

CONTRASTES DE HIPÓTESIS

El objetivo de esta práctica es ilustrar las diferentes técnicas de contraste de hipótesis mediante el programa SPSS. Comenzaremos con el problema de contrastar si unos datos provienen de una población normal, para posteriormente estudiar algunos ejemplos de contrastes de medias. Utilizaremos los datos del fichero `ei-x3.txt` que contiene los tiempo de ejecución (en segundos) de 40 trabajos procesados en un centro de cálculo.

1 ¿Cómo introducir datos en SPSS/PC?

A continuación se ilustra como introducir los datos, para el ejemplo `ei-x3.txt`:

1. `Archivo` ↔ `Abrir` ↔ `Datos`.
2. En el cuadro de diálogo que aparece, en `Archivos de tipo`, elegir la extensión `Texto (*.txt)`, seleccionar el archivo `ei-x3.txt` ↔ `Abrir`.
3. En el cuadro de `Asistente para importación de texto`, elegimos las opciones teniendo en cuenta que los datos están almacenado: 1 caso por línea, y en el fichero no aparece el nombre de las variables.
4. Para cambiar el nombre a las variables, seleccionamos `Ver` ↔ `Variabes`, y cambiamos `v1` por `tiempo`.
5. Por último, `Archivo` ↔ `Guardar`.

2 Contraste de Kolmogorov - Smirnov

En este apartado estudiamos cómo contrastar la hipótesis de que los datos provienen de una población normal, i.e. $H_0 : X \sim \mathcal{N}(\mu, \sigma^2)$ mediante el contraste de Kolmogorov - Smirnov. 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` ↔ `Pruebas no paramétricas` ↔ `KS de 1 muestra ...`

- Pasamos la variable tiempo al cuadro **Contrastar variables**. Notemos que en el cuadro **Distribución de contraste** podemos seleccionar Normal, Uniforme, Binomial o Poisson.

Si pulsamos **Aceptar**, la salida del SPSS es la siguiente:

Prueba de Kolmogorov-Smirnov para una muestra

		TIEMPO
N		40
Parámetros normales ^{a,b}	Media	23,950
	Desviación típica	19,779
Diferencias más extremas	Absoluta	,162
	Positiva	,162
	Negativa	-,123
Z de Kolmogorov-Smirnov		1,027
Sig. asintót. (bilateral)		,242

- a. La distribución de contraste es la Normal.
 b. Se han calculado a partir de los datos.

El estadístico de contraste Z de Kolmogorov - Smirnov¹, cuya expresión es $Z = \sup_{x \in \mathbb{R}} |F_n(x) - \Phi_{\mu=\bar{x}, \sigma^2=s^2}(x)|$, donde F_n es la función de distribución empírica de la muestra, y $\Phi_{\mu=\bar{x}, \sigma^2=s^2}$ es la función de distribución de la normal $\mathcal{N}(\mu = \bar{x}, \sigma^2 = s^2)$, queda en el ejemplo $Z = 1.027$ con un p -valor igual a 0.242, de manera que no rechazamos la hipótesis nula de normalidad.

3 Contraste de hipótesis de la media

En este apartado estudiamos cómo contrastar si la media de los tiempos de ejecución es igual a un valor dado, por ejemplo 25 segundos, es decir $H_0 : \mu = 25$, suponiendo que la población es normal:

- Se selecciona **Analizar** \leftrightarrow **Comparar medias** \leftrightarrow **Prueba T para una muestra ...**
- Pasamos la variable tiempo al cuadro **Variabes a contrastar** y en el cuadro **Valor de prueba** escribimos 25.
- El nivel de confianza utilizado por defecto es del 95%. Para cambiarlo, hay que seleccionar el botón **Opciones**.

Si pulsamos **Aceptar**, la salida del SPSS es la siguiente:

¹Recordemos que el contraste de Kolmogorov - Smirnov necesita especificar completamente la distribución bajo H_0 , es decir deberíamos especificar el valor poblacional de la media, μ , y de la varianza, σ^2 . En SPSS los valores de la media y la varianza son estimados a partir de los datos, con lo cual la distribución del estadístico de contraste se modifica y se utilizan los ajustes empíricos de Lilliefors.

Estadísticos para una muestra

	N	Media	Desviación típ.	Error típ. de la media
TIEMPO	40	23,950	19,779	3,127

Prueba para una muestra

	Valor de prueba = 25					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
TIEMPO	-,336	39	,739	-1,050	-7,376	5,276

La fórmula del estadístico de contraste $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$, donde μ_0 es el valor de prueba. La columna `gl` da los grados de libertad, en este caso `gl = 39`. La columna `sig(bilateral)` da el p -valor = $\Pr\{|t_{39}| \geq 0.336\}$, en este caso obtenemos 0.739, de manera que no rechazaremos la hipótesis nula a los niveles usuales ($\alpha = 0.01, 0.05$ ó 0.1). El programa calcula un intervalo de confianza para $\mu - \mu_0$, es decir, si $\mu_0 = 0$, calcula un intervalo de confianza para μ .

4 Intervalos de confianza para la diferencia de medias

Supongamos que los primeros 20 tiempos corresponden a trabajos de alumnos de primer ciclo, y los restantes a alumnos de segundo ciclo, y queremos contrastar si los tiempos medios de ejecución son iguales en ambos tipos de alumno. Para realizar el contraste de medias se procede de la siguiente forma:

- Las mediciones de ambos grupos se han de incluir en la misma variable, que podemos denominar `tiempo`. A continuación hay que definir una variable de agrupación o codificación (a la que llamaremos `alumno`) que tenga, 20 unos en las 20 primeras observaciones y doses en los lugares restantes. El papel de esta variable es identificar los datos que pertenecen a cada uno de los dos grupos.
- 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 `tiempo` y la variable de agrupación `alumno`. En `Definir grupos` se indica el código asignado a cada muestra (en nuestro caso, 1 para la primera y 2 para la segunda) ↔ `Continuar` ↔ `Aceptar`.

Estadísticos de grupo

ALUMNO	N	Media	Desviación típ.	Error típ. de la media
TIEMPO 1	20	21,250	19,518	4,364
2	20	26,650	20,168	4,510

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
TIEMPO	Se han asumido varianzas iguales	1,227	,275	-,860	38	,395	-5,400	6,276	-18,105	7,305
	No se han asumido varianzas iguales			-,860	37,959	,395	-5,400	6,276	-18,105	7,305

En http://www.uam.es/personal_pdi/ciencias/aalonso/Docencia/e-ii.html se encuentran los ficheros de datos ASCII: ei-x3.txt a utilizar en esta práctica.