Puedes no imprimir este archivo y consultarlo en formato digital, ahorrarás papel y tinta. Si decides imprimirlo, por favor hazlo en papel reciclado, a doble cara y con poca tinta. Sé ecológico. Muchas gracias.

Prácticas de Estadística con *Statgraphics*

Generación de números (pseudo)aleatorios

Ejercicio Varios1

Generar 4 muestras de 100 elementos, una de cada una de las siguientes distribuciones: Binomial, Uniforme, Normal y Poisson.

a) Cambiar los parámetros de las distribuciones y de las muestras.

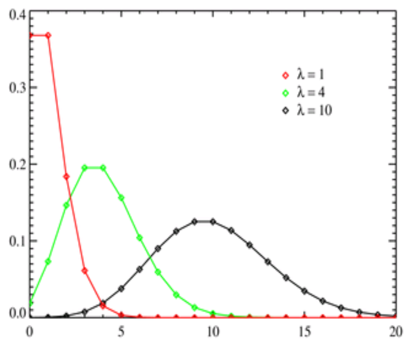
b) Analizar los gráficos de la función de densidad.

Objetivos

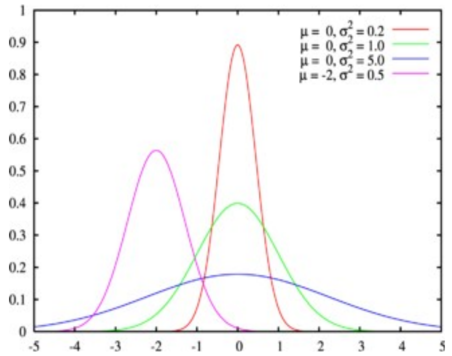
Aprender a generar números aleatorios de distintas distribuciones, para distintos valores de sus parámetros. Aprender a generar e interpretar los gráficos de las funciones de densidad, así como su dependencia del valor de los parámetros.

Teoría

http://en.wikipedia.org/wiki/Probability_distribution



Poisson probability mass function
(http://en.wikipedia.org/wiki/Poisson_distribution)



Normal density function
(http://en.wikipedia.org/wiki/Normal_distribution)

Menús

i) Para **generar una muestra aleatoria simple** de una distribución, tenemos que utilizar el menú:

Plot --> *Probability Distributions...* --> A continuación seleccionar la distribución deseada.

ii) Para **cambiar los parámetros de la distribución**:

Colocarse encima del resultado *Probability Distributions* que ha mostrado el programa en la mitad izquierda --> Pulsar el botón secundario del ratón --> Elegir *Analysis Options...* (Nota: Observad que se pueden generar de una vez muestras de distintas combinaciones de parámetros de esta misma distribución).

iii) Para c**ambiar el tamaño de la muestra**, que es 100 por defecto:

Pulsar el botón *Tabular options* (el segundo de la última fila) --> Marcar el resultado *Random Numbers* --> Pulsar el botón secundario del ratón encima del cuadro nuevo que aparece --> Elegir *Pane Options...*(Nota: También se pueden cambiar los parámetros de la distribución al pulsar aquí el botón secundario y *Analysis Options...*).

iv) En la mitad de la derecha aparece por defecto el gráfico *Density/Mass Function*. Pero recordemos que podemos **elegir los gráficos** que se muestran en esa parte desde el botón

Graphical options (tercer botón de la última fila).

v) Para **guardar una muestra** que hayamos generado:

Pulsar el botón *Save results* (cuarto botón de la última fila) --> Elegir qué variables queremos guardar y los nombres que les queremos poner (Si hemos generado de una vez muestras correspondientes a varias combinaciones de parámetros de una misma distribución, aparecerán también aquí esas opciones para ellas) --> Para ver la muestra tenemos que minimizar esta ventana principal y maximizar la hoja de cálculo en que *Statgraphics* presenta los datos.

Teorema del Límite Central: convergencia al modelo normal

<u>Ejercicio Binomial1</u>

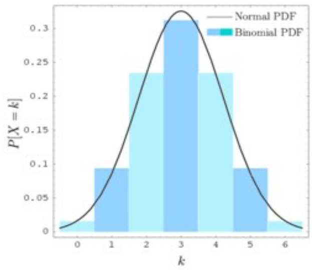
*Enunciar el Teorema del Límite Central y aplicarlo al caso de la distribución binomial Bin(*p = *0,3;* n*) para un* n *pequeño (*n = *10) y para un* n *grande (*n = *10000). Para ello, tómese una muestra de 100 observaciones.*

<u>Objetivos</u>

Comprobar empíricamente la convergencia a la distribución normal (Teorema del Límite Central). Aprender a transformar muestras (en este caso tipificándolas). Aprender a dibujar el histograma de una muestra.

<u>Teoría</u>

Una variable aleatoria *B(p,n)* tipificada tiende a la distribución *N(0,1)* cuando *n* crece. http://en.wikipedia.org/wiki/Central_Limit_Theorem



Binomial probability density function and normal approximation for $n = 6$ and $p = 0.5$.
(http://en.wikipedia.org/wiki/Binomial_distribution)

Nota: Es necesario distinguir entre el resultado teórico que dice que, cuando X_i se distribuyen según una <u>normal</u> $N(\mu,\sigma^2)$, el cociente $(\bar{X}-\mu)/\sqrt{\frac{\sigma^2}{n}}$ <u>sigue exactamente</u> una N(0, 1) y este teorema, en el que X_i pueden seguir <u>cualquier distribución</u> y el cociente, cuando se cumplen las hipótesis y *n* crece, sigue <u>aproximadamente</u> –tiende a– una N(0,1).

<u>Menús</u>

i) Para generar las muestras, elegir los parámetros y guardar estos cálculos:

Como en el ejercicio 1, generamos las muestras de la distribución binomial, en un caso para $n = 10$ y en el otro para $n = 100000$ (mejor que el 10000 que viene en el enunciado); recordad que se pueden generar las dos muestras a la vez. --> Guardamos las muestras, de nuevo como en el ejercicio 1.

Ahora tenemos que tipificar esas variables. Para ello generaremos una nueva columna por cada muestra; *Statgraphics* se encarga de aplicar la operación que le indicamos a cada elemento de las columnas que hemos guardado.

ii) Para **transformar las variables**:

En la hoja de los datos, pulsamos en la parte superior de una columna vacía, de forma que se selecciona entera --> Pulsar el botón secundario del ratón y elegir *Generate Data...* --> Introducir la expresión de la transformación, que es $(X - p*n)/sqrt(p*(1-p)*n)$, donde X es el nombre de la columna antigua, y p y n son los parámetros de la binomial.

Hay que hacer este paso anterior para cada una de las dos columnas guardadas, y hay que cambiar tanto el nombre de la variable como el valor de n al hacerlo para la segunda columna.

iii) Para **dibujar el histograma**:

Podemos dibujar el histograma, que es un estimador de la función de densidad, pulsando el botón *Histogram*, que está en la fila de botones de arriba, el que tiene unas barritas rojas, e indicando el nombre de la variable (¡las nuevas!). (Otra opción sería hacer *Describe* --> *Numerical Data* --> *One Variable Analysis...* --> Pulsar *Graphical options* --> Seleccionar *Frequency Histogram*. Este segundo camino tiene la ventaja de que no sólo ofrece el histograma).

El histograma de la segunda distribución (n grande) debe aproximarse más a la función de densidad de la distribución límite, es decir, a la de una normal estándar.

Estimación puntual: una población de Poisson

Ejercicio Poisson1

Se supone que el número de erratas en un libro de texto, X, sigue una distribución de Poisson de parámetro $\lambda = 10$.

a) Generar una muestra de tamaño 1000.
b) Hallar teóricamente el estimador de máxima verosimilitud para el parámetro λ, y calcularlo para esta muestra.

Objetivos

Aprender a generar números aleatorios de un modelo (en este caso Poisson), para distintos valores de sus parámetros. Calcular teóricamente el estimador máximo-verosímil del parámetro λ, que en este caso coincide con la media muestral, y evaluarlo en la muestra que se ha generado.

Menús

i) Para **generar una muestra aleatoria simple** de una distribución, tenemos que utilizar el menú:

Plot --> *Probability Distributions...* --> A continuación seleccionar la distribución deseada.

ii) Para **cambiar los parámetros de la distribución**: Por defecto el programa utiliza el valor

10 para esta distribución, por tanto no necesitamos cambiarlo. Si quisiésemos cambiar este valor, lo haríamos como en la práctica 1:

Colocarse encima del resultado *Probability Distributions* que ha mostrado el programa en la mitad izquierda --> Pulsar el botón secundario del ratón --> Elegir *Analysis Options...* (Nota: Observad que se pueden generar de una vez muestras de distintos valores del parámetro de esta distribución).

iii) Para c**ambiar el tamaño de la muestra**, que es 100 por defecto:

Pulsar el botón *Tabular options* (el segundo de la última fila) --> Marcar el resultado *Random Numbers* --> Pulsar el botón secundario del ratón encima del cuadro nuevo que aparece --> Elegir *Pane Options... --> Introducir el valor 1000.* (Nota: También se pueden cambiar los parámetros de la distribución al pulsar aquí el botón secundario y *Analysis Options...*).

iv) En la mitad de la derecha aparece por defecto el gráfico *Density/Mass Function*. Pero recordemos que podemos **elegir los gráficos** que se muestran en esa parte desde el botón *Graphical options* (tercer botón de la última fila).

v) Para **guardar una muestra** que hayamos generado:

Pulsar el botón *Save results* (cuarto botón de la última fila) --> Elegir qué variables queremos guardar y los nombres que les queremos poner (Si hemos generado de una vez muestras correspondientes a varias combinaciones de parámetros de una misma distribución, aparecerán también aquí esas opciones para ellas) --> Para ver la muestra tenemos que minimizar esta ventana principal y maximizar la hoja de cálculo en que *Statgraphics* presenta los datos.

vi) Para **calcular el estimador de máxima verosimilitud (media muestral, en este caso)**:

Entrar en el menú *Describe* --> *Numerical Data* --> Elegir *One-Variable Analysis...* --> Pulsar el botón *Tabular Options* (segundo) y marcar *Summary Statistic.* Aparece un resumen con varios estadísticos de la muestra, el que nos interesa es *Average*. (Nota: Pulsando el botón secundario del ratón encima de este resumen, y eligiendo *Pane Options...* podemos elegir que se muestren más o menos estadísticos).

Estimación puntual: una población normal

<u>Ejercicio Normal1</u>

Se quiere probar la efectividad de un antitérmico en reducir la temperatura. Para ello se tomó la temperatura de 9 niños de 4 años de edad afectados de gripe, antes y después de haberles suministrado el antitérmico, y se obtuvieron las siguientes reducciones de temperatura:

1,2; 1,7; 1,6; 1,7; 1; 1; 2,6; 3 y 1

Suponiendo que la variable reducción de temperatura *es normal, hallar un estimador insesgado, consistente y de varianza mínima para la media y para la varianza, respectivamente.*

<u>Objetivos</u>

Calcular, para el modelo normal, estimadores de la media y de la varianza poblacionales.

<u>Menús</u>

i) En este caso no generamos la muestra, sino que está dada, y la tenemos que introducir a mano en la hoja del programa.

Como sabemos por la teoría, los estimadores que cumplen las propiedades que se piden en el enunciado son: la media muestral para estimar la media poblacional, y la cuasivarianza muestral (o varianza muestral corregida) para estimar la varianza poblacional.

ii) Para **calcular la media muestral y la varianza muestral (corregida)**:

Entrar en el menú *Describe* --> *Numerical Data* --> Elegir *One-Variable Analysis...* --> Pulsar el botón *Tabular Options* (segundo) y marcar *Summary Statistic.* Aparece un resumen con varios estadísticos de la muestra, los que nos interesan ahora son *Average* y *Variance*. (Nota: Pulsando el botón secundario del ratón encima de este resumen, y elegiendo *Pane Options...* podemos elegir que se muestren más o menos estadísticos). *Statgraphics*, como la mayoría de programas, proporciona realmente la cuasivarianza muestral o varianza muestral corregida, porque es insesgada, en vez de la varianza muestral.

Intervalos de confianza: una población normal

Ejercicio Normal1

Se quiere probar la efectividad de un antitérmico en reducir la temperatura. Para ello se tomó la temperatura de 9 niños de 4 años de edad afectados de gripe, antes y después de haberles suministrado el antitérmico, y se obtuvieron las siguientes reducciones de temperatura:

1,2; 1,7; 1,6; 1,7; 1; 1; 2,6; 3; 1

Suponiendo que la variable reducción de temperatura *es normal, hallar un intervalo de confianza del 95% para la media y para la desviación típica.*

Objetivos

Calcular, para el modelo normal, intervalos de confianza, en este caso para la media y para la desviación típica poblacionales.

Menús

En este caso no generamos la muestra, sino que está dada, y la tenemos que introducir a mano en la hoja del programa. Para **calcular el intervalo de confianza**:

Si partimos de la situación del menú anterior, pulsamos el botón *Tabular Options* (segundo) y marcamos *Confidence Intervals.* El programa proporciona los intervalos de confianza para la media y para la desviación típica poblacionales, y utiliza por defecto un nivel de confianza del 95%. Para cambiar el nivel de confianza, sólo tenemos que pulsar el botón secundario del ratón y elegir *Pane Options...* ¿Qué le sucederá a la longitud del intervalo al variar el nivel de confianza?

Ejercicio Normal2

Los valores del pH de una solución en 10 determinaciones diferentes son los siguientes:

6,8; 6,78; 6,77; 6,80; 6,78; 6,80; 6,82; 6,81; 6,80 y 6,79

Suponiendo normal la distribución de la población de todas las determinaciones del pH de esa solución:

a) Hallar un intervalo de confianza al 95% para la media y la varianza poblacionales.
b) Hallar un intervalo de confianza al 65% para la media y la varianza poblacionales.

<u>Objetivos</u>

Obtener intervalos de confianza, con distintos niveles de significación, para la media, la desviación típica y la varianza poblacionales a partir de una muestra.

<u>Menús</u>

i) Introducimos a mano la muestra en la hoja de datos del programa.

ii) Tanto para calcular los intervalos de confianza como para contrastar la hipótesis, tenemos que **hacer** primero **el análisis de la muestra/variable**:

Entrar en el menú *Describe* --> *Numerical Data* --> Elegir *One-Variable Analysis...* --> Indicamos el nombre de la columna que queremos analizar.

iii) Para **calcular los intervalos de confianza** que nos piden:

Sobre los resultados del apartado ii), pulsamos el botón *Tabular Options* (segundo) y marcamos *Confidence Intervals.* El programa proporciona los intervalos de confianza para la media y para la desviación típica poblacionales, utiliza por defecto un nivel de confianza del 95%. Para cambiar el nivel de confianza, sólo tenemos que pulsar el botón secundario del ratón y elegir *Pane Options...*

Obsérvese que el programa no proporciona el intervalo de confianza para la varianza, sino para la desviación típica. Como la varianza es el cuadrado de la desviación, su intervalo se obtiene elevando al cuadrado los extremos del intervalo que hemos obtenido. **Debajo de los resultados, *Statgraphics* siempre proporciona una ayuda para interpretarlos**. En este caso, al final de la ayuda podemos leer: *While the confidence interval for the mean is quite robust and not very sensitive to violations of this assumption* [la de normalidad], *the confidence interval for the standard deviation is quite sensitive. If the data do not come from a normal distribution, the interval for the standard deviation may be incorrect. To check whether the data come from a normal distribution, select Summary Statistics from the list of Tabular Options, or choose Normal Probability Plot from the list of Graphical Options*. Vamos a comprobar que efectivamente podíamos suponer que la variable subyacente sigue una distribución normal. Para ello:

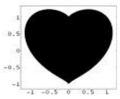
Pulsamos el botón *Tabular Options* (segundo) y marcamos *Summary Statistics* --> Encima de la ventana nueva que aparece pulsamos el botón secundario del ratón y elegimos *Pane Options...* --> Marcamos tanto *Skewness* (asimetría) como *Kurtosis* (curtosis), que son dos coeficientes cuyos valores caracterizan a la distribución normal. La teoría dice que para poder suponer que la población es normal, es necesario que ambos coeficientes estén dentro del intervalo [-2, +2].

Intervalos de confianza: dos poblaciones normales

<u>Ejercicio Normal3</u>

Las presiones sistólicas de dos grupos independientes de niños, para el primero de los cuales sus padres son hipertensos y para el segundo normales, dan los siguientes valores:

Grupo 1º: 100; 102; 96; 106; 110; 110; 120; 112; 112; 90
Grupo 2º: 104; 88; 100; 98; 102; 92; 96; 100; 96; 96



Suponiendo que las dos poblaciones son normales y de varianzas iguales y desconocidas, calcular un intervalo de confianza del 95% para la diferencia de medias.

Objetivos

Aprender a calcular intervalos de confianza para la diferencia de medias de dos poblaciones normales, cuando las varianzas son iguales y desconocidas.

Menús:

i) En este caso tampoco generamos las muestras, las tenemos que introducir a mano.

ii) Para **calcular el intervalo de confianza de la diferencia de medias**:

Entrar en el menú *Compare* --> *Two Samples* --> *Two-Sample Comparison...* --> Pulsar el botón *Tabular Options* (segundo) y marcar *Comparison of Means.* Aparecen los intervalos de confianza tanto para cada una de las medias como para su diferencia, con un nivel de confianza del 95%. Antes de cerrar esta ventana, es necesario copiar la semilongitud del intervalo (cantidad que se suma y se resta a la media estimada para construirlo).

Contrastes de hipótesis: una población binomial

Ejercicio Binomial2

Se ignora la proporción de familias numerosas en una región, y con el fin de determinar dicha proporción, se toma una muestra de 800 familias, siendo la proporción observada 0,18. Formulamos la hipótesis nula p=0,2 frente a la alternativa p≠0,2, y queremos contrastarla para un nivel de significación de 0,05. Se pide además la curva de pontencia del contraste.

Objetivos

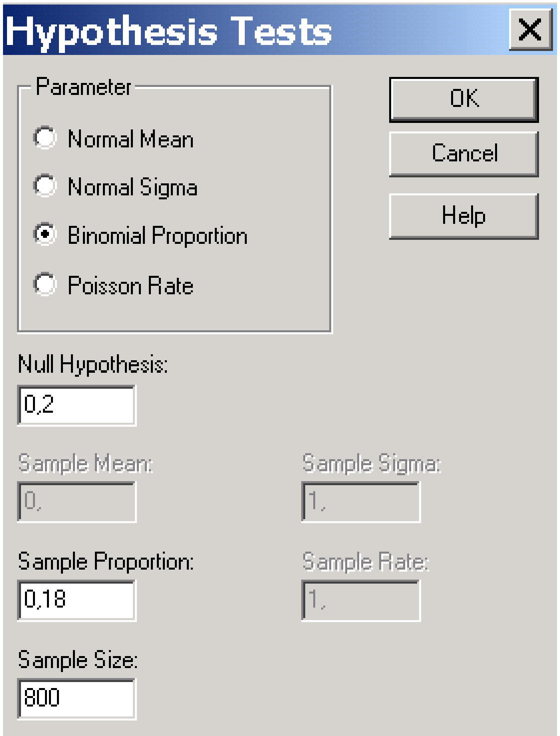
Hacer contrastes de hipótesis sobre una proporción, es decir, sobre el parámetro p que utilizan las distribuciones Bernoulli y Binomial.

Menús

i) En este caso tampoco conocemos la muestra, sólo el valor de la proporción muestral, que es un estimador del parámetro poblacional.

ii) Para **contrastar la hipótesis nula** $p=0,2$ frente a $p\neq 0,2$ con un nivel de significación de $\alpha = 0,05$:

Entramos en *Describe* --> *Hypothesis Tests...* --> Ahora tenemos que elegir la distribución binomial, e introducir los datos que nos dan en el enunciado según se muestra en el dibujo



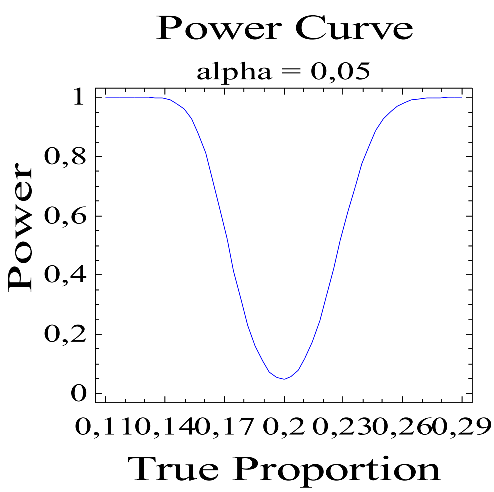
Los resultados ahora son:

```
Hypothesis Tests
----------------
Sample proportion = 0,18
Sample size = 800
```

`Approximate 95,0% confidence interval for p: [0,156004;0,205363]`

```
Null Hypothesis: proportion = 0,2
Alternative: not equal
P-Value = 0,17068
```

`Do not reject the null hypothesis for alpha = 0,05.`



Contrastes de hipótesis: una población normal

Ejercicio Normal2

Los valores del pH de una solución en 10 determinaciones diferentes son los siguientes:

6,8; 6,78; 6,77; 6,80; 6,78; 6,80; 6,82; 6,81; 6,80 y 6,79

Suponiendo normal la distribución de la población de todas las determinaciones del pH de esa solución. Contrastar la hipótesis nula de que la media poblacional es 6,8 contra la alternativa, con un nivel de significación $\alpha = 0{,}05$.

Objetivos

Hacer un contraste de hipótesis para un valor concreto de la media poblacional.

Menús

i) Introducimos a mano la muestra en la hoja de datos del programa.

ii) Tanto para calcular los intervalos de confianza como para contrastar la hipótesis, tenemos que **hacer** primero **el análisis de la muestra/variable**:

Entrar en el menú *Describe* --> *Numerical Data* --> Elegir *One-Variable Analysis...* --> Indicamos el nombre de la columna que queremos analizar.

Para **contrastar la hipótesis nula** de que la media es *6,8* con un nivel de significación de $\alpha = 0{,}05$:

Sobre los resultados del apartado ii), pulsamos el botón *Tabular Options* (segundo) y marcamos *Hypothesis Tests*. El programa proporciona los resultados de tres contrastes de hipótesis; por defecto para la hipótesis nula de que la media es *0* y valor de $\alpha = 0{,}05$. Sólo queremos cambiar la hipótesis nula, ambos valores se cambian pulsando el botón secundario del ratón y elegiendo *Pane Options...*

El programa proporciona tres contrastes de hipótesis: *t-test*, *sign test* y *signed rank test*. El primero de los tres es un contraste paramétrico, esto quiere decir que supone que la población sigue un determinado modelo de distribución (en este caso, normal); los dos últimos contrastes son no paramétricos, por lo que serían también válidos para cuando los datos no provienen de una distribución normal. En este caso hemos visto que podemos suponer que nuestros datos provienen de una distribución normal.

Ejercicio Normal4

En un preparado alimenticio infantil se especifica que según análisis garantizados el contenido mínimo de proteínas es del 42%. Tratamos de comprobar esta especificación tomando 10 preparados, los analizamos para determinar su contenido en proteínas y obtenemos una media del 40% y una desviación típica del 3,5%. Suponiendo normal la variable contenido de proteínas:
a) ¿Es correcta la especificación citada al nivel $\alpha = 0{,}05$?

Objetivos

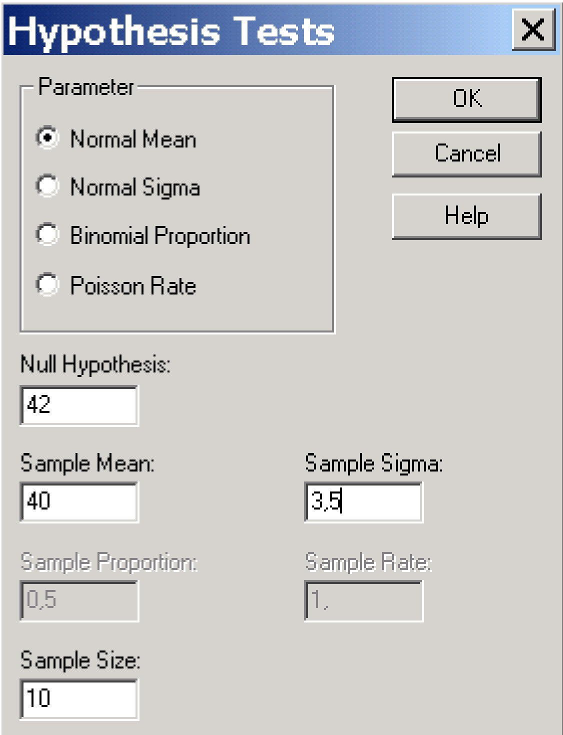
Realizar contrastes de hipótesis cuando no conocemos la muestra sino algunos valores «medidos» en ella.

Menús

i) En este caso no conocemos siquiera la muestra, sólo el valor de dos estadísticos evaluados en ella.

ii) Para **contrastar la hipótesis nula** de que la media es mayor o iqual al *42%* con un nivel de significación de $\alpha = 0{,}05$:

Entramos en *Describe* --> *Hypothesis Tests...* --> Por defecto está marcado el parámetro que queremos contrastar, así que sólo tenemos que introducir los datos que nos dan en el enunciado según se muestra en el dibujo (obsérvese que estamos haciendo contrastes paramétricos de cuatro posibles distribuciones)




```
Hypothesis Tests
----------------
Sample mean = 40,0
Sample standard deviation = 3,5
Sample size = 10

95,0% confidence interval for mean: 40,0 +/- 2,50376   [37,4962;42,5038]

Null Hypothesis: mean = 42,0
Alternative: not equal
Computed t statistic = -1,80702
P-Value = 0,104227
Do not reject the null hypothesis for alpha = 0,05.
```

Contraste de hipótesis: dos poblaciones normales

Ejercicio Normal5

Un instituto de dietética requiere comparar dos dietas. Se selecciona una muestra de 50 individuos al azar de una población de personas comparables con exceso de peso. A 25 personas se les suministra la dieta A y a los 25 restantes la dieta B. Las pérdidas de peso expresadas en kg al cabo de una semana ofrecen los siguientes datos muestrales:

Dieta A:	$\bar{x}_A = 4{,}3$ *kg*	$\sigma_A = 1{,}4$ *kg*
Dieta B:	$\bar{x}_B = 3{,}6$ *kg*	$\sigma_B = 1{,}1$ *kg*

Contrastar la hipótesis de igualdad de medias con un nivel de significación de 0,05 suponiendo normal la variable pérdida de peso.

Objetivos

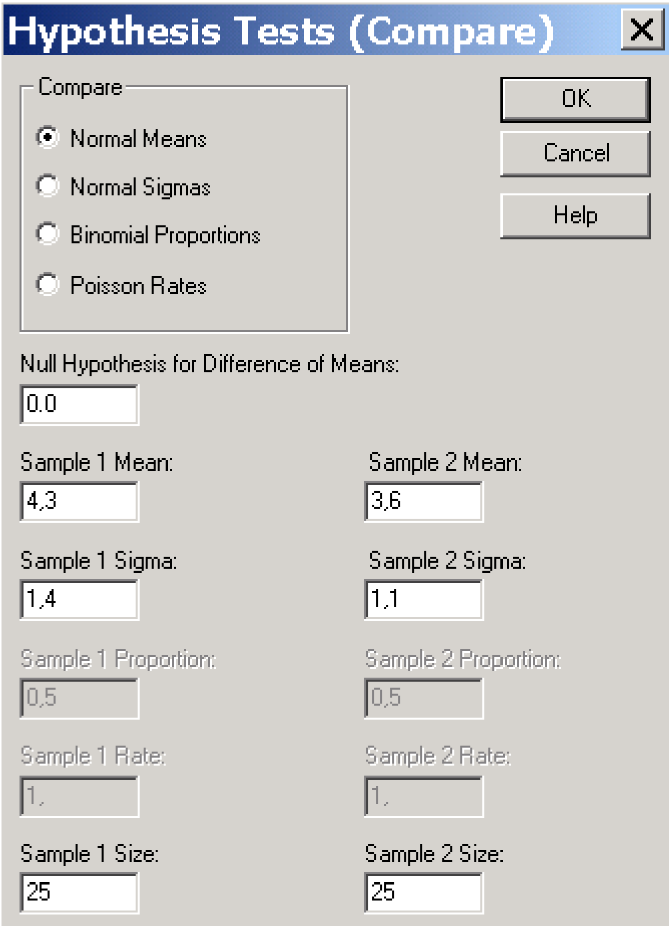
Contrastar si hay diferencia o no entre las medias de dos poblaciones.

Menús

i) No conocemos las muestras, sólo dos estimadores de ellas.

ii) Para **contrastar la hipótesis nula** de que las medias son iguales con un nivel de significación de $\alpha = 0{,}05$:

Entramos en *Compare* --> *Two samples* --> *Hypothesis Tests...* --> Introducimos los valores como se muestra en el dibujo




```
Hypothesis Tests
----------------
Sample means = 4,3 and 3,6
```

```
Sample standard deviations = 1,4 and 1,1
Sample sizes = 25 and 25

95,0% confidence interval for difference between means: 0,7 +/- 0,715968
[-0,0159681;1,41597]

Null Hypothesis: difference between means = 0,0
Alternative: not equal
Computed t statistic = 1,9658
P-Value = 0,0551213
Do not reject the null hypothesis for alpha = 0,05.
(Equal variances assumed).
```

Contraste de hipótesis: bondad de ajuste al modelo de Poisson

Ejercicio Poisson2

En el transcurso de 2 horas el número de llamadas por minuto solicitadas a una centralita telefónica presenta la siguiente distribución:

nº llamadas/minuto	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
frecuencia absoluta	*6*	*18*	*32*	*35*	*17*	*10*	*2*	*0*

a) Compruébese, a un nivel 0,01, si la variable nº llamadas/minuto *se distribuye o no según una ley de Poisson.*
b) Complementar de modo gráfico la respuesta al apartado anterior.

Objetivos

Hacer contrastes de bondad de ajuste a la distribución de Poisson, tanto numérica como gráficamente.

Menús

En este caso no generamos los datos ni los introducimos a mano, como en las prácticas anteriores, sino que los vamos a cargar («importar») desde un archivo. Los mismos datos están en dos formatos: en texto plano y en el propio formato de *StatGraphics*. Conviene practicar cómo cargar ambos:

http://www.casado-d.org/edu/datosPracticasEstadistica--Statgraphics.txt

ó, en el formato de *Statgraphics*,

http://www.casado-d.org/edu/datosPracticasEstadistica--Statgraphics.sf3

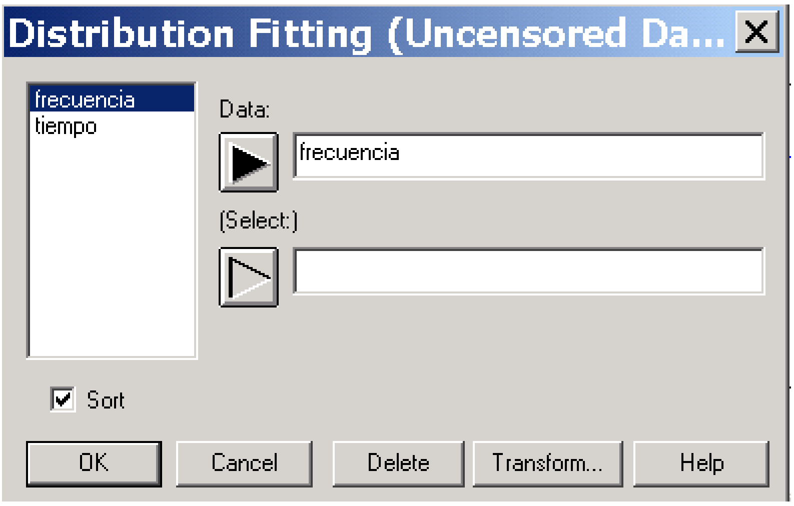
i) Para **cargar el archivo de datos**, una vez que lo tenemos grabado en nuestro ordenador, tenemos que ir al menú:

Entrar en el menú *File* --> *Open* --> *Open Data File...*

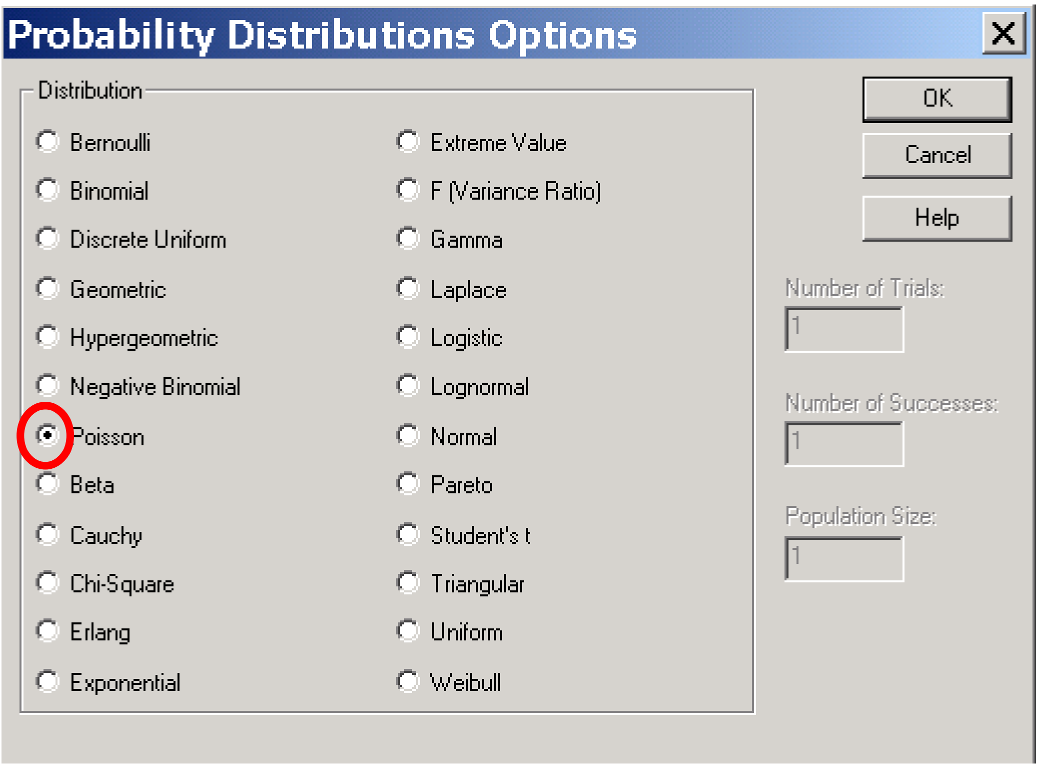
Al cargar el formato de texto, debemos informar al programa de cómo están los datos en el archivo. En este caso los datos están colocados en columnas, separados por tabulaciones y con el nombre de las variables en la primera fila.

ii) Para **hacer el contraste de bondad de ajuste**:

Entrar en el menú *Describe* --> *Distributions* --> *Distribution Fitting (Uncensored Data)*. --> Elegimos la columna a la que le queremos aplicar el contraste.



En la ventana del análisis, para hacer aparecer el contraste de bondad de ajuste, tenemos que pulsar el botón *Tabular options* (botón amarillo) y marcar *Goodness-of-Fit Tests*. El programa contrasta por defecto con una distribución normal, para cambiarla pulsamos el botón secundario del ratón, seleccionamos *Analysis Options...* y marcamos la distribución de *Poisson*.



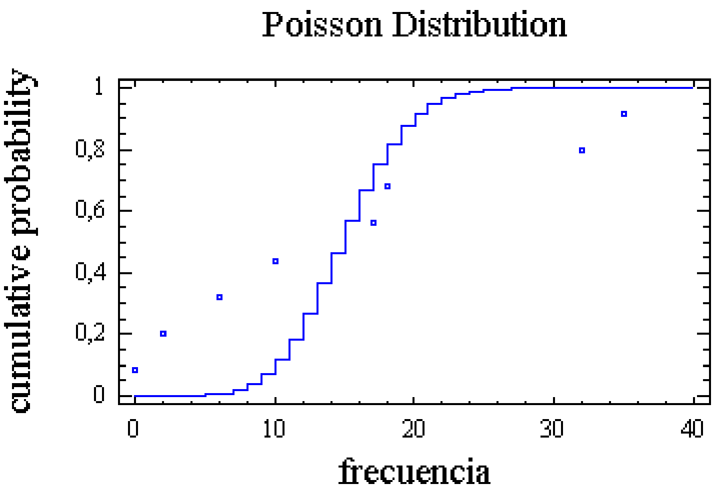

```
Goodness-of-Fit Tests for frecuencia

                              Chi-Square Test
--------------------------------------------------------------------------------
                Lower          Upper       Observed       Expected
                Limit          Limit      Frequency      Frequency      Chi-Square
--------------------------------------------------------------------------------
         at or below           12,5             4           2,14           1,61
                12,5           15,5             0           2,40           2,40
above           15,5                            4           3,46           0,09
--------------------------------------------------------------------------------
```

```
Chi-Square = 4,10412 with 1 d.f.   P-Value = 0,0427754
```

Según el contraste χ^2 : el p-valor es 0,0427754, quiere decir que para un valor $\alpha >$ 0,0428 rechazamos el ajuste a Poisson. Si trabajamos al 1%, estaríamos aceptando el ajuste. Observamos que el modelo χ^2 tiene un 1 grado de libertad (k-r-1, con k=3 categorías y r=1 parámetro de la Poisson). El estadístico es 4.10412.

iii) Para **evaluar gráficamente el contraste de bondad de ajuste**:

Pulsar el botón *Graphical options* (tercer botón) y seleccionar *Quantile Plot*.



En este gráfico se muestran las probabilidades acumuladas (cuantiles) para los datos reales y para el modelo teórico.

Contraste de hipótesis: bondad de ajuste al modelo exponencial

Ejercicio Exponencial1

Una compañía que fabrica productos de electrónica desea contrastar si la duración de sus bombillas sigue una distribución exponencial. Se han obtenido los siguientes datos, considerados independientes:

3,16	*6,55*	*36,87*	*0,71*	*20,33*	*3,58*	*2*	*0,78*	*5,2*	*23,53*
37,27	15,39	10,58	26,19	2,95	13,23	4,18	8,96	15,42	1,16

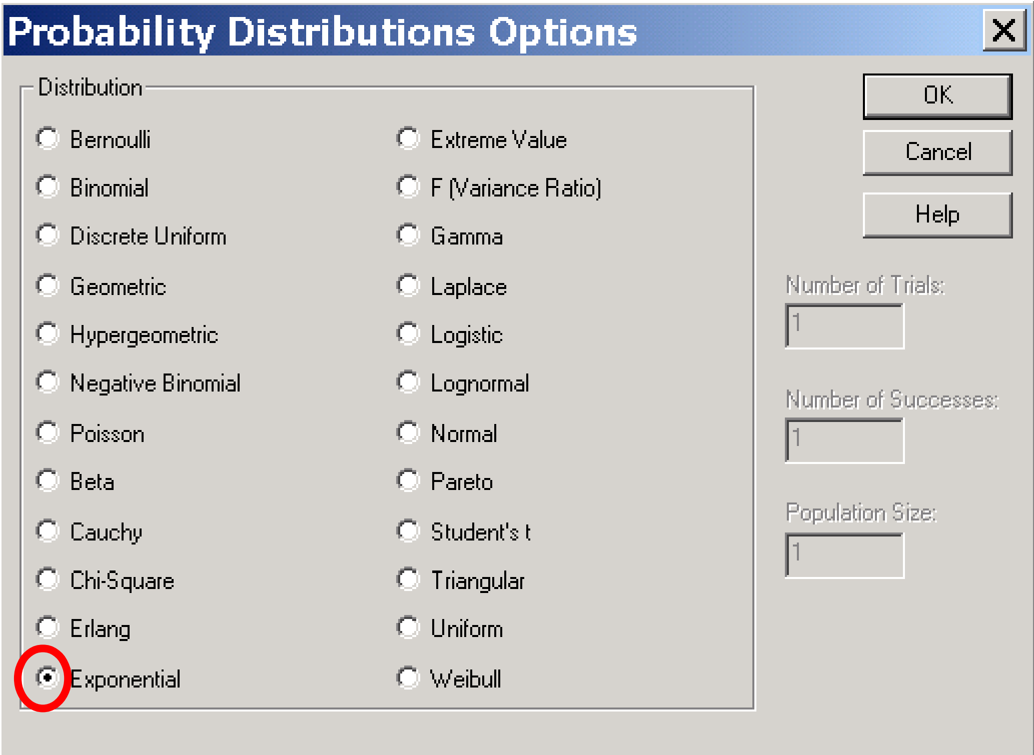
a) Constrastar si estos datos siguen una distribución exponencial.
b) Generar datos de precipitaciones con el mismo patrón que los anteriores datos.
c) Hacer otra vez el ejercicio generando otros 20 datos de otra distribución o con otros parámetros.

Objetivos

Hacer contrastes de bondad de ajuste a la distribución exponencial. Reproducir el modelo que siguen los datos, es decir, generar más datos del mismo modelo que los dados.

Menús

i) Para hacer el contraste de bondad de ajuste, seguir los pasos indicándole al programa que ahora queremos aplicar el contraste a la variable *tiempo*, y seleccionando en el siguiente paso la distribución *Exponencial*.



Goodness-of-Fit Tests for tiempo

Chi-Square Test

	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
	at or below	1,8347	3	2,86	0,01
	1,8347	4,00469	4	2,86	0,46
	4,00469	6,66055	3	2,86	0,01
	6,66055	10,0845	1	2,86	1,21
	10,0845	14,9104	2	2,86	0,26
	14,9104	23,1602	3	2,86	0,01
above	23,1602		4	2,86	0,46

Chi-Square = 2,4 with **5 d.f.** **P-Value = 0,791474**

Estimated Kolmogorov statistic DPLUS = 0,103842
Estimated Kolmogorov statistic DMINUS = 0,0755703
Estimated overall statistic DN = 0,103842
Approximate **P-Value = 0,982306**

Los datos aparecen agrupados en 7 clases, por lo que los grados de libertad son 5 (7 – 1 – 1). El estadístico χ^2 es 2,4 y el p-valor es 0,791474. Por ello, para un α > 0,791474 rechazaremos el ajuste a la exponencial. <u>Si trabajamos, por ejemplo, con</u> α <u>= 0,05 estaríamos aceptando el ajuste a un modelo exponencial</u>.

Situación parecida tenemos con el p-valor del contraste de Kolmogorov, que es todavía mayor 0.982306, y que diría que casi siempre aceptaríamos la hipótesis nula de ajuste a la exponencial.

ii) Para **generar datos que sigan el mismo modelo que los de la muestra** necesitamos conocer al valor del parámetro del modelo exponencial, o en su defecto una estimación. Como para este modelo la esperanza teórica es el valor inverso del parámetro, utilizaríamos el inverso de la media muestral como estimador del parámetro (en algunos libros el parámetro aparece como λ , y en otros como $1/\lambda$, en las expresiones del modelo; hay que tener cuidado con esto). Para generar una muestra aleatoria simple de una distribución, tenemos que utilizar el menú:

Plot --> *Probability Distributions...* --> A continuación seleccionar la distribución deseada.

Tenemos que cambiar el valor del parámetro que el programa utiliza por defecto por el valor estimado, pero como podemos observar en el paso anterior, el programa trabaja —y nos pide— directamente la media de la distribución, así que no tenemos siquiera que preocuparnos del valor del parámetro:

```
Analysis Summary

Data variable: tiempo

20 values ranging from 0,71 to 37,27

Fitted exponential distribution:
     mean = 11,902
```

iii) Para **cambiar los parámetros de la distribución**:

Colocarse encima del resultado *Probability Distributions* que ha mostrado el programa en la mitad izquierda --> Pulsar el botón secundario del ratón --> Elegir *Analysis Options...*

Determinación del tamaño muestral mínimo: una población binomial

Ejercicio Binomial 2

Se ignora la proporción de familias numerosas en una región, y con el fin de determinar dicha proporción, se toma una muestra de 800 familias, siendo la proporción observada 0,18. Formulamos la hipótesis nula p=0,2 frente a la alternativa p≠0,2, y queremos contrastarla para un nivel de significación de 0,05. ¿Cuál sería el tamaño de muestra necesario en el contraste anterior para α = 0,01?

Objetivos

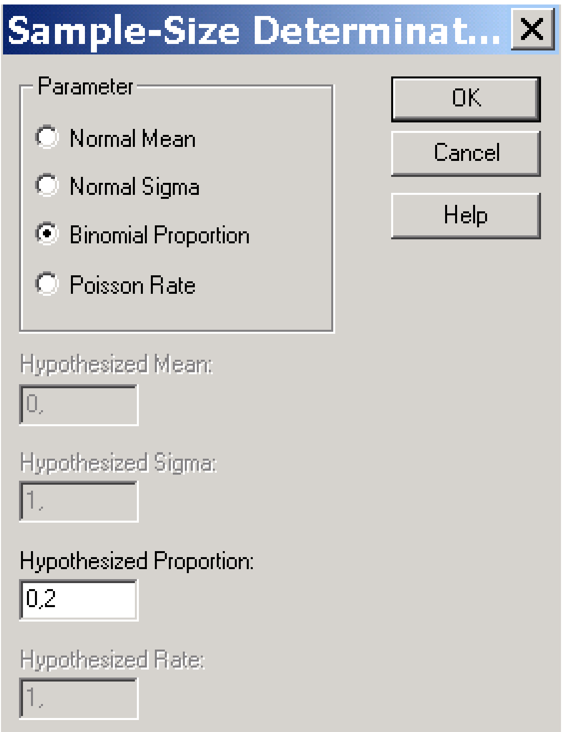
Calcular el tamaño muestral necesario para hacer un contraste, con una significación determinada, sobre esta proporción.

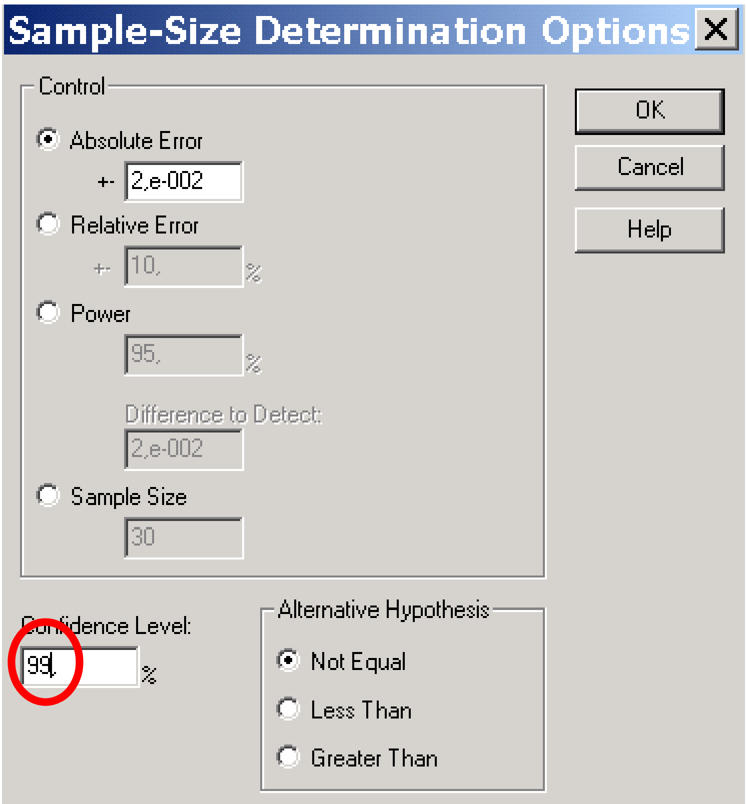
Menús

i) En este caso tampoco conocemos la muestra, sólo el valor de la proporción muestral, que es un estimador del parámetro poblacional.

ii) Para **calcular el tamaño muestral necesario**:

Entrar en el menú *Describe* --> *Sample-Size Determination...* --> Se introducen los valores en los campos --> A continuación aparece el cuadro en el que tenemos que indicarle al programa el nivel de significación que queremos utilizar.






```
Sample-Size Determination
-------------------------

Parameter to be estimated: binomial parameter
Desired tolerance: +- 0,02 when proportion = 0,2
Confidence level: 99,0%
```

`The required sample size is n=2654 observations.`

Determinación del tamaño muestral mínimo: una población normal

Ejercicio Normal3

Las presiones sistólicas de dos grupos independientes de niños, para el primero de los cuales sus padres son hipertensos y para el segundo normales, dan los siguientes valores:

Grupo 1º: 100; 102; 96; 106; 110; 110; 120; 112; 112; 90
Grupo 2º: 104; 88; 100; 98; 102; 92; 96; 100; 96; 96

¿Qué tamaños de muestras se necesitan para que, al nivel de confianza del 99%, el intervalo de confianza para la diferencia de las medias sea el obtenido en el apartado anterior, suponiendo que las desviaciones típicas poblacionales valen 7 ambas y que la diferencia de medias vale 5?

Objetivo:

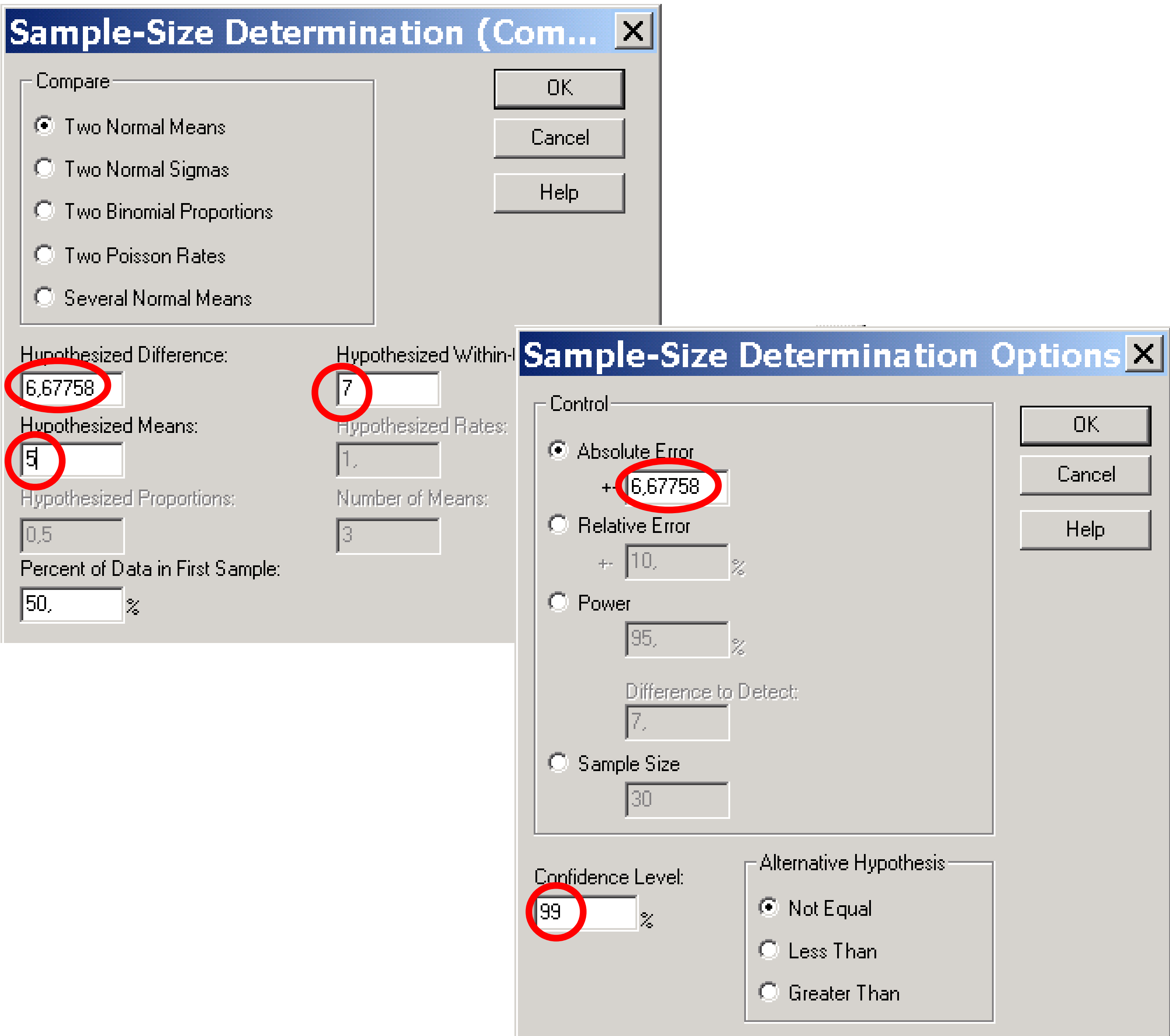
Aprenda a calcular el tamaño muestral necesario para conseguir intervalos de confianza del nivel de confianza deseado.

Menús:

i) En este caso tampoco generamos las muestras, las tenemos que introducir a mano.

ii) Para **calcular el tamaño muestral necesario**:

Entrar en el menú *Compare* --> *Two Samples* --> *Sample-Size Determination...* --> Se introducen los valores en los campos como indican las siguientes figuras. No os olvidéis del 99%...



Ejercicio Normal4

En un preparado alimenticio infantil se especifica que según análisis garantizados el contenido mínimo de proteínas es del 42%. Tratamos de comprobar esta especificación tomando 10 preparados, los analizamos para determinar su contenido en proteínas y obtenemos una media del 40% y una desviación *típica del 3,5%. Suponiendo normal la variable* contenido de proteínas. *¿Es correcta la especificación citada al nivel α = 0,05?*

<u>Objetivos</u>

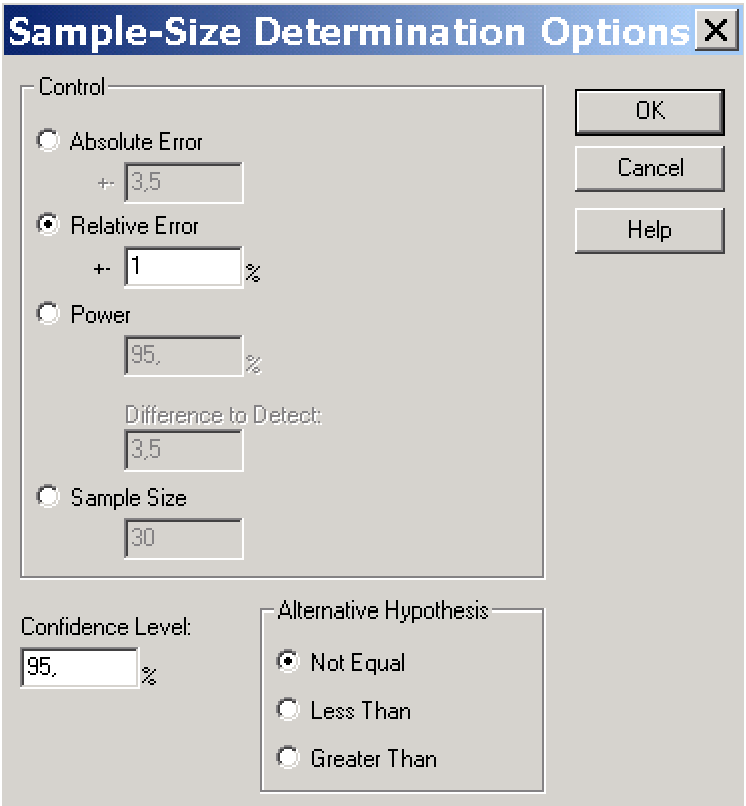
Calcular el tamaño muestral necesario cuando no conocemos la muestra sino algunos valores «medidos» en ella.

<u>Menús</u>

i) En este caso no conocemos siquiera la muestra, sólo el valor de dos estadísticos evaluados en ella.

ii) Para **calcular el tamaño muestral necesario**:

Entrar en el menú *Describe* --> *Sample-Size Determination...* --> Se introducen los valores en los campos como en el apartado anterior. --> A continuación aparece el cuadro en el que tenemos que indicarle al programa que queremos utilizar un error relativo del 1%



Éstos son los resultados que se obtienen:

```
Sample-Size Determination
-------------------------

Parameter to be estimated: normal mean
Desired tolerance: +- 1,0% when mean = 42,0
Confidence level: 95,0%
Assumed sigma: 3,5
```

`The required sample size is n=270 observations.`

Universidad Complutense de Madrid
∟ Facultad de Ciencias Económicas y Empresariales
∟ Departamento de Estadística e Investigación Operativa II
∟ David Casado de Lucas

20 de febrero del 2012