

Prueba Kolmogorov-Smirnov para una muestra:

Es una prueba de bondad de ajuste.

Se emplea en una muestra independiente.

El tipo de variable es cuantitativa continua (debe ser medida en escala al menos ordinal).

Esta prueba responde a la pregunta: ¿Ajusta la distribución empírica de datos muestrales de una variable ordinal o cuantitativa a una distribución teórica conocida?

Esta prueba no requiere que los datos sean agrupados, lo que permite que ésta haga uso de toda la información del conjunto de datos. Puede utilizarse con muestras de cualquier tamaño (mientras que la X^2 requiere que las muestras tengan un tamaño mínimo).

Hipótesis:

$H_0: F(x) = F_T(x)$ para toda x desde $-\infty$ hasta $+\infty$

$H_1: F(x) \neq F_T(x)$ para al menos una x

Como es una prueba de bondad de ajuste aquí interesa no rechazar la hipótesis nula, es decir, interesa que el valor de p sea mayor de 0,05 para no rechazar la hipótesis nula (queremos que $p > 0,05$).

Ejemplo: Se efectuaron mediciones del nivel de glucemia de 36 hombres adultos en ayuno, no obesos y aparentemente sanos. Estas mediciones se muestran en la tabla que se presenta. Se pretende saber si es posible concluir que tales datos no pertenecen a una población que sigue una distribución normal, con una media de 80 y una desviación típica de 6. Emplee un $\alpha = 0,05$.

Valores de glucemia en 36 varones sanos					
75	92	80	80	84	72
84	77	81	77	75	81
80	92	72	77	78	76
77	86	77	92	80	78
68	78	92	68	80	81
87	76	80	87	77	86

Respuesta:

Supuestos: La muestra disponible es una muestra aleatoria simple que se extrajo de una población que sigue una distribución continua.

Hipótesis:

$H_0: F(x) = FT(x)$ para toda x desde $-\infty$ hasta $+\infty$

$H_1: F(x) \neq FT(x)$ para al menos una x

Ahora abrimos el programa SPSS y vamos a la Vista de datos para introducir los datos del ejemplo. Debe quedarles así:

The screenshot shows the SPSS 'Data View' window. The title bar reads '1-Base de datos. Prueba K-S para1 mue'. The menu bar includes 'Archivo', 'Edición', 'Ver', 'Datos', and 'Trans'. Below the menu is a toolbar with icons for file operations and data management. The main area displays a table with 23 rows and 2 columns. The first column is labeled '1' and the second column is labeled 'glucemia'. The values in the 'glucemia' column are: 75, 92, 80, 80, 84, 72, 84, 77, 81, 77, 75, 81, 80, 92, 72, 77, 78, 76, 77, 86, 77, 92, 80. At the bottom, there are two buttons: 'Vista de datos' (highlighted in yellow) and 'Vista de variables'.

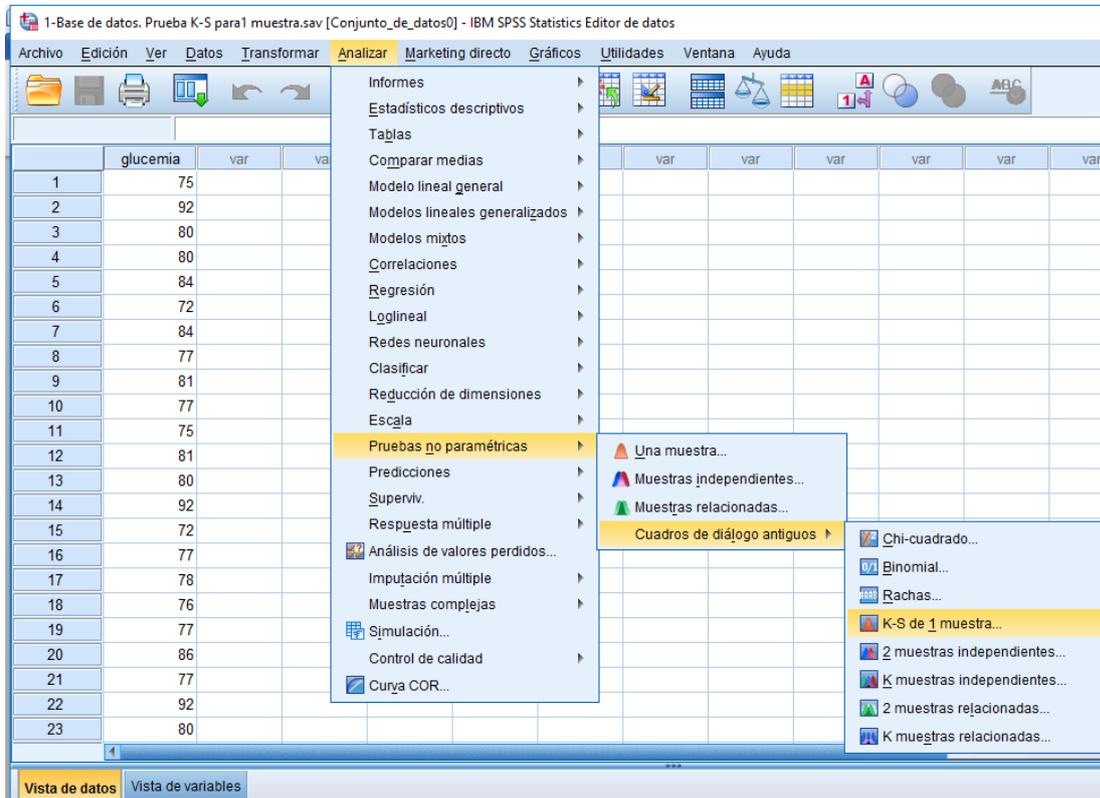
	glucemia	var
1	75	
2	92	
3	80	
4	80	
5	84	
6	72	
7	84	
8	77	
9	81	
10	77	
11	75	
12	81	
13	80	
14	92	
15	72	
16	77	
17	78	
18	76	
19	77	
20	86	
21	77	
22	92	
23	80	

Ahora en la Vista de variables introducimos el nombre de la variable, su etiqueta (que es el nombre real de nuestra variable) y ponemos en Medida la Escala pues se trata de una variable cuantitativa continua. Debe quedarles así:

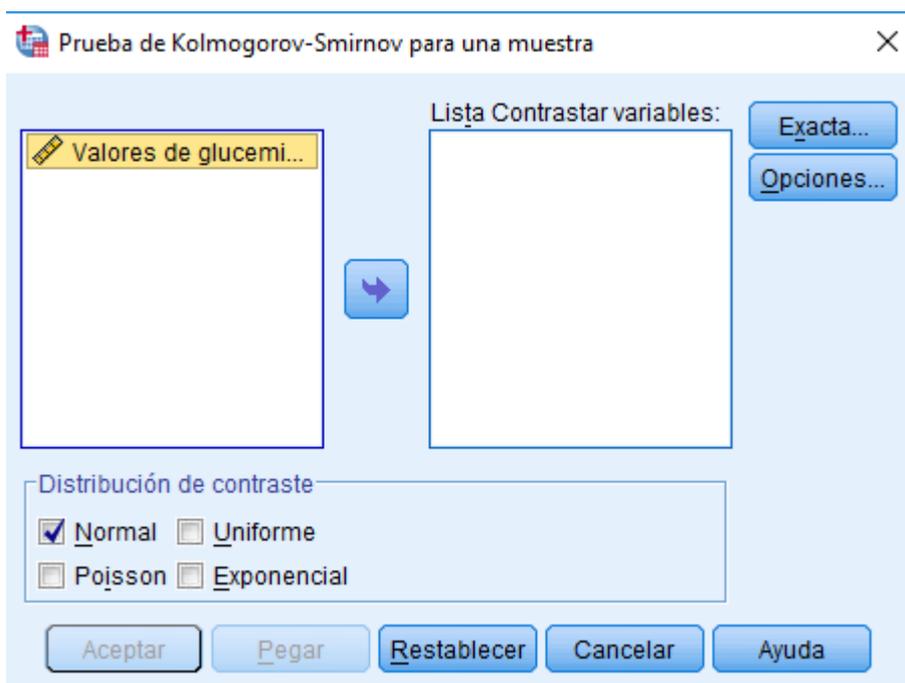
The screenshot shows the SPSS 'Variable View' window. The title bar reads '1-Base de datos. Prueba K-S para1 muestra.sav [Conjunto_de_datos0] - IBM SPSS Statistics Editor de datos'. The menu bar includes 'Archivo', 'Edición', 'Ver', 'Datos', 'Transformar', 'Analizar', 'Marketing directo', 'Gráficos', 'Utilidades', 'Ventana', and 'Ayuda'. Below the menu is a toolbar with various icons. The main area displays a table with 2 columns: 'Nombre' and 'glucemia'. The values in the 'glucemia' column are: Numérico, 8, 0, Valores de glucemia, Ninguna, Ninguna, 8, Derecha, Escala, Entrada. At the bottom, there are two buttons: 'Vista de datos' and 'Vista de variables'.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	glucemia	Numérico	8	0	Valores de glucemia	Ninguna	Ninguna	8	Derecha	Escala	Entrada

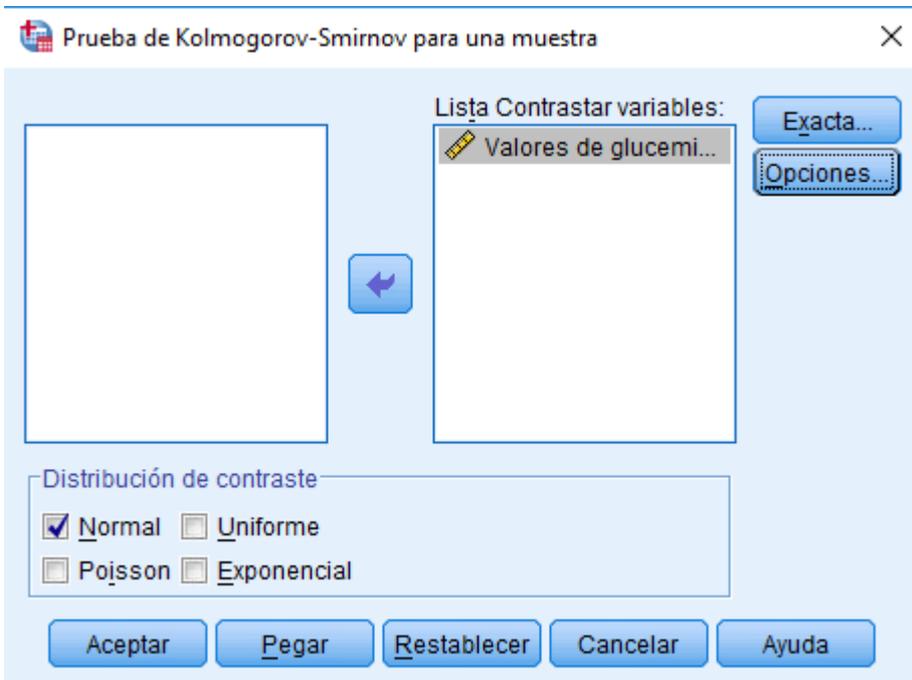
Ahora vamos al menú Analizar y damos un clic ahí, luego saldrá un menú desplegable y nos desplazamos hacia debajo de ese menú con el mouse y donde dice Pruebas no paramétricas nos paramos y saldrá otro menú desplegable ahora hacia la derecha y con el mouse vamos hacia donde dice Cuadros de diálogo antiguos y nos paramos ahí y entonces saldrá otro menú desplegable que dice K-S de 1 muestra y ahí damos un clic. Debe quedarles así:



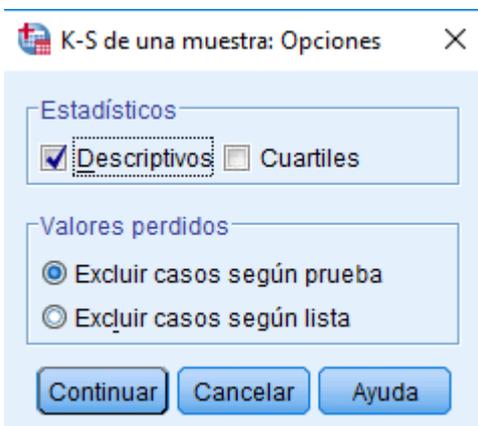
Ahora saldrá una ventana como la siguiente:



Ahora activaremos la variable nuestra (Valores de glucemi...) dándole un clic sobre ella y la pondremos en el cuadro de la derecha debajo donde dice Lista Contrastar variables:. Deben dejar activada la opción Normal, es decir, no desmarcarla. Debe quedarles así:



Ahora en el botón Opciones... pueden darle un clic y marcar con otro clic donde dice Descriptivos:



Luego dan clic en el botón Continuar. Ahora dan clic en Aceptar y saldrán los siguientes resultados:

Pruebas no paramétricas

[Conjunto_de_datos0] C:\Users\Silvia\Desktop\Manual de ejercicios de SPSS\Pruebas no paramétricas\1-Para variables cuantitativas\1-Una sola muestra\Kolmogorov-Smirnov\1-Base de datos. Prueba K-S para muestra.sav

Estadísticos descriptivos					
	N	Media	Desviación típica	Mínimo	Máximo
Valores de glucemia	36	80,08	6,199	68	92

Prueba de Kolmogorov-Smirnov para una muestra

		Valores de glucemia
N		36
Parámetros normales ^{a,b}	Media	80,08
	Desviación típica	6,199
	Absoluta	,163
Diferencias más extremas	Positiva	,163
	Negativa	-,095
Z de Kolmogorov-Smirnov		,981
Sig. asintót. (bilateral)		,291

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Interpretación: en la tabla titulada Estadísticos descriptivos vemos que la media de los valores de glucemia de los pacientes estudiados fue de 80,08 (no se les olvide poner aquí la unidad de medida de la variable), con una desviación estándar de 6,199 ((no se les olvide poner la unidad de medida de la variable).

En la tabla titulada Prueba de Kolmogorov-Smirnov para una muestra puede verse el valor del estadígrafo que es Z de Kolmogorov-Smirnov (cuyo valor fue de 0,981). Ahora vemos el valor de p (Sig. asintót. (bilateral)) fue de 0,291.

Como el valor de p fue mayor que 0,05 no se rechaza la hipótesis nula y se concluye que hay evidencias suficientes para pensar que la muestra proviene de la distribución especificada, con un nivel de significación del 5%.