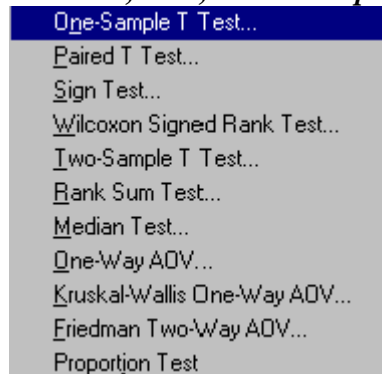


## Práctica 8: Contrastes paramétricos.

El programa Statistix puede realizar los cálculos correspondientes a varios contrastes paramétricos de hipótesis. Se encuentran los comandos en el menú:

*Statistics->One, Two, Multi Sample tests :*



Las instrucciones que nos pueden interesar de momento son:

- a.- **One sample t test:** test de Student para una muestra, contraste sobre la media.
- b.- **Two sample t test:** test de Student para dos muestras, contraste de igualdad de medias.

### Procedimiento para One sample t test :

**Ejemplo:** Se obtuvieron los siguientes datos en la medición de la intensidad de una corriente :

3,823 3,844 3,762 3,871 3,762

Se supone que el error que se comete durante la medición sigue una distribución normal de media 0 de varianza desconocida  $\sigma^2$ .

*Construir un intervalo de confianza al nivel de 95% para la intensidad real y realizar el contraste de hipótesis para comprobar si la intensidad real es significativamente menor que 3.90.*

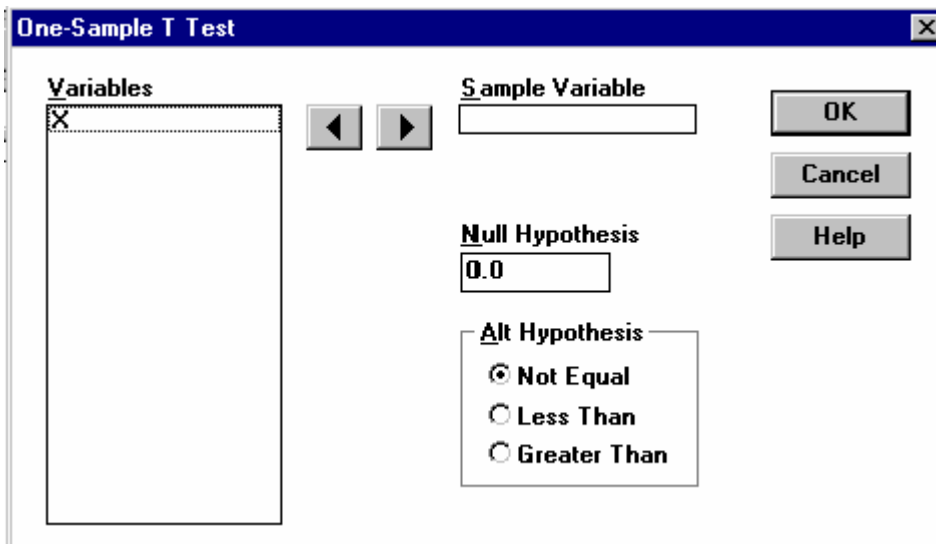
### Resolución:

Empezamos por introducir los datos (menú **Data->insert variable**). Una vez que los tengamos en nuestra hoja de cálculo, podemos obtener de manera simultánea el intervalo de confianza y el contraste de hipótesis. En nuestro caso las hipótesis de interés son:

$$H_0 : \mu = 3.9$$

$$H_1 : \mu < 3.9$$

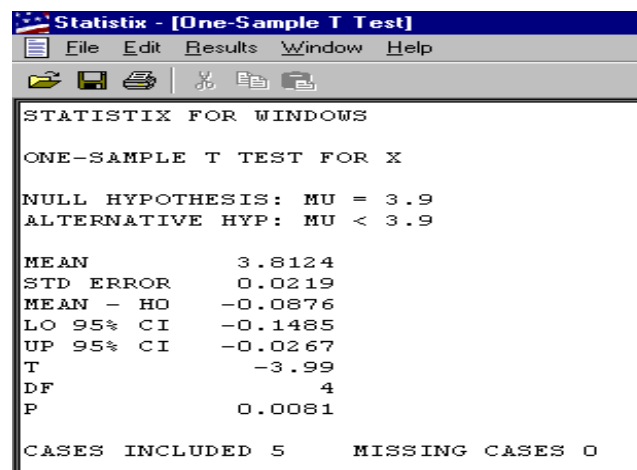
En el menú, **Statistics->One, Two, Multi Sample tests**, activamos la opción **One Sample t test** y aparece una ventana de diálogo



En esa ventana, encontramos, como es usual con Statistix, una lista de las variables que ya tenemos definidas y entre las cuales hay que escoger la variable cuyo parámetro queremos contrastar. Una vez seleccionada en esa lista la variable, la pasamos al cuadro " *Sample variable*" con las flechas. En el cuadro " *Null hypothesis*" indicamos el valor  $\mu_0$  que queremos contrastar, es decir el que entra en la definición de  $H_0 : \mu = \mu_0$ , mientras que en el cuadro " *Alt hypothesis*" debemos seleccionar el tipo hipótesis alternativa:

- not equal*  $H_1 : \mu \neq \mu_0$
- less than*  $H_1 : \mu < \mu_0$
- greater than*  $H_1 : \mu > \mu_0$

Al pinchar en OK, obtenemos una ventana que recoge los resultados:



donde:

- *Mean, Std error*: representan la media y desviación muestrales.
- *Mean-H<sub>0</sub>, LO 95% CI, UP 95% CI* : representan la media, extremo inferior y superior del intervalo de confianza al 95 % de la media  $\mu - \mu_0$ .
- *T* : valor del estadístico de Student.
- *D.F* : grados de libertad ( Degrees of Freedom).
- *P* : p-valor.

En este caso obtenemos un p-valor de 0.0081, por lo tanto rechazamos  $H_0$  con gran confianza y afirmamos que la intensidad real es significativamente menor que 3.900.

### Procedimiento para Two-sample t test :

**Ejemplo:** En el departamento de control de calidad de una empresa, se quiere determinar si habido un descenso significativo de la calidad de su producto entre las producciones de dos semanas consecutivas a consecuencia de un incidente ocurrido durante el fin de semana. Deciden tomar una muestra de la producción de cada semana, si la calidad de cada artículo se mide en una escala de 100, obtienen los resultados siguientes:

**Semana 1 :** 93 86 90 90 94 91 92 96

**Semana 2 :** 93 87 97 90 88 87 84 93

Construya un intervalo de confianza para la diferencia de medias al nivel de 95%. A continuación realice el contraste de hipótesis para probar si ha habido un descenso significativo de la calidad en la semana 1 y la semana 2.

### Resolución:

Empezamos por introducir las variables, tenemos dos posibilidades:

- (a) definimos una variable llamada "**Calidad**" que contiene los 16 datos, y otra variable "**Semana**" que valga 1 si el dato de calidad corresponde a la primera semana y 2 si proviene de la segunda. Esto corresponde a una presentación categórica de los datos.
- (b) definimos dos variables "**Calidad1**" y "**Calidad2**" que contienen los 8 datos de calidad de cada una de las dos semanas. Corresponde a una presentación de los datos en tabla.

Al activar la instrucción **Two-sample t test**, aparece una ventana, con una lista de las variables definidas. Debemos escoger en el cuadro "**Model specification**" la manera en la que están presentadas nuestras variables. Si los datos que queremos analizar se presentan en dos variables distintas (opción (b) anterior), escogemos la opción "**Table**". En cambio si hemos escogido la opción (a) anterior activamos la opción "**Categorical**"

- **Con la opción (b):** Después de especificar el modelo, pasamos las dos variables que queremos analizar desde la lista de variables al cuadro "**Table variables**".
- **Con la opción (a):** pasamos la variable "**Calidad**" al cuadro "**Dependent variable**", y la variable "**Semana**" al cuadro "**Categorical variable**".

Para rellenar los cuadros "**Null Hypothesis**" y "**Alt Hypothesis**", debemos considerar que contrastamos la diferencia entre las dos medias, (por ejemplo, para igualdad de medias, el cuadro "**Null Hypothesis**" debe contener 0). En nuestro problema las hipótesis son:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

La ventana de resultados es:

```

TWO-SAMPLE T TESTS FOR CALIDAD BY SEMANA
-----
SEMANA      MEAN      SAMPLE      S.D.      S.E.
-----
1           91.500    8           3.0237    1.0690
2           89.875    8           4.2237    1.4933
DIFFERENCE  1.6250

NULL HYPOTHESIS: DIFFERENCE = 0
ALTERNATIVE HYP: DIFFERENCE > 0

ASSUMPTION      T      DF      P      95% CI FOR DIFFERENCE
-----
EQUAL VARIANCES  0.88   14     0.1956  (-2.3139, 5.5639)
UNEQUAL VARIANCES 0.88   12.7   0.1964  (-2.3527, 5.6027)

TESTS FOR EQUALITY      F      NUM DF      DEN DF      P
OF VARIANCES           1.95      7           7           0.1988

CASES INCLUDED 16      MISSING CASES 0
    
```

Las primeras líneas contienen un resumen descriptivo de las dos variables de la muestra. A continuación aparecen las hipótesis que hemos contrastado (**difference=0** contra **difference <> 0**). El programa ha realizado el cálculo de dos estadísticos diferentes:

- El primero corresponde al caso en que suponemos que las dos variables tienen la misma varianza poblacional.
- Para el cálculo del segundo, las dos varianzas pueden ser distintas.

Para cada uno de esos dos casos, tenemos el valor del estadístico T, el número de grados de libertad, el p valor así como el intervalo de confianza al 95% para la diferencia de medias. **¿Cuál es su conclusión para el problema que nos interesa?**

**Ejercicio 1:** Cierta fabricante suministra determinado material siendo una característica de interés de este la resistencia en  $\text{kg/cm}^2$ . El fabricante afirma que la media de resistencia es de  $220 \text{ kg/cm}^2$ . Una empresa que quiere comprar el material decide contrastar la afirmación anterior. Para ello toma una muestra de 9 elementos de ese material y obtiene los siguientes datos sobre la resistencia:

203 229 215 220 223 233 208 228 209

*¿Puede concluirse a partir de los datos que la media es significativamente diferente de la que afirma el fabricante?*

**Ejercicio 2:** En un estudio sobre el tiempo empleado a la hora de almorzar entre los empleados de una determinada empresa, se seleccionan 16 mujeres y 16 hombres, obteniéndose los siguientes resultados en minutos:

**Mujeres** 54 61 44 50 50 54 59 54 22 58 45 30 25 29 24 38

**Hombres** 61 46 50 17 45 31 20 54 27 38 30 42 58 44 58 38

*¿Podemos concluir que la duración empleada es la misma en hombres y mujeres?*