = E(X) = \sum [xP(x)]$$

donde P(x) es la probabilidad de valores posibles de la variable aleatoria x. Es decir, se multiplica cada valor de x por la probabilidad de que ocurra, y luego se suman estos productos.
Para una variable aleatoria continua:

$$\mu = E[X] = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

La **varianza** describirá la dispersión de la distribución.
Para una variable aleatoria discreta:

$$\sigma^2 = \sum \left[(x-\mu)^2 P(x)\right]$$

Para una variable aleatoria continua:

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2\; f(x)\, dx$$

Óbviamente, la **desviación estándar** σ la calcularemos al extraer la raíz cuadrada de la varianza.

- **La distribución Binomial.**
Consideremos una variable aleatoria X que da el número de éxitos que aparecen al repetir n veces de forma independiente un experimento en idénticas condiciones. En esta situación diremos que X sigue una distribución Binomial.

Ejemplos:

X= número de huevos defectuosos en un paquete de 12.
Y= número de 2 al tirar 10 veces un dado.

Las características principales de este modelo de distribución son:

1. Repetir *n* pruebas independientes unas de otras.

2. Para cada una de las pruebas sólo pueden darse dos resultados: éxito o fracaso

3. La probabilidad de éxito en cada prueba es de *p.*

En tales condiciones, diremos que la v.a. *X* = *"nº de éxitos en las n pruebas"* sigue una **distribución Binomial** de parámetros *n* y *p*, y lo escribiremos como $X \sim B(n,p)$.

Observamos que la v.a. *X* sólo puede tomar los valores *0, 1, 2, 3, …* , n siendo por tanto una v.a. discreta.

Así pues, las **funciones de probabilidad** y de **distribución** de una distribución binomial son las siguientes:

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{para } x=0,1,2,3\ldots n$$

donde $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$ y

$$F(x) = P(X \le x) = \sum_{i=0}^{n} P(X = i)$$

De la misma manera, la **media** y la **desviación estándar** de una distribución binomial son:

$$\mu = n * p \quad , \qquad \sigma = \sqrt{n * p * (1-p)}$$

La **distribución de Bernoulli** es un caso particular de la binomial cuando *n=1* .

Veamos unos ejemplos que muestran cómo aplicar la distribución Binomial:

Ejemplos:

1. Una empresa industrial que fabrica componentes mecánicos para aviones dispone de dos distribuidores por Europa, uno situado en Francia y otro en Alemania. Ambos tienen el 20% de posibilidades de cerrar un pedido con un consorcio industrial de farbicación de aviones.

   Si el distribuidor francés contacta con 5 consorcios:

a) ¿Cuál es la probabilidad de que el distribuidor francés consiga a lo sumo 2 acuerdos de distribución?

   Sea X="Número de acuerdos de distribución del distribuidor francés a 5 consorcios"

   p = probabilidad de éxito = P(cerrar un acuerdo) = 0,2
   n = número de clientes = 5

   X sigue una distribución Binomial, X ~ B(5 , 0,2)

   Nuestro objetivo es calcular P(X < = 2).

   P(X<=2) = P(X=2)+P(X=1)+P(X=0)=0,94208

Por su parte,

$$P(X=0)=\frac{5!}{0!(5-0)!}0.2^{0}\cdot(1-0.2)^{5}=0.32768$$

$$P(X=1)=\frac{5!}{1!(5-1)!}0.2^{1}\cdot(1-0.2)^{4}=0{,}4096$$

$$P(X=2)=\frac{5!}{2!(5-2)!}0.2^{2}\cdot(1-0.2)^{3}=0.2048$$

Por lo tanto la probabilidad de que el distribuidor francés cierre a lo sumo dos acuerdos es igual a 0,94208

b) ¿Cuál sería el número medio esperado de acuerdos que conseguiría cerrar el distribuidor francés?

Para calcular cual será el número medio esperado de acuerdos de distribución más probable que cierre el distribuidor calculamos la media de una distribución binomial que nos da el número medio de éxitos, en este caso sería, n*p= 5*0,2=1. Por lo tanto el número medio esperado de acuerdos logrados por el distribuidor francés será de 1.

2. El presidente de una compañía planea contactar con otras 18 compañías en busca de nuevos socios para su negocio. Sus analistas han estimado que la probabilidad de que una firma contactada al azar acepte incorporarse como socio es de 0,6. ¿Cuál es la probabilidad de que acabe reclutando 5 o más socios de entre las 18 compañías contactadas? ¿Cuál es el número medio esperado de socios que se incorporarán al proyecto?

Sabemos que X ~ B(18, 0.6). Nos piden hallar P(X>=5).

P(X>=5) = 1-P(X<5)=1- P(X=4)-P(X=3)-P(X=2)-P(X=1)-P(X=0)= 1-0,00127=0,9987

Ya que, $P(X=0)=\frac{18!}{0!(18-0)!}0.6^{0}\cdot(1-0.6)^{18}=0.0000000687$

$$P(X=1)=\frac{18!}{1!(18-1)!}0.6^{1}\cdot(1-0.6)^{17}=0.00000185$$

$$P(X=2)=\frac{18!}{2!(18-2)!}0.6^{2}\cdot(1-0.6)^{16}=0.0000236$$

$$P(X=3)=\frac{18!}{3!(18-3)!}0.6^{3}\cdot(1-0.6)^{15}=0.000189$$

$$P(X=4)=\frac{18!}{4!(18-4)!}0.6^{4}\cdot(1-0.6)^{14}=0.00106$$

Por tanto, P(X>=5) = 0.9987128, así pues la probabilidad de que se incorporen al proyecto más de cinco socios es de 0,9987.

Para calcular cual será el número de socios medio esperado que se incorpore al proyecto calculamos la media de una distribución Binomial que nos da el número medio de éxitos, en este caso sería, n*p= 18*0,6=10,8 que redondeando sería 11. Por tanto, el número medio esperado de socios que se incorporen al proyecto será de 11.

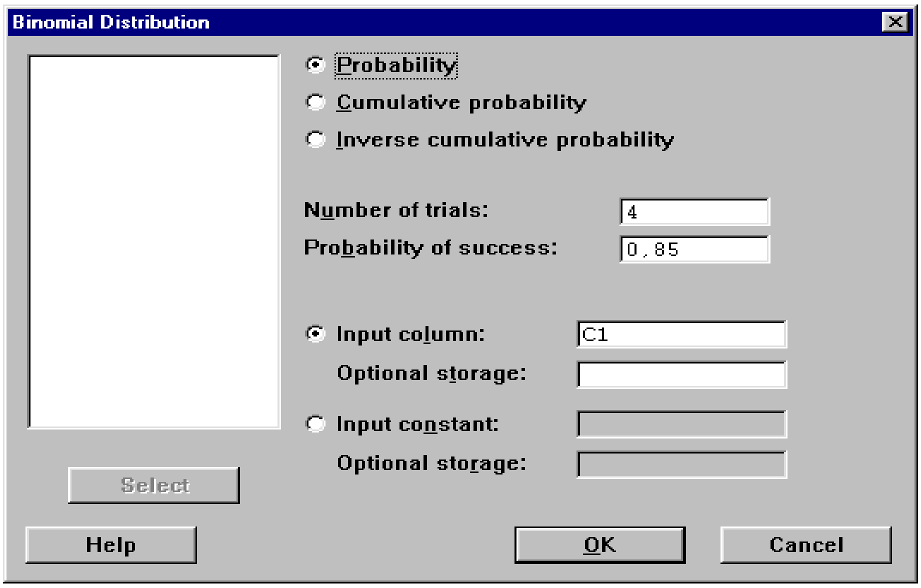
Ejemplos con Minitab:

1. Supongamos que *X* es una variable aleatoria (v.a.) que sigue una distribución binomial de parámetros $n = 4$ y $p = 0{,}85$.

   Veamos cómo podemos calcular la función de probabilidad de esta v.a.:

   En primer lugar, en la columna C1 colocaremos los posibles valores que esta v.a. puede tomar, i.e., 0, 1, 2, 3 y 4.

   Seleccionamos *Calc* > *Probability Distributions* > *Binomial* y completamos los campos como se indica en la imagen inferior:



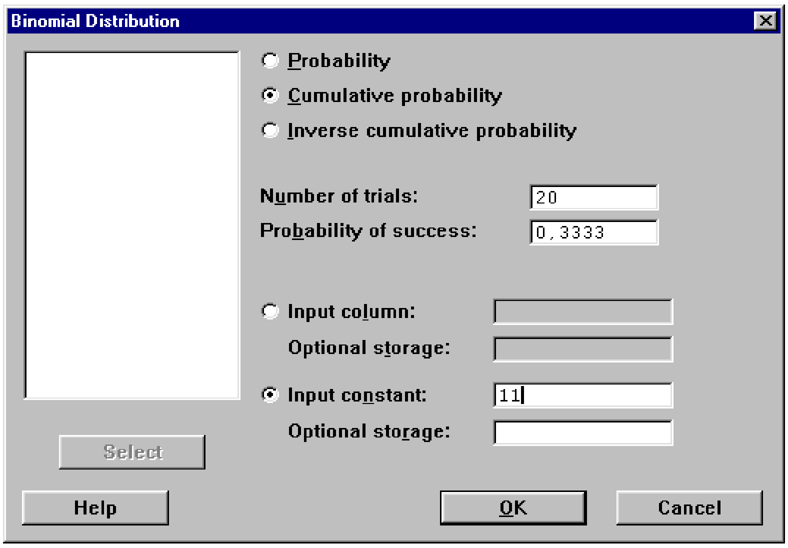

```
Probability Density Function

Binomial with n = 4 and p = 0,850000

x        P( X = x)
0,00         0,0005
1,00         0,0115
2,00         0,0975
3,00         0,3685
4,00         0,5220
```

Análogamente, el siguiente ejemplo nos muestra cómo calcular la función de distribución:

2. Supongamos que X sigue una distribución Binomial de n=20 y cuya probabilidad de éxito es 0.3333, es decir $X \sim B(20\,,\,0{,}3333)$.

Queremos calcular la probabilidad de que *X* tome un valor menor o igual a 11, i.e., P(X<=11):

Seleccionamos *Calc* > *Probability Distributions* > *Binomial* y completamos los campos como se indica en la imagen inferior:



El resultado es el siguiente:

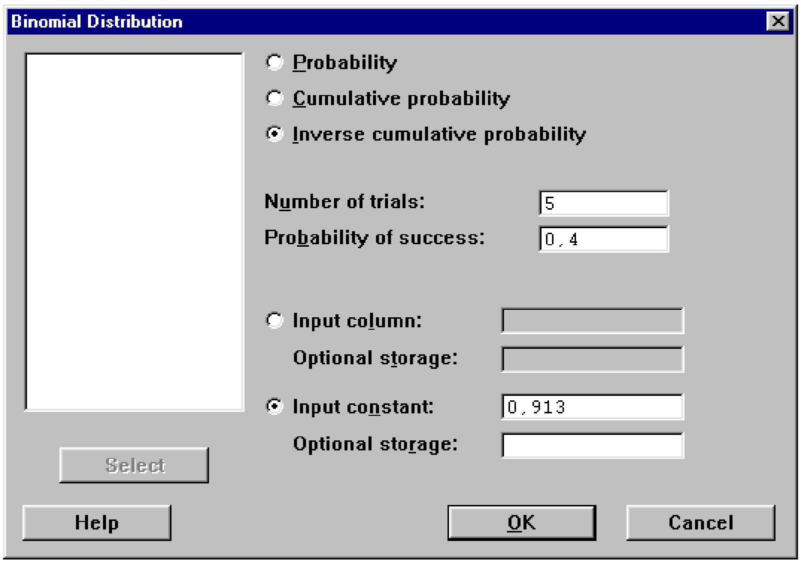
```
Cumulative Distribution Function

Binomial with n = 20 and p = 0,333300

x      P( X <= x)
11,00        0,9870
```

Por tanto, P(X<=11)=0.9870

Veamos un ejemplo de cómo aplicar la función de distribución inversa:

3. Sea $X \sim B(5{,}0{,}4)$. En esta ocasión, queremos saber cuál será el valor *c* de *X* tal que $P(X \leq c) = 0{,}913$:

   Seleccionar *Calc* > *Probability Distributions* > *Binomial* y completamos los campos como se indica en la imagen inferior:



El output que obtenemos nos dice que *c = 3* es el valor que deja a su izquierda el 91,3% de la distribución de *X* .

```
Inverse Cumulative Distribution Function

Binomial with n = 5 and p = 0,400000

x      P( X <= x)                 x      P( X <= x)
3         0,9130                  4         0,9898
```

Por último, aplicaremos todo lo anterior a un caso real:

4. La compañía aérea "Viajar" ofrece, entre otros, 10 vuelos diarios de Barcelona a Madrid. Se ha estudiado, que la probabilidad de que alguno de ellos se retrase es de 0.25.
   ¿Cuál es la probabilidad de que hoy no haya ningún vuelo que se retrase? ¿y la probabilidad de que no se retrasen más de dos vuelos?

   Sea X = "número de vuelos retrasados" y sabemos que X~B(10,0.25)

   Para calcular cuál es la probabilidad de que no haya ningún vuelo que se retrase, seleccionamos *Calc > Probability Distributions > Binomial.*

   El resultado es el siguiente:

```
Probability Density Function

Binomial with n = 10 and p = 0.250000

x       P( X = x)
0.00        0.0563
```

Por tanto, P(X=0) = 0.056, es decir, la probabilidad de que hoy no se retrase ninguno de los 10 vuelos es muy baja, aprox. 0.056

Ahora, para calcular la probabilidad de que no se retrasen más de dos vuelos, es decir, P(X<=2), seleccionaremos *Calc > Probability Distributions > Binomial*, y activaremos la opción de *Cumulative Probability,* obteniendo el siguiente resultado:

```
Cumulative Distribution Function

Binomial with n = 10 and p = 0.250000

x      P( X <= x)
2.00         0.5256
```

Por tanto, la probabilidad de que menos de 2 vuelos se retrasen es de aproximadamente el 0.53.

La distribución de Poisson

Consideremos X una variable que da el número de individuos que presentan una cierta característica por unidad de tiempo, volumen, superficie,… Entonces diremos que X sigue una distribución de Poisson.

Ejemplos:
X= Número de coches que cruzan un cruce en una hora.
Y= Número de enfermos de Sida por año y por Comunidad Autónoma.

La función de probabilidad de la distribución de Poisson es:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \text{para } x=0,1,2,3,\ldots.$$

donde λ es el número medio de ocurrencias durante un intervalo específico de tiempo, superficie, .. *e* es la constante exponencial y *x* es el número de ocurrencias (éxitos).

Observamos de la expresión de la función de probabilidad que el parámetro λ caracteriza las variables con distribuciónde Poisson.
Otra característica de la Poisson es que su media es igual a su varianza y ambas son igual al parámetro λ:

$$\mu = \lambda \quad , \qquad \sigma = \lambda$$

Observamos además que una variable con distribución Poisson toma infinitos valores, 0,1,… Ahora bien, las probabilidades van disminuyendo cada vez más rápidamente cuando el valor es alto, haciéndose prácticamente nulas a partir de un valor. Por esto muchas veces la distribución de Poisson también se la llama distribución de los sucesos "raros" o poco probables.

Aproximación de la Binomial a la Poisson.

Una distribución Binomial con una probabilidad de éxito p muy pequeña y n grande se aproxima a una distribución de Poisson con λ= n*p.
Algunas referencias utilizan esta aproximación cuando n>30 y p>0.1 y/o np<5.

Veamos un ejemplo que muestra cómo aplicar la distribución de Poisson haciendo uso de Minitab:

Ejemplo con Minitab:

Siguiendo con el ejemplo anterior, supongamos que tomamos una muestra aleatoria de 1000 vuelos y observamos que se perdieron 240 maletas. Esto indica que el número medio de maletas perdidas por vuelo es 0.24.
Si el número de maletas perdidas por vuelo sigue una distribución de Poisson de media 0.24, ¿cuál es la probabilidad de no perder ninguna maleta?

Sea X = "número de maletas perdidas" y sabemos que X ~ Po(0.24)

Seleccionamos *Calc* > *Probability Distributions* > *Poisson* y obtenemos:

```
Probability Density Function

Poisson with mu = 0.240000

x       P( X = x)
0.00         0.7866
```

Por tanto, se espera que aproximadamente el 79% de los vuelos no tengan ningún problema con la pérdida de equipaje.

La distribución normal

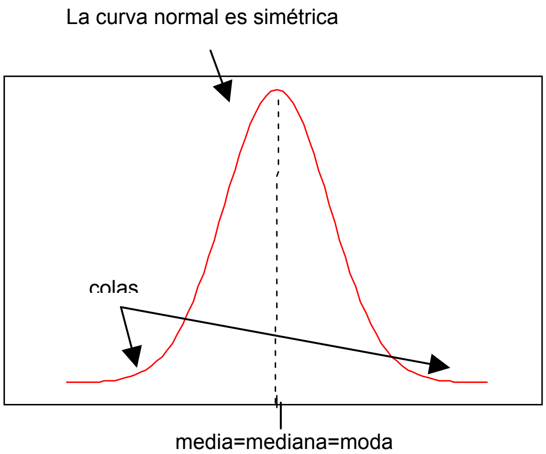
La distribución normal es la distribución de probabilidad continua más importante. Multitud de variables aleatorias continuas siguen una distribución normal o aproximadamente normal.
Una de sus características más importantes es que cualquier distribución de probabilidad, tanto discreta como continua, se puede aproximar por una normal bajo ciertas condiciones.

La distribución de probabilidad normal y la curva normal que la representa, tienen las siguientes características:

- La curva normal tiene forma de campana y un solo pico en el centro de la distribución. De esta manera, la media aritmética, la mediana y la moda de la distribución son iguales y se localizan en el pico. Así, la mitad del área bajo la curva se encuentra a la derecha de este punto central y la otra mitad está a la izquierda de dicho punto.

- La distribución de probabilidad normal es simétrica alrededor de su media.

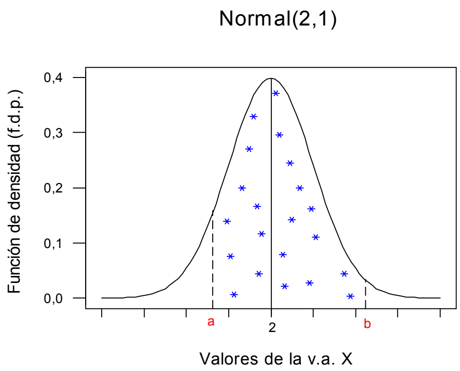
- La curva normal desciende suavemente en ambas direcciones a partir del valor central. Es asintótica, lo que quiere decir que la curva se acerca cada vez más al eje X pero jamás llega a tocarlo. Es decir, las "colas" de la curva se extienden de manera indefinida en ambas direcciones.



Para indicar que una variable aleatoria (v.a.) sigue una distribución normal de media μ y desviación estándar σ usaremos la expresión: $X \sim N(\mu,\sigma)$.

La probabilidad de que una variable aleatoria (v.a.) X tome un valor determinado entre dos números reales *a* y *b* coincide con el área encerrada por la función $f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$ (**función de densidad de probabilidad**) entre los puntos *a* y *b*, es decir :

$$P(a \le X \le b) = \int_a^b f(x)dx$$



Como hemos comentado anteriormente, observar que:

- La distribución normal es simétrica respecto de su media μ .
- El área total encerrada por *f(x)* vale 1, i.e.: $\int_{-\infty}^{+\infty} f(x)dx = P(-\infty < X < +\infty) = 1$.
- Al ser *X* v.a. continua, $P(X=a) = \int_a^a f(x)dx = 0$, $\forall\, a \in R \;\Rightarrow\; P(X \le a) = P(X<a)$.

La distribución normal estándar:

Se observó que no existe una sola distribución de probabilidad normal, sino una “familia” de ellas. Como sabemos, cada una de las distribuciones puede tener una media (μ) o una desviación estándar distinta (σ). Por tanto, el número de distribuciones normales es ilimitado y sería imposible proporcionar una tabla de probabilidades para cada combinación de μ y σ.

Para resolver este problema, se utiliza un solo “miembro” de la familia de distribuciones normales, aquella cuya media es 0 y desviación estándar 1 que es la que se conoce como **distribución estándar normal**, de forma que todas las distribuciones normales pueden convertirse a la estándar, restando la media de cada observación y dividiendo por la desviación estándar.

Primero, convertiremos la distribución real en una distribución normal estándar utilizando un valor llamado Z, o **estadístico Z** que será la distancia entre un valor seleccionado, designado X, y la media μ, dividida por la desviación estándar σ.

Formalmente, si $X \sim N(\mu,\sigma)$, entonces la v.a. $Z = \frac{X-\mu}{\sigma}$ se distribuye según una normal de media 0 y desviación estándar 1, i.e.: $Z \sim N(0,1)$, que es la distribución llamada **normal estándar o tipificada**.

De esta manera, un valor Z mide la distancia entre un valor especificado de X y la media aritmética, en las unidades de la desviación estándar. Al determinar el valor Z utilizando la expresión anterior, es posible encontrar el área de probabilidad bajo cualquier curva normal haciendo referencia a la distribución normal estándar en las tablas correspondientes.

Así pues, para averiguar el área encerrada bajo la curva utilizaremos la tabla que encontraremos al final de este apartado. Dicha tabla nos proporciona la probabilidad de que la v.a. normal estándar *Z* tome un valor situado a la izquierda de un número *c*, i.e.: *P(Z<c)*. En otras palabras, esta tabla nos da el valor del área encerrada por *f(x)* entre -∞ y *c*.

Distribución muestral de la media de las muestras:

Consistiría en una distribución de probabilidad de todas las medias posibles de las muestras de un tamaño de muestra dado.

Así pues, dada una población (a la cual representaremos por la v.a. *X*), podemos extraer de la misma *k* muestras, cada una de ellas de tamaño *n*. Para cada una de las *k* muestras podemos calcular un estadístico, p.e., la media de las *n* observaciones que la componen.

Así tendremos un total de *k* nuevos valores $\bar{x}_i, i = 1,\ldots,k$. Podemos asociar estos valores a una nueva v.a. $\bar{X}$, cuya distribución llamaremos **distribución muestral**.

Una de las propiedades más importantes es la siguiente:

Teorema (Distribución de las Medias Muestrales):

Sea *X* una v.a. **cualquiera** de media μ y desviación típica σ, entonces:

- Si consideramos **todas** las muestras aleatorias posibles, cada una de ellas de tamaño *n*, se cumplirá que $\mu_{\bar{x}} = \mu$ y $\sigma_{\bar{x}} = \sigma / \sqrt{n}$.
- Además, si *X* sigue una distribución normal, $\bar{X}$ también será normal.

Teorema Central del Límite:

Sea *X* una v.a. **cualquiera** de media μ y desviación típica σ, entonces:

Si el tamaño muestral *n* es “suficientemente grande” (en la práctica suele valer *n>30*), la distribución de las medias muestrales se aproxima a la de una normal, i.e.:

$$\bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

La importancia del TCL radica en que **sea cuál sea** la distribución de la población original (v.a. *X*), conforme el tamaño de las muestras (*n*) aumenta, la distribución de las medias se va aproximando a la de una normal (de la cual conocemos muchas propiedades).
Así, si la población tiene una distribución de probabilidad normal, entonces, para cualquier tamaño de muestra la distribución del muestreode la media también tendrá una distribución normal. Si la distribución de la población es simétrica (pero no normal), se verá que surge la forma normal como lo establece el TCL aún con muestras de al menos 30 para observar las características de normalidad.

Aproximación de la Binomial a la distribución Normal.
(una aplicación del teorema Central del límite)

Si $X \approx B(n,p)$ y el nº de pruebas *n* es “muy grande” (en la práctica es suficiente con verificar: *n*p≥5* y *n*(1-p)≥5*), entonces podemos aproximar la distribución binomial anterior a una normal, en concreto: $X \approx N\left(n*p, \sqrt{n*p*(1-p)}\right)$. Esta aproximación será tanto mejor cuanto mayor sea *n*.

Hay que tener en cuenta que, antes de aplicar la distribución normal, es necesario asegurarse de que la distribución que queremos aproximar es, efectivamente, binomial.

Para ello, hay que comprobar:

- Que un experimento sólo puede tener dos resultados posibles y mutuamente excluyentes: un “éxito” y un “fracaso”.
- La distribución es consecuencia de contar el número de éxitos de un número fijo de pruebas.
- Cada prueba es independiente.
- La probabilidad, p, permanece igual de una prueba a la siguiente.

En el caso de una v.a. discreta, tiene sentido preguntarse por la probabilidad de que ésta tome un determinado valor. Sin embargo, si consideramos que la v.a. *X* es continua, entonces *P(X=a) = 0,* $\forall a \in R$. Por este motivo tendremos que aplicar el llamado **factor de corrección por continuidad** que veremos a continuación, es decir, en el caso anterior calcularemos *P(a-0,5<X<a+0,5).*

Ejemplos:

1. El PER de una acción que cotiza en bolsa indica el número de veces que su precio es mayor que el beneficio por acción y este ratio es uno de los más importantes que utilizan habitualmente los inversores. Supongamos que tenemos la población de todos los PER que tiene una media de 10,5 y una desviación estándar de 4,5. ¿Cuál es la probabilidad de que, en una muestra de 40 acciones, el PER medio sea menor que 9?

   Por el teorema del Límite central, como n=40 y es mayor que 30 podemos afirmar que la distribución muestral de la media de los PER se aproximará a una distribución normal.

   $$P(\overline{X} < 9) = P(\overline{X} < 9) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{9-10{,}5}{4{,}5/\sqrt{40}}\right) = P(Z < -2{,}11) \xrightarrow{TABLAS} 0{,}0174$$

   Por lo tanto existe una probabilidad del 1,74% de que la media de los PER de la muestra sea menor que 9.

2. El Presidente de una multinacional de telecomunicaciones, está preocupado por el número de teléfonos móviles producidos por su empresa que tienen algún defecto. En promedio, 110 teléfonos al día son devueltos por este problema, con una desviación estándar de 64. El presidente de esta empresa ha decidido que a menos que pueda estar un 80% seguro de que, en promedio, no se devolverán más de 120 teléfonos al día durante lo siguientes 48 días, ordenará una reorganización general del proceso productivo. ¿se ordenará el reajuste decidido por el Presidente?

Para que se ordene la reorganización del proceso productivo, la probabilidad d que la media de teléfonos devueltos al día durante os próximos 48 días sea menor que 0,8. Entonces debemos calcular la probabilidad de que la media no se mayor que 120,

$$P(\overline{X} \leq 120)$$

$$P\left(\overline{X} \leq 120\right) = P(\overline{X} \leq 120) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{120-110}{64/48}\right) = P\left(Z \leq 1{,}08\right) \xrightarrow{TABLAS} 0{,}8599 = 0{,}8599$$

Por lo tanto existe una probabilidad de 0,8599 de que no de devuelva en promedio más de 120 teléfonos al día durante los próximos 48 días, por lo que como esta probabilidad es mayor que 0,8, no habrá reajuste del proceso productivo comentado por el Presidente.

3. El director de Recursos Humanos de una empresa, desea estudiar el nivel de precisión de las 70 secretarías de su compañía. Anteriormente, el número diario de errores de procesamiento de palabras cometido por cada secretaria había sido aproximadamente normal con un promedio de 18 y una desviación estándar de 4. El director de Recursos Humanos inspecciona actualmente a 15 secretarias elegidas aleatoriamente. ¿Cuál es la probabilidad de que el número promedio de errores por secretaria sea mayor de 20?

$$P\left(\overline{X} > 20\right) = 1 - P(\overline{X} \leq 20) = 1 - P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{20-18}{4/\sqrt{15}}\right) = 1 - P\left(Z \leq 1{,}94\right)$$

$$\xrightarrow{TABLAS} 1 - 0{,}9738 = 0{,}0262$$

Por lo tanto existe una probabilidad de 0,0262 de que el número medio de errores por secretaria sea mayor que 20.

Ejemplo con minitab:

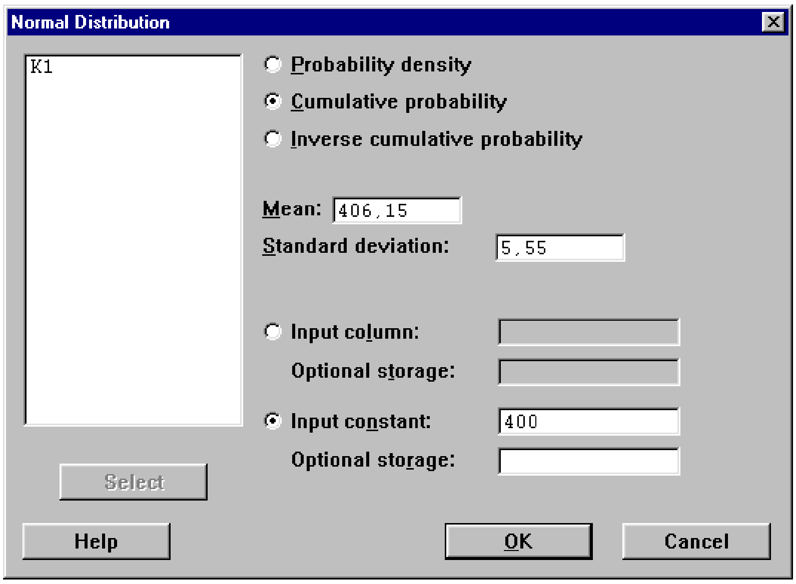
Según viene publicado en una prestigiosa revista de economía, el salario semanal medio de los profesores universitarios europeos es de 406,15 €. Se estima además que la desviación estándar de dichos salarios es de 55,50 €. Supongamos ahora que pretendemos tomar una muestra aleatoria de 100 profesores para estudiar sus salarios. Calcular las siguientes probabilidades referentes a la media de dicha muestra:

1. La probabilidad de que la media de la muestra sea menor de 400 €.

   En primer lugar, observar lo siguiente: como n = 100 >> 30, por el Teorema Central del Límite tendremos que la distribución de las medias muestrales $\overline{X}$ se podrá aproximar por una normal con media 406,15 y desviación estándar 5,50.

   Hemos de hallar $P(\overline{X} < 400)$:

   Seleccionamos: *Calc > Probability Distributions > Normal :*




```
Cumulative Distribution Function

Normal with mean = 406,150 and standard deviation = 5,55000

          x       P( X <= x)
  400,0000          0,1339
```

Por lo tanto existe una probabilidad del 13;39 % de que el salario medio se menor de 400.

2. La probabilidad de que la media de la muestra esté entre 400 y 410 € .

Sabemos que $P(400 < \overline{X} < 410) = P(\overline{X} < 410) - P(\overline{X} < 400)$. La segunda de éstas probabilidades ya la hemos calculado en el apartado anterior.

Para calcular la primera se razona análogamente, obteniendo que:

```
Cumulative Distribution Function

Normal with mean = 406,150 and standard deviation = 5,55000

          x       P( X <= x)
  410,0000          0,7561
```

Por tanto, tendremos: $P(400 < \overline{X} < 410) = P(\overline{X} < 410) - P(\overline{X} < 400) = 0{,}6222$

3. La probabilidad de que la media de la muestra sea mayor de 415 € .

En este caso, $P(\overline{X} > 415) = 1 - P(\overline{X} < 415)$. Hemos de calcular pues esta última probabilidad, lo cual haremos de forma análoga a los apartados anteriores.

Obtendremos lo siguiente:

```
Cumulative Distribution Function
```

```
        Normal with mean = 406,150 and standard deviation = 5,55000

                 x      P( X <= x)
          415,0000         0,9446
```

Por consiguiente, $P(\overline{X} > 415) = 1 - P(\overline{X} < 415) = 0{,}0554$

4. Hallar el valor del salario medio c tal que $P(X < c) = 0{,}95$.

   Seleccionamos nuevamente: *Calc* > *Probability Distributions* > *Normal* , pero ahora elegiremos la opción *Inverse Cumulative Probability* , con lo que obtendremos :

```
        Inverse Cumulative Distribution Function

        Normal with mean = 406,150 and standard deviation = 5,55000

         P( X <= x)           x
            0,9500      415,2789
```

NOTA.- En la siguiente dirección: http://huizen.dds.nl/~berrie/ encontraréis algunos vídeos que ilustran las distintas distribuciones que se han descrito en este apartado.

CASOS PRÁCTICOS CON SOFTWARE

1) Una compañía de seguros tiene una cartera de 2.000 pólizas que cubren la asistencia psicológica en caso de accidente. La empresa estima que este siniestro tiene una probabilidad de ocurrencia del 2 por mil en un año, y un coste medio de 100000 u.m. por siniestro. Calcular:

   a) La probabilidad de afrontar más de 3 siniestros en el año.

      Sea X="Número de siniestros"
      Como el número de pruebas es muy grande y la probabilidad de éxito es muy pequeña, vemos que X sigue una distribución de Poisson, donde μ=2000.2/1000, es decir, X~Po(4)

      Para calcular la probabilidad de P(X>3), seleccionamos *Calc* > *Probability Distributions* > *Poisson*:

```
Cumulative Distribution Function

Poisson with mu = 4.00000

         x      P( X <= x)
      3.00         0.4335
```

      Por tanto, P(X>3) = 1 - P(X<3) = 1- 0.433 = 0.567

   b) La reserva que ha guardado la compañía para los siniestros del año, nos asegura que tiene una probabilidad del 99,2% de poder afrontar todos los siniestros que ocurran.

En este caso, queremos calcular el número máximo de siniestros para poder cubrir los gastos que ellos suponen.

Para ello, seleccionamos *Calc* > *Probability Distributions* > *Poisson* y rellenamos como sigue:

```
Inverse Cumulative Distribution Function

Poisson with mu = 4.00000

         x      P( X <= x)              x      P( X <= x)
         9        0.9919               10        0.9972
```

Por tanto, la empresa habrá dotado una reserva para afrontar un máximo de 9 siniestros, ya que, a un coste de 100.000 u.m. supondrá 900.000 u.m.

2) En un estudio reciente se demostró que el 64% de catalanes, con estudios universitarios acabados, hacen uso de internet a diario en su trabajo. Si seleccionamos una muestra de 60 catalanes con estas características:

a) ¿Cuál es la probabillidad de que 32 o más sean usuarios de internet?

Definimos X=”Catalanes con estudios universitarios usuarios de internet”.

Observamos que $X \sim B(60,0.64)$

Como n*p= 60 * 0.64 = 38.40 >> 5, y n*p*(1-p) = 60 * 0.64 * 0.36 = 13.82 , podemos aproximar la distribución binomial a una distribución normal N(38.40, 3.72).

Por tanto, seleccionamos *Calc* > *Probability Distributions* > *Normal*:

```
Cumulative Distribution Function

Normal with mean = 38.4000 and standard deviation = 3.72000

         x      P( X <= x)
   32.0000        0.0427
```

P(X>=32) = 1 – P(X<32) = 1 - 0.0427 = 0.9678

b) ¿Y cuál es la probabilidad de que más de 32 y menos de 44 hagan uso de internet?

Queremos calcular P(32<X<44) = P(X<44) – P(X<32) = 0.9339 – 0.0427= 0.8912

Seleccionamos *Calc* > *Probability Distributions* > *Normal*:

```
Cumulative Distribution Function

Normal with mean = 38.4000 and standard deviation = 3.72000

          x     P( X <= x)
    44.0000        0.9339
```

c) Si queremos conseguir que un porcentaje de un 90%, ¿qué número de usuarios necesitaríamos?

Seleccionamos *Calc* > *Probability Distributions* > *Normal*:

```
Inverse Cumulative Distribution Function

Normal with mean = 38.4000 and standard deviation = 3.72000

 P( X <= x)          x
     0.9000    43.1674
```

Así, pues, haría falta unos 43 catalanes para llegar al porcentaje del 90%

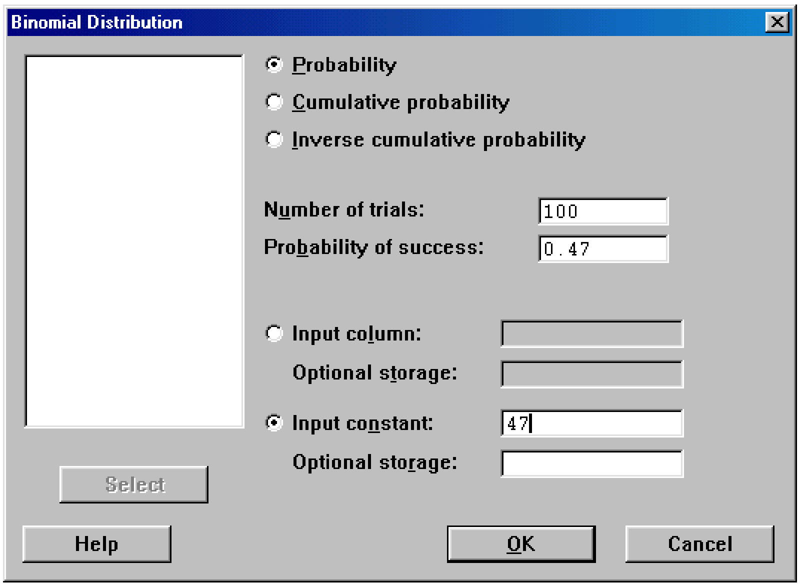
3) Supongamos que en una población, sólo el 47% de los habitantes son favorables a las gestiones municipales realizadas por la alcaldía. Se selecciona aleatoriamente una muestra de 100 personas y se les pasa un cuestionario, de manera independiente a cada una.

a) ¿Cuál es la probabilidad de que, en la muestra, aparezcan exactamente 47 personas favorables al Ayuntamiento?

Sea X=”Número de personas favorables al Ayuntamiento ”

Además, X seguirá una distribución binomial con n=100 y cuya probabilidad de éxito será 0.47, es decir, $X \sim B(100,0.47)$

Para calcular P(X=47), seleccionaremos *Calc* > *Probability Distribution* > *Binomial*:



Y obtenemos....

Probability Density Function

Binomial with n = 100 and p = 0.470000

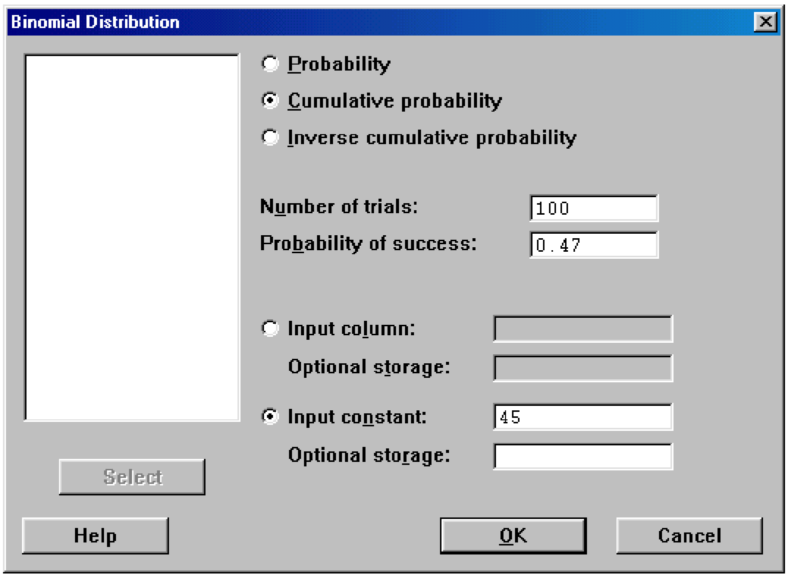
x	P(X = x)
47.00	0.0797

La probabilidad de que haya exactamente 47 personas que estén a favor del alcalde es 0.08, es decir, el 8%

b) ¿Cuál es la probabilidad de que entre 45 y 50 personas (incluidos estos valores), se muestren a favor de las gestiones del Ayuntamiento?

Queremos calcular P(45<=X<=50), es decir, P(X<=50)-P(X<=45)

Para ello, seleccionamos *Calc > Probability Distributions > Binomial*, activando la opción *Cumulative Probability*:



Cumulative Distribution Function Binomial with n = 100 and p = 0.470000 x P(X <= x) 45.00 0.3827

Razonamos análogamente para P(X<=50), obteniendo:

Cumulative Distribution Function Binomial with n = 100 and p = 0.470000 x P(X <= x) 50.00 0.7587

OJO!! También tenemos que calcular el valor para P(X=45), ya que P(X>=45)=1-P(X<45):

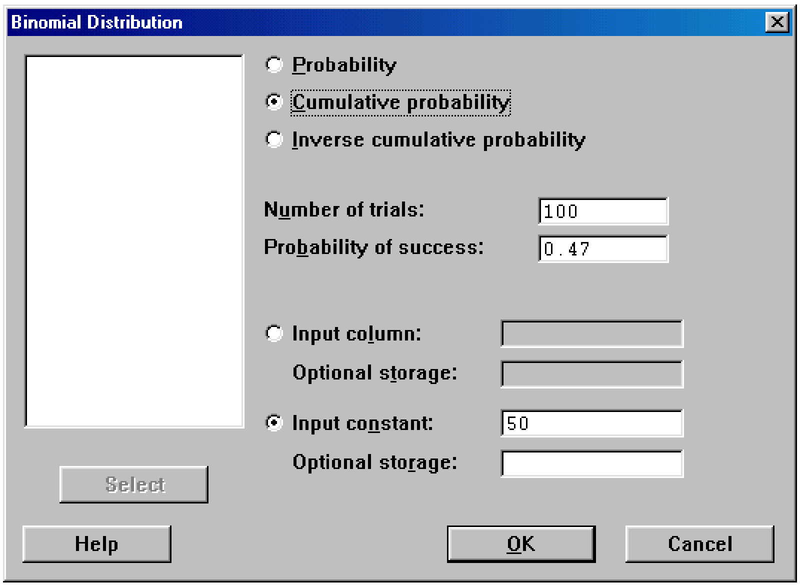
Probability Density Function Binomial with n = 100 and p = 0.470000 x P(X = x) 45.00 0.0738

P(45<=X<=50) = P(X<=50) - P(X<45) = 0.7587 – 0.3827 + 0.0738 = 0.4498

Así pues, la probabillidad de que entre 45 y 50 personas se muestren a favor de las gestiones del Ayuntamiento es aproximadamente de 0.45

c) ¿Es probable que en la muestra, si está bien escogida, puedan salir una mayoría de personas favorables al alcalde?

Para calcular la probabilidad de que, en una muestra de 100 personas, la mayoría esté a favor del alcalde, calcularemos P(X>50).
Para ello, seleccionamos *Calc > Probability Distributions > Binomial*, y activamos la opción *Cumulative Probability*:



Cumulative Distribution Function
Binomial with n = 100 and p = 0.470000

x	P(X <= x)
50.00	0.7587

Por tanto, P(X>50) = 1 - P(X<=50) = 1 - 0.7587 = 0.2413
Así pues, el porcentaje esperado de que la mayoría de las personas de la muestra estén a favor del alcalde es el 24%.

d) Si hacemos la predicción “a la muestra saldrán entre 45 y 50 personas favorables”, ¿cuál es el riesgo de equivocarnos?

Como hemos visto en el apartado b), existe una probabilidad de 0.45 de que entre 45 y 50 personas sean favorables al Ayuntamiento, por tanto, el riesgo de equivocarnos sería:

1 - P(45<=X<=50)= 0.55

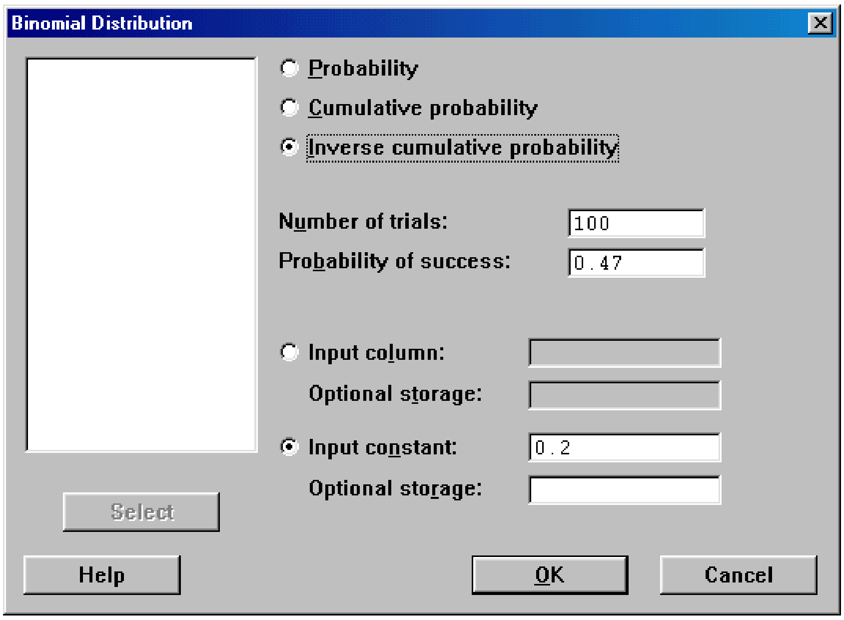
e) Si el ayuntamiento pretende hacer una previsión a partir de la muestra diciendo: “La mayoría de votantes estarán a nuestro favor”, ¿qué riesgo tiene esta predicción?

Análogamente, en el apartado c), hemos visto que existe una probabilidad de 0.24 de que la mayoría de los encuestados estén a favor de las gestiones del ayuntamiento, por tanto, el riesgo de esta predicción por parte del ayuntamiento seria 0.76

f) ¿Qué previsión podemos hacer para poder decir "el número habitual de personas favorables a la gestión municipal que saldran en la muestra...", si queremos asumir un riesgo máximo de 0.2 ?

Queremos calcular c, tal que, $P(X<c)=0.2$

Seleccionamos *Calc* > *Probability Distributions* > *Binomial* y activamos *Inverse Cumulative Probability*:



Inverse Cumulative Distribution Function

Binomial with n = 100 and p = 0.470000

x	P(X <= x)	x	P(X <= x)
42	0.1838	43	0.2420

Por tanto, el número habitual de personas que aparecerán en la muestra favorables a la gestión municipal es aproximadamente de 42.

BIBLIOGRAFÍA

[1] F. Moya Anegón, J. López Gijón, C. García Caro(1996): "Técnicas cuantitativas aplicadas a la biblioteconomia y documentación". Ed. Síntesis.

[2] D. A. Lind, R.D. Mason, W. G. Marchal (2001): "Estadística para Administración y Economía". Ed. Irwin McGraw-Hill.

[3] R. Johnson (1996): "Elementary Statistics". Ed. Duxbury.

[4] E. Farber (1995): "A guide to Minitab". Ed. McGraw-Hill.

ENLACES

- http://www.unalmed.edu.co/~estadist/binomial/binomial.htm : Descripción y representación de la distribución binomial.
- http://es.geocities.com/riotorto/tabl/tabl_binomial/tabl_binomial.htm : Descripción de la distribución binomial.
- http://huizen.dds.nl/~berrie/ : Vídeos de conceptos de estadística
- http://www.udc.es/dep/mate/recursos.html : Recursos de internet para la enseñanza y aprendizaje de la estadística.

Financiado por la Secretaría de Estado de Educación y Universidades (MECD)