

Contraste de hipótesis con Statgraphics Centurion

Ficheros empleados: Transistor.sf3, Estaturas.sf3

1. Introducción

Una forma habitual de hacer inferencia acerca de uno o más parámetros de una población consiste en formular una conjetura o hipótesis sobre los parámetros poblacionales y verificar la validez de dicha hipótesis con la información muestral obtenida. Si los datos muestran unos valores que se alejan mucho de lo que sostiene la hipótesis formulada, consideraremos que nuestra hipótesis es incompatible con los datos. Por tanto, deberemos rechazarla. Por el contrario, si el resultado muestral observado es similar a lo que sostiene la hipótesis, podremos aceptarla. La teoría estadística sobre contraste de hipótesis nos proporciona la forma de cuantificar la discrepancia la hipótesis planteada y lo que se observa en la muestra.

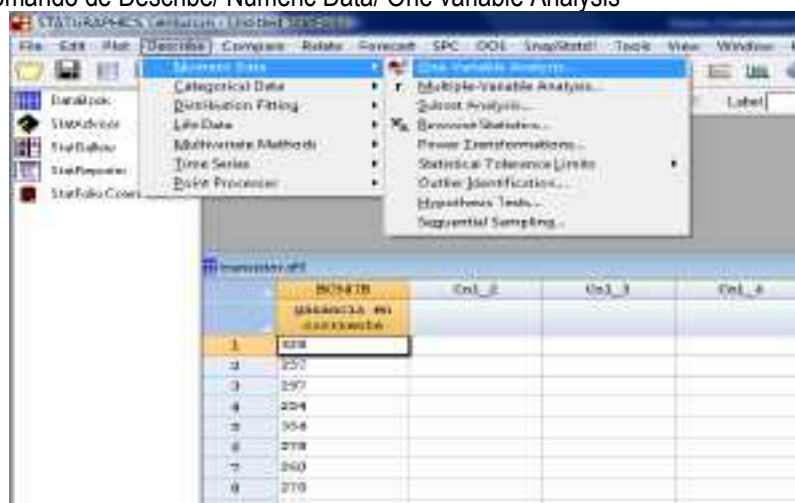
2. Objetivos

- ◆ El investigador debe plantear la hipótesis nula acerca del parámetro poblacional objeto de estudio. (contrastes unilaterales o bilaterales).
- ◆ Seleccionar el Estadístico de Contraste basándonos, tanto en la población de partida, como en la muestra seleccionada (tamaño muestral).
- ◆ Establecer los valores críticos que delimitan las regiones de rechazo y aceptación. (Regla de decisión) o calcular el p-valor.
- ◆ Tomar una decisión sobre la Hipótesis Nula a la vista del p-valor o el valor crítico.

3. Contraste Bilateral para la media poblacional (μ) en el caso de muestras grandes

Ejemplo: Un fabricante de transistores del tipo BC547B (*transistor.sf3*) desea saber si su producción sigue manteniendo el mismo nivel de calidad. Más concretamente, desea saber si la media de la distribución en ganancia de corriente de los transistores se mantiene en 290. Para ello toma una muestra de su producción de transistores de tamaño 100 y usará un nivel de significación del 5%. Los pasos a seguir son los siguientes

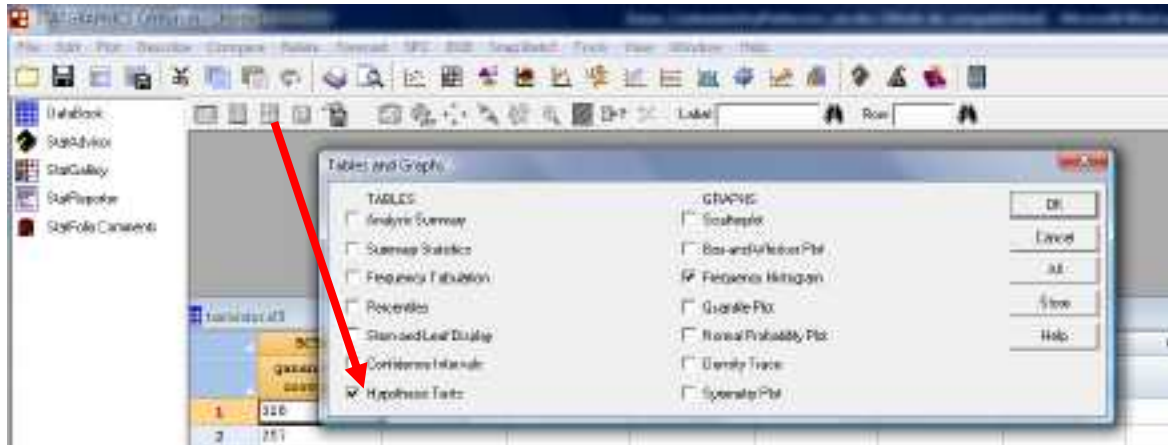
1. En primer lugar vamos a abrir la ventana para trabajar con la muestra recogida en el archivo *transistor.sf3*. Utilizamos los comando de Describe/ Numeric Data/ One variable Analysis



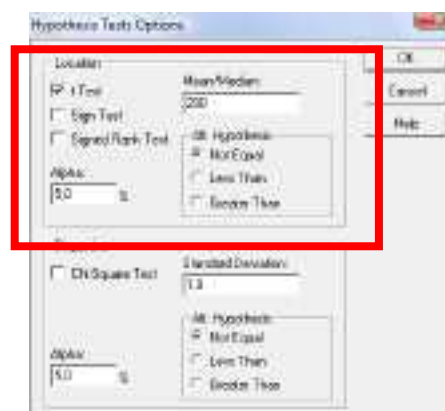
2. Seleccionamos la muestra (es de tamaño 100) con la que vamos a trabajar, que es la variable BC547B



3. En la ventana de Panel and Graphs seleccionamos la opción de *Hypothesis* (damos por supuesto que el analista ha realizado un estudio descriptivo de la variable antes de efectuar este análisis de inferencia)



4. El programa realiza, en este apartado, varios contrastes tanto para la media como para la varianza. Para simplificar seleccionamos únicamente el test que nos interesa, que es el test basado en el estadístico t para contrastes de la media. El programa utiliza un Estadístico de Contraste que sólo es válido si la variable de interés es normal, o bien si el tamaño muestral es suficientemente grande (>30 datos), en cuyo caso nos es indiferente la distribución de la variable de interés. En nuestro caso, tenemos un tamaño muestral de 100 observaciones, que son suficientes para poder realizar este contraste sin preocuparnos mucho de la distribución de la variable. El Statgraphics plantea por defecto la hipótesis nula de $\mu = 0$ a un nivel de significación de un 0,05. En el caso de que se quiera modificar estos datos, pulsando el botón derecho del ratón en Pane Options introducimos los datos de nuestro ejemplo, de la siguiente forma:



Es en la Hipótesis Alternativa donde podemos plantear un Contraste unilateral o bilateral. Tenemos las opciones de:

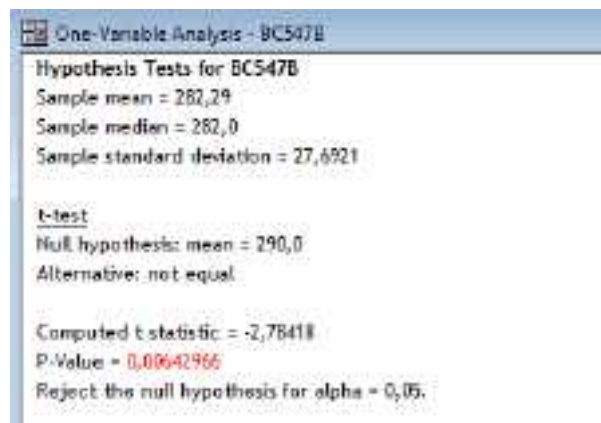
- ◆ Plantear un Contraste Bilateral: Not equal
- ◆ Plantear un Contraste Unilateral a la derecha: Greater than.
- ◆ Plantear un Contraste Unilateral a la Izquierda: Less Than.

En nuestro ejemplo debemos marcar la opción de un contraste bilateral ya que es el fabricante de transistores quiere plantear las siguientes hipótesis:

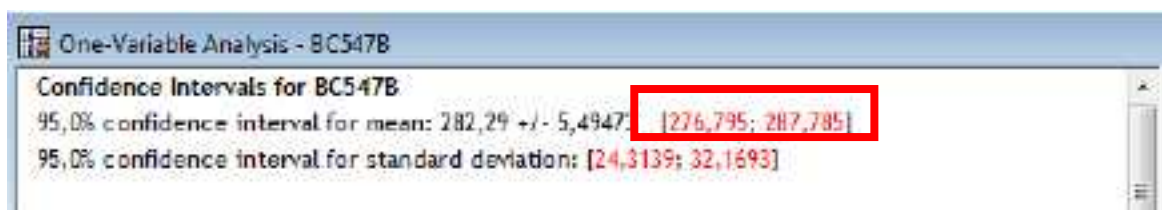
$$H_0: \mu=290$$

$$H_1: \mu \neq 290$$

5. Como resultado del análisis, el programa informa de los resultado muestrales necesarios para realizar el contraste: el valor de Estadístico de Contraste (-2,78418), el p-valor (0,006429) y la decisión tomada sobre la Hipótesis Nula (en este caso rechazar la hipótesis nula). Como el p-valor es $< 0,05$ rechazamos la hip. Nula. Es decir, se ha encontrado suficiente evidencia muestral para no considerar como verdadero el valor de $\mu=290$ planteado en la hip. Nula con un 5% de significación. Esto nos llevaría a pensar que, o bien la media ha cambiado, o que el aparato de medida está descalibrado. La salida que proporciona el Statgraphics se muestra a continuación.



Como se ha visto en el curso, la decisión que tomemos sobre la hipótesis nula en un Contraste Bilateral es posible realizarla utilizando un Intervalo de Confianza para el mismo nivel de significación. En nuestro caso, si seleccionamos *Confidence Intervals* en *Tabular Options* observamos que en el intervalo obtenido **no** incluye el valor de $\mu=290$, lo que nos lleva a rechazar la hipótesis nula (ver guión de prácticas sobre Intervalos de Confianza).



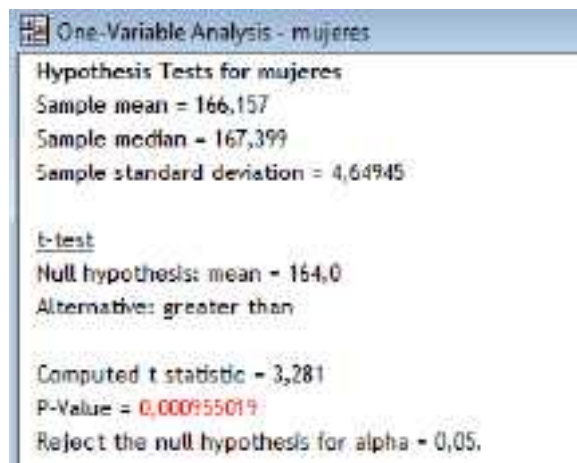
4. Contraste Unilateral para la media poblacional en caso de muestras grandes.

Ejemplo: En el fichero **estaturas.sf3** se recogen datos de alturas de 50 estudiantes masculinos y 50 femeninos entre 18 y 25 años, tomadas en campus universitarios madrileños. Según estudios antropométricos los jóvenes de este rango de edad tienen una estatura media de 164 cm y 177 cm, respectivamente. Se quiere comprobar si, por término medio, los estudiantes madrileños son más altos que la media de altura del resto de los estudiantes españoles. ¿Para qué nivel de significación mínimo se rechazarían dichas hipótesis?

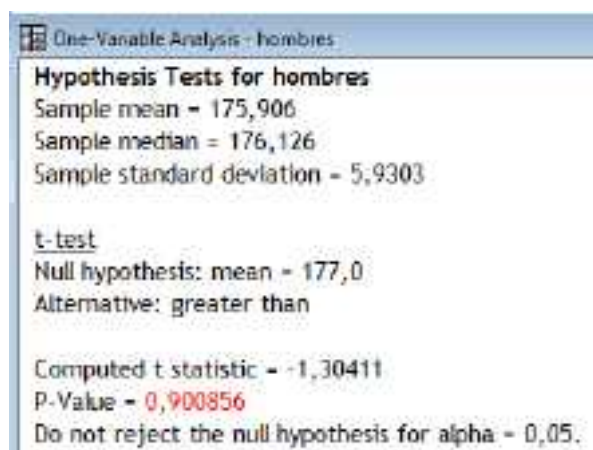
1. En primer lugar, plantearíamos las hipótesis nulas de:

$$\begin{array}{ll} H_0: \mu(\text{mujer}) \leq 164 & H_1: \mu(\text{mujer}) > 164 \\ H_0: \mu(\text{hombre}) \leq 177 & H_1: \mu(\text{hombre}) > 177 \end{array}$$

Como cada muestra es suficientemente grande, no nos preocupa mucho qué distribución sigue la población. Visualizamos los datos con un histograma para ver si hay algún dato atípico que pueda sesgar los resultados. Para hacer el contraste vamos a Describe/ Numeric Data/ One variable Analysis... Primero hacemos el Contraste para los Datos de mujeres.



Lo mismo hacemos para la muestra de hombres.



2. En base a los p-valores obtenidos podemos concluir lo siguiente:

- ◆ En el caso de las mujeres, el p-valor es 0,00095 por lo que para los valores de α habituales rechazaríamos la hipótesis nula planteada. Al ser un p-valor tan bajo esta hipótesis se rechaza con bastante claridad. Podemos decir que los datos nos proporcionan mucha confianza para afirmar que las estudiantes madrileñas son más altas por término medio que la altura media española.

- ♦ En el caso de los hombres tenemos un p-valor muy alto (0,90085) por lo que no podemos rechazar la hipótesis nula con los niveles de significación habituales. No hay por tanto evidencia suficiente para afirmar que los jóvenes madrileños tengan una estatura media mayor que la media española. En este caso, no habría hecho falta realizar el test, pues los datos no contradicen lo que dice H_0 . Se podía ya anticipar que el p-valor iba a ser grande. Sólo rechazamos cuando los datos hacen lo que dice H_1 (es decir, si se tuviese una media muestral mayor que 177) de forma muy evidente.

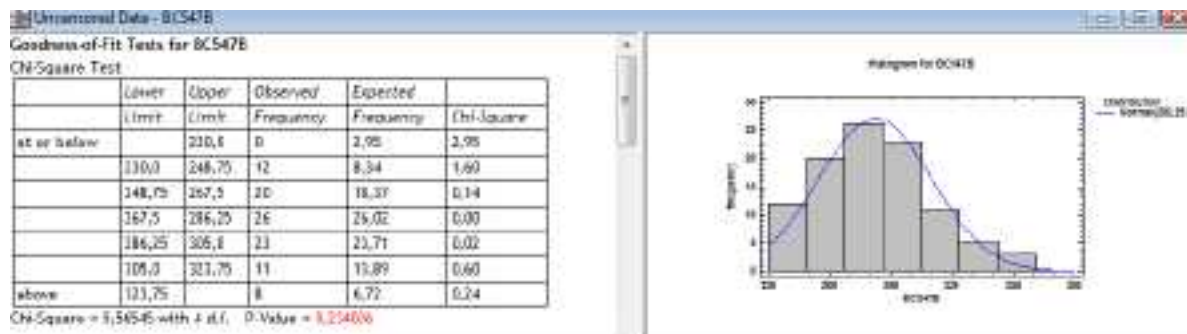
5. Contraste sobre la varianza poblacional en poblaciones normales.

Utilizamos nuevamente el ejemplo de los transistores del tipo BC547B (*transistor.sf3*), en el que se desea saber si su producción sigue manteniendo el mismo nivel de calidad. Más concretamente, se desea saber si la varianza de la producción no ha aumentado respecto al valor histórico de 760 ($\sigma_0^2=760$; $\sigma_0=27,568$). El contraste que se quiere realizar es si la varianza ha aumentado o no, que expresado como un contraste de hipótesis sería:

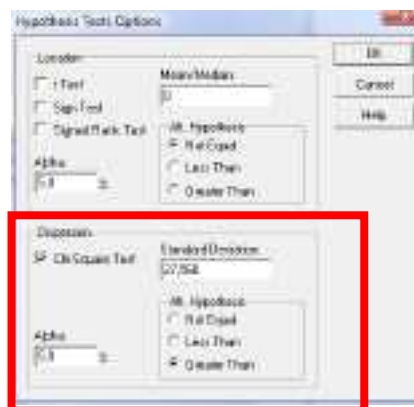
$$H_0: \sigma^2 \leq 760 \text{ (ó } \sigma \leq 27,568)$$

$$H_1: \sigma^2 > 760 \text{ (ó } \sigma > 27,568)$$

Este contraste se basa en la distribución chi-cuadrado y sólo es válido si la población es normal. Realizamos entonces un test de normalidad antes de continuar con el contraste (ver guión de prácticas de Ajuste de Distribuciones). El resultado que se obtiene haciendo un histograma con 10 clases y sin usar clases equiprobables en el test chi-cuadrado es:



Podemos entonces asumir normalidad (no podemos rechazar la hipótesis nula de normalidad). El test de la varianza se hace en el mismo lugar que el test de medias. Seleccionamos en Pane Options el test Chi-cuadrado y los detalles de nuestro contraste (hay que introducir la desviación típica en lugar de la varianza:



El resultado que se obtiene es

Hypothesis Tests for BC547B

Sample mean = 282,29

Sample median = 282,0

Sample standard deviation = 27,6921

chi-square test

Null hypothesis: sigma = 27,568

Alternative: greater than

Computed chi-square statistic = 99,8936

P-Value = 0,455935

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

El p-valor del contraste es mucho mayor que 0.05, por lo que no podemos rechazar la hipótesis nula de que la varianza del proceso no ha aumentado.