

ESTADÍSTICA – 1º AMBIENTALES

TERCERA PRÁCTICA: INTERVALOS DE CONFIANZA & CONTRASTES DE HIPÓTESIS (GUIADA & RESUELTA)

El objetivo de esta práctica es calcular Intervalos de Confianza (IC) y Contrastes de Hipótesis (CH) con el programa SPSS. Para ello, se plantean tres ejercicios: Primeramente se plantea el problema de calcular IC & CH para la media de una población normal (con varianza desconocida) y posteriormente se estudian los problemas relacionados con la comparación de dos muestras.

Antes de comenzar dos observaciones importantes:

1. El programa SPSS resuelve simultáneamente y en la misma tabla el IC y el CH.

2. El programa SPSS resuelve contrastes de hipótesis bilaterales usando el p-valor.

Descripción de los datos: Para la resolución de los ejercicios se utilizarán los datos que aparecen en los dos siguientes ficheros:

metabolismo.sav: las variables contienen la tasa metabólica y la masa corporal magra de una muestra de 7 hombres y 12 mujeres.

Ozono.sav: Las variables Enero y Febrero corresponden al grosor de la capa de ozono en cierta zona en los meses de enero y febrero en 15 años seleccionados aleatoriamente.

Descarga los dos ficheros de Moodle al disco duro del ordenador (por ejemplo, al escritorio). Para ello, y una vez en moodle, posicionar el ratón sobre cada uno de ellos, pinchar con el botón derecho y marcar *Guardar enlace...*

PRIMERA PARTE: INTERVALO DE CONFIANZA Y CONTRASTES DE HIPÓTESIS PARA LA MEDIA DE UNA POBLACIÓN

Se utiliza el archivo **metabolismo.sav**.

Queremos saber qué información podemos deducir sobre la *Tasa Metabólica media* poblacional a partir de los datos muestrales. Para ello calcularemos **Intervalos** y **Contrastes** de dicho parámetro, basados en los datos que disponemos y haciendo la hipótesis de que la población es normal:

EJERCICIO 1:

a) Obtén la tabla con el IC y el CH al 95% (cópiala) y a partir de ella responde:

1. Escribe cuál es el IC para la *Tasa metabólica media* facilitado por la tabla.
2. ¿Cuánto vale el margen de error del IC?
3. Especifica el CH, es decir, escribe la hipótesis nula H_0 que aparece en la tabla.
4. ¿Qué p-valor has obtenido?
5. En base al p-valor, ¿aceptamos o rechazamos la hipótesis nula? ¿por qué?

b) Repite el apartado anterior al 99% y especificando $\mu_0=1300$.

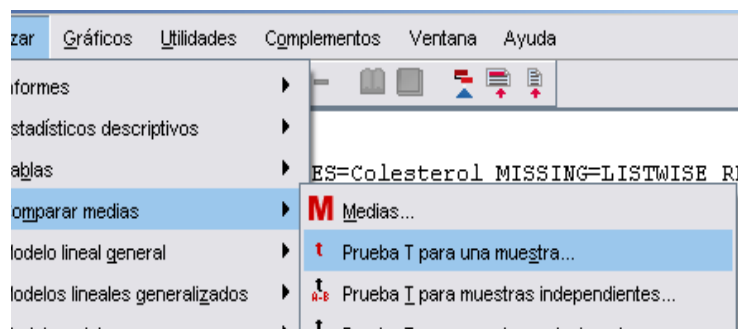
GUÍA PARA REALIZAR EL EJERCICIO 1:

1. Pinchar sobre el fichero **metabolismo.sav** para abrirlo. Si así no se abre directamente, se hace desde el desplegable “Archivo”.

Recordar de las prácticas 1 y 2 que en la parte inferior del editor aparecen dos pestañas: *Vista de datos* y *Vista de variables*. La pestaña *Vista de datos* corresponde a la pantalla anterior y nos muestra los nombres y valores de las variables. La pestaña *Vista de variables* muestra las características de cada variable (Nombre, Tipo, Anchura,...). Además aparece la segunda pantalla habitual en SPSS, el *Visor de resultados*, en el que irán apareciendo los resultados de nuestro análisis estadístico (tablas, gráficos,...). Los resultados se pueden guardar al final en un fichero de extensión .spv.

2. SPSS facilita en la misma tabla el IC y el CH. Para obtener dicha tabla seleccionamos bien en el Editor de datos o bien en el Visor de resultados (da lo mismo):

ANALIZAR → COMPARAR MEDIAS → PRUEBA T PARA UNA MUESTRA



En el cuadro VARIABLES PARA CONTRASTAR seleccionamos la variable *Tasa metabólica...* que contiene las tasas metabólicas de los individuos seleccionados.

Para especificar el **nivel de confianza** hay que seleccionar el botón OPCIONES. El utilizado por defecto es del 95 %, si queremos otro nivel lo cambiamos.

En VALOR DE PRUEBA se especifica el dato a contrastar: μ_0 ; por defecto SPSS toma el valor 0, es decir, por defecto, el programa resuelve el contraste $H_0: \mu_0=0$. **Para cambiarlo** hay que escribir el valor con el que queremos realizar el contraste

Indicaciones:

- Si sólo queréis resolver un IC debéis mantener este valor de prueba.
- Si queréis resolver un CH, en una gran mayoría de los problemas este valor no tiene ningún sentido y hay que cambiarlo.

Si pulsamos ACEPTAR, obtenemos la salida deseada:

Estadísticos para una muestra

	N	Media	Desviación típ.	Error típ. de la media
Tasa	19	1369,53	257,504	59,075

Prueba para una muestra

	Valor de prueba = 0					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
Tasa	23,183	18	,000	1369,526	1245,41	1493,64

Los datos que aparecen en la tabla Prueba para una muestra son:

$$t = (x - \mu_0) / (s / \sqrt{n}) \text{ (¡no confundir con la } t \text{ de Student!)}$$

La columna “gl” da los grados de libertad ($n-1$).

La diferencia de medias es $x - \mu_0$.

La columna “sig(bilateral)” da el p-valor del contraste $H_0 : \mu = \mu_0 ; H_1 : \mu \neq \mu_0$.

Conclusión: en general, el programa

1. Calcula el IC($\mu - \mu_0$), es decir, si $\mu_0 = 0$, calcula el IC(μ)
2. Resuelve el CH bilateral $H_0: \mu = \mu_0$ (por defecto $H_0: \mu = 0$).

La tabla Prueba para una muestra nos facilita la siguiente información sobre el apartado a):

1. **Intervalos de Confianza:** es la columna correspondiente a 95% INTERVALO DE CONFIANZA PARA LA DIFERENCIA, de donde deducimos que, a un nivel de confianza del 95% el Intervalo para la media es: $IC_{95\%}(\mu) = (1245,41; 1493,64)$

Prueba para una muestra

	Valor de prueba = 0					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
Tasa	23,183	18	,000	1369,526	1245,41	1493,64

2. El margen de error =124,116

3. **Contrastes de Hipótesis:** debemos mirar el VALOR DE LA PRUEBA que nos proporciona el μ_0 . En este caso, el programa ha resuelto, al nivel 95%, el contraste $H_0 : \mu = 0$; es decir, nos planteamos si la tasa metabólica media es de 0 (lo que no tiene ningún sentido)

Prueba para una muestra

	Valor de prueba = 0					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
Tasa	23,183	18	,000	1369,526	1245,41	1493,64

4. Debemos mirar la columna SIG. que nos proporciona el p-valor = 0,000

5. Como p-valor = $0 < \alpha = 0,05$, la respuesta a este contraste sería que rechazamos la hipótesis nula: claramente la tasa media metabólica es distinta a 0 ¡obviamente!

Para resolver el apartado b) la tabla que debéis obtener es:

Prueba para una muestra

	Valor de prueba = 1300					
	t	gl	Sig. (bilateral)	Diferencia de medias	99% Intervalo de confianza para la diferencia	
					Inferior	Superior
Tasa	1,177	18	,255	69,526	-100,52	239,57

1. $IC_{99\%}(\mu-1300)=(-100,52; 239,57)$ $IC_{99\%}(\mu)=(1300-100,52; 1300+239,57)=(1199,48; 1539,57)$

2. Margen de error=170,045

3. $H_0 : \mu = 1300$; es decir, nos planteamos si la tasa metabólica media es de 1300 (tiene más sentido que la del apartado anterior)
4. P-valor= 0,255
5. Como p-valor = 0,255 > $\alpha=0,01$ aceptamos la hipótesis nula: hay evidencias estadísticas para afirmar que la tasa media metabólica es de 1300.

Indicación: Si queremos resolver un contraste unilateral tenemos que dividir entre 2 el p-valor calculado por SPSS (y recordar que H_0 debe de ser la contraria de la que queremos “probar”).)

SEGUNDA PARTE: INTERVALOS DE CONFIANZA Y CONTRASTES DE HIPÓTESIS PARA LA DIFERENCIA DE MEDIAS DE DOS POBLACIONES NORMALES

Cuando queremos comparar las medias de dos muestras procedentes de poblaciones normales es importante distinguir si las muestras son independientes o si los datos son emparejados.

Veamos cómo llevar a cabo el análisis en cada uno de estos casos.

MUESTRAS INDEPENDIENTES:

Con los datos disponibles (datos muestrales) queremos comparar la *Tasa metabólica media poblacional de los hombres*, μ_1 , con la *Tasa metabólica media poblacional de las mujeres*, μ_2 . Obsérvese que tenemos claramente dos muestras independientes: Hombres (M) y Mujeres (F)

Ya vimos en la Práctica anterior (a través del Box-plot) que la Tasa Metabólica media **muestral** de los hombres, 1600, es superior a la Tasa Metabólica media **muestral** de las mujeres, 1235,08.

Preguntas naturales a este respecto son:

- ¿Aportan los datos muestrales evidencia suficiente para afirmar que μ_1 es siempre mayor que μ_2 ?
- ¿Es la diferencia significativa para poder afirmar lo mismo en toda la población?

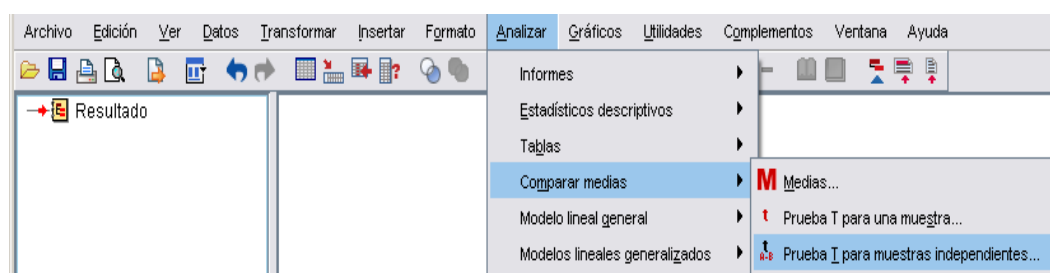
EJERCICIO 2:

- a) Obtén la tabla que da tanto el IC como el CH para la diferencia de medias al 95%
 1. Escribe cuál es el IC para la diferencia si asumimos igualdad de varianzas.
 2. Escribe cuál es el IC para la diferencia si no asumimos igualdad de varianzas.
 3. Según los IC ¿podemos afirmar que la Tasa metabólica media poblacional de los hombres es superior a la de las mujeres?
 4. Especifica el CH, es decir, escribe la hipótesis nula H_0 que aparecen la tabla.
 5. ¿Qué p-valor has obtenido?
 6. En base al p-valor, ¿aceptamos o rechazamos la hipótesis nula del contraste? ¿por qué?
- b) Repite el apartado (a) al nivel de confianza del 99%.

GUÍA PARA REALIZAR EL EJERCICIO 2

1. Para hallar el IC($\mu_1 - \mu_2$) y el CH: $\mu_1 = \mu_2$, de muestras independientes seleccionamos, en el Editor de datos o en el Visor de resultados (da lo mismo):

ANALIZAR → COMPARAR MEDIAS → PRUEBA T PARA MUESTRAS INDEPENDIENTES



En el cuadro de diálogo que aparece se selecciona la VARIABLE A CONTRASTAR: *Tasa...*; y la VARIABLE DE AGRUPACIÓN: *Sexo...*

En DEFINIR GRUPOS se indica el código asignado a cada muestra, en nuestro caso, M para la primera (los hombres) y F para la segunda (las mujeres).

Se pulsa ACEPTAR y obtenemos la salida deseada.

Prueba de muestras independientes

	Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
	F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
								Inferior	Superior
Tasa metabólica del individuo									
Se han asumido varianzas iguales	,001	,972	4,068	17	,001	364,917	89,707	175,651	554,182
No se han asumido varianzas iguales			4,062	12,632	,001	364,917	89,834	170,266	559,567

1. $IC_{95\%}(\mu_1 - \mu_2) = (175'651, 554'182)$

2. $IC_{95\%}(\mu_1 - \mu_2) = (170'266, 559'5)$

3. Las diferencias de medias, en ambas situaciones, se encuentran en un intervalo positivo, por lo tanto $\mu_1 > \mu_2$ y, efectivamente, con una confianza del 95% y acorde los datos muestrales, podemos afirmar que la tasa metabólica media de los hombres es superior a la de las mujeres.

4. $H_0: \mu_1 = \mu_2$ (el contraste plantea si tasa media metabólica de los hombres es igual a la de las mujeres)

5. P-valor=0.001

6. Como p-valor < α la respuesta a este contraste sería que rechazamos la hipótesis nula: con una confianza del 95% podemos afirmar que la tasa media metabólica de los hombres es distinta a la de las mujeres

DATOS EMPAREJADOS:

Utilizaremos el fichero **ozono.sav**.

Se quieren comparar los datos corresponden al grosor de la capa de ozono (en unidades Dobson) en cierta zona en los meses de enero y febrero en 15 años seleccionados aleatoriamente.

Obsérvese que el estudio se está haciendo sobre la misma zona los mismos 15 años, por lo tanto no podemos suponer que las muestras sean independientes.

Una unidad Dobson corresponde a una capa de ozono de 10^{-5} metros de espesor bajo una temperatura y una presión estándar

EJERCICIO 3:

a) Obtén la tabla que da tanto el IC como el CH para la diferencia de medias al 95% (cópiala)

1. Escribe cuál es el IC para la diferencia.

2. Según el IC ¿podemos afirmar que en febrero la capa de ozono tiene en media más grosor que en enero?

3. Especifica el CH, es decir, escribe la hipótesis nula H_0 que aparecen la tabla.

4. ¿Qué p-valor has obtenido?

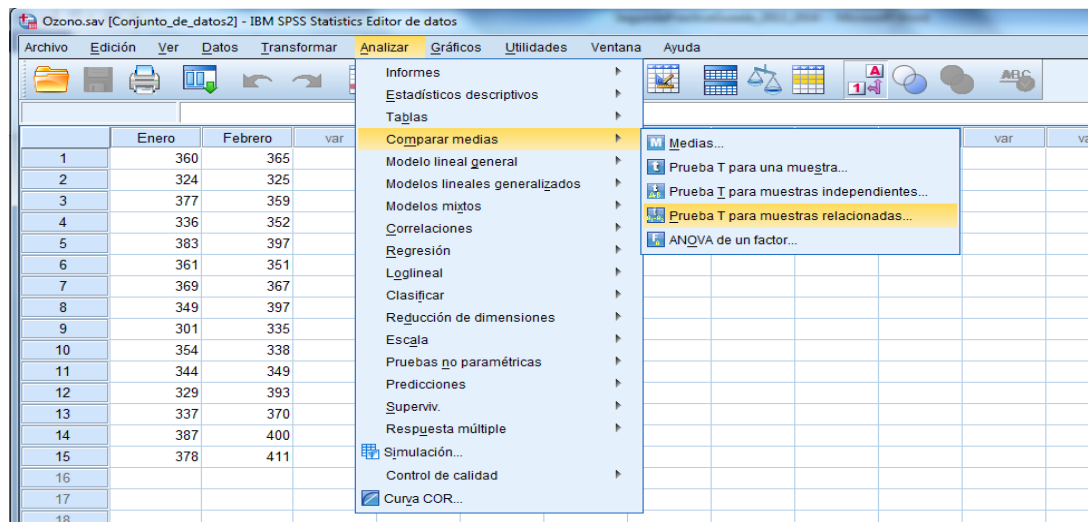
5. En base al p-valor, ¿aceptamos o rechazamos la hipótesis nula del contraste? ¿por qué?

b) Repite el apartado (a) al nivel de confianza del 99%.

GUÍA PARA REALIZAR EL EJERCICIO 3

1. Para abrir los datos de **Ozono.sav** en el fichero SPSS repetir exactamente los mismos pasos que se indicaron en el Ejercicio 1.
2. Para hallar el Intervalo de Confianza para la diferencia de medias de muestras no independientes seleccionamos, en el Editor de datos o en el Visor de resultados (da lo mismo):

ANALIZAR → COMPARAR MEDIAS → PRUEBA T PARA MUESTRAS RELACIONADAS



En el cuadro de diálogo que aparece se seleccionan las dos variables y se pasan a la primera fila del cuadro VARIABLES EMPAREJADAS.

Se pulsa ACEPTAR debemos obtener la salida deseada:

Prueba de muestras relacionadas

		Diferencias relacionadas							
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	Enero - Febrero	-14,667	23,700	6,119	-27,791	-1,542	-2,397	14	,031

1. $IC_{95\%}(\mu) = IC_{95\%}(\mu_1 - \mu_2) = (-27'791, -1'542)$.

2. La diferencia se encuentra en un intervalo negativo, por lo tanto $\mu_1 < \mu_2$ y, efectivamente, con una confianza del 95% y acorde los datos muestrales podemos afirmar que en febrero hay más grosor en la capa de ozono.

3. $H_0: \mu_1 = \mu_2$ transformado a $H_0: \mu = 0$ (el contraste plantea si el grosor de la capa de ozono en enero es el misma que en febrero)

4. P-valor=0.031

5. Como p-valor < α la respuesta a este contraste sería que rechazamos la hipótesis nula: con una confianza del 95% podemos afirmar que el grosor medio de la capa de ozono en los meses de enero y febrero son diferentes