

# Índice general

<b>3. Distribuciones de Probabilidad. Estimación</b>	<b>3</b>
3.1. Modelos de distribución	3
3.1.1. Modelos de distribución para variables discretas	3
3.1.2. Modelos de distribución para variables continuas	4
3.2. Distribuciones de Probabilidad en Statgraphics	5
3.2.1. Algunos Ejemplos de Distribuciones de Probabilidad en Statgraphics	9
3.2.2. Gráficos de normalidad	19
3.3. Estimación y Contraste de Hipótesis	22
3.4. Inferencia de una muestra en Statgraphics	24
3.4.1. Cuando partimos de los estadísticos muestrales	25
3.4.2. Cuando tenemos los datos muestrales	27
3.4.3. Ejemplos sobre Estimación y Contraste de Hipótesis	29



## Capítulo 3

# Distribuciones de Probabilidad. Estimación

### 3.1. Modelos de distribución

Uno de los objetivos del cálculo de probabilidades es obtener modelos generales de comportamiento de los distintos fenómenos aleatorios. En este tema vamos a estudiar algunos de los modelos de distribución más importantes para variables discretas (Bernoulli, Binomial y Poisson) y continuas (Normal, t-Student, Chi-cuadrado y F-Snedecor).

#### 3.1.1. Modelos de distribución para variables discretas

##### Distribución Bernoulli

Este modelo se aplica a un experimento aleatorio que presenta sólo dos resultados posibles e incompatibles a los que se denomina éxito y fracaso. A este experimento aleatorio se le asocia una variable aleatoria  $X$  definida como

$X =$  número de éxitos al realizar el experimento

cuya distribución viene dada en la siguiente tabla

$x_i$	$p_i$
0	$1 - p$
1	$p$

donde  $p$  es la probabilidad de éxito y  $1 - p$  es la probabilidad de fracaso que se suele representar por  $q$ , así  $p + q = 1$ .

Esta distribución tiene media  $E[X] = p$  y varianza  $Var[X] = pq$ .

##### Distribución Binomial

Supongamos, como en el caso anterior, que tenemos un experimento aleatorio que presenta sólo dos resultados posibles e incompatibles a los que se denomina éxito y fracaso; y representaremos por  $p$  la probabilidad de éxito y por  $q = 1 - p$  la probabilidad de fracaso. Repetimos  $n$  veces el experimento de forma independiente y en las mismas condiciones. Definimos la variable aleatoria

$X$  = número de éxitos en las  $n$  realizaciones del experimento

En esta situación diremos que la variable  $X$  tiene una distribución Binomial,  $X \sim B(n, p)$ . La función masa de probabilidad es

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

la media es  $E[X] = np$  y la varianza es  $Var(X) = npq$ .

### Distribución Poisson

Consideremos ahora un experimento que consiste en observar el número de ocurrencias de un determinado suceso en un intervalo de longitud fija de una magnitud continua, como puede ser un intervalo de tiempo, una longitud, una superficie, un volumen, etc., de forma que el número medio de ocurrencias por unidad de magnitud es constante y las ocurrencias en dos intervalos disjuntos son independientes.

En estas condiciones la variable

$X$  = número de ocurrencias del suceso en ese intervalo de longitud fija

sigue una distribución de Poisson de parámetro  $\lambda$ ,  $X \sim P(\lambda)$ , donde  $\lambda$  es el número medio de ocurrencias del suceso en ese intervalo. La función masa de probabilidad viene dada por

$$P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

y tanto su media y como su varianza es  $\lambda$ ,  $E[X] = Var(X) = \lambda$ .

### 3.1.2. Modelos de distribución para variables continuas

#### Distribución Normal

Una variable aleatoria continua se dice que tiene una distribución normal de parámetros  $\mu$  y  $\sigma$ , con  $-\infty < \mu < +\infty$  y  $\sigma > 0$ , y se representa como  $X \sim \mathcal{N}(\mu, \sigma)$ , si su función de densidad es de la forma

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \quad -\infty < x < +\infty$$

Tiene media  $E[X] = \mu$  y varianza  $Var(X) = \sigma^2$ . Es la distribución continua más importante, tanto por su aplicación, ya que modeliza gran cantidad de fenómenos, como por su utilidad para obtener, a partir de ella, otras distribuciones también de gran importancia en la inferencia estadística y que estudiaremos en los siguientes apartados.

#### Distribución Chi-cuadrado

Esta distribución se obtiene como suma de los cuadrados de variables normales estándar independientes, es decir:

Sean  $Z_1, \dots, Z_\nu$  variables aleatorias independientes e idénticamente distribuidas según una  $N(0, 1)$ , diremos que la variable  $X = Z_1^2 + \dots + Z_\nu^2$  sigue una distribución  $\chi^2$  con  $\nu$  grados de libertad, y lo representaremos como  $X \sim \chi_\nu^2$ .

Se obtiene así una variable aleatoria continua asimétrica y positiva con media  $E[X] = \nu$  y varianza  $Var(X) = 2\nu$ .

### Distribución $t$ de Student

Se obtiene a partir del cociente entre una variable normal estándar y la raíz cuadrada de una variable chi-cuadrado dividida entre sus grados de libertad, ambas independientes, es decir:

Sean  $X \sim N(0, 1)$  e  $Y \sim \chi^2_\nu$  variables aleatorias independientes, diremos que la variable

$$T = \frac{X}{\sqrt{Y/\nu}}$$

sigue una distribución  $t$  de Student con  $\nu$  grados de libertad,  $T \sim t_\nu$ .

Se obtiene una variable aleatoria continua simétrica con media  $E[X] = 0$  y varianza  $Var(X) = \frac{\nu}{\nu - 2}$ .

### Distribución $F$ de Snedecor

Se obtiene como el cociente entre dos distribuciones chi-cuadrado independientes divididas por sus correspondientes grados de libertad, es decir:

Sean  $X \sim \chi^2_{\nu_1}$  e  $Y \sim \chi^2_{\nu_2}$  variables aleatorias independientes, diremos que la variable

$$F = \frac{X/\nu_1}{Y/\nu_2}$$

sigue una distribución  $F$  de Snedecor con  $\nu_1$  y  $\nu_2$  grados de libertad, y se representa como  $F \sim F_{\nu_1, \nu_2}$ .

Se obtiene una variable aleatoria continua asimétrica y positiva.

## 3.2. Distribuciones de Probabilidad en Statgraphics

El estudio de las distribuciones de probabilidad en Statgraphics se puede realizar en el menú **Descripción/Distribuciones/Distribuciones de Probabilidad...** (Figura 3.1, pag. 5) o bien en el menú **Gráficos/Distribuciones de Probabilidad...** (Figura 3.2, pag. 5).



Figura 3.1: Distribuciones de probabilidad en el Menú Descripción

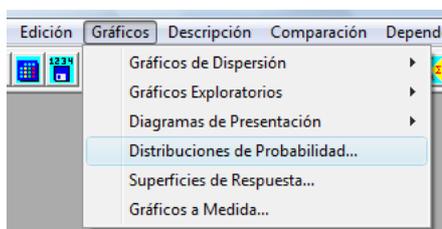


Figura 3.2: Distribuciones de probabilidad en el Menú Gráficos

Por defecto está seleccionada la distribución normal <sup>1</sup> (Figura 3.3, pag. 6). Así, por ejemplo, si dejamos marcada esta distribución y pulsamos **Aceptar**, obtenemos la ventana de la Figura 3.4 (pag. 7)

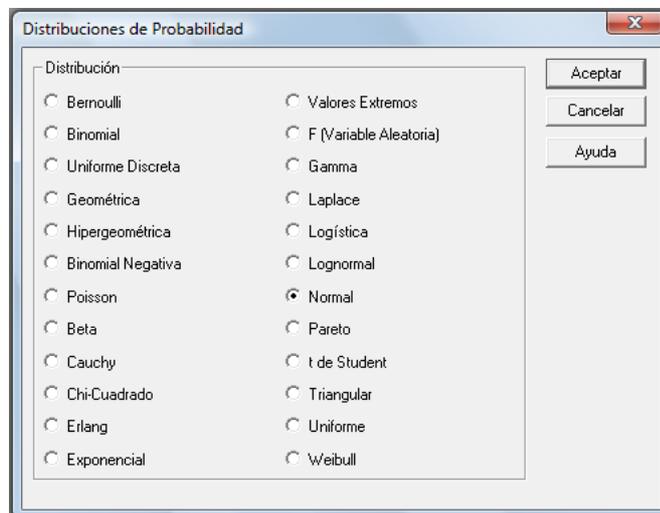


Figura 3.3: Distribuciones de probabilidad

Veamos la utilidad de los cuatro primeros botones que aparecen en la parte superior izquierda de la ventana de la Figura 3.4 (pag. 7):

- El primer botón, con fondo de color rojo , nos permite cambiar de distribución. Pulsando sobre él nos vuelve a salir la ventana de la Figura 3.3 (pag. 6).
- El segundo botón, con fondo de color amarillo , se utiliza para indicar las opciones tabulares. Los resultados de las distintas opciones se mostrarán en las ventanas de la columna izquierda del análisis. Pulsando sobre él obtenemos la ventana de la Figura 3.5 (pag. 7), cuyas opciones son las siguientes
  - **Resumen del Análisis:** Marcando esta opción se obtiene una ventana que muestra información sobre la distribución seleccionada y sus parámetros. Se pueden seleccionar hasta 5 variables con la misma distribución y diferentes parámetros, para indicar estos parámetros se pulsa con el botón derecho del ratón sobre cualquier ventana del análisis (se obtiene un menú igual o similar al de la Figura 3.6) (pag. 7) y se selecciona **Opciones de Análisis...** (se obtiene así la Figura 3.7, pag. 8)
  - **Distribución Acumulada:** Se obtiene una ventana que se utiliza para obtener las probabilidades de que las variables sean menor o mayor que un valor concreto, además calcula la función de densidad (caso continuo) o la función masa de probabilidad (caso discreto) en ese valor concreto. Se pueden seleccionar hasta 5 valores, pulsando con el botón derecho sobre la ventana que resulte y seleccionando **Opciones de Ventana...** se obtiene la ventana de la Figura 3.8 (pag. 8) donde indicaremos esos valores.

<sup>1</sup>Para ilustrar todo lo que sigue a continuación se utilizará la distribución normal, para cualquier otra distribución las ventanas que se van obteniendo son similares.

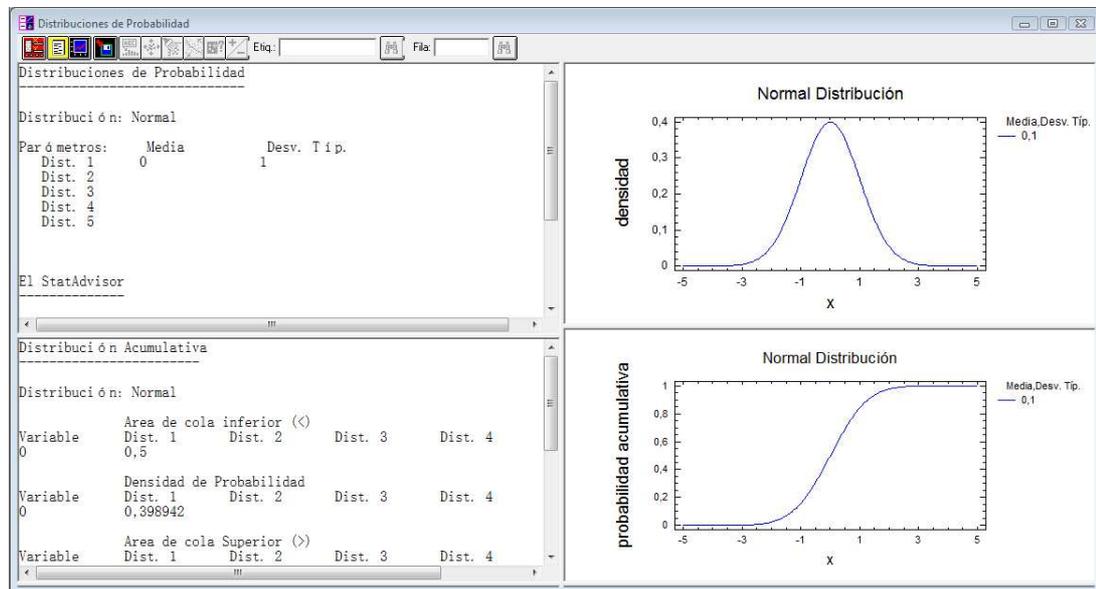


Figura 3.4: Ejemplo de ventana de Distribuciones de probabilidad (Normal)



Figura 3.5: Opciones tabulares

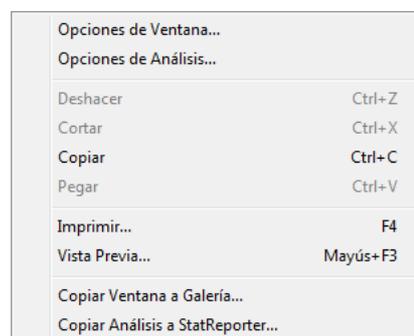


Figura 3.6: Menú del botón derecho

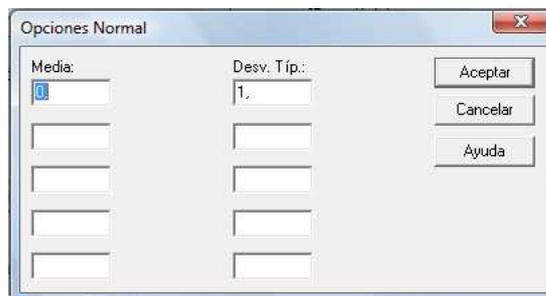


Figura 3.7: Opciones de Análisis

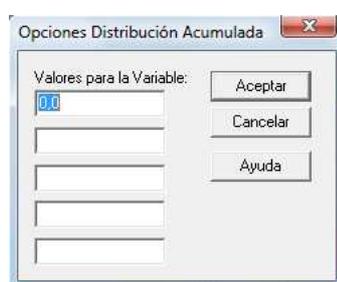


Figura 3.8: Distribución Acumulada

- **CDF Inverso:** Se obtiene una ventana que se utiliza para obtener los cuantiles o valores críticos de las distribuciones. Introducimos la probabilidad  $p$  y obtendremos un valor  $a$  tal que  $P[X \leq a] = p$ . Se pueden calcular hasta 5 cuantiles, para ello se pulsa con el botón derecho sobre la ventana correspondiente y seleccionando **Opciones de Ventana...**, obtenemos la ventana de la Figura 3.9 (pag. 8), donde indicaremos los distintos valores de  $p$ .

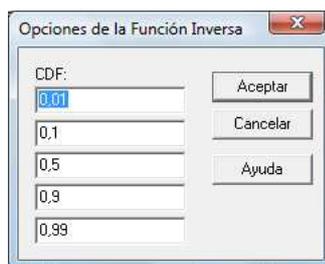


Figura 3.9: CDF Inversa

- **Números Aleatorios:** Se utiliza para simular una muestra aleatoria de la distribución que estamos estudiando. Para indicar el tamaño muestral se pulsa con el botón derecho sobre la correspondiente ventana que aparezca y se selecciona **Opciones de Ventana...** (Figura 3.10, pag. 9).  
Para guardar esta muestra aleatoria tendremos que pulsar sobre el cuarto botón, guardar

resultados .

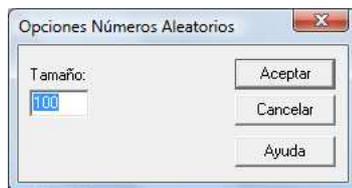


Figura 3.10: Números Aleatorios



- El tercer botón, opciones gráficas , se utiliza para seleccionar distintos gráficos de la distribución, pulsando sobre él tenemos ventana de la Figura 3.11 (pag. 9).



Figura 3.11: Opciones Gráficas

Los gráficos que se pueden obtener son:

- Función de Densidad (caso continuo) o Función Masa de Probabilidad (caso discreto).
- Función de Distribución (CDF).
- Función de Supervivencia.
- Función Log Supervivencia.
- Función de Riesgo.



- El cuarto botón, guardar los resultados , sirve para guardar números aleatorios de las diferentes variables que hayamos indicado en el análisis, pulsando sobre él obtenemos la ventana de la Figura 3.12 (pag. 10), donde deberemos marcar que números aleatorios queremos guardar y el nombre de la variable destino.

### 3.2.1. Algunos Ejemplos de Distribuciones de Probabilidad en Statgraphics

#### Representación gráfica de distribuciones

**Ejemplo:** Representar en un mismo gráfico las funciones de densidad de las siguientes distribuciones normales:  $N(0,1)$ ,  $N(0,3)$  y  $N(8,1)$ .

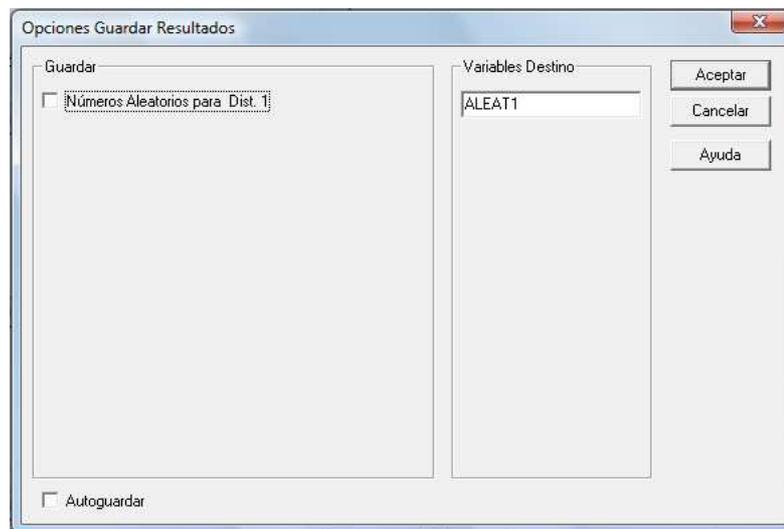


Figura 3.12: Guardar Resultados

### Solución:

Para ello seleccionamos en el menú principal **Gráficos/Distribuciones de Probabilidad...** (Ver Figura 3.2, pag. 5) y en la ventana de las distribuciones (Figura 3.3, pag. 6) dejamos seleccionada la distribución normal y pulsamos **Aceptar**.

Pulsamos con el botón derecho del ratón sobre cualquiera de las ventanas de este análisis de Distribuciones de Probabilidad y seleccionamos **Opciones de Análisis...**, así se muestra la ventana de la Figura 3.13 (pag. 10), en esta ventana pondremos los parámetros de las tres variables que vamos a representar (Figura 3.14, pag. 11) y pulsamos **Aceptar**, en la ventana superior derecha obtenemos el gráfico (Figura 3.15, pag. 11).

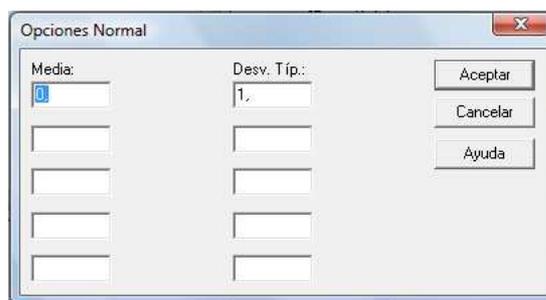


Figura 3.13: Media y desviación típica

### Obtención de probabilidades y Percentiles

**Ejemplo 1.** Una compañía de seguros descubre que hay un 5% de la población que tiene un cierto tipo de accidente cada año. Si se seleccionan 8 asegurados al azar en la población

- ¿Cuál es la probabilidad de que no más de dos de ellos tengan un accidente de este tipo el próximo año?

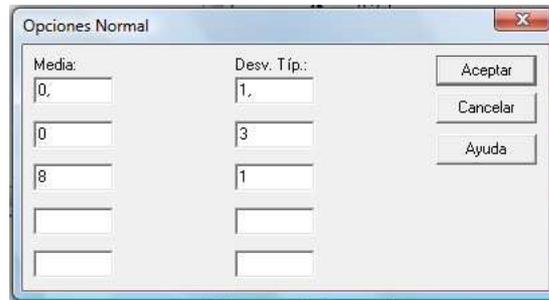


Figura 3.14: Media y desviación típica

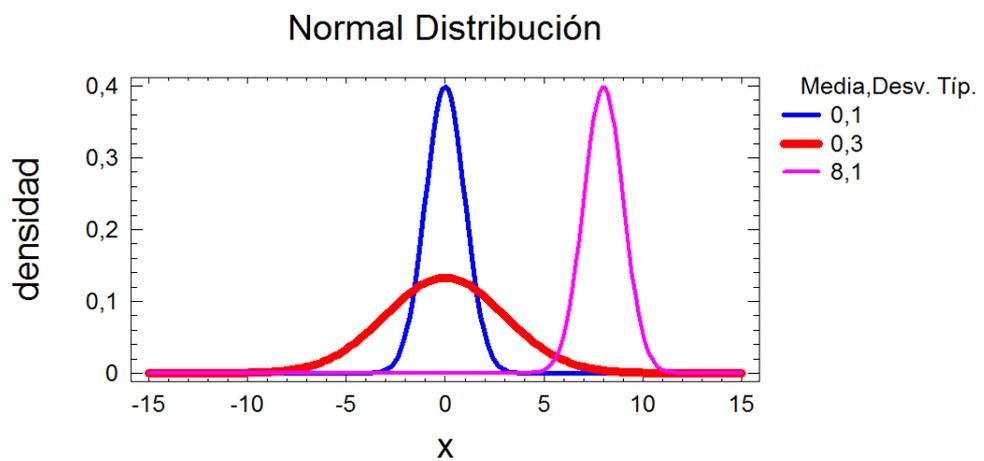


Figura 3.15: Gráfico las funciones de densidad

- b) ¿Cuál es la probabilidad de que no haya accidentes?
- c) ¿Cuál es la probabilidad de que haya más de tres accidentes?

**Solución:**

Definimos  $X \equiv$  número de asegurados que tienen un accidente de ese tipo en el próximo año de entre los 8 seleccionados, esta variable sigue una distribución Binomial con  $n = 8$  y  $p = 0.05$ ,  $X \sim B(8, 0.05)$ . Hay que calcular:

- a)  $P[X \leq 2] = P[X < 3]$
- b)  $P[X = 0]$
- c)  $P[X > 3]$

Para ello seleccionamos en el menú principal **Gráficos/Distribuciones de Probabilidad...** (Ver Figura 3.2, pag. 5) y en la ventana de las distribuciones (Figura 3.3, pag. 6) seleccionamos la distribución Binomial y pulsamos **Aceptar**.

Pulsamos con el botón derecho sobre cualquier ventana del análisis y seleccionamos **Opciones de Análisis...** En la ventana que aparece habrá que indicar la **Probabilidad del Evento**,  $p = 0.05$ , y el número de **Ensayos**,  $n = 8$ , tal y como se muestra en la Figura 3.16 (pag. 12) y pulsamos **Aceptar**.

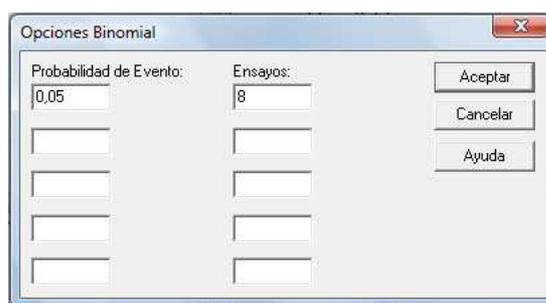


Figura 3.16: Opciones Binomial

Pulsaremos con el botón derecho del ratón sobre la ventana de **Distribución Acumulativa** y seleccionaremos **Opciones de Ventana** (ver Figura 3.17, pag. 13), y en la ventana que sale tendremos que indicar los **Valores para la Variable** que en nuestro caso son 0 y 3 (Ver Figura 3.18, pag. 13) y pulsamos **Aceptar**.

Miramos los resultados en la ventana de **Distribución Acumulativa** (ver Figura 3.19, pag. 14), de donde sacamos que:

- a)  $P[X \leq 2] = P[X < 3] = 0,994212$
- b)  $P[X = 0] = 0,66342$
- c)  $P[X > 3] = 0,000371695$

**Ejemplo 2.** Obtener los percentiles  $P_5$ ,  $P_{30}$ ,  $P_{80}$  y  $P_{95}$  de una variable con distribución  $\chi_7^2$ .

**Solución:**

Entramos en el menú **Gráficos/Distribuciones de Probabilidad...**(Figura 3.2, pag. 5). En la ventana de distribuciones seleccionamos la distribución **Chi-Cuadrado** (Figura 3.3, pag. 6) y pulsamos **Aceptar**.

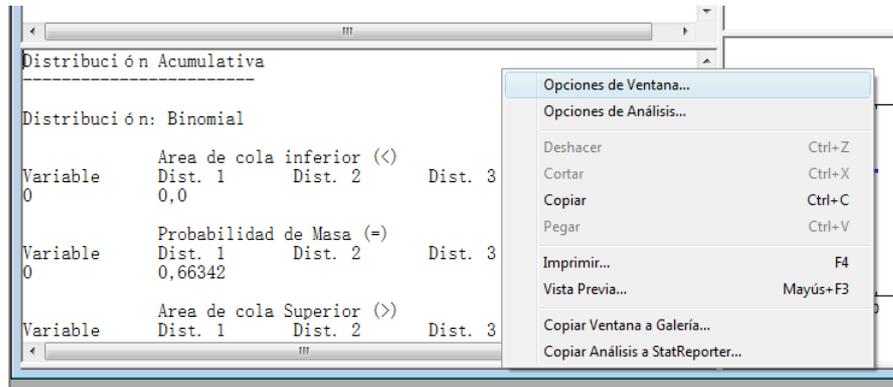


Figura 3.17: Ventana de Distribución Acumulativa

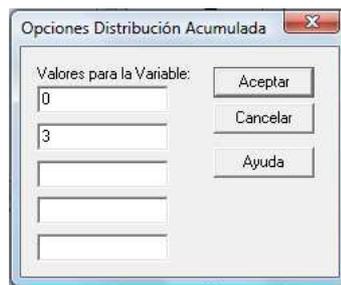


Figura 3.18: Valores para la Variable

Figura 3.19: Resultados

Distribución Acumulativa					
-----					
Distribución: Binomial					
	Area de cola inferior (<)				
Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,0				
3	0,994212				
	Probabilidad de Masa (=)				
Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,66342				
3	0,00541647				
	Area de cola Superior (>)				
Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,33658				
3	0,000371695				
El StatAdvisor					
-----					
<p>Este cuadro evalúa la distribución binomial acumulativa. Calculará las áreas de cola para hasta 5 valores críticos de la distribución. También calculará la función de densidad o función de masa. Por ejemplo, el resultado indica que, para la primera distribución especificada, la probabilidad de obtener un valor menor que 0,0 es 0,0. También, la probabilidad de obtener un valor mayor que 0,0 es 0,33658. La probabilidad de obtener un valor exactamente igual a 0,0 es 0,66342.</p>					
<					

Pulsamos con el botón derecho sobre cualquier ventana del análisis y seleccionamos **Opciones de Análisis....** En la ventana que aparece habrá que indicar los **Grados de Libertad**, tal y como se muestra en la Figura 3.20 (pag. 15) y pulsamos **Aceptar**.

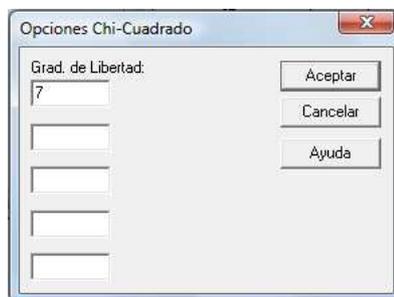


Figura 3.20: Opciones Chi-Cuadrado



Pulsamos el botón de **Opciones tabulares**, seleccionamos **CDF Inverso** (Figura 3.21, pag. 15) y pulsamos **Aceptar**.



Figura 3.21: Opciones tabulares: Marcar CDF Inverso

Pulsaremos con el botón derecho del ratón sobre la ventana de **Inversa CDF** y seleccionaremos **Opciones de Ventana** (ver Figura 3.17, pag. 13), y en la ventana que sale tendremos que indicar el valor de  $p$  para obtener un valor  $a$  tal que  $P[X \leq a]$ , en nuestro caso como tenemos que calcular los percentiles  $P_5$ ,  $P_{30}$ ,  $P_{80}$  y  $P_{95}$  tendremos que introducir 0.05, 0.30, 0.80 y 0.95 respectivamente (Ver Figura 3.23, pag. 16) y pulsamos **Aceptar**.

Miramos los resultados en la ventana de **Inversa CDF** (ver Figura 3.24, pag. 16), de donde sacamos que:

- $P_5 = 2,16735$
- $P_{30} = 4,67133$
- $P_{80} = 9,80325$
- $P_{95} = 14,0671$

### Generación de números aleatorios

**Ejemplo.** Generar una muestra de 500 valores de una distribución de Poisson de media 3.

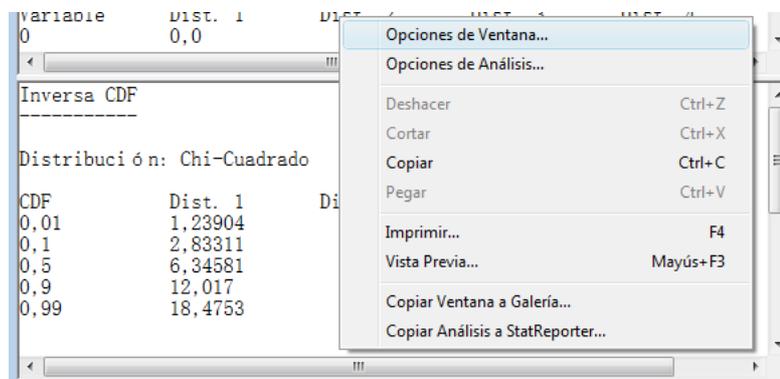


Figura 3.22: Ventana Inverse CDF

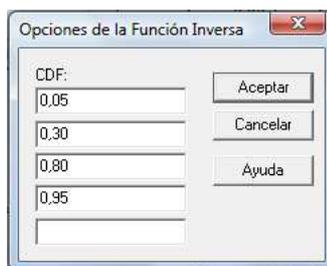


Figura 3.23: Opciones de Inverse CDF

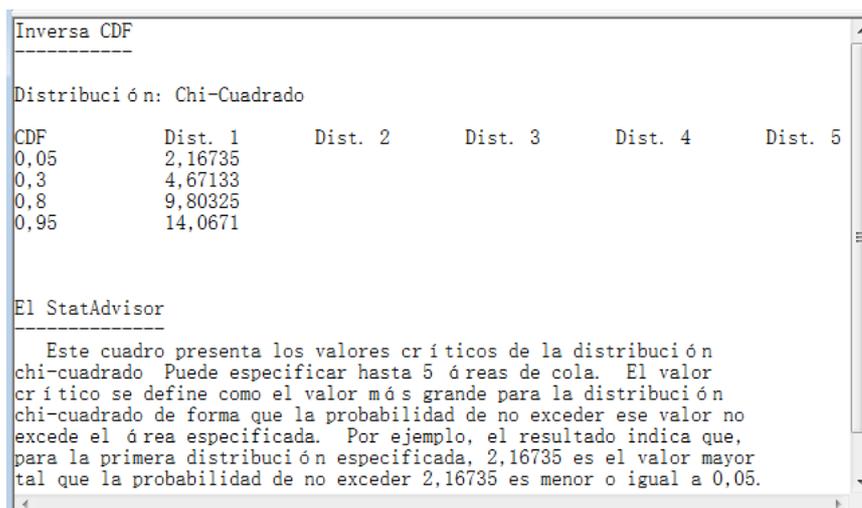


Figura 3.24: Resultados

**Solución:**

Entramos en el menú **Gráficos/Distribuciones de Probabilidad...**(Figura 3.2, pag. 5). En la ventana de distribuciones seleccionamos la distribución de **Poisson** (Figura 3.3, pag. 6) y pulsamos **Aceptar**. Pulsamos con el botón derecho sobre cualquier ventana del análisis y seleccionamos **Opciones de Análisis...** En la ventana que se obtiene habrá que indicar la **Media**,  $\lambda = 3$ , tal y como se muestra en la Figura 3.25 (pag. 17) y pulsamos **Aceptar**.

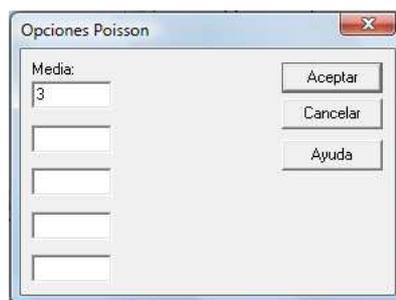


Figura 3.25: Opciones Poisson

Ahora pulsamos el botón de **Opciones tabulares**, , seleccionamos **Números Aleatorios** (Figura 3.26, pag. 17) y pulsamos **Aceptar**.

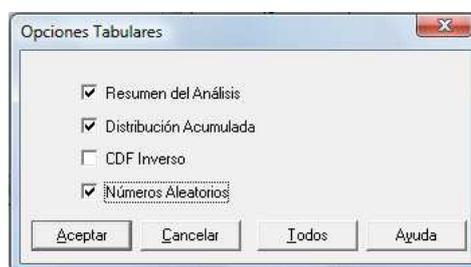


Figura 3.26: Opciones tabulares: Marcar Números Aleatorios

Pulsaremos con el botón derecho del ratón sobre la ventana de **Números Aleatorios** y seleccionaremos **Opciones de Ventana** (ver Figura 3.27, pag. 18), y en la ventana que se obtiene tendremos que indicar el **Tamaño** de la muestra que vamos a generar, cambiando el número 100 que hay por defecto por 500 (Ver Figura 3.28, pag. 18) y pulsamos **Aceptar**.

Finalmente pulsaremos sobre el botón **Guardar resultados**, , donde en el recuadro **Guardar** marcaremos la opción **Números Aleatorios para Dist. 1** y en el recuadro **Variables destino** indicaremos el nombre que se asignará a la muestra, por ejemplo **Poisson**, como se muestra en la Figura 3.29 (pag. 18).

Pulsamos **Aceptar** y en la ventana de datos se crea una columna llamada **Poisson** que contiene la muestra aleatoria (Figura 3.30, pag. 19).

Figura 3.27: Ventana Números Aleatorios

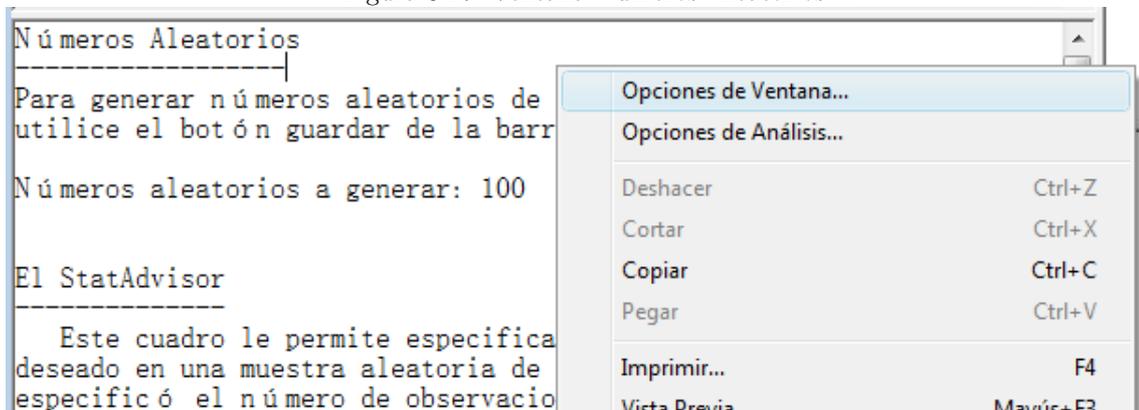


Figura 3.28: Opciones Números Aleatorios

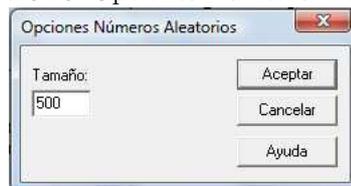


Figura 3.29: Guardar Números Alatorios

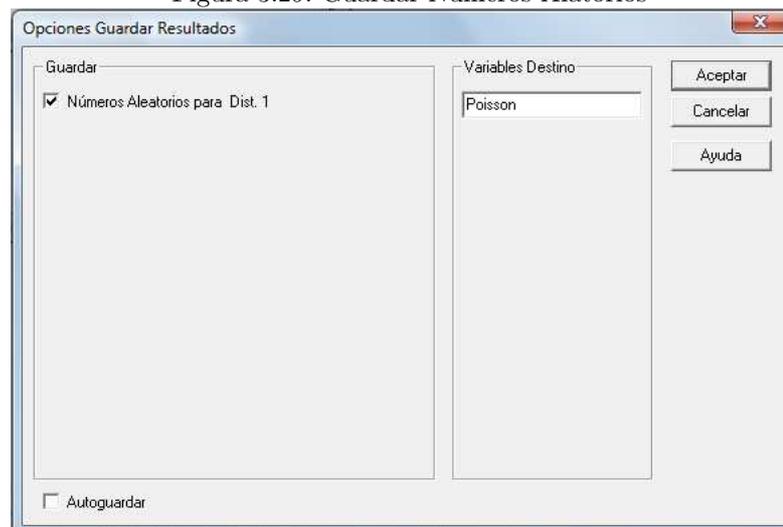


Figura 3.30: Números Aleatorios

<Sin Nombre>	
	Poisson
1	4
2	9
3	4
4	8
5	4
6	6
7	1
8	2
9	2
10	2
11	5
12	1
13	3
14	3
15	3
16	4
17	4
18	4
19	4
20	7

### 3.2.2. Gráficos de normalidad

Los gráficos de normalidad permiten comprobar rápidamente si los datos que estamos estudiando proceden de una distribución normal. Veamos con un ejemplo como se obtienen estos gráficos y cual es la interpretación del resultado.

**Ejemplo:** Supongamos que deseamos realizar un estudio estadístico de una determinada población, para ello obtenemos una muestra aleatoria de tamaño 10 cuyos resultados fueron los siguientes:

5,1 6,7 3,9 5,3 5,3 4,1 4,5 6,2 6,6 6,5

estudiar mediante un gráfico de normalidad si pueden proceder de una población normal.

**Solución:**

Crearemos una variable, que por ejemplo llamaremos **Muestra** y de **Tipo Numérica** (Figura 3.31, pag. 20), e introducimos los datos (Figura 3.32, pag. 20)

Para obtener un gráfico de normalidad deberemos entrar en el menú **Gráficos/Gráficos Exploratorios/Gráfico Probabilístico...** (Figura 3.33, pag. 21)

Obtendremos la ventana de la Figura 3.34 (pag. 21) y pasaremos la variable **Muestra** al campo de **Datos**, tal como se observa en la Figura 3.35 (pag. 21) y pulsaremos **Aceptar**.

El resultado será el gráfico de la Figura 3.36 (pag. 22)

Hemos obtenido un gráfico con una recta que representa la distribución normal y unos puntos que representan las observaciones de la muestra, si estos puntos están próximos a la recta indicará que las observaciones proceden de una población normal, tal y como ocurre en este ejemplo.

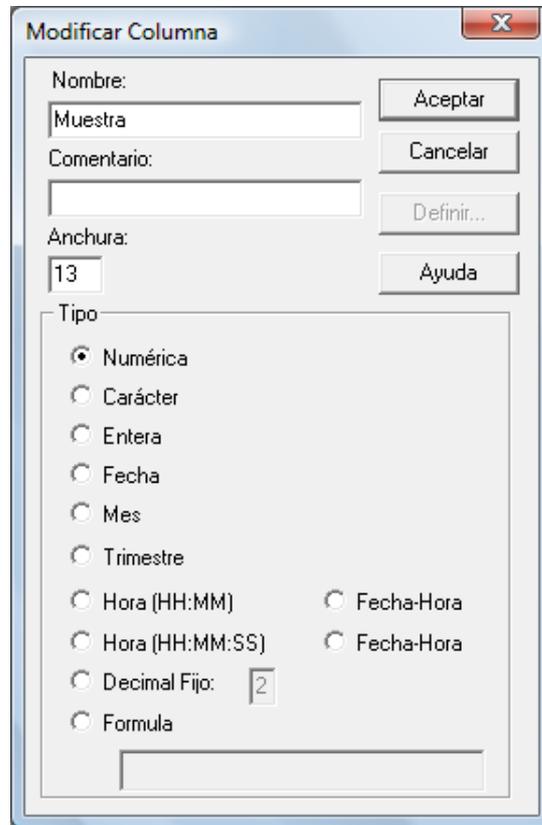


Figura 3.31: Definir Variable

	<b>Muestra</b>
1	5,1
2	6,7
3	3,9
4	5,3
5	5,3
6	4,1
7	4,5
8	6,2
9	6,6
10	6,5
11	

Figura 3.32: Introducir Datos

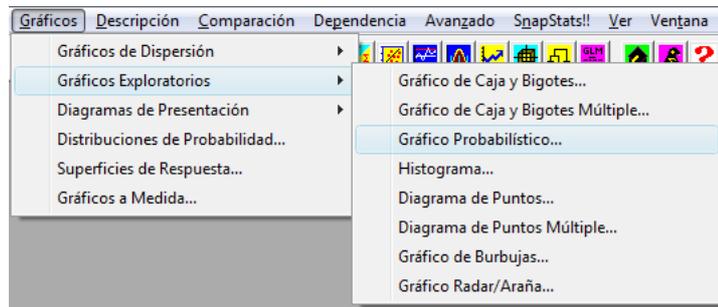


Figura 3.33: Menú del Gráfico de Normalidad



Figura 3.34: Grafico de Probabilidad: Entrada de datos



Figura 3.35: Datos: Muestra

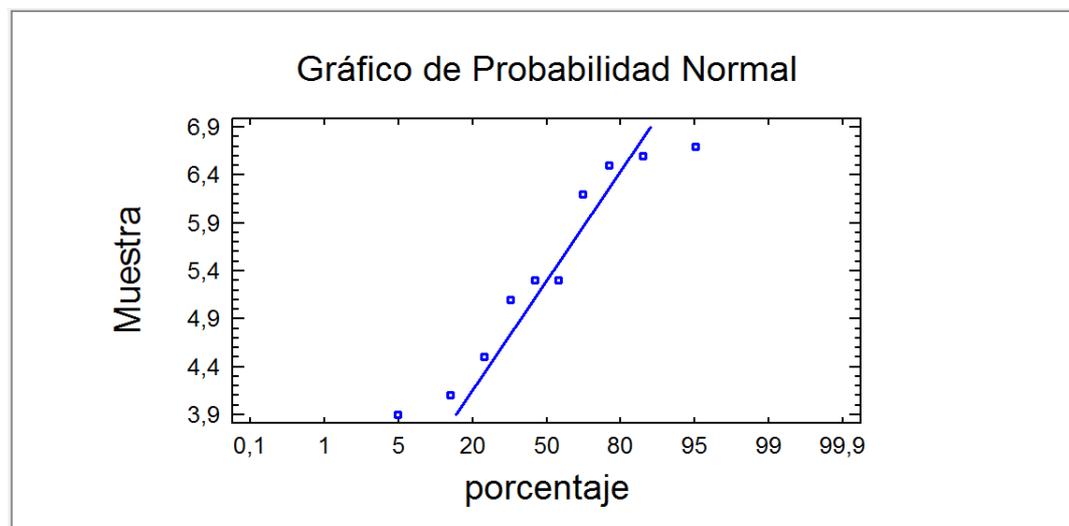


Figura 3.36: Grafico de Normalidad

### 3.3. Estimación y Contraste de Hipótesis

La Inferencia Estadística pretende obtener conclusiones, a partir de una muestra, sobre alguna característica de la población. Supongamos que estamos interesados en conocer un determinado parámetro de la población, dependiendo del enfoque que le demos, podemos clasificar la Inferencia Estadística de la siguiente forma

- **Estimación:** Consiste en buscar un valor o un intervalo de valores que aproxime el verdadero valor del parámetro, podemos distinguir dos tipos
  - **Estimación puntual:** Se busca un único valor a partir de la muestra que aproxime al verdadero valor del parámetro.
  - **Estimación por intervalos de confianza:** Se busca un intervalo de valores donde el parámetro estará con una determinada probabilidad, a esta probabilidad se le llama nivel de confianza  $(1 - \alpha)$ , donde  $\alpha$  es un valor entre 0 y 1 que se llama nivel de significación.
- **Contraste de hipótesis:** Consiste en establecer una hipótesis acerca del valor del parámetro (llamada hipótesis nula  $H_0$ ) y contrastar si eso es posible, frente a otra hipótesis alternativa ( $H_1$ ). Para ello se define un estadístico, cuyo valor en la muestra se denomina valor experimental, y se compara con un valor teórico, que se suele determinar a partir de alguna de las distribuciones definidas anteriormente (Normal,  $t$  de Student, chi-cuadrado y  $F$  de Snedecor). En un contraste de hipótesis se denomina nivel de significación ( $\alpha$ ) a la probabilidad de rechazar  $H_0$  cuando ésta es cierta, así deberemos elegir éste valor próximo a 0. Generalmente, los paquetes estadísticos como Statgraphics, obtienen lo que se denomina  $p$ -valor, que representa el mínimo valor de  $\alpha$  para el cual se debería rechazar la hipótesis nula, así si el  $p$ -valor que se obtiene es menor que el nivel de significación  $\alpha$  elegido para realizar el contraste, se rechazará la hipótesis nula.

#### Inferencia sobre la media $\mu$ de una población normal

Supongamos que estamos estudiando una población normal de media  $\mu$  y varianza  $\sigma^2$  (ambas desconocidas),  $X \sim N(\mu, \sigma)$ , y seleccionamos una muestra cuyos valores observados son  $x_1, \dots, x_n$  con el fin de estimar el parámetro  $\mu$ .

Estimación puntual:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Intervalo de confianza:

$$\left( \bar{x} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \right)$$

Contrastes de hipótesis:

Contraste	Estadístico	Rechazo $H_0$
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$t_{exp} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$	Si $t_{exp} > t_{n-1, \alpha/2}$
$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$	$t_{exp} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$	Si $t_{exp} > t_{n-1, \alpha}$
$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$t_{exp} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$	Si $t_{exp} < -t_{n-1, \alpha}$

### Inferencia sobre la varianza $\sigma^2$ de una población normal

Supongamos que estamos estudiando una población normal de media  $\mu$  y varianza  $\sigma^2$  (ambas desconocidas),  $X \sim N(\mu, \sigma)$ , y seleccionamos una muestra cuyos valores observados son  $x_1, \dots, x_n$  con el fin de estimar el parámetro  $\sigma^2$ .

Estimación puntual:

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Intervalo de confianza:

$$\left( \frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2}, \frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2} \right)$$

Contraste de hipótesis:

Contraste	Estadístico	Rechazo $H_0$
$\left. \begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{array} \right\}$	$\chi_{exp}^2 = \frac{(n-1)S^2}{\sigma_0^2}$	Si $\chi_{exp}^2 > \chi_{n-1, \alpha/2}^2$ ó $\chi_{exp}^2 < \chi_{n-1, 1-\alpha/2}^2$
$\left. \begin{array}{l} H_0 : \sigma^2 \leq \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{array} \right\}$	$\chi_{exp}^2 = \frac{(n-1)S^2}{\sigma_0^2}$	Si $\chi_{exp}^2 > \chi_{n-1, \alpha}^2$
$\left. \begin{array}{l} H_0 : \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{array} \right\}$	$\chi_{exp}^2 = \frac{(n-1)S^2}{\sigma_0^2}$	Si $\chi_{exp}^2 < \chi_{n-1, 1-\alpha}^2$

### Inferencia sobre una proporción $p$

Supongamos que estamos estudiando una determinada característica ( $A$ ) en una población, los individuos por lo tanto se pueden clasificar en dos clases, los que cumplan esa característica ( $A$ ) y los que no la cumplen ( $\bar{A}$ ). Representamos por  $p$  la proporción de individuos de la población que cumplen la característica. Con el fin de estimar esta proporción  $p$  tomamos una muestra (de tamaño suficientemente grande,  $n \geq 30$ ) y representaremos por  $u_p$  la proporción de individuos en la muestra que cumplen la característica  $A$ .

Estimación puntual:

$$\hat{p} = u_p$$

Intervalo de confianza:

$$\left( \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Contraste de hipótesis:

Contraste	Estadístico	Rechazo $H_0$
$\left. \begin{array}{l} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{array} \right\}$	$z_{exp} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	Si $z_{exp} > z_{\alpha/2}$
$\left. \begin{array}{l} H_0 : p \leq p_0 \\ H_1 : p > p_0 \end{array} \right\}$	$z_{exp} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	Si $z_{exp} > z_{\alpha}$
$\left. \begin{array}{l} H_0 : p \geq p_0 \\ H_1 : p < p_0 \end{array} \right\}$	$z_{exp} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	Si $z_{exp} < -z_{\alpha}$

### 3.4. Inferencia de una muestra en Statgraphics

Vamos a distinguir dos casos: cuando partimos de los estadísticos muestrales y cuando tenemos los datos muestrales.

### 3.4.1. Cuando partimos de los estadísticos muestrales

Entramos en el menú **Descripción/Contraste de Hipótesis...** (Figura 3.37, pag. 25) y obtenemos la ventana de la Figura 3.38 (pag. 25).

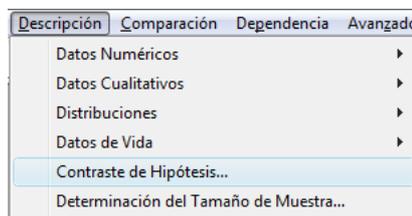


Figura 3.37: Menú Contraste de Hipótesis

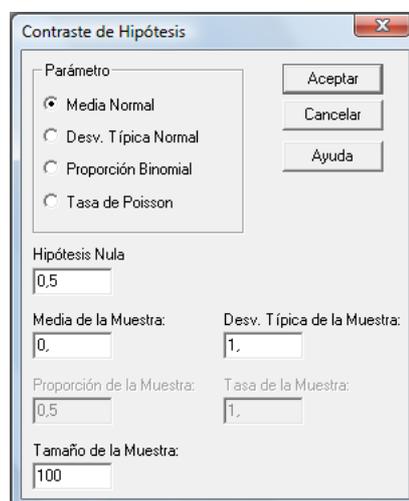


Figura 3.38: Contraste de Hipótesis. Media Normal

En esta ventana deberemos indicar el parámetro sobre el que deseamos hacer el análisis, de los cuales nos interesa:

- **Media de una Normal:** En este caso habrá que indicar el valor  $\mu_0$  para realizar el contraste (Hipótesis Nula,  $H_0$ ), media muestral, cuasidesviación típica muestral y tamaño muestral (Figura 3.38, pag. 25).
- **Desviación Típica de una Normal:** Tendremos que indicar el valor  $\sigma_0$  del contraste (Hipótesis Nula,  $H_0$ ), cuasidesviación típica muestral y tamaño muestral (Figura 3.39, pag. 26).
- **Proporción Binomial:** Donde indicaremos  $p_0$  para el contraste de hipótesis (Hipótesis Nula,  $H_0$ ), proporción muestral y tamaño muestral (Figura 3.40, pag. 26).

Pulsamos **Aceptar** y obtendremos los resultados en una ventana similar a la que muestra la Figura 3.41 (pag. 27).

Finalmente, pulsando con el botón derecho del ratón sobre la ventana de resultados y seleccionamos **Opciones de Análisis...** podremos indicar el nivel  $\alpha$  al que realizamos la inferencia y el sentido de la

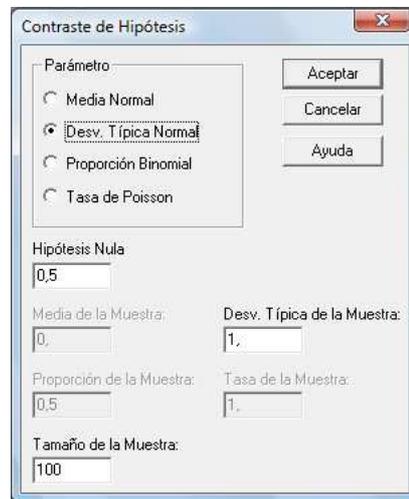


Figura 3.39: Contraste de Hipótesis. Desv. Típica Normal

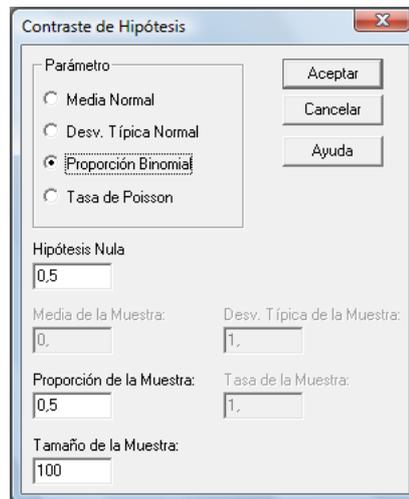


Figura 3.40: Contraste de Hipótesis. Proporción

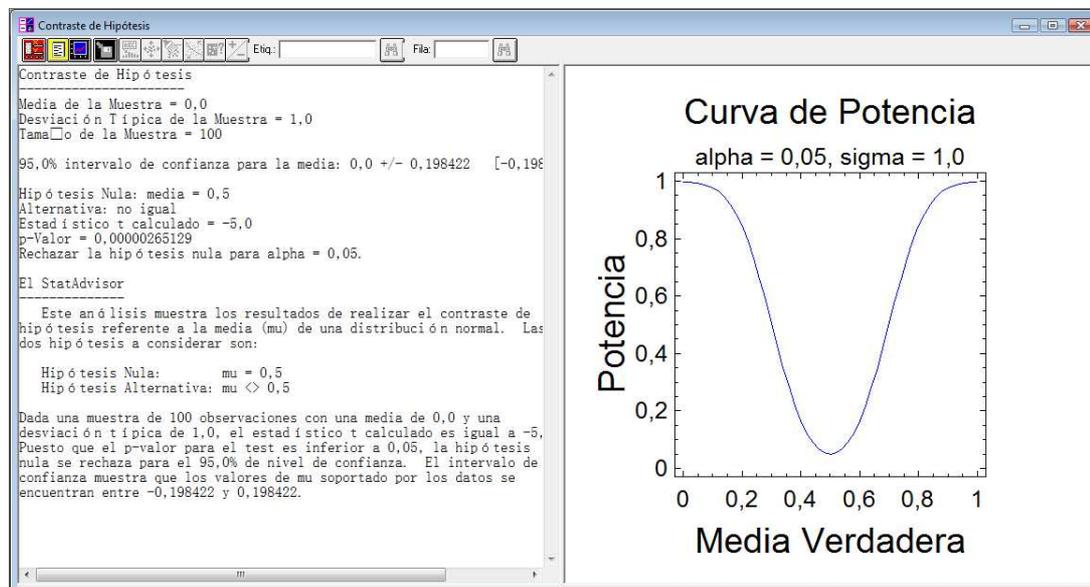


Figura 3.41: Contraste de Hipótesis

Hipótesis Alternativa ( $H_1$ ) del contraste: No Igual ( $\neq$ ), Menor que ( $<$ ) y Mayor que ( $>$ ), Figura 3.42 de la página 27.

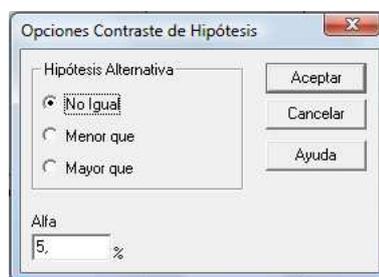


Figura 3.42: Opciones del Contraste de Hipótesis

### 3.4.2. Cuando tenemos los datos muestrales

Una vez introducidos los datos, entraremos en el menú **Descripción/Datos Numéricos/Análisis Unidimensional...** (Figura 3.43, pág. 28)

En la ventana que se obtiene se selecciona la variable, se pasa al campo **Datos** pulsando sobre el triángulo, tal como muestra la Figura 3.44 (pág. 28), y se pulsa **Aceptar**.

Se obtendrá una ventana similar a la que tenemos en la Figura 3.45 (pág. 28)

En la ventana superior izquierda muestra el número de observación de la variable y en la ventana inferior izquierda tenemos un resumen estadístico de la variable, entre los que podemos encontrar la media muestral y la cuasidesviación típica muestral, así con estos datos podríamos realizar el estudio utilizando lo explicado en el apartado anterior o bien proceder de la forma que se expone a continuación.

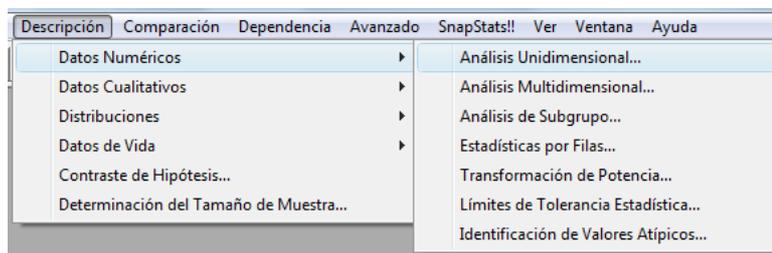


Figura 3.43: Menú Análisis Unidimensional



Figura 3.44: Seleccionar variable de los datos

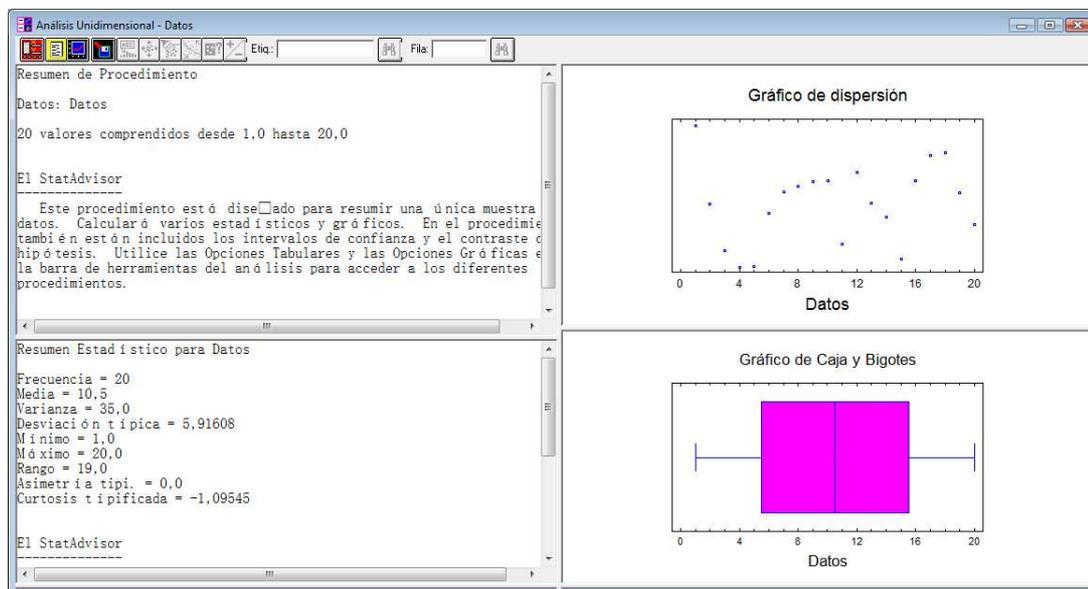


Figura 3.45: Resultado del Análisis Unidimensional



Pulsaremos sobre el botón de Opciones tabulares, , y marcaremos las opciones **Intervalos de Confianza** y de **Contraste de Hipótesis** (Figura 3.46, pág. 29), al pulsar **Aceptar** tendremos las dos ventanas correspondientes.



Figura 3.46: Opciones Tabulares

En la ventana de **Intervalos de Confianza** obtenemos los intervalos para la media y para la desviación típica a un nivel del 95 %, para modificar este nivel de confianza pulsaremos con el botón derecho sobre la ventana y seleccionaremos **Opciones de Ventana...**, obteniendo la ventana de la Figura 3.47 (pág. 29)

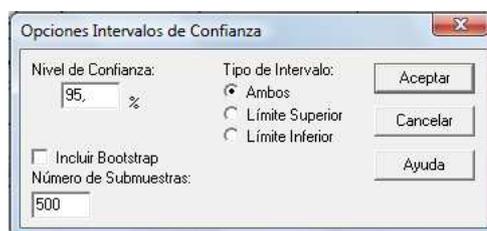


Figura 3.47: Opciones de Intervalos de Confianza

Sobre la ventana de **Contastes de Hipótesis** pulsaremos con el botón derecho del ratón y seleccionamos **Opciones de Análisis...** para indicar el valor  $\mu_0$  para realizar el contraste sobre la media (Hipótesis Nula,  $H_0$ ), el nivel de significación  $\alpha$  y el sentido de la Hipótesis Alternativa ( $H_1$ ) del contraste: No Igual ( $\neq$ ), Menor que ( $<$ ) y Mayor que ( $>$ ), Figura 3.48 de la página 30.

### 3.4.3. Ejemplos sobre Estimación y Contraste de Hipótesis

**Ejemplo 1** De una población normal se selecciona una muestra aleatoria de tamaño 10 cuyos valores son los siguientes:

5,1 6,7 3,9 5,3 5,3 4,1 4,5 6,2 6,6 6,5

Obtener:

- Obtener estimaciones puntuales para la media y la varianza.
- Obtener un intervalo de confianza para la media a un nivel del 99 %.

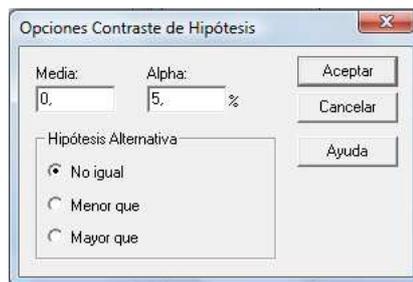


Figura 3.48: Opciones de Contraste de Hipótesis

- c) Obtener un intervalo de confianza para la varianza a un nivel del 98 %.
- d) Realizar el siguiente contraste a un nivel de significación del 10 %.

$$\begin{cases} H_0 : \mu \geq 7 \\ H_1 : \mu < 7 \end{cases}$$

- e) Realizar el siguiente contraste a un nivel de significación del 3 %.

$$\begin{cases} H_0 : \sigma^2 = 4 \\ H_1 : \sigma^2 \neq 4 \end{cases}$$

**Solución:**

Crearemos una variable, que llamamos **Muestra** y de **Tipo Numérica** (Figura 3.49, pag. 30), e introducimos los datos (Figura 3.50, pag. 31)

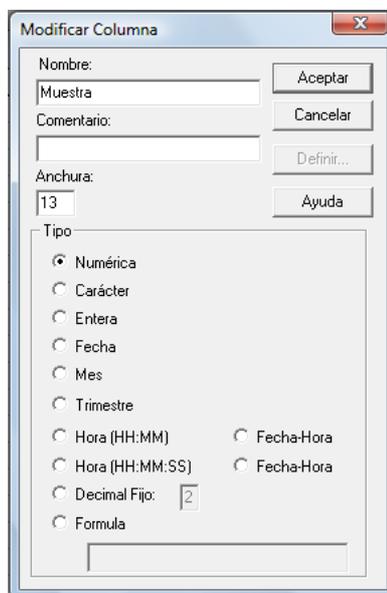


Figura 3.49: Definir Variable

	Muestra
1	5,1
2	6,7
3	3,9
4	5,3
5	5,3
6	4,1
7	4,5
8	6,2
9	6,6
10	6,5
11	

Figura 3.50: Introducir Datos

Seleccionamos en el menú principal **Descripción/Datos Numéricos/Análisis Unidimensional...** (Figura 3.51, pag. 31), pasamos la variable **Muestra** al campo de datos (Figura 3.52, pag. 31) y pulsamos **Aceptar**.

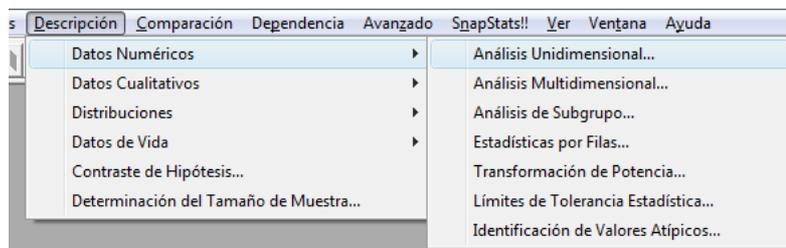


Figura 3.51: Menú Análisis Unidimensional



Figura 3.52: Seleccionar variable

En la ventana **Resumen Estadístico para Muestra** (Figura 3.53, pag. 32) podemos observar la media muestral (Media = 5,42) y la cuasivarianza muestral (Varianza = 1,09289).

Así las estimaciones puntuales para la media y para la varianza son:

$$\hat{\mu} = 5,42$$

$$\hat{\sigma}^2 = 1,09289$$

```
Resumen Estadístico para Muestra
Frecuencia = 10
Media = 5,42
Varianza = 1,09289
Desviación típica = 1,04541
Mínimo = 3,9
Máximo = 6,7
Rango = 2,8
Asimetría tipi. = -0,18083
Curtosis típificada = -0,978055
```

Figura 3.53: Resumen estadístico para la muestra

A continuación pinchamos en el botón de Opciones tabulares, , seleccionamos las opciones de **Intervalos de Confianza** y **Contraste de Hipótesis** (Figura 3.54, pag. 32) y pulsamos **Aceptar**.



Figura 3.54: Seleccionar Intervalos y Contrastes

Como los intervalos que salen son a un nivel de confianza del 95 % pulsamos sobre la ventana de **Intervalos de Confianza** con el botón derecho del ratón y entramos en **Opciones de Ventana...** para modificar el nivel, primero al 99 % (Figura 3.55, pag. 32) y después al 98 % (Figura 3.56, pag. 33).

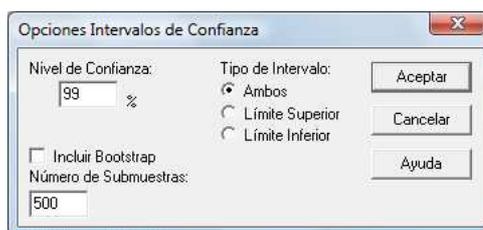


Figura 3.55: Nivel de confianza 99 %

Se obtiene como resultado las imágenes de las Figuras 3.57 (pag. 33) y 3.58 (pag. 33) respectivamente, de donde se tiene que:

- El I.C. para la media al 99 % es: [4,34564; 6,49436]
- El I.C. para la desviación típica al 98 % es: [0,673783; 2,17047]

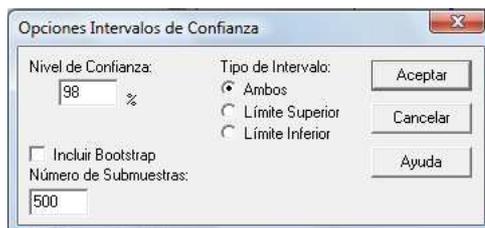


Figura 3.56: Nivel de confianza 98 %

- El I.C. para la varianza al 98 % es:

$$[(0,673783)^2; (2,17047)^2] \equiv [0,453983531089; 4,7109400209]$$

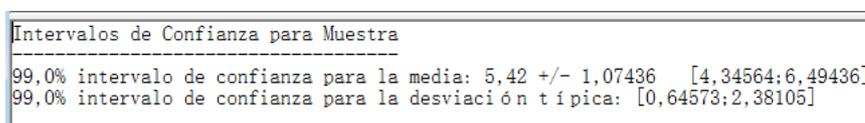


Figura 3.57: I.C. al 99 %

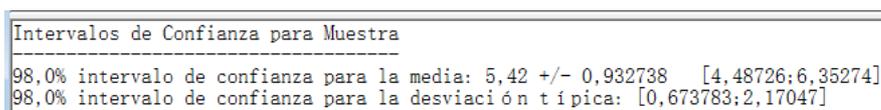


Figura 3.58: I.C. al 98 %

Para realizar el contraste sobre la media a un nivel de significación del 10 %

$$\begin{cases} H_0 : \mu \geq 7 \\ H_1 : \mu < 7 \end{cases}$$

pulsaremos con el botón derecho sobre la ventana del **Contraste de Hipótesis** y seleccionamos **Opciones de Ventana...** En esta ventana habrá que indicar el valor  $\mu_0 = 7$  en el campo **Media**, el nivel de significación  $\alpha$  expresado en % en el campo **Alpha**, en nuestro caso 10, y en el recuadro de **Hipótesis Alternativa** se marca **Menor que** (Figura 3.59, pag. 34).

Se obtienen los resultados de la Figura 3.60 (pag. 34). Tenemos que el valor del estadístico es  $t_{exp} = -4,77935$  con un  $p$ -valor de 0,00050109, como este  $p$ -valor es menor que  $\alpha = 0,1$  se tiene que se rechaza la hipótesis nula a ese nivel de significación.

Para realizar el contraste para la varianza a un nivel de significación del 3 %

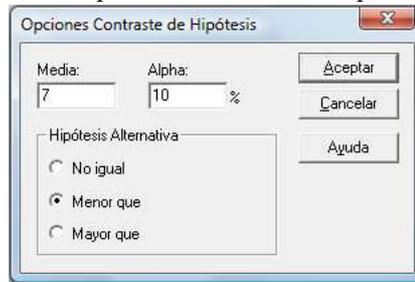
$$\begin{cases} H_0 : \sigma^2 = 4 \\ H_1 : \sigma^2 \neq 4 \end{cases}$$

se selecciona en el menú principal **Descripción/Contraste de Hipótesis...** (Figura 3.61, pag. 34).

La ventana que se obtiene se completa como indica la (Figura 3.62, pag. 35), es decir:

- Se marca como parámetro la opción de **Desv. Típica Normal**.
- Se indica el valor  $\sigma_0 = \sqrt{4} = 2$  en el campo **Hipótesis Nula**.

Figura 3.59: Opciones del Contraste para la media



```
contraste t
-----
Hipótesis nula: media = 7,0
Alternativa: menor que
Estadístico t = -4,77935
P-valor = 0,00050109
Se rechaza la hipótesis nula para alpha = 0,1.
```

Figura 3.60: Resultados del contraste para la media

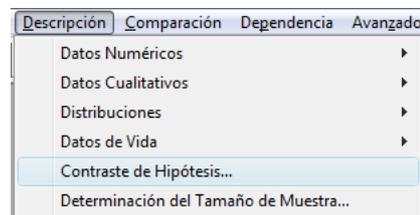


Figura 3.61: Menú Contraste

- En el campo **Desv. Típica de la Muestra** se pone el valor de la cuasidesviación típica muestral obtenida anteriormente (Figura 3.53, pag. 32), **Desviación típica = 1,04541**.
- Como **Tamaño de la Muestra** pondremos el valor **10**.

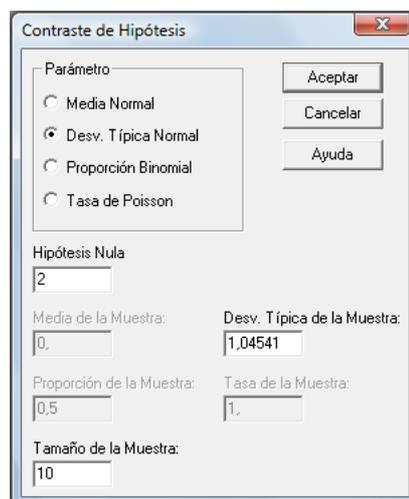


Figura 3.62: Contraste para la varianza

Para indicar el nivel de significación pulsamos con el botón derecho sobre la ventana de **Contraste de Hipótesis** y seleccionamos **Opciones del Análisis...** En el campo correspondiente (**Alpha**) se pone el nivel de significación en %, en nuestro caso 3 (Figura 3.63, pag. 35), y se pulsa **Aceptar** para obtener el resultado.

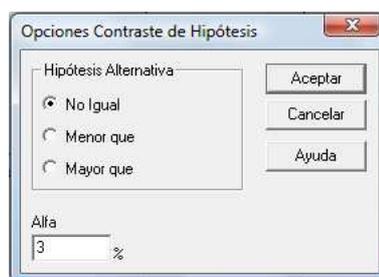


Figura 3.63: Opciones del Contraste para la varianza

Se obtiene la imagen de la Figura 3.64 (pag. 36), de donde sacamos los siguientes resultados:

$\chi_{exp}^2 = 2,45898$
$p - valor = 0,0360675$
$\alpha = 0,03$

Como el p-valor es mayor que  $\alpha$  se tiene que el resultado es **No rechazar la hipótesis nula** para ese nivel de significación.

```

Contraste de Hipótesis
-----
Desviación Típica de la Muestra = 1,04541
Tamaño de la Muestra = 10

97,0% intervalo de confianza para sigma: [0,692466;2,05247]

Hipótesis Nula: desviación típica = 2,0
Alternativa: no igual
Estadístico Chi-cuadrado calculado = 2,45898
p-Valor = 0,0360675
No rechazar la hipótesis nula para alpha = 0,03.

El StatAdvisor
-----
Este análisis muestra los resultados de realizar el contraste de
hipótesis referente a la desviación típica (sigma) de una distribución
normal. Las dos hipótesis a considerar son:

Hipótesis Nula:      sigma = 2,0
Hipótesis Alternativa: sigma > 2,0

Dada una muestra de 10 observaciones con una desviación típica de
1,04541, el estadístico chi-cuadrado calculado es igual a 2,45898.
Puesto que el p-valor para el test es superior o igual a 0,03, la
hipótesis nula no puede rechazarse para el 97,0% de nivel de
confianza. El intervalo de confianza muestra que los valores de sigma
soportado por los datos se encuentran entre 0,692466 y 2,05247.
    
```

Figura 3.64: Resultados del Contraste para la varianza

**Ejemplo 2** En una ciudad se desea estimar la proporción de individuos que han votado a un determinado partido político, para ello se selecciona una muestra de 1000 votantes, de los cuales 400 han votado a dicho partido. Obtener un intervalo de confianza del 95 % para la proporción de votantes de la ciudad que han votado a ese partido. Realizar el siguiente test a un nivel de significación del 5 %:

$$\begin{cases} H_0 : p \geq 0,5 \\ H_1 : p < 0,5 \end{cases}$$

**Solución:**

La proporción muestral se calcula dividiendo 400 entre 1000, así

$$\hat{p} = \frac{400}{1000} = 0,4.$$

A continuación entramos en el menú principal **Descripción/Contraste de Hipótesis...** (Figura 3.65, pag. 36).

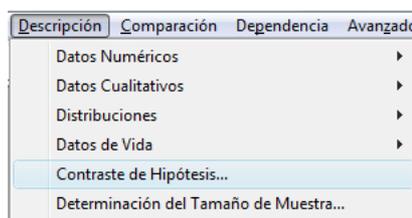


Figura 3.65: Menú Contraste de Hipótesis

La ventana que se obtiene se completa como indica la (Figura 3.66, pag. 37), es decir:

- Se marca como parámetro la opción de **Proporción Binomial**.
- Se indica el valor  $p_0 = 0,5$  en el campo **Hipótesis Nula**.
- En el campo **Proporción de la Muestra** se pone el valor de  $u_p = \hat{p} = 0,4$ .
- Como **Tamaño de la Muestra** pondremos el valor **1000**.

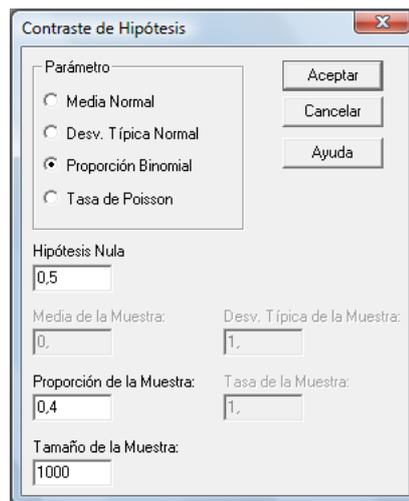


Figura 3.66: Contraste de hipótesis para una proporción

El nivel de confianza es 0,95 por lo que no hay que modificar nada, así mirando la ventana de resultados (Figura 3.67, pag. 37) tenemos que el intervalo de confianza para  $p$  es:  $[0,36799; 0,429933]$ .

```
Contraste de Hipótesis
-----
Proporción de la Muestra = 0,4
Tamaño de la Muestra = 1000

Aproximado 95,0% intervalo de confianza para p: [0,36799;0,429933]
```

Figura 3.67: Resultado del Intervalo de confianza

Sin embargo para el contraste hay que indicar el sentido de la hipótesis alternativa en el contraste, para ello pulsamos con el botón derecho del ratón sobre la ventana de resultados del **Contraste de Hipótesis** y seleccionamos **Opciones de Análisis...**, donde marcaremos como **Hipótesis Alternativa** la opción de **Menor que** y pulsamos el botón **Aceptar** (Figura 3.68, pag. 38).

Los resultados se muestran en la Figura 3.69 (pag. 38), ya que el  $p$ -valor es menor que  $\alpha = 0,05$  se rechaza la hipótesis nula a ese nivel.

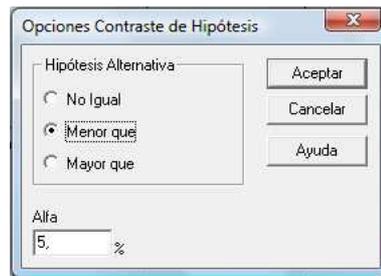


Figura 3.68: Seleccionar Menor que

```
Hipótesis Nula: proporción = 0,5  
Alternativa: menor que  
p-Valor = 1,56438E-10  
Rechazar la hipótesis nula para alpha = 0,05.
```

Figura 3.69: Resultado del Contraste de Hipótesis