Nombre: Javier Rivilla Arredondo
DNI: 53247378D
Email: jra48@alu.ua.es

Práctica 5

Ejercicio 5.1: El precio de venta (en euros) de un ordenador tipo multicore en diez establecimientos de informática es 3100, 2700, 2800, 3400, 2600, 2750, 3270, 3105, 3800, 2970. Suponiendo normalidad para esta distribución de precios, obtén razonadamente el intervalo de confianza para la media poblacional con un nivel de confianza del 98 por ciento. Interpreta el resultado.

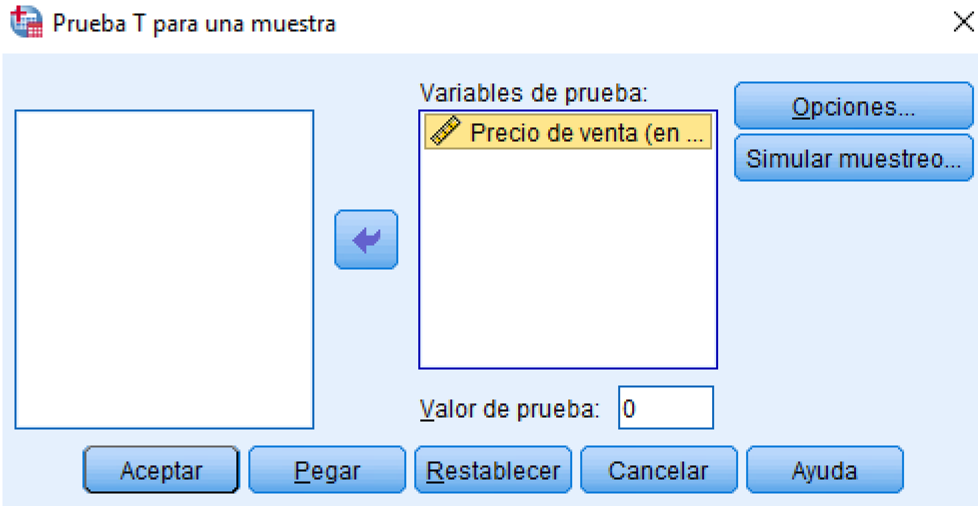
Solución:

En el ejercicio nos piden el intervalo de confianza para la media poblacional, como desconocemos la desviación típica poblacional haremos lo siguiente con el SPSS:

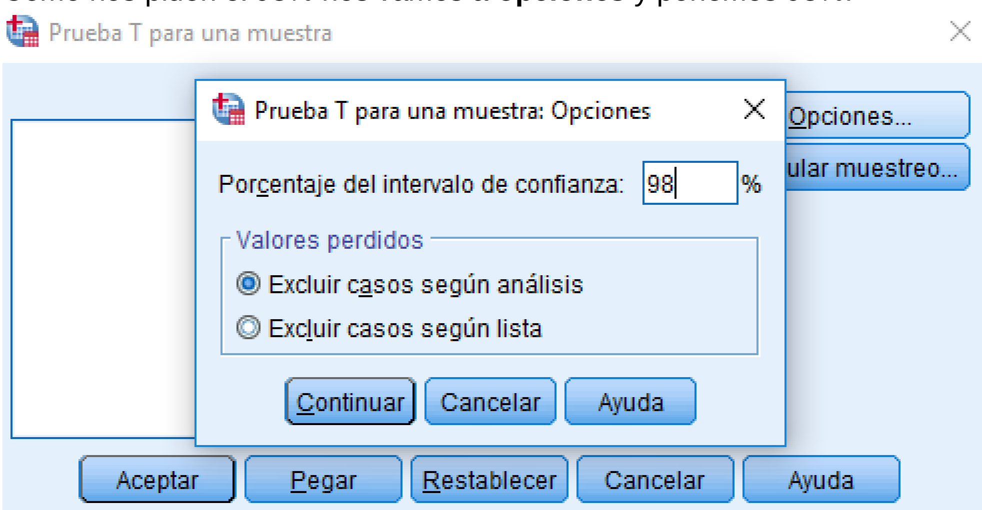
Primero añadimos los datos al SPSS:

	Precio
1	3100
2	2700
3	2800
4	3400
5	2600
6	2750
7	3270
8	3105
9	3800
10	2970

Analizar → Comparar medias → Prueba T para una muestra



Como nos piden el 98% nos vamos a **opciones** y ponemos 98%:



El resultado obtenido será el siguiente (marcado en azul):

Prueba para una muestra

	Valor de prueba = 0					
					98% de intervalo de confianza de la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
Precio de venta (en euros)	26,162	9	,000	3049,500	2720,62	3378,38

De esta forma el intervalo de confianza que obtenemos es {**2720.62, 3370.38**}.

Podemos deducir la media poblacional tiene un intervalo de confianza entre {**2720.62, 3378.38**} de precios.

Ejercicio 5.2: Se ha medido la velocidad (Km. /h) sobre una muestra de 12 coches que circulan por una carretera, obteniéndose 98, 85, 77, 92, 112, 106, 89, 101, 121, 77, 120, 90. Sabiendo que la desviación típica de la velocidad de los coches se asume que es igual a 14 Km. /h, calcula razonadamente un intervalo de confianza al 96 por ciento para la velocidad media de los vehículos en dicha carretera suponiendo que la muestra ha sido obtenida por muestreo aleatorio simple sobre una población normal. Interpreta el resultado.

Solución:

$n = 12$
$\dot{X} = 97,33$
$\sigma = 14$ Km/h
¿Intervalo de confianza al 96%?

Sacamos la media muestral para ellos vamos a **Analizar → Comparar medias:**

Y obtendremos los siguientes valores:

Informe

Velocidad en Km/h

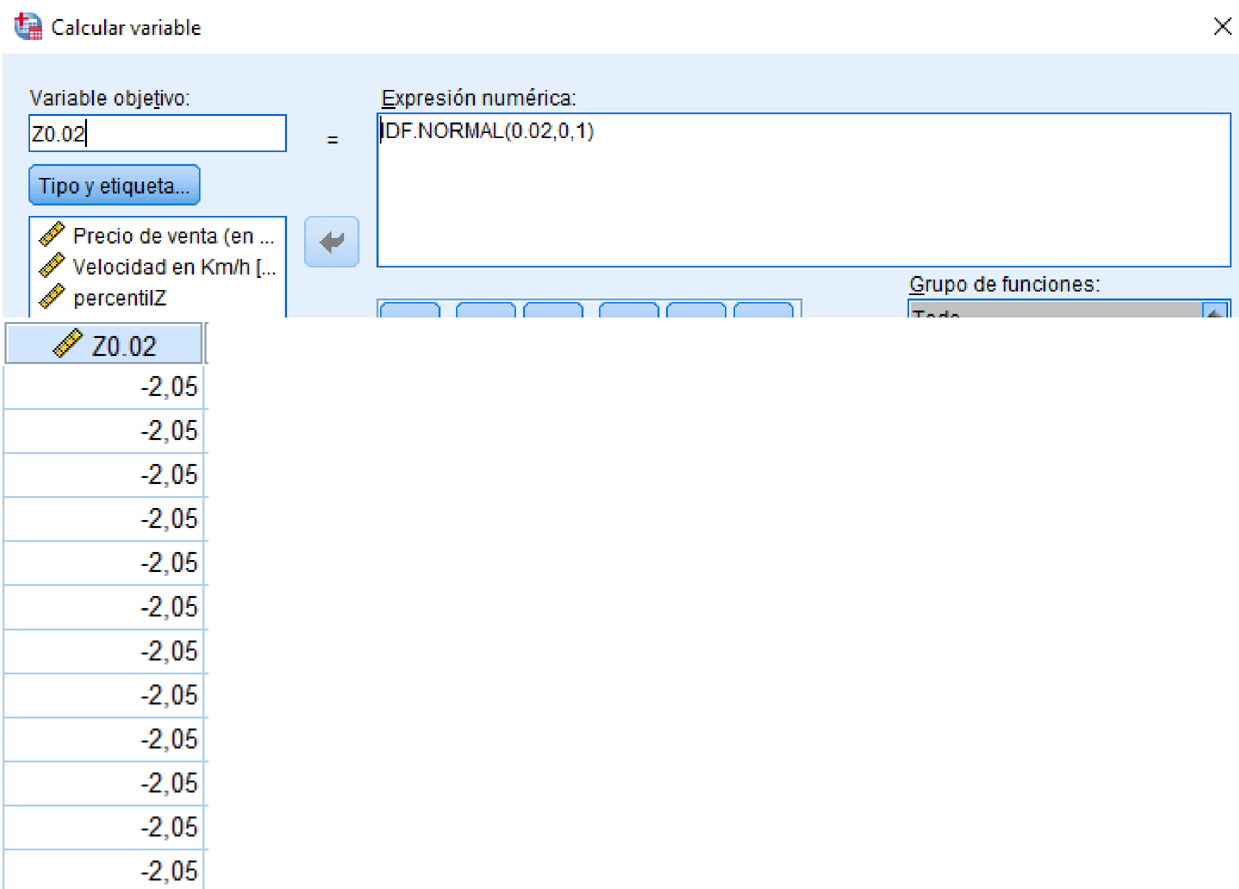
Media	N	Desv. Desviación
97,33	12	15,102

Por lo que la fórmula que debemos utilizar es la siguiente:

$[\dot{X} - \sigma/\sqrt{(n)} * Z((1-0.96)/2), \dot{X} + \sigma/\sqrt{(n)} * Z((1-0.96)/2)]$

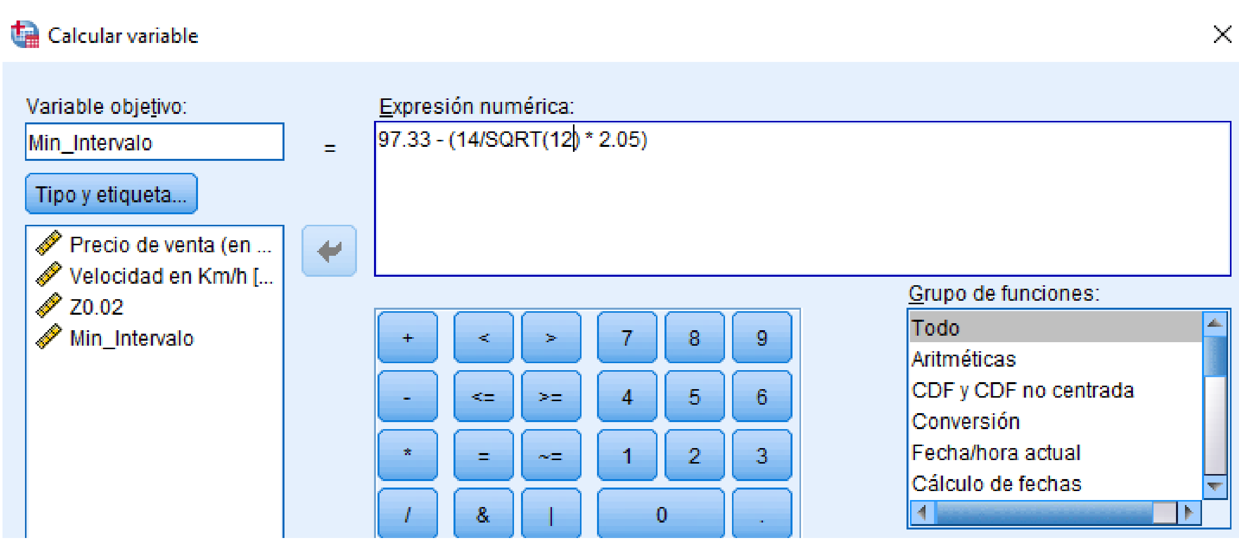
$[97,33 - 14/\sqrt{(14)} * Z0.02, 97,33 + 14/\sqrt{(14)} * Z0.02]$

Obtenemos el valor del percentil *Z0.02* →



Como podemos observar los valores de $Z_{0.02} \rightarrow 2{,}05$

$[97{,}33 - 14/\sqrt{14} * Z_{0.02},\ 97{,}33 + 14/\sqrt{14} * Z_{0.02}]$



Calculamos de esta forma la parte mínima del intervalo de confianza.

Min_Intervalo	Max_Intervalo
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61
89,05	105,61

Y posteriormente calculamos la parte máxima del intervalo de confianza:

Por lo que el intervalo de confianza del 96% será: **[89.05, 105.61]**

Ejercicio 5.3: La velocidad media de una muestra de 200 conexiones a Internet es 3.6 Mb/sg. Sabiendo que la desviación típica de la velocidad de las conexiones es 0.7 Mb/sg. Halla el intervalo de confianza para la media poblacional para un nivel de confianza del 99 por ciento. Interpreta el resultado.

$\dot{X}$= 3,6 Mb/sg
n = 200
σ = 0.7Mb/sg

– Si σ es conocida, entonces

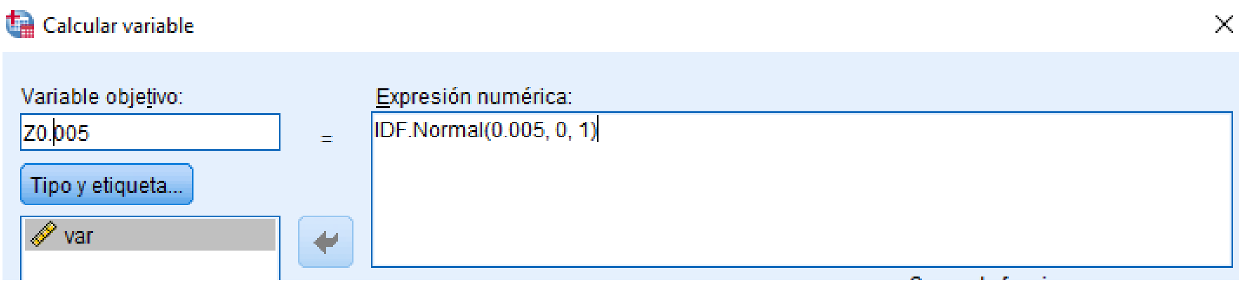
$$\left[\overline{x} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}, \overline{x} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}\right],$$

$Z_{0.01/2} = Z_{0.005}$

Sabemos que cuando conocemos el desviación utilizamos:

[$\dot{X}$ - σ /$\sqrt{(n)}$ * $Z_{0.005}$, $\dot{X}$ + σ /$\sqrt{(n)}$ * $Z_{0.005}$]
[3,6 - 0.7 / $\sqrt{200}$ * $Z_{0.005}$, 3.6 + 0.7 / $\sqrt{200}$ * $Z_{0.005}$]

Calculamos el percentil 99 para la normal:
Para ello nos vamos al SPSS y utilizamos el **IDF.Normal(0.005, 0, 1)**



Z0.005
-2,58

Por lo que tal rango será el siguiente:

Rango_minimo	Rango_maxima
3,47	3,73

Por lo que el intervalo de confianza del 99% será **[3.47, 3.73]** de la velocidad media de las conexiones.

Ejercicio 5.4: Una asociación empresarial desea estimar el porcentaje de empresas que utilizan sistemas basados en software libre. Se admite un error máximo del 3 por ciento con una confianza del 98 por ciento. ¿Qué tamaño mínimo de muestra se necesita en la investigación?

Ejercicio realizado en clase de teoría

Ejercicio 5.5: Una universidad quiere hacer un sondeo acerca del porcentaje de sus estudiantes que tienen Internet en casa. El censo es de 30000 estudiantes y deciden realizar un muestreo aleatorio de tamaño 1500. Para una confianza del 98%, ¿qué error se estará cometiendo al extrapolar los datos obtenidos por la muestra a la población?

n = **30000** estudiantes

- Como podemos comprobar se trata de una muestra grande, por tanto, tendremos que utilizar la siguiente fórmula, ya que necesitamos el error y tenemos los datos.

$$\boxed{n_0 = \left(\frac{Z_{\frac{\alpha}{2}}}{\epsilon}\right)^2 p(1-p),}$$

Donde p es la proporción:

n0 = **1500**
p = ½ porque no lo tenemos
σ = 1 - 0.98 = **0.02**
$Z_{0.01}$ = **2.33**

Sustituimos en la fórmula y despejamos:

1500 = (2.33/error) ^ 2 * ½ *(1 - 1/2)
1500 = 2,33/error ^ 2 * ¼
error = √(2.33^2 /1500 * ¼) = **0.030**

Por lo tanto **0.030** es el error que se está cometiendo al extrapolar los datos obtenidos por la muestra de la población.
Resuelve el apartado anterior sabiendo además que en un sondeo previo se ha estimado que el 75% de los estudiantes disponen en casa de Internet.

$$n_0 = \left(\frac{Z_{\frac{\alpha}{2}}}{\epsilon}\right)^2 p(1-p),$$

Los estudiantes que disponen de casa son 30000 * 0.75 = 22500 son los estudiantes que disponen de casa de internet, por lo que p = 22500 / 30000 = **0.75**

1500 = (2.33 / error) ^ 2 * 0.75 *(1 - 0.75)
1500 = (2,33 / error) ^ 2 * 0.1875
1500 = (5.428 / error ^ 2) * 0.1875
1500/0.1875 = (5.428 / error^2)
8000 = 5.428 / error ^ 2
Error = $\sqrt{6.78 * 10^{-4}}$ = **0.02604**

El error obtenido después del sondeo previo es de **0.02604.**

Ejercicio 5.6: Un centro de investigación ha diseñado un programa de fisioterapia con la Wii, para que los pacientes de rehabilitación puedan hacer los ejercicios en casa. El centro desea saber si con el uso de dicho programa el tiempo de recuperación es menor. Para ello utilizaron el programa sobre 2450 pacientes y se obtuvo que 1800 pacientes necesitaron menos de dos semanas de rehabilitación. Los datos estimados hasta el momento sobre este tipo de rehabilitaciones indicaban que en el 60 por ciento de los casos la rehabilitación era inferior a dos semanas y en el 40 por ciento de los casos restantes superior o igual a dos semanas.

Calculando un intervalo de confianza al 99 por ciento explica si los datos obtenidos con el uso de la Wii mejoran los datos estimados previamente.

n = 2450
p = 1800/2450
α = 1 - 0.99 = 0.01 / 2 = 0.005
$Z_{0.005}$ = 2.58

$$\left[\overline{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}, \overline{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right].$$

[1800 / 2450 - (2.58 * √(1800/2450 * (1 - 1800/2450) / 2450)), …. + ….]

[**0.7116**, **0.7577]**

Por lo que según la estimación de los resultados de la wii era un 60% mientras que a nosotros nos sale que mejoran entre un 71,16% y un 75,76%.

Luego para saber cuantos son superiores o iguales a 2 semanas será la diferencia, es decir, la proporción será de 650/2450 y los porcentajes serán menores que el 40% citados en el enunciado. Entre el 24.2% y el 26%.

Ejercicio 5.7: Un centro de investigación está trabajando en el diseño de un sistema de seguridad para el reconocimiento de personas que entran en cierta entidad bancaria. Para esto están utilizando dos algoritmos distintos. El algoritmo 1 se aplicó en 12 sucursales bancarias y el algoritmo 2 en 7 sucursales bancarias.
La tasa de error de dichos algoritmos en cada caso, expresada en tanto por ciento, fue:
Primer algoritmo: 17, 14, 12, 10, 10, 13, 23, 23, 20, 11, 25, 17.
Segundo algoritmo: 21, 10, 16, 17, 13, 12, 13.
Suponiendo normalidad, encuentra los dos intervalos de confianza al 97.5 por ciento de confianza (varianzas iguales o distintas) para la diferencia de tasas medias de error. Interpreta el resultado.

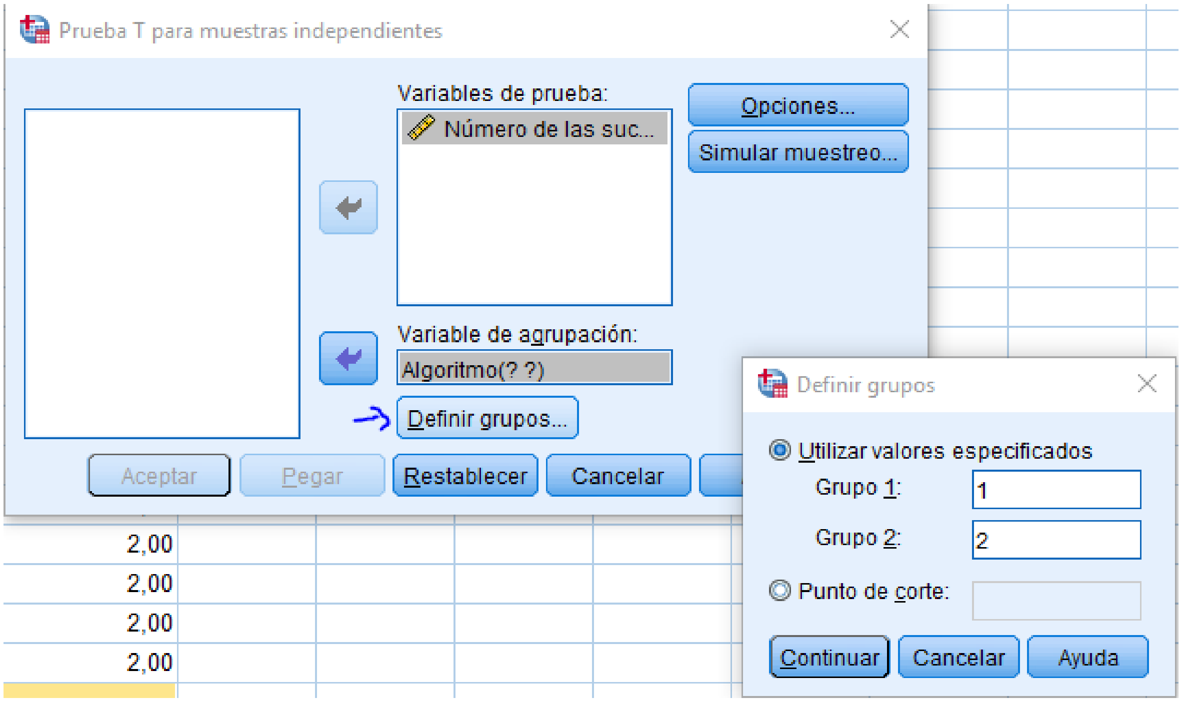
Primero añadiremos los valores al SPSS:

- En la columna de sucursal me refiero a sucursal bancaria donde se verán definidas por un algoritmo.

- En la columna de algoritmo me refiero a los algoritmos que corresponden a cada sucursal bancaria.
- Tomaremos muestra independiente ya que las sucursales son independiente entre sí.

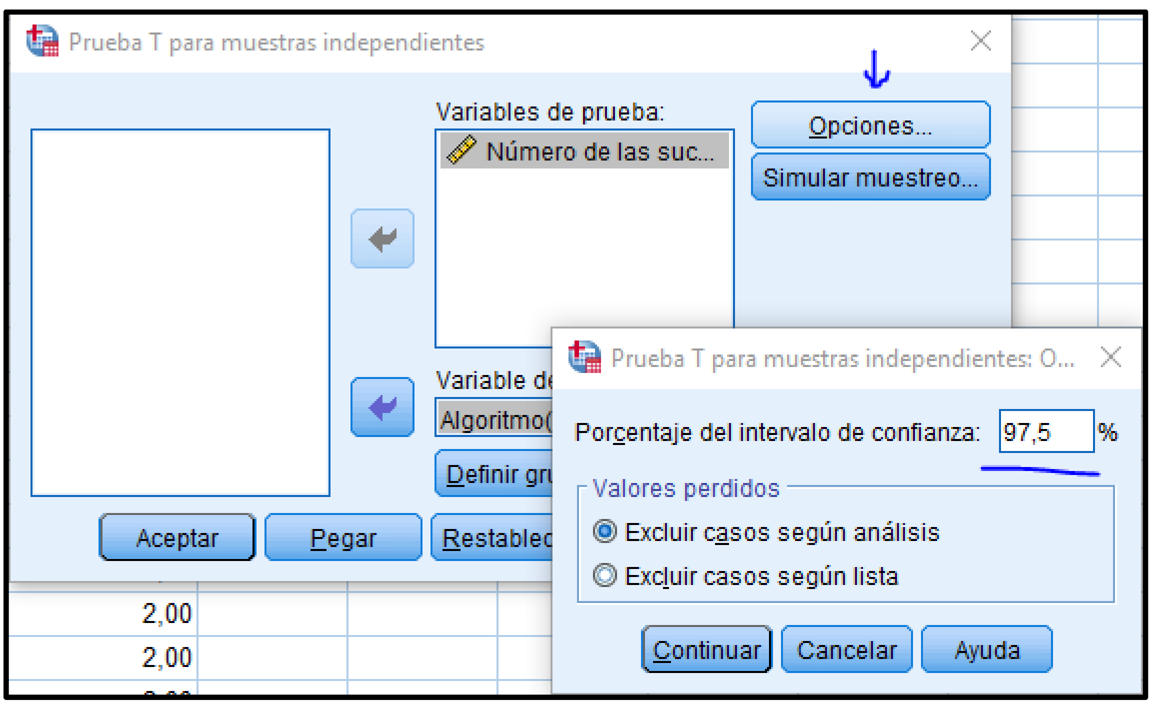
	Sucursales	Algoritmo
1	17,00	1,00
2	14,00	1,00
3	12,00	1,00
4	10,00	1,00
5	10,00	1,00
6	13,00	1,00
7	23,00	1,00
8	23,00	1,00
9	20,00	1,00
10	11,00	1,00
11	25,00	1,00
12	17,00	1,00
13	21,00	2,00
14	10,00	2,00
15	16,00	2,00
16	17,00	2,00
17	13,00	2,00
18	12,00	2,00
19	13,00	2,00

Ahora vamos a **Analizar** → **Comparar medias** → **Prueba T para muestras independientes**



Añadimos en variables de prueba los número de sucursales bancarias y agrupamos en algoritmos. A continuación, definimos los grupos pertenecientes, algoritmo 1 y algoritmo 2.

Una vez hacemos lo anterior nos vamos a opciones:



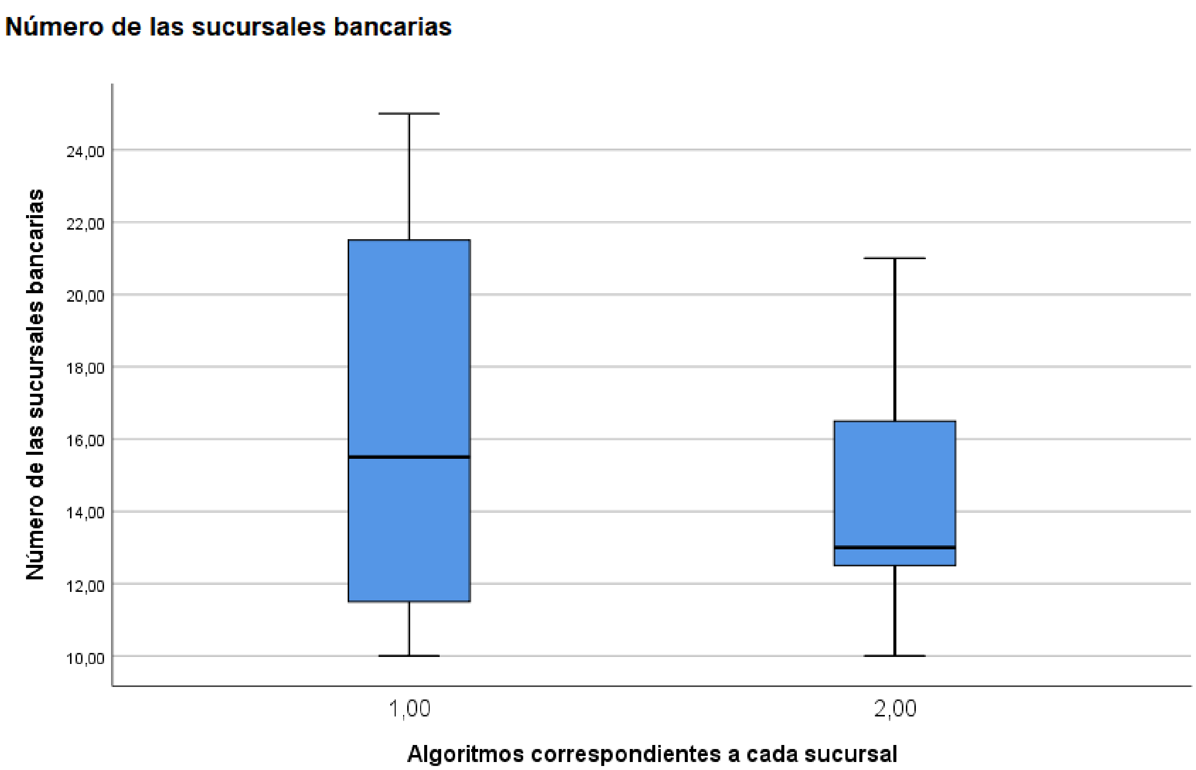
Donde añadiremos el valor del porcentaje del intervalo de confianza, le damos a continuar y el resultado será el siguiente:

97,5% de intervalo de confianza de la diferencia	
Inferior	Superior
-4,02012	7,37726
-3,48758	6,84472

El intervalo de confianza será por tanto para dos casos:

- Se han asumido varianzas iguales: {-4.02012, 7.37726}
- No se han asumido varianzas iguales: {-3.48758, 6.84472}

Bien ahora nos vamos al diagrama de cajas:



Como podemos observar en ambos algoritmos, en el algoritmo 1 las tasas de errores de las sucursales comprendidas entre el 25% y el 50% están más dispersas que entre el 50% y el 75% pero no con tanta diferencia como en el algoritmo 2. Por lo que yo elegiría el algoritmo 1 ya que la tasa de errores en el algoritmo 2 es bastante más predominante que en el algoritmo 1.

Ejercicio 5.8: El Instituto Nacional de Estadística ha estimado que el 15 por ciento de las empresas españolas realizan compras por Internet. En la Comunidad Valenciana se muestrearon 2500 empresas, de las cuales 320 habían realizado compras por Internet. Calculando un intervalo de confianza al 95 por ciento, explica si las cifras obtenidas en la Comunidad Valenciana están en concordancia con el porcentaje estimado por el Instituto Nacional de Estadística.

n = 2500

p = 320/2500 = 0.128

α = 1 - 0.95 = 0.05 / 2 = 0.025

$Z_{0.025}$ = 1.96

$$\left[\overline{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}, \overline{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right].$$

[0.128 - (1.96 * √(0.128 * (1 - 0.128) / 2500)), …. + ….]

Una vez resolvemos los cálculos anteriores el intervalo de confianza será:

{0.11490, 0.141096}

Esto quiere decir que como mínimo el 11,5% de las empresas españolas habían realizado compras por internet y como máximo un 14%. Por lo que el instituto Nacional de Estadística lo ha estimado un poquito de más diciendo que es un 15%.

Ejercicio 5.9: Basándote en el fichero de datos car.sav, calcula los intervalos de confianza al 95 y 98 por ciento para el porcentaje de coches del año 78 que son originarios de Japón.

Para ello filtramos por fecha y sacamos el número total de coches que son originarios de japón:

País de origen

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	EE.UU.	22	61,1	61,1	61,1
	Europa	6	16,7	16,7	77,8
	Japón	8	22,2	22,2	100,0
	Total	36	100,0	100,0	

n = 36
p = 8/36 = 0.222
α = 1 - 0.95 = 0.05
Z0.025 = 1.96

$$\left[\overline{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}, \overline{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right].$$

[0.222 - 1.96 * √(0.22 * (1 - 0.222) / 36, … + …]

Para el intervalo de confianza del 98%:
α = 1 - 0.98 = 0.02 / 2 = 0.01
Z0.01 = 2,33

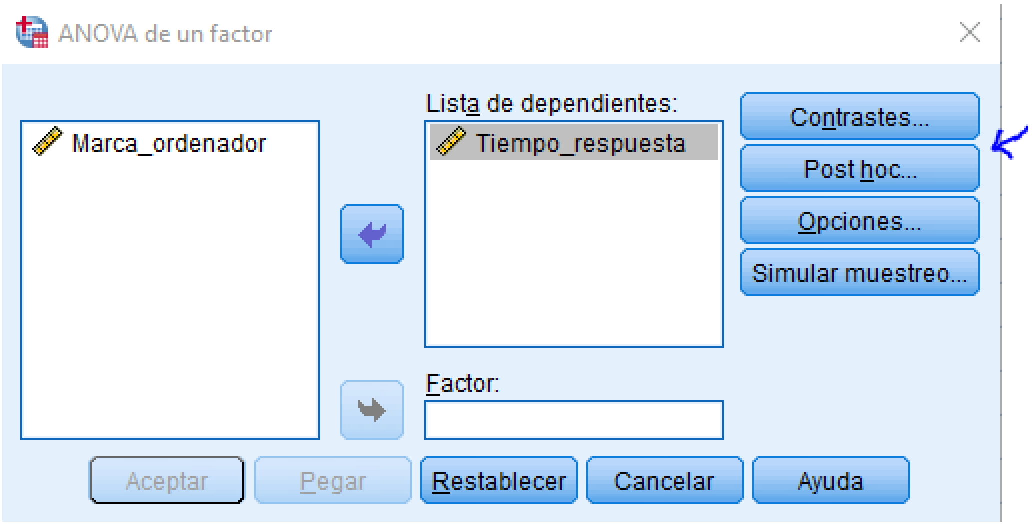
[0.222 - 2.33 * √(0.22 * (1 - 0.222) / 36, … + …]

- Para el intervalo de confianza del 95%: {0.08467, 0.35532}
- Para el intervalo de confianza del 98%: {0.059134, 0.38086}

El intervalo 1 es **0.27065** de diferencia y el intervalo 2 es **0.3217** por lo que el intervalo dos tiene mayor amplitud que el 1. Esto es debido a que para un intervalo del 98% de confianza hay más porcentaje de coches originados de japón.

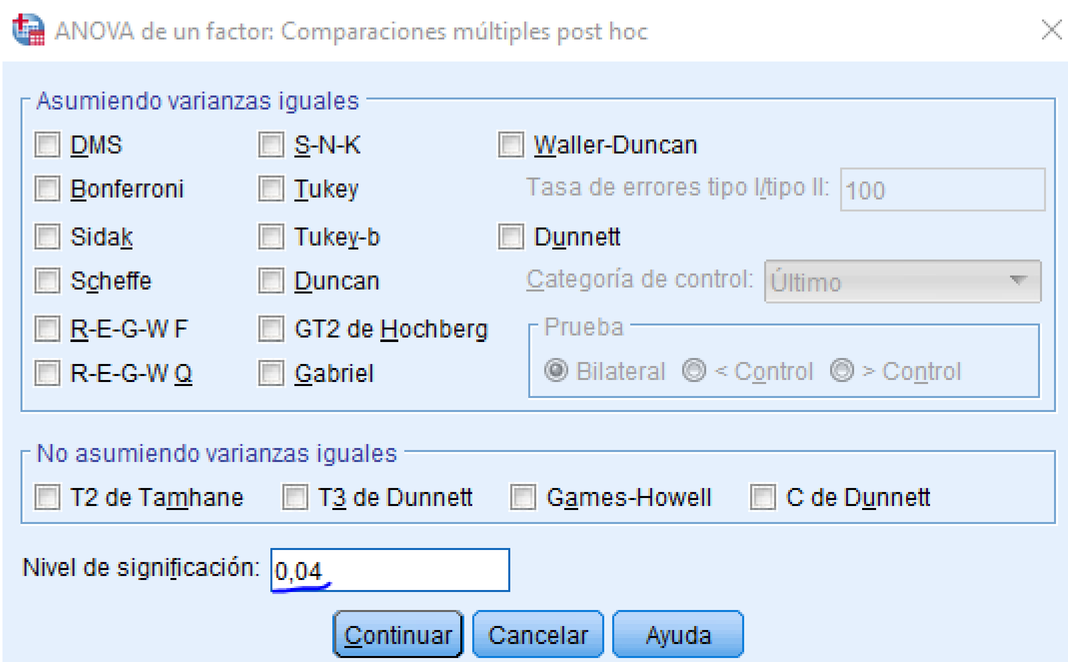
Ejercicio 5.10: Basándote en el fichero de datos tiemporespuesta.sav y bajo el supuesto de normalidad calcula los intervalos de confianza al 96 por ciento para el tiempo medio de respuesta de cada una de las marcas de los ordenadores. Interpreta y compara los resultados obtenidos.

Para ello primero seleccionamos **Analizar → Comparar medias → Anova de un factor**

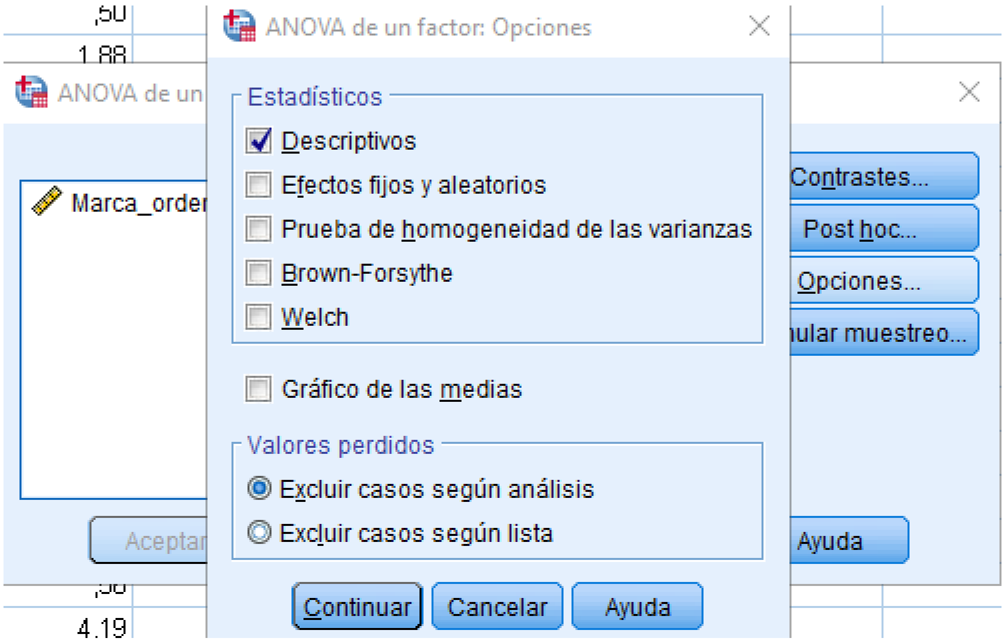


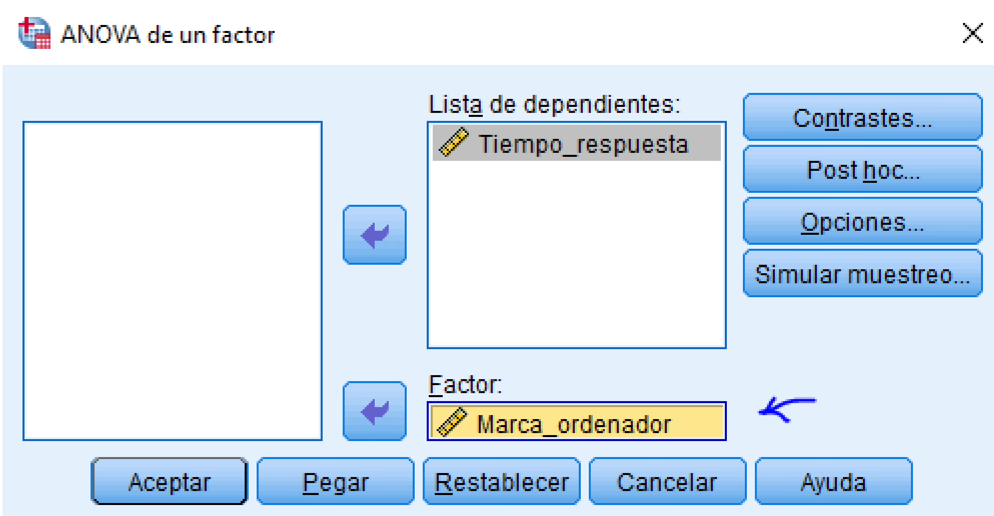
A continuación, seleccionamos **Post hoc…**

Como nos piden el nivel de significación 1-0.96 = 0.04:



Para un intervalo de confianza del 96 y nos vamos a opciones donde seleccionaremos descriptivos:





Continuamos y aceptamos para que el SPSS nos genere el estadístico descriptivo de lo que estamos buscando:

	N	Media	Desv. Desviación	Desv. Error	Límite inferior	Límite superior	Mínimo	Máximo
marca A	26	2,3373	2,56973	,50397	1,2994	3,3752	,13	10,26
Marca B	24	2,5421	2,34343	,47835	1,5525	3,5316	,18	9,15
Marca C	30	2,7167	2,77445	,50654	1,6807	3,7527	,13	13,51
Marca D	20	4,6375	2,66850	,59670	3,3886	5,8864	1,59	9,99
Total	100	2,9603	2,70102	,27010	2,4244	3,4962	,13	13,51

Y como hemos puesto por marca (factor) nos dará para cada marca el intervalo de confianza buscado.

Con esto podemos observar que la *marca A* consta con un intervalo de confianza del { 1.2994 , 3.3752 }. Por tanto, tenemos una confianza del 96% por ciento de que entre el 1,3 y el 3.375 está el tiempo de respuesta medio de la *marca A*.

Lo mismo pasa con las demás marcas:

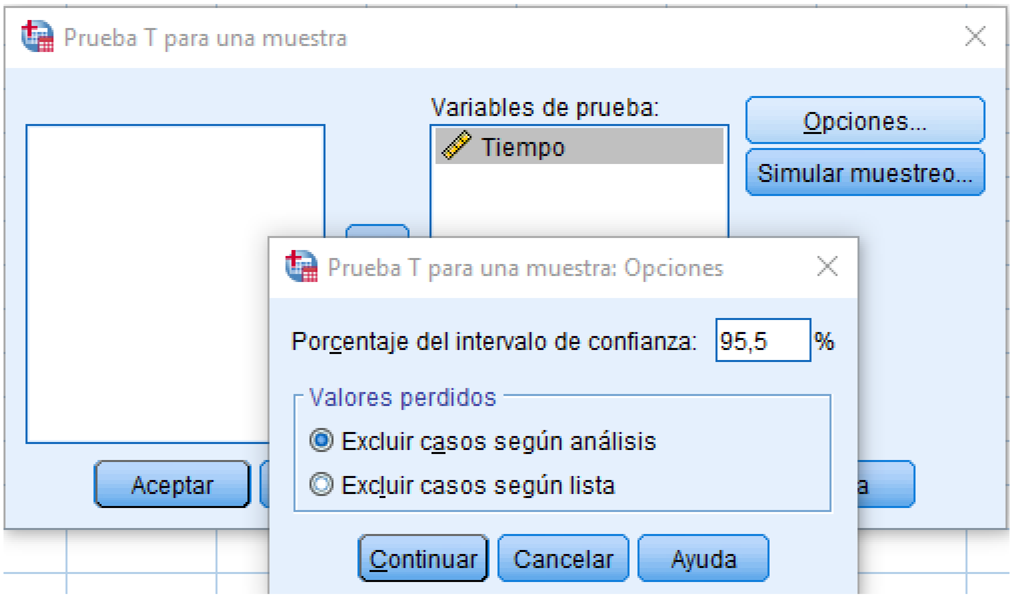
- *Marca B*: tenemos una confianza del 96% de que entre el 1,55 y el 3,53 está el tiempo de respuesta medio de la marca B.
- *Marca C*: tenemos una confianza del 96% de que entre el 1,68 y el 3,75 está el tiempo de respuesta medio de la marca C.
- *Marca D*: tenemos una confianza del 96% de que entre el 3,38 y el 5,88 está el tiempo de respuesta medio de la marca D.

Ejercicio 5.11: Para estimar la duración de un trayecto se realiza diez veces en distintas circunstancias, obteniéndose los siguientes tiempos en minutos: 28, 40, 30, 20, 32, 34, 19, 28, 23, 32. Construir un intervalo de confianza al 95.5% para el tiempo medio que requiere este trayecto.

Añadimos los datos al SPSS de la siguiente forma:

	Tiempo
1	28,00
2	40,00
3	30,00
4	20,00
5	32,00
6	34,00
7	19,00
8	28,00
9	23,00
10	32,00

Una vez añadimos los datos como desconocemos la desviación tenemos que ir a **Analizar** → **Comparar medias** → **Prueba T para una muestra**:



Nos dará el siguiente intervalo de confianza:

Prueba para una muestra

	Valor de prueba = 0					
					95,5% de intervalo de confianza de la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
Tiempo	13,875	9	,000	28,60000	23,8043	33,3957

Por lo que para un intervalo de confianza del 95.5% el tiempo medio del trayecto estará entre 23,80 y 33,39 minutos.

Ejercicio 5.12: En una muestra de n personas seleccionada por muestreo aleatorio simple de una población, se ha obtenido que el número medio de asistencias al cine durante un mes es 3.5, con desviación típica muestral de 3 veces. Se desea construir un intervalo de confianza del 95% para el parámetro asistencia media mensual al cine de la población. Tomando:
a) n= 100
b) n=20
c) n=6

Ejercicio hecho en la clase de teoría

Ejercicio 5.13: Una firma comercial encuesta a 100 individuos para conocer sus opiniones sobre la elección de dos productos alternativos A y B recientemente fabricados. El resultado de la encuesta arroja que el producto A lo han elegido 55 individuos y el producto B 45. Hallar un intervalo de confianza al 95% para la proporción de los que eligen cada producto

Ejercicio hecho en la clase de teoría

Ejercicio 5.14: En una muestra de tamaño 64, tomada de una determinada población, se estudia una característica X y se sabe que su media es 1012. La característica X en la población sigue una distribución normal con desviación típica 25. Hallar intervalos de confianza para el valor medio de la característica X con coeficientes de confianza del 95% y del 90%.

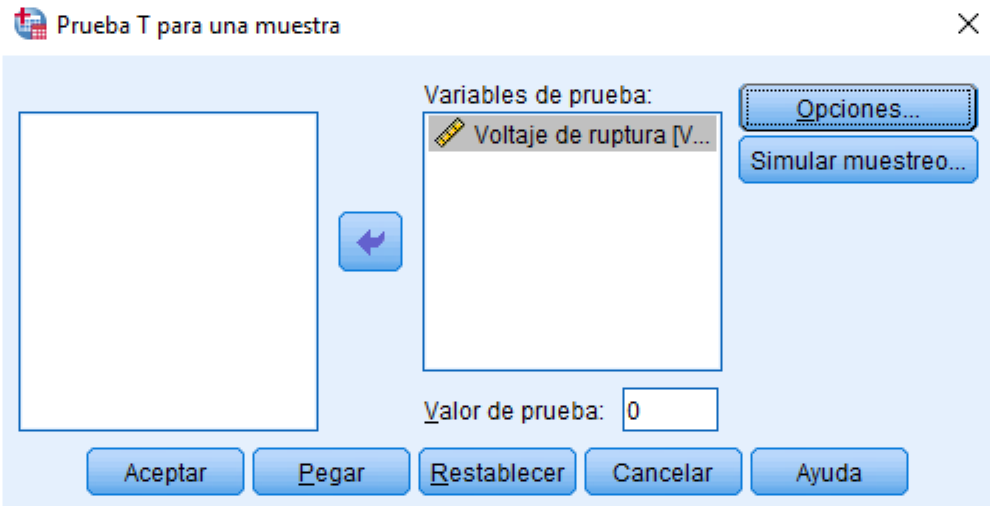
Ejercicio hecho en la clase de teoría

Ejercicio 5.15: Se leyó la siguiente información sobre el voltaje de ruptura de circuitos eléctricamente cargados de una distribución normal:
1470, 1510, 1690, 1740, 1900, 2000, 2030, 2100, 2190, 2200, 2290, 2380, 2390, 2480, 2500, 2580, 2700
Calcular un intervalo de confianza al 95% para la variabilidad del voltaje de ruptura.

Añadimos los datos de la muestra en el SPSS:

	Voltaje
1	1470
2	1510
3	1690
4	1740
5	1900
6	2000
7	2030
8	2100
9	2190
10	2200
11	2290
12	2380
13	2390
14	2480
15	2500
16	2580
17	2700

Una vez tenemos añadidos los datos como es una muestra solamente nos vamos a **Analizar → Comparar medias → Prueba T para una muestra**



Nos metemos en opciones para ver si está correctamente el intervalo de confianza que queremos calcular (predeterminado pone 95%).

Una vez hemos aceptado nos saldrá el siguiente intervalo de confianza del 95%;

	Valor de prueba = 0					
					95% de intervalo de confianza de la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
Voltaje de ruptura	23,660	16	,000	2126,471	1935,94	2317,00

Esto quiere decir que el voltaje de ruptura de un intervalo de confianza del 95% está entre el 1936 y el 2317.