Práctica 2: Intervalos de confianza y contrastes con SPSS[1]

El objetivo de esta práctica es aprender a calcular intervalos de confianza y a contrastar hipótesis con el programa SPSS. Comenzaremos con el problema de calcular un intervalo de confianza y contrastar hipótesis relativas a la media de una población normal (con varianza desconocida) y posteriormente estudiaremos los problemas relacionados con la comparación de dos muestras.

Utilizaremos los datos del fichero `metabolismo.sav` que ya conocemos. Recordemos que las variables contienen la tasa metabólica y la masa corporal magra de una muestra de 7 hombres y 12 mujeres.

1. IC y contrastes para la media

En este apartado estudiamos cómo calcular intervalos y contrastes para la tasa metabólica media, basados en los datos de los que disponemos y haciendo la hipótesis de que la población es normal.

Para ello, una vez que hemos abierto el fichero de datos que queremos analizar y lo tenemos a la vista en el "Editor de datos":

- Se selecciona `Analizar` ↪ `Comparar medias` ↪ `Prueba T para una muestra...`
- En el cuadro `Variables para contrastar` seleccionamos la variable que contiene las tasas metabólicas.
- El nivel de confianza utilizado por defecto es del 95 %. Para cambiarlo, hay que seleccionar el botón `Opciones`.
- El `Valor de prueba` corresponde a μ_0 en el contraste de hipótesis. Por defecto toma el valor cero.

Si pulsamos `Aceptar`, obtenemos la siguiente salida:

[1] Para escribir estas notas se ha utilizado la versión SPSS 17

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

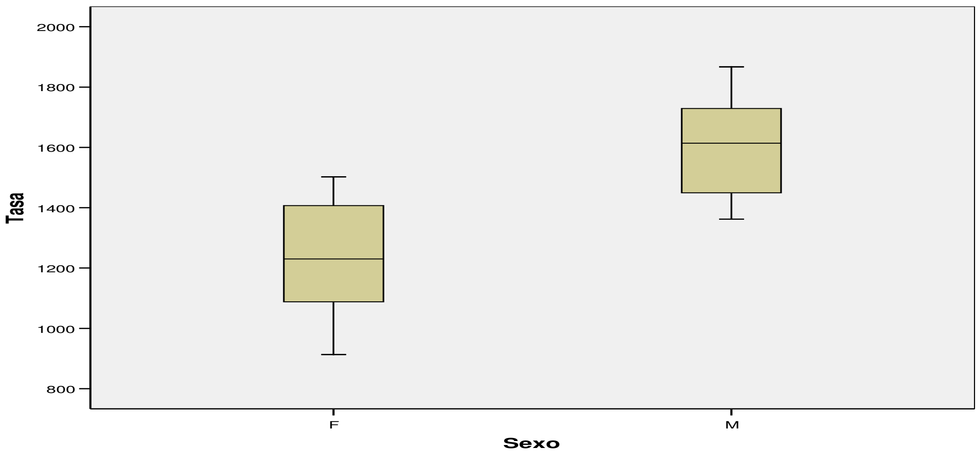
En la tabla anterior, $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$, donde μ_0 es el valor de prueba. La columna "gl" da los grados de libertad $(n-1)$. La diferencia de medias es $\bar{x} - \mu_0$. La columna "sig(bilateral)" da el p-valor del contraste $H_0: \ \mu = \mu_0$. En general, el programa calcula un intervalo de confianza para $\mu - \mu_0$, es decir, si $\mu_0 = 0$, calcula un intervalo de confianza para μ. En este caso, el intervalo es [1245,41, 1493,64].

2. Intervalos de confianza y contrastes para la diferencia de medias

Cuando se dispone de dos muestras procedentes de poblaciones normales y queremos comparar las medias, es conveniente primero representar diagramas de cajas para hacernos una idea aproximada de las diferencias y semejanzas entre las dos muestras. Posteriormente, se puede calcular un intervalo de confianza para la diferencia de medias $\mu_1 - \mu_2$ o contrastar cualquier hipótesis para compararlas. Es importante distinguir si las dos muestras son independientes o si los datos son emparejados. Veamos cómo llevar a cabo el análisis en cada uno de estos casos.

2.1. Muestras independientes

Supongamos que queremos, con los datos disponibles, comparar la tasa metabólica media de los hombres μ_1 con la de las mujeres μ_2. ¿Aportan los datos evidencia suficiente para afirmar que μ_1 es mayor que μ_2? Si representamos un diagrama de cajas, vemos que en la muestra disponible la media de los hombres es superior a la de las mujeres. ¿Es la diferencia significativa para poder afirmar lo mismo en la población?



- Una vez que tenemos los datos preparados, se selecciona

`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 y F para la segunda).

- Se pulsa `Aceptar`.

Debemos obtener la salida siguiente:

Estadísticos de grupo

	Sexo	N	Media	Desviación típ.	Error típ. de la media
Tasa	M	7	1600,00	189,240	71,526
	F	12	1235,08	188,283	54,353

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Tasa	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

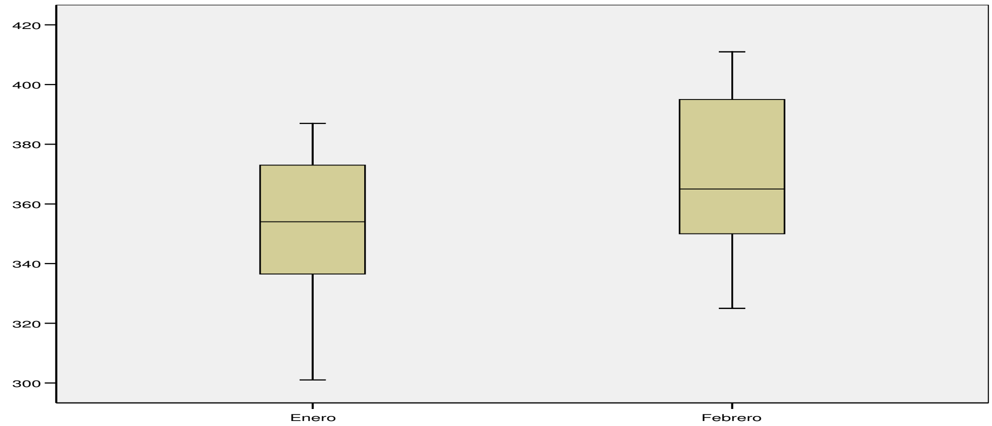
2.2. Datos emparejados

Los siguientes datos corresponden al grosor de la capa de ozono (en unidades Dobson [2]) en cierta zona en los meses de enero y febrero en 15 años seleccionados aleatoriamente.

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

Enero	360	324	377	336	383	361	369	349	301	354	344	329	337	387	378
Febrero	365	325	359	352	397	351	367	397	335	338	349	393	370	400	411

Si representamos diagramas de cajas de estos datos, obtenemos:



En esta muestra se obtiene un mayor grosor de la capa de ozono en febrero que en enero. ¿Es la diferencia significativa como para afirmar a nivel $\alpha = 0{,}05$ que en febrero la capa de ozono tiene en media más grosor que en enero en general?

Como las medidas de los dos meses corresponden al mismo año, los datos son emparejados y no podemos suponer que las muestras sean independientes. Para llevar a cabo el contraste hay que proceder de la siguiente forma:

- Creamos dos variables, llamadas `Enero` y `Febrero` que contengan los datos.
- Una vez que tenemos los datos preparados, se selecciona

  `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 siguiente:

Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. de la media
Par 1	Enero	352,6000	15	24,47681	6,31989
	Febrero	367,2667	15	26,74073	6,90443

Correlaciones de muestras relacionadas

		N	Correlación	Sig.
Par 1	Enero y Febrero	15	,575	,025

Prueba de muestras relacionadas

		Diferencias relacionadas					t	gl	Sig. (bilateral)
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior	Superior			
Par 1	Enero - Febrero	-14,66667	23,69951	6,11919	-27,79102	-1,54232	-2,397	14	,031

3. Cuestiones

1. Calcula dos intervalos de confianza para la tasa metabólica media (con niveles de confianza del 90 % y del 99 %). ¿Cuánto valen los márgenes de error de ambos intervalos?

2. Contrasta si los datos aportan suficiente evidencia para afirmar que la tasa metabólica media es mayor que 1300.

3. A nivel $\alpha = 0{,}05$, ¿podemos afirmar que la tasa metabólica media de los hombres es superior a la de las mujeres?

4. ¿Cuál es el p-valor del contraste de la cuestión anterior?

5. Para los datos utilizados, ¿cuánto vale el estimador combinado de la varianza (que se suele denotar por S_p^2)?

6. Repite los cálculos realizados para comparar las tasas metabólicas de hombres y mujeres, pero ahora fijando un nivel de confianza del 90 %. ¿Qué es lo que cambia y qué permanece constante en la salida respecto al caso anterior?

7. ¿Cuánto vale el coeficiente de correlación entre el grosor de la capa de ozono en enero y en febrero? ¿Apoya este valor el supuesto de que las muestras no son independientes?

8. Si d_i es la diferencia entre el grosor de la capa de ozono de enero y febrero en el año i, ¿cuánto valen la media y la desviación típica de $d_1, \ldots, d_{15}$?

9. ¿Podemos afirmar, a nivel $\alpha = 0{,}05$, que en febrero la capa de ozono tiene en media más grosor que en enero? ¿Y a nivel $\alpha = 0{,}02$?