

(Mini)Manual de PSPP, alternativa libre a SPSS

Román Salmerón Gómez

18 de mayo de 2015

Resumen

PSPP es una aplicación de software libre para el análisis de datos. Se presenta en modo gráfico y está escrita en el lenguaje de programación C. Usa la biblioteca científica GNU para sus rutinas matemáticas y gráficas para la generación de gráficos. Es un reemplazo libre para el software propietario SPSS. Proporciona funcionalidades básicas como: frecuencias, tablas cruzadas, comparación de media, regresión lineal, fiabilidad, reordenamiento de datos, pruebas no paramétricas, factor de análisis entre otras características. Los formatos de salida pueden ser en ASCII, PDF, PostScript o HTML así como algunos gráficos estadísticos (histogramas, gráficos circulares o gráficos de la distribución Normal). PSPP puede importar formatos de Gnumeric, OpenDocument, hojas de Excel, bases de datos Postgres, valores separados por coma y archivos ASCII. Puede exportar archivos en formato SPSS y archivos ASCII. Algunas de las bibliotecas usadas por PSPP pueden ser accedidas vía programación (WIKIPEDIA).

1. Introducción

En el presente documento se pretende proporcionar al lector los pasos básicos para poder trabajar con el software libre PSPP con el objetivo de realizar un análisis descriptivo de una base de datos y de realizar inferencia en la misma.

Debido a su similitud con SPSS, cualquier manual sobre dicho paquete estadístico es útil para aprender a usar PSPP. En cualquier caso, en la página web de PSPP (<https://www.gnu.org/software/pspp/>) se puede descargar un manual de uso (<https://www.gnu.org/software/pspp/manual/pspp.pdf>). Otros manuales interesantes que se pueden encontrar en internet son:

- Manual básico de análisis de datos con PSPP (<http://educonest.bligoo.cl/media/users/29/1493467/files/552749/JMoncadaH-PSPP.pdf>) de Juan Moncada Herrera.
- Manual introductorio al uso del programa PSPP para el análisis de datos (<http://www.index-f.com/lascasas/documentos/lc0773.pdf>) de Pedro Jorge Araujo.
- PSPP, A Purdue STAT 582 User Manual (http://www.stat.purdue.edu/~jennings/stat582/software/pspp_manual.pdf) de Kyle Bernis.

A continuación, en primer lugar se muestra cómo descargar e instalar el programa. En segundo lugar cómo introducir los datos o incorporarlos desde otro formato. Finalmente se aborda el análisis de los datos introducidos mediante un estudio descriptivo, de inferencia o no paramétrico.

2. Descarga e instalación

La descarga del programa para el sistema operativo Windows¹ se puede hacer desde la página web <http://pspp.awardspace.com/>. Dependiendo de si se dispone de un sistema operativo de 32 o 64 bits²

¹Para otros sistemas operativos consultar la página web <https://www.gnu.org/software/pspp/get.html>.

²Esta información se puede consultar pulsando el botón derecho del ratón sobre el icono de *Equipo*.

Highlights of the current PSPP-for-MSWindows setup

PSPP info:		Package info:	
Current version:	Master version = 0.8.4	MSWindows version:	MSWindows XP and newer
Questions/Suggestions:	pspp-users@gnu.org	Package Size:	32 Mb
Information about PSPP:	http://www.gnu.org/software/pspp	Size on disk:	72 Mb
PSPP Manual:	PDF or HTML (current version will be installed on your PC by the installer package)	Technical:	MinGW based Cross-compiled on openSUSE 13.1

Downloads:

Version	Installer for 32bits version	Installer for 64bits version
	Will work on 32 and 64bits MSWindows	Works only on a 64bits MSWindows version
0.8.4-g267362	PSPP 0.8.4 2015-04-27 32bits	PSPP 0.8.4 2015-04-27 64bits
0.8.4-g6036c7	PSPP 0.8.4 2015-03-27 32bits	PSPP 0.8.4 2015-03-27 64bits
0.8.4-ge9f1ea	PSPP 0.8.4 2015-02-24 32bits	PSPP 0.8.4 2015-02-24 64bits
0.8.4-ga5d5cf	PSPP 0.8.4 2015-02-10 32bits	PSPP 0.8.4 2015-02-10 64bits

- There are no known issues with this release. Multi-user or Single-user installation is automatically detected by this build.
- Remove versions dated before 2014-12-20 before installing the newest version.
- These are development versions, build from daily snapshots. If you encounter a problem or have suggestions for improvement please report it in pspp-users@gnu.org. Don't forget to check the [pspp-users list](#) for answers.
- If necessary you can always delete this version and reinstall the previous version.

Figura 1: Página web de descarga de PSPP para Windows

se debe descargar un instalador u otro. Si se desconoce por completo esta característica se recomienda la descarga de la versión de 32 bits ya que funcionará aunque el sistema operativo sea de 64 bits (en cambio la otra sólo funcionará si es de 64 bits).

En el momento de elaborar el presente documento la última versión disponible es la 0.8.4 correspondiente al 27 de abril de 2015 (ver Figura 1). Al pulsar sobre la versión correspondiente (en este caso aparece en la imagen anterior en otro color el correspondiente enlace) se nos redirecciona a una página web donde de forma automática se descargará el instalador del programa³.

Pulsando sobre el archivo descargado (en este caso *pspp-084-20150427-64bits-Setup.exe*) se iniciará la instalación (ver Figura 2). Dejando las opciones por defecto el programa se instalará en dos pasos (pulsar *Siguiente*, *Instalar* y *Cerrar*).

3. Datos

Para iniciar el programa simplemente hay que pinchar sobre el icono que se ha creado en el escritorio (buscar PSPP) o en el botón de *Inicio de Windows*. La pantalla inicial del mismo corresponde a la que aparece en la Figura 3. Como se puede observar disponemos de dos pestañas: *Vista de Datos* y *Vista de Variables*. En la primera se visualizarán (y se podrán modificar) los datos y en la segunda las características de las variables. En ambos casos se dispone de un menú de opciones en la parte superior que nos permitirá realizar distintas acciones sobre los datos.

Sin más que situarse sobre la casilla vacía situada al lado del *Caso 1* se puede empezar a introducir los datos (ver Figura 4). De forma automática se irán creando más filas (observaciones) y columnas (variables).

Como se puede observar, las variables se nombran de forma automática: *Var0001*, *Var0002*, etc. Si se desea cambiar el nombre de las mismas y otras características relacionadas con ellas se ha de ir a la pestaña correspondiente a la vista de variables. Algunas características a modificar son:

³Si esto no ocurre porque tengamos bloqueada esta opción, se nos ofrece la opción de pulsar sobre un enlace para descargar el instalador. Buscar el siguiente texto: *Problems with the download? Please use this direct link, or try another mirror.*

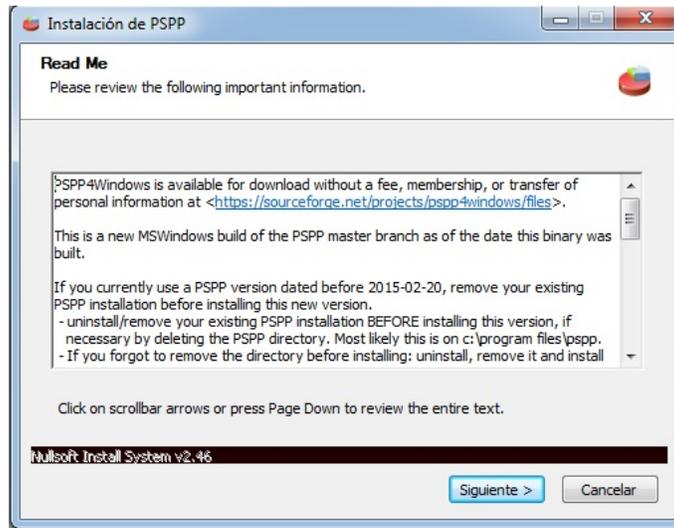


Figura 2: Primer paso de la instalación de PSPP para Windows

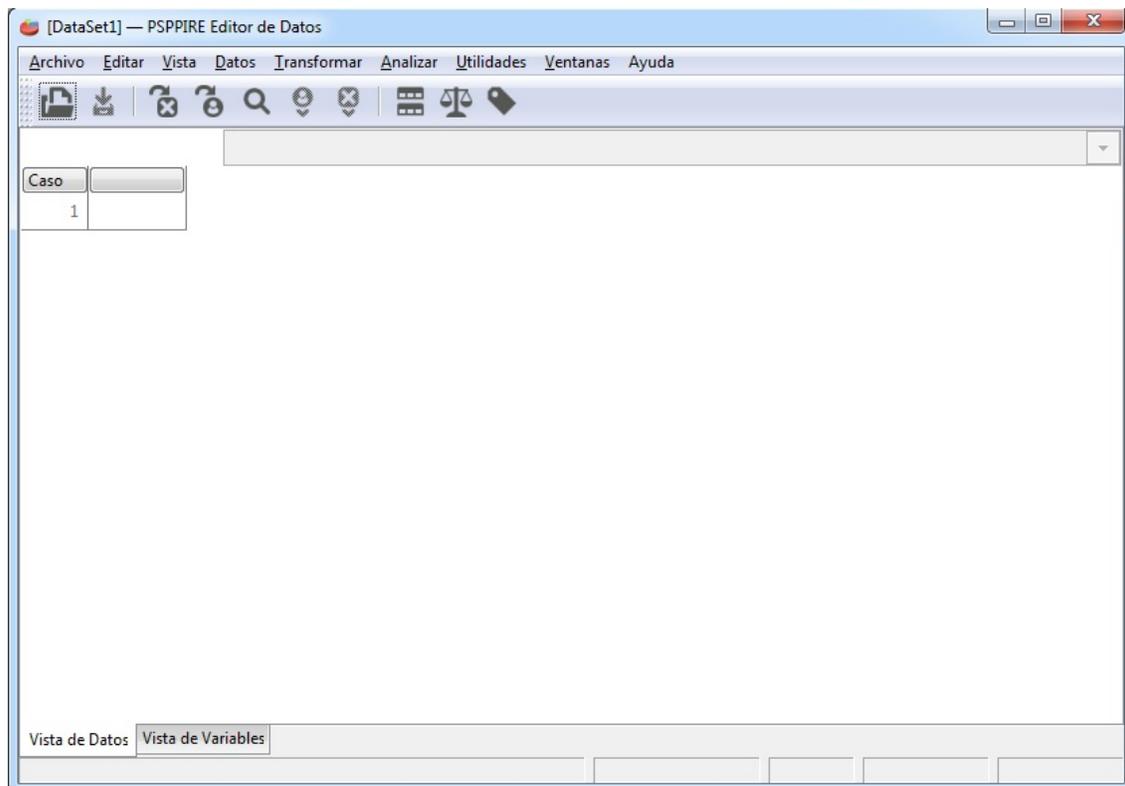


Figura 3: Pantalla de inicio de PSPP para Windows

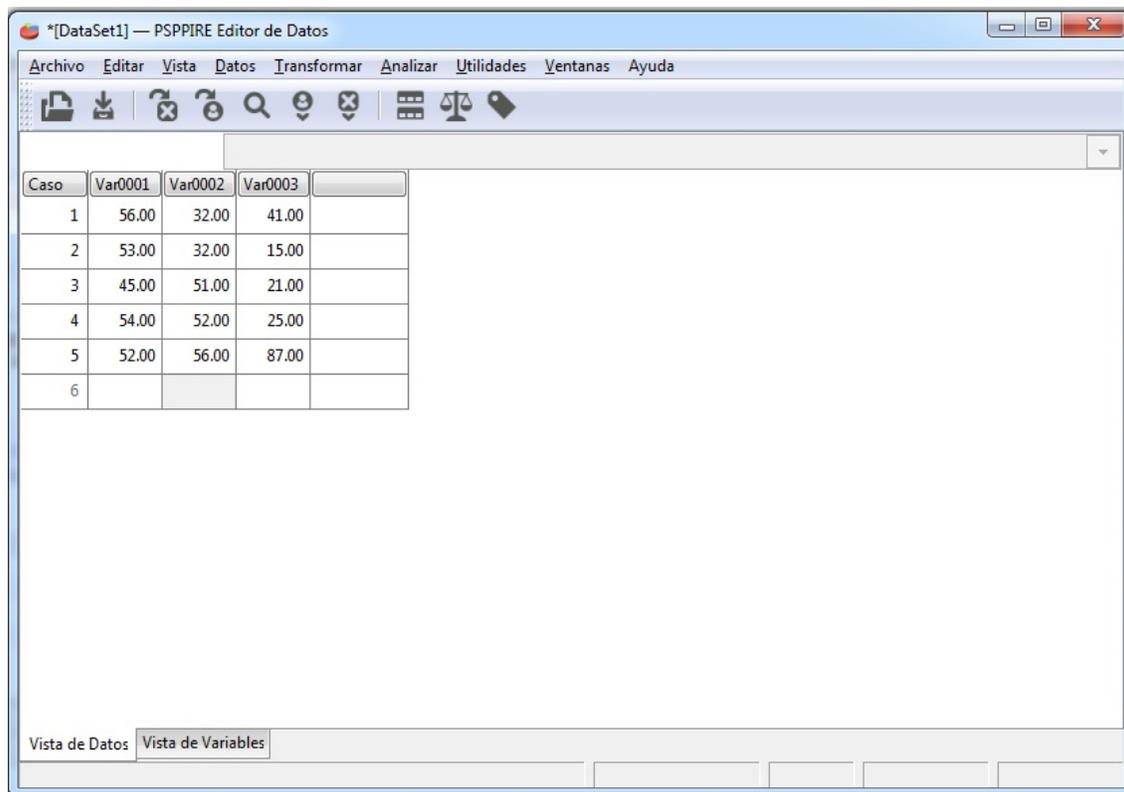


Figura 4: Introducción de datos en PSPP para Windows

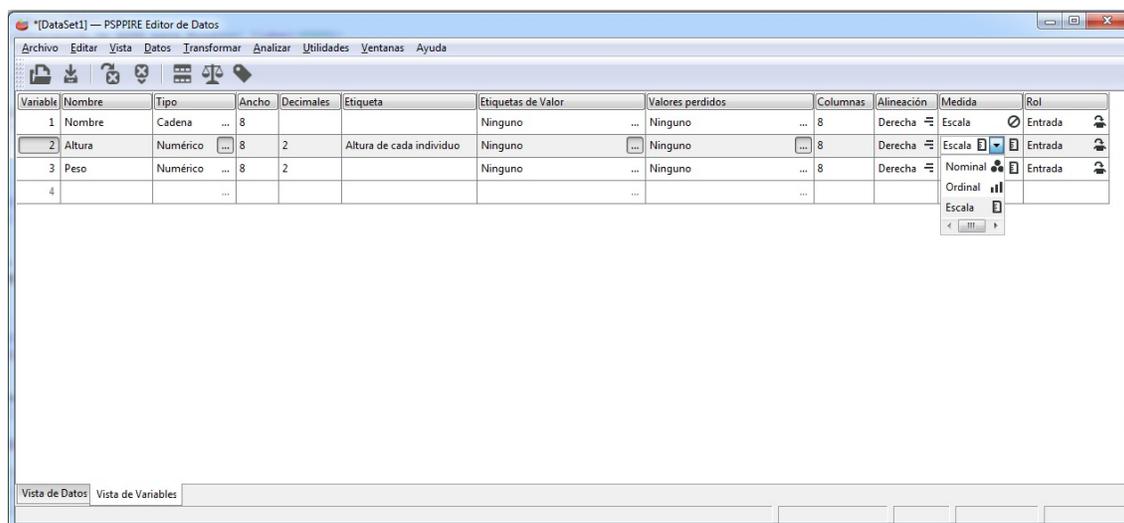


Figura 5: Modificación de las características de las variables en PSPP para Windows

Nombre: Permite especificar un nombre corto para cada variable (evitar acentos y demás caracteres *exóticos*).

Tipo: Permite especificar la naturaleza de la variable: de cadena (es decir, de texto), numérica, fecha, etc.

Decimales: Permite especificar el número de decimales que deseamos visualizar para cada variable.

Etiqueta: Permite especificar un nombre largo para cada variable de forma que se pueda aportar cierta información sobre la misma (como, por ejemplo, la unidad de medida).

Etiqueta de valor: Permite asignar un valor a un número concreto. Muy útil cuando se tiene variables cualitativas ya que permite aclarar, por ejemplo, que el 0 corresponde a *Hombre* y el 1 a *Mujer*.

Medida: Permite establecer la unidad de medida de la variable. A saber, nominal (variable cualitativa sin orden previo, como puede ser el estado civil), ordinal (variable cualitativa donde existe un orden, como puede ser una respuesta clasificada como *mala* o *bueno*) y de escala (variable cuantitativa donde hay una relación de orden y magnitud, como el peso donde 2kg es menor que 4kg y además es la mitad).

En cualquier caso, es muy común disponer los datos en un formato previo (por ejemplo, en una hoja de cálculo o en un archivo de texto) distinto al correspondiente a PSPP. Por tal motivo, es interesante incorporarlos directamente de dicho formato evitando introducir los datos directamente. Siguiendo la secuencia *Archivo* → *Importar datos...* del menú superior se obtiene la ventana que aparece en la Figura 6. Dicha ventana nos permitirá seleccionar el archivo que tiene la información que queremos incorporar.

Puesto que nos centraremos en la importación de datos procedente de archivos de texto plano⁴ debemos seleccionar la opción *Archivos de Texto (*.txt)*. A continuación nos solicitan cierta información:

- La cantidad de datos a importar: normalmente todos los casos (pulsar *Adelante*).
- El siguiente paso (ver Figura 7) es muy importante ya que hay que indicar la primera línea del archivo que contiene datos (en este caso la segunda ya que en la primera se tienen los nombres de las variables). Al seleccionar la segunda fila se activa una pestaña en la parte inferior izquierda que debemos seleccionar también ya que entonces especificamos que la línea que está por encima de la seleccionada inicialmente contiene el nombre de las variables (pulsar *Adelante*).
- En el siguiente paso hay que indicar cómo están separados los datos. En este caso seleccionaremos por *Tabulador*. En la parte inferior se puede visualizar si se separan correctamente las observaciones correspondientes a cada variable (pulsar *Adelante*).
- En el último paso podemos comprobar que la importación se ha realizado correctamente. Si hay algún error se puede pulsar sobre *Atrás* para corregirlo, mientras que en caso contrario pulsar sobre *Aplicar*.

Si el proceso se ha realizado correctamente han de aparecer dos nuevas ventanas de PSPP, una con los datos y otra con el código usado para incorporar los datos. No debemos olvidar que aunque nosotros realizamos las tareas mediante un menú de ventanas, todas las acciones que realicemos corresponden a un código concreto escrito en el lenguaje de programación de PSPP. Dichas órdenes son las que aparecen en la segunda ventana y será una constante que se irá repitiendo cada vez que ejecutemos alguna orden. Puesto que nosotros no estamos interesados en este (mini)manual en dicho código cerraremos dicha ventana.

⁴He encontrado algún problema en la incorporación de datos desde una hoja Excel, por tal motivo nos centraremos en el caso de la incorporación de información desde un archivo de texto. Si se tiene la información como hoja Excel simplemente hay que guardarla como *Texto (delimitado por tabulaciones) (*.txt)*.

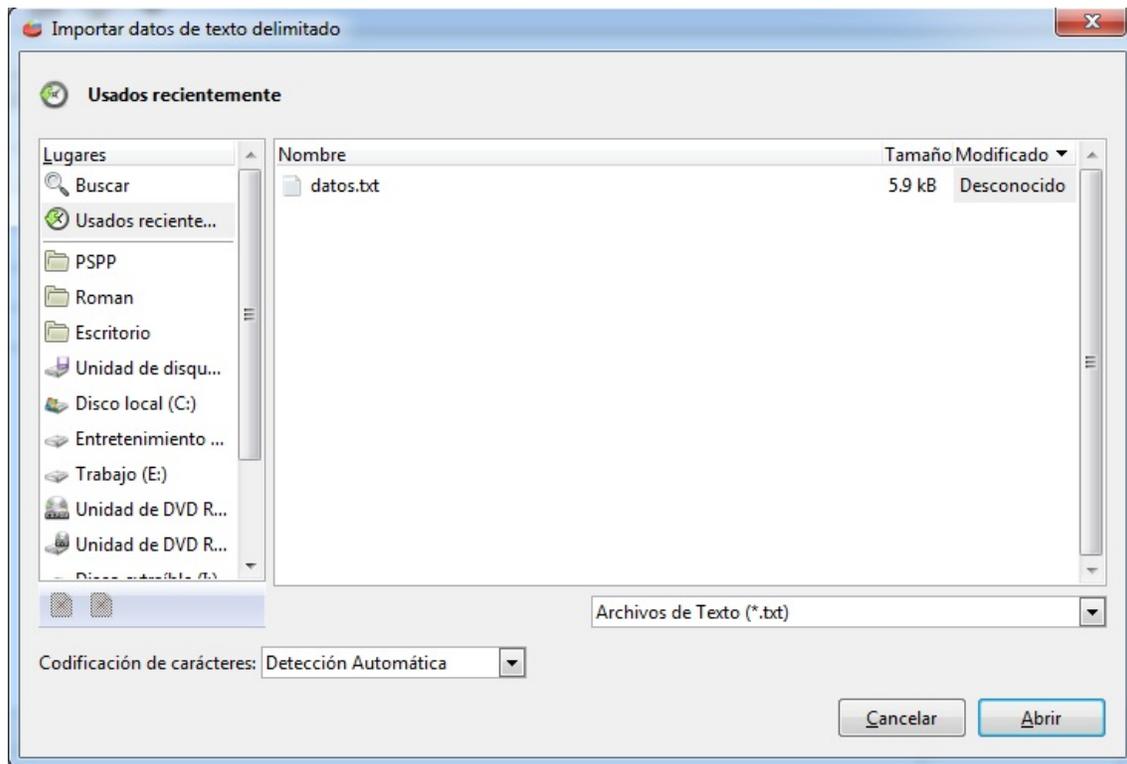


Figura 6: Importar datos con PSPP para Windows (1)

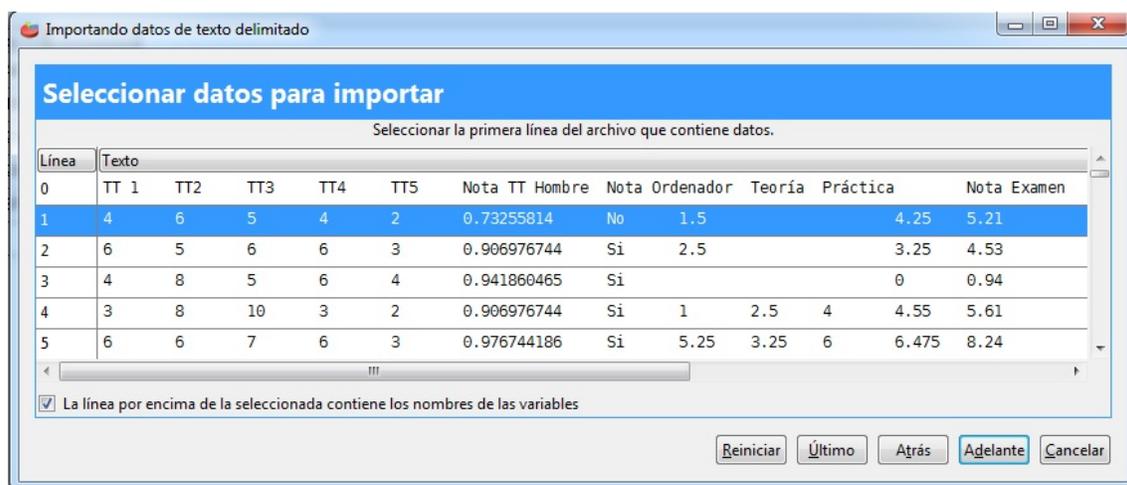


Figura 7: Importar datos con PSPP para Windows (2)

Para finalizar, guardaremos los datos en el formato propio de PSPP (que es *.sav*, al igual que en SPSS). Siguiendo la secuencia *Archivo -> Guardar como...* del menú superior nos aparecerá una nueva ventana donde simplemente tenemos que especificar el nombre del archivo y la carpeta donde deseamos que se guarde.

4. Estadística Descriptiva

En este apartado abordaremos el análisis descriptivo de un conjunto de datos. Concretamente, analizaremos los datos disponibles en la siguiente dirección web <http://www.ugr.es/~romansg/material/datos.sav>. Dichos datos corresponden a las calificaciones obtenidas por 89 alumnos en distintas pruebas (se puede ver la descripción de las variables en el visor de variables).

Siguiendo la secuencia *Analizar -> Estadística Descriptiva* aparecen cuatro opciones:

Frecuencias: Mediante esta opción se pueden calcular los estadísticos descriptivos más comunes⁵ para cada una de las variables seleccionadas. También permite hacer las siguientes representaciones gráficas para las variables seleccionadas: histograma, diagrama de barras y diagrama de sectores (ver Figura 8). Al pulsar sobre *Aceptar* se obtendrán los resultados requeridos en una nueva ventana.

Descriptivos: Mediante esta opción se pueden calcular los estadísticos descriptivos más comunes. La principal diferencia con respecto a la opción anterior es que esta está pensada para variables cuantitativas mientras que la anterior para variables cualitativas (en tal caso no tiene sentido calcular ciertos estadísticos descriptivos) o cuantitativas con pocos casos.

Explorar: Mediante esta opción se pueden calcular los principales estadísticos descriptivos y percentiles (sin más que seleccionar estas opciones en la ventana que surge al pulsar sobre *Estadísticos...*) de una variable en función de los valores de otra (normalmente cualitativa, llamada *Factor*). Ver Figura 9. Así, por ejemplo, se podría obtener la calificación media (o cualquier otra característica) distinguiendo entre el género (hombre o mujer).

Tablas Cruzadas: Mediante esta opción se pueden obtener las tablas de contingencia correspondientes a dos variables (para que tenga sentido han de ser cualitativas). Además, proporciona el valor del estadístico chi-cuadrado sobre independencia de ambas características.

En todos los casos, las ventanas de resultados obtenidas tienen un menú superior donde, entre otras acciones, se pueden guardar los resultados obtenidos en diversos formatos (PDF, HTML, TXT, etc) sin más que especificar el nombre del archivo, la carpeta donde se desea guardar y el formato (menú desplegable en la parte inferior).

Así, por ejemplo, en la Figura 10 se muestran los principales estadísticos descriptivos para las variables consideradas y en la Figura 11 los de la variable *Calificación final* distinguiendo en función del género (en el primer caso se ha optado por exportar los resultados como HTML mientras que en el segundo como texto plano). Podemos ver que la calificación media es de 4.76 puntos, si bien para los hombres es de 5.85 y para las mujeres de 3.01.

Otro análisis interesante puede ser calcular la correlación existente entre distintas variables. Esto se consigue siguiendo la secuencia *Analizar -> Correlación Bivariada...* y seleccionando las variables a analizar (seleccionar las opciones de test a dos colas y marcar las correlaciones significativas). En los resultados obtenidos (ver Figura 12) se obtienen, en este caso para las calificaciones de teoría, práctica y total del examen adelantado, las correlaciones de Pearson entre dichas variables. Atendiendo al p-valor asociado a cada correlación, al ser menor que 0.05, se rechaza la hipótesis nula de que las correlaciones sean cero. Es decir, se puede decir al nivel de significación del 5% que las correlaciones entre las tres variables son significativamente distintas de cero.

⁵Media, desviación estándar, mínimo, máximo, varianza, asimetría, rango, moda, curtosis, mediana, suma.

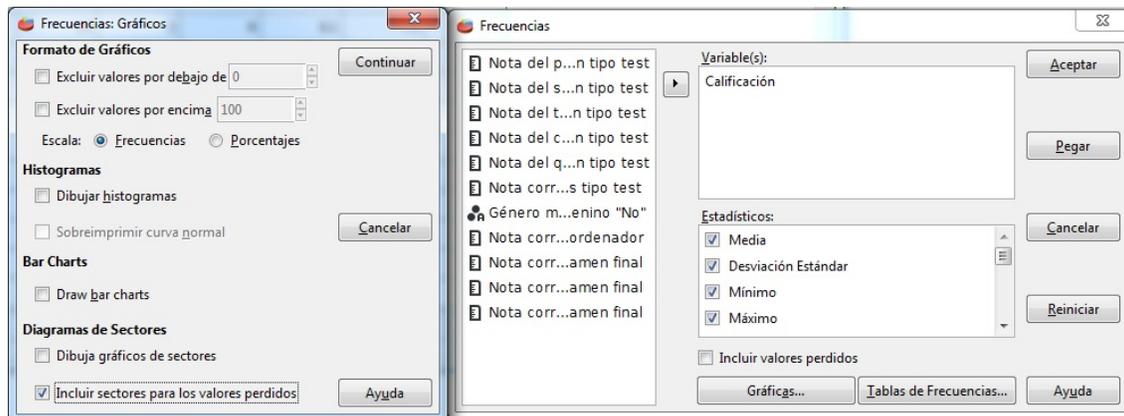


Figura 8: Menú *Frecuencias* en PSPP para Windows

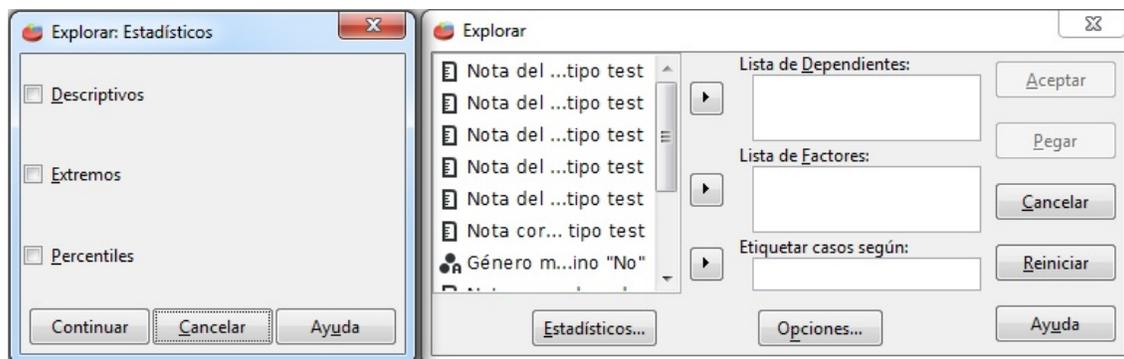


Figura 9: Menú *Explorar* en PSPP para Windows

DESCRIPTIVES

DESCRIPTIVES
 /VARIABLES= TT1 TT2 TT3 TT4 TT5 Nota_TT Nota_Ordenador Teoría Práctica Nota_Examen Calificación
 /STATISTICS=DEFAULT RANGE SUM VARIANCE KURTOSIS SKEWNESS.

Casos válidos = 89; casos con valor(es) perdido(s) = 61.

Variable	N	Media	Desv Std	Varianza	Kurtosis	Err.Est.Curt.	Asimetría	Err.Est.Asim.	Intervalo	Mínimo	Máximo	Suma
Nota del primer examen tipo test	80	5.41	1.81	3.26	-.77	.53	-.25	.27	7.00	1.00	8.00	433.00
Nota del segundo examen tipo test	84	7.50	2.03	4.13	-.57	.52	-.27	.26	8.00	3.00	11.00	630.00
Nota del tercer examen tipo test	70	5.90	1.96	3.86	-.05	.57	.43	.29	8.00	2.00	10.00	413.00
Nota del cuarto examen tipo test	79	4.77	1.78	3.15	-.63	.53	-.04	.27	7.00	1.00	8.00	377.00
Nota del quinto examen tipo test	75	3.52	1.44	2.06	-.65	.55	-.14	.28	6.00	.00	6.00	264.00
Nota correspondiente a todos los exámenes tipo test	89	.83	.30	.09	.71	.51	-.82	.26	1.36	.00	1.36	73.85
Nota correspondiente al examen de ordenador	76	3.46	2.22	4.93	-.97	.54	.35	.28	8.25	.00	8.25	262.65
Nota correspondiente a la parte de teoría del examen final	32	3.07	.90	.82	-.63	.81	-.65	.41	3.00	1.00	4.00	98.25
Nota correspondiente a la parte de práctica del examen final	32	4.79	.94	.88	.88	.81	-.88	.41	4.00	2.00	6.00	153.25
Nota correspondiente al examen final	89	3.49	2.26	5.11	-1.14	.51	-.25	.26	7.00	.00	7.00	310.80
Nota final de toda la asignatura (engloba evaluación continua y examen final)	89	4.76	2.69	7.25	-.95	.51	-.16	.26	9.60	.00	9.60	424.01

Figura 10: Principales estadísticos descriptivos de los datos considerados

Descriptivos

		#	#	#	
Género de los alumnos: 1 Hombre, 0 Mujer		#	#	#	
		#	#	#	
Nota final de toda la asignatura (engloba evaluación continua y examen final)	Mujer	Media	3.01	.39	
		Intervalo de Confianza 95% para la Media	Límite Inferior	2.22	
			Límite Superior	3.79	
		Media recortada al 5%		2.95	
		Mediana		3.04	
		Varianza		5.06	
		Desviación Estándar		2.25	
		Mínimo		.00	
		Máximo		7.20	
		Intervalo		7.20	
		Intervalo intercuartílico		4.09	
		Asimetría		.21	
		Curtosis		-1.25	
	Hombre		Media	5.85	.32
			Intervalo de Confianza 95% para la Media	Límite Inferior	5.21
				Límite Superior	6.49
			Media recortada al 5%		5.92
		Mediana		5.68	
		Varianza		5.57	
		Desviación Estándar		2.36	
		Mínimo		.94	
		Máximo		9.60	
		Intervalo		8.66	
		Intervalo intercuartílico		3.13	
		Asimetría		-.48	
		Curtosis		-.40	

Figura 11: Principales estadísticos descriptivos de la variable *Calificación final* por género

Objetivo — PSPPIRE Visor de resultados

Archivo Editar Ventanas Ayuda

CORRELATIONS

CORRELATION
/VARIABLES = Teoría Práctica Nota_Examen
/PRINT = TWOTAIL NOSIG.

Correlaciones

		Nota correspondiente a la parte de teoría del examen final	Nota correspondiente a la parte de práctica del examen final	Nota correspondiente al examen final
Nota correspondiente a la parte de teoría del examen final	Correlación de Pearson	1.00	.55	.88
	Sign. (2-colas)		.001	.000
	N	32	32	32
Nota correspondiente a la parte de práctica del examen final	Correlación de Pearson	.55	1.00	.89
	Sign. (2-colas)	.001		.000
	N	32	32	32
Nota correspondiente al examen final	Correlación de Pearson	.88	.89	1.00
	Sign. (2-colas)	.000	.000	
	N	32	32	89

Figura 12: Resultados del procedimiento *Correlación Bivariada...*

5. Inferencia Estadística

En este apartado tiene por objetivo mostrar cómo hacer inferencia estadística sobre un conjunto de datos. Por tanto, se centra en las opciones disponibles en la secuencia *Analizar* → *Comparar Medias*. A saber:

Medias: Mediante esta opción se puede obtener las medias y varianzas en subgrupos (categorías de una o más variables independientes, anteriormente denominadas *factores* y que reciben el nombre de *Lista de Independientes*).

Prueba T para una muestra: Mediante esta opción se puede contrastar si la media de una sola variable difiere de una constante especificada. También se obtiene el intervalo de confianza correspondiente.

Prueba T para muestras independientes: Mediante esta opción se puede comparar (mediante un contraste o intervalo) las medias de dos grupos (especificados en la variable de agrupación). En este caso, mediante la prueba de Levene, además se contrasta la igualdad de varianzas.

Prueba T para muestras relacionadas Mediante esta opción se puede comparar las medias de dos variables de un solo grupo: se trata de una única muestra a la cual se le miden variables antes y después de un evento.

ANOVA de un factor: Mediante esta opción se podrá contrastar la hipótesis de que varias medias son iguales. Esta técnica es una extensión de la prueba t para dos muestras (evidentemente esta es un caso particular de esta opción).

A partir de la primera opción se podría calcular, por ejemplo, la calificación final de los alumnos en función de su edad (ver Figura 13). En tal caso se puede observar que la calificación media máxima, 5.24, corresponde a los 37 alumnos que tienen 20 años, mientras que la calificación mínima, 4.3, a los 6 alumnos que tienen 22 años. En el caso de querer saber si las 4 medias obtenidas son significativamente distintas desde un punto de vista estadístico (numéricamente es claro que lo son) se puede recurrir a la última

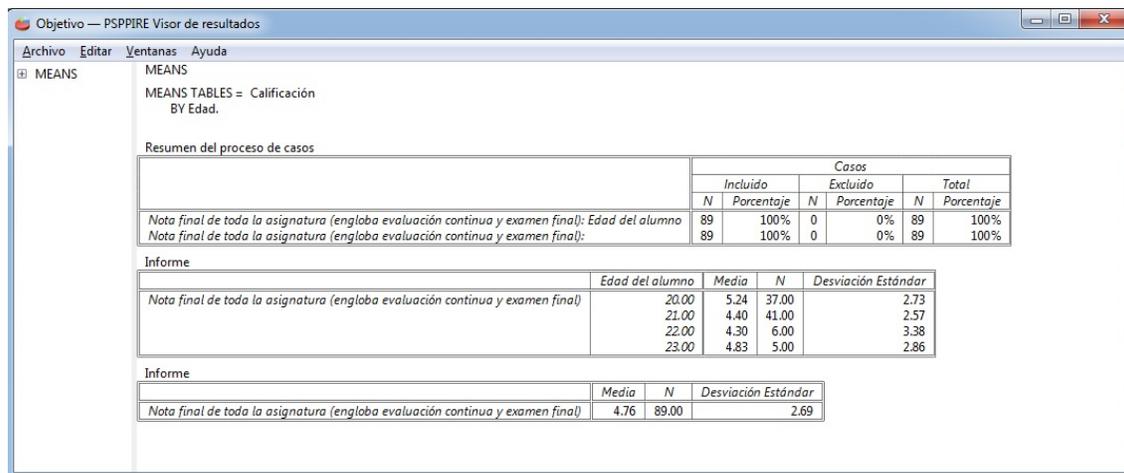


Figura 13: Resultados del procedimiento *Medias* de *Comparar Medias*

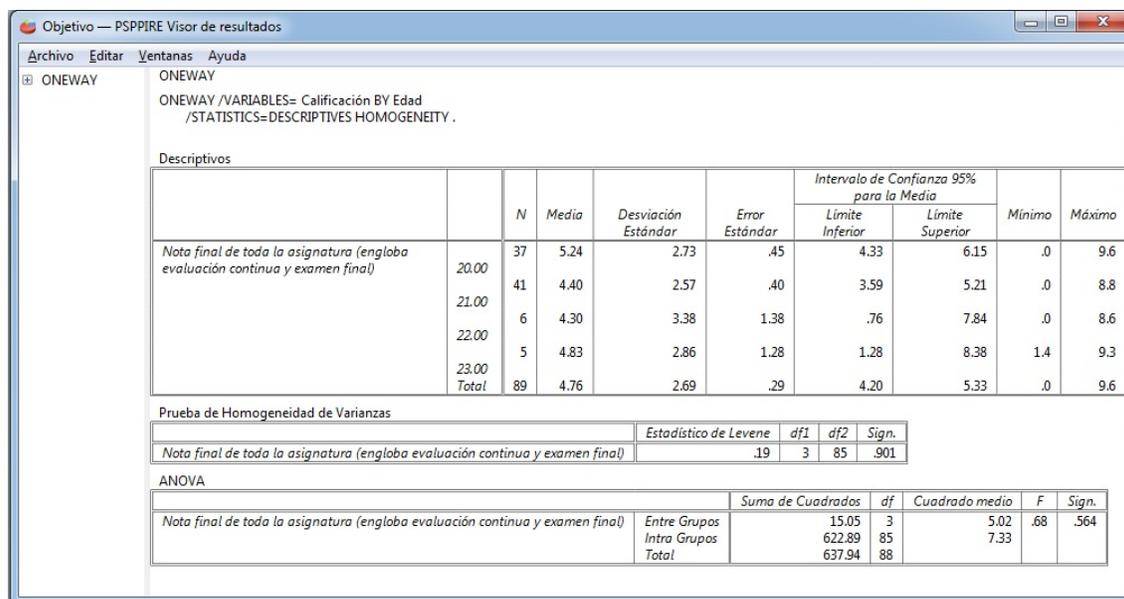


Figura 14: Resultados del procedimiento *ANOVA* de *Comparar Medias*

opción. Seleccionando las opciones *Descriptivos* y *Homogeneidad*, los resultados⁶ se pueden observar en la Figura 14. Puesto que el p-valor asociado a la *Prueba de Homogeneidad de Varianzas*, 0.901, es mayor que 0.05, no se rechaza la hipótesis nula de que todas las varianzas son iguales. De igual forma, puesto que el p-valor asociado al *ANOVA*, 0.564, también es mayor que 0.05, tampoco se rechaza la hipótesis nula de que las 4 medias son iguales de forma simultánea.

Si se selecciona la opción *Prueba T para una muestra*, simplemente hay que añadir las variables sobre las que se desea realizar la inferencia y el valor de prueba con el que se desea comparar la media de la variable (ver Figura 15). En *Opciones* se puede modificar el nivel de confianza al que se desea trabajar. Los resultados (ver Figura 16) proporcionan los valores medios y desviaciones típicas de cada una de las variables, en este caso, de las notas correspondientes al examen final y a la calificación final de la asignatura. También proporciona el p-valor del contraste realizado y el intervalo de confianza asociado.

⁶Adviértase que las pruebas devuelven el p-valor del contraste realizado, el cual se define como el mínimo valor de significación a partir del cual se rechaza la hipótesis nula. De manera que si p-valor es menor que 0.05 se rechaza la hipótesis nula al nivel de significación del 5%.

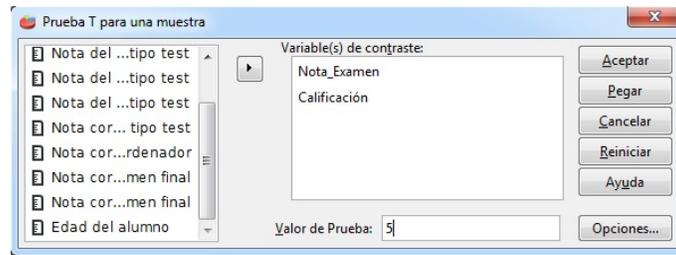


Figura 15: Procedimiento *Prueba T para una muestra* de *Comparar Medias*

Objetivo — PSPPIRE Visor de resultados

Archivo Editar Ventanas Ayuda

T-TEST

T-TEST /TESTVAL=5
/VARIABLES= Nota_Examen Calificación /MISSING=ANALYSIS
/CRITERIA=CIN(0.95).

Estadísticas de una muestra

	N	Media	Desviación Estándar	Err.Est.Media
Nota correspondiente al examen final	89	3.49	2.26	.24
Nota final de toda la asignatura (engloba evaluación continua y examen final)	89	4.76	2.69	.29

Prueba de una muestra

Valor de prueba = 5.000000

	t	df	Sign. (2- colas)	Diferencia Media	Intervalo de confianza 95% de la Diferencia	
					Inferior	Superior
Nota correspondiente al examen final	-6.29	88	.000	-1.51	-1.98	-1.03
Nota final de toda la asignatura (engloba evaluación continua y examen final)	-.83	88	.411	-.24	-.80	.33

Figura 16: Resultados del procedimiento *Prueba T para una muestra* de *Comparar Medias*

Se observa que la nota media correspondiente al examen final es de 3.49 mientras que la de la asignatura es de 4.76. Realizado el contraste correspondiente, para la primera variable se obtiene un p-valor menor a 10^{-3} , luego menor que 0.05, por lo que se rechaza la hipótesis nula de que la nota media del examen final sea igual a 5 al 5% de significación. Mientras que para la segunda variable no se rechaza la hipótesis nula, al 5% de significación, de que su media sea igual a 5 ya que el p-valor, 0.411, es mayor que 0.05.

Además, nos proporcionan también el intervalo de confianza para la diferencia $\mu - \mu_0$. Es decir, para la primera variable se estima que, al 95% de confianza, la media pertenece al intervalo dado por (3,02, 3,97) y para la segunda al (4,2, 5,33). Viendo si el valor de prueba, en este caso el 5, pertenece o no a dichos intervalos de confianza se llegan a las mismas conclusiones que a partir de los contrastes. Sin embargo, para la primera variable se podría añadir más información aparte de que la calificación media es distinta de 5. Puesto que el extremo superior es menor que 4, se tendría al 95% de confianza, que la nota media del examen final es inferior a 4.

Finalmente, con la opción *Prueba T para muestras independientes* podremos contrastar si las medias de dos variables (en realidad de dos subgrupos de una misma variable) se pueden considerar iguales o no. La introducción de datos es análoga (ver Figura 17) al caso de una muestra con la salvedad de que en este caso hay que especificar los dos subgrupos de la variable de agrupación a partir de la opción *Definir Grupos* (mediante el menú desplegable se pueden definir dichos grupos de forma sencilla). Los resultados (ver Figura 18) ofrecen en primer lugar la media y varianza de cada subgrupo, en este caso, de la nota final de la asignatura en función de si tiene género masculino o femenino. Y a continuación la información correspondiente a los contrastes:

- El primero, *Prueba de Levene*, comprueba si las varianzas poblacionales (que son desconocidas⁷) pueden considerarse iguales. En el ejemplo considerado, puesto que el p-valor, 0.74, es mayor que

⁷En la práctica es difícil encontrarse con situaciones donde se conozcan las características poblacionales, por lo que el paquete informático está enfocado al caso en el que sólo se disponen de las observaciones de las variables.

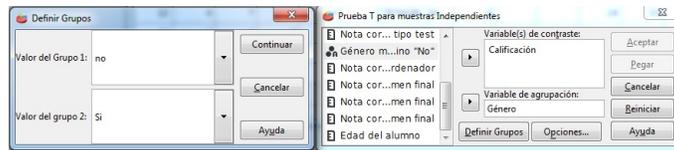


Figura 17: Procedimiento *Prueba T para muestras independientes de Comparar Medias*

T-TEST
T-TEST /VARIABLES= Calificación
/GROUPS=Género("No","Si") /MISSING=ANALYSIS
/CRITERIA=CIN(0.95).

Estadísticas de grupo

	Género masculino "Si" y femenino "No"	N	Media	Desviación Estándar	Err.Est.Media
Nota final de toda la asignatura (engloba evaluación continua y examen final)	No	34	3.01	2.25	.39
	Si	55	5.85	2.36	.32

Prueba para muestras independientes

	Prueba de Levene para la igualdad de varianzas	Prueba T para la Igualdad de Medias								
		F	Sign.	t	df	Sign. (2-colas)	Diferencia Media	Err.Est. de la Diferencia	Intervalo de confianza 95% de la Diferencia	
									Inferior	Superior
Nota final de toda la asignatura (engloba evaluación continua y examen final)	Se asume igualdad de varianzas	.11	.740	-5.62	87.00	.000	-2.85	.51	-3.85	-1.84
	Igualdad de varianzas no asumida			-5.69	72.64	.000	-2.85	.50	-3.84	-1.85

Figura 18: Resultados del procedimiento *Prueba T para muestras independientes de Comparar Medias*

0.05 no se rechaza al 5 % de significación que las varianzas poblacionales sean iguales.

- El segundo, *Prueba T para Igualdad de Medias*, comprueba si las medias pueden considerarse o no iguales. En este caso se barajan dos posibilidades en función de los resultados obtenidos en la prueba anterior: la primera fila corresponde al caso en el que no se rechaza la hipótesis nula de igualdad de varianzas, mientras que la segunda corresponde al caso en el que si se rechaza. La interpretación de los resultados es análoga al caso de una muestra: puesto que el p-valor, menor que 10^{-3} , es menor que 0.05, se rechaza la hipótesis nula de que las calificaciones medias en hombres y mujeres sean iguales al 5 % de significación. Además, puesto que el intervalo de confianza asociado, $(-3,85, -1,84)$, es negativo se tiene que la calificación final media de los hombres es mayor que la de las mujeres (para esta interpretación es importante tener claro cómo se han definido los grupos) al nivel de confianza del 95 %.

6. Estadística no paramétrica

Para finalizar, a continuación, veremos como se estudia si un conjunto de datos se distribuye como una normal. Siguiendo la secuencia *Analizar -> Pruebas No-Paramétricas -> K-S para Una Muestra...* se obtiene una nueva ventana donde simplemente hay que introducir la variable que se quiere comprobar si procede de una Normal y seleccionar esta distribución de entre las distribuciones de prueba (otras opciones ofrecidas son la Poisson, uniforme y exponencial).

En la Figura 19 se muestran los resultados obtenidos para la variable *Calificación*. Teniendo en cuenta que el p-valor asociado al contraste, 0.515, es mayor que 0.05, no se rechaza la hipótesis nula de que la muestra considerada proceda de una población Normal.

Prueba Kolmogorov_Smirnov para una muestra		Nota final de toda la asignatura (engloba evaluación continua y examen final!)
N		89
Parámetros Normal	Media	4.76
	Desviación Estándar	2.69
Diferencias Más Extremas	Absoluto	.09
	Positivo	.09
	Negativo	-.08
Z de Kolmogorov-Smirnov		.82
Sig. Asint. (2-colas)		.515

Figura 19: Resultados del procedimiento *K-S para Una Muestra...* de *Pruebas No-Paramétricas*

Si se desea, por ejemplo, comprobar si la *Calificación* para los subgrupos de hombres y mujeres se distribuye según una Normal, previamente hay que dividir la muestra en función de la variable *Género*. Esto se consigue siguiendo la secuencia *Datos -> Dividir Archivo...* de forma que en la opción *Organizar los resultados por grupos* hay que especificar la variable *Género* (ver Figura 20). Siguiendo los mismos pasos de antes, en la Figura 21 se muestran los resultados obtenidos. Puesto que el p-valor obtenido para la submuestra formada por los hombres, 0.79, es mayor que 0.05, no se rechaza la hipótesis nula de que el conjunto de datos especificado proceda de una distribución Normal. Las mismas conclusiones se obtienen para la submuestra de las mujeres ya que el p-valor en este caso es de 0.586.

Si no se desea hacer más análisis distinguiendo entre hombres y mujeres, antes de realizar cualquier otra operación es muy importante deshacer la selección realizada. Así, volviendo a la secuencia *Datos -> Dividir Archivo...* habrá que seleccionar la opción *Analizar todos los casos. No crear grupos*.

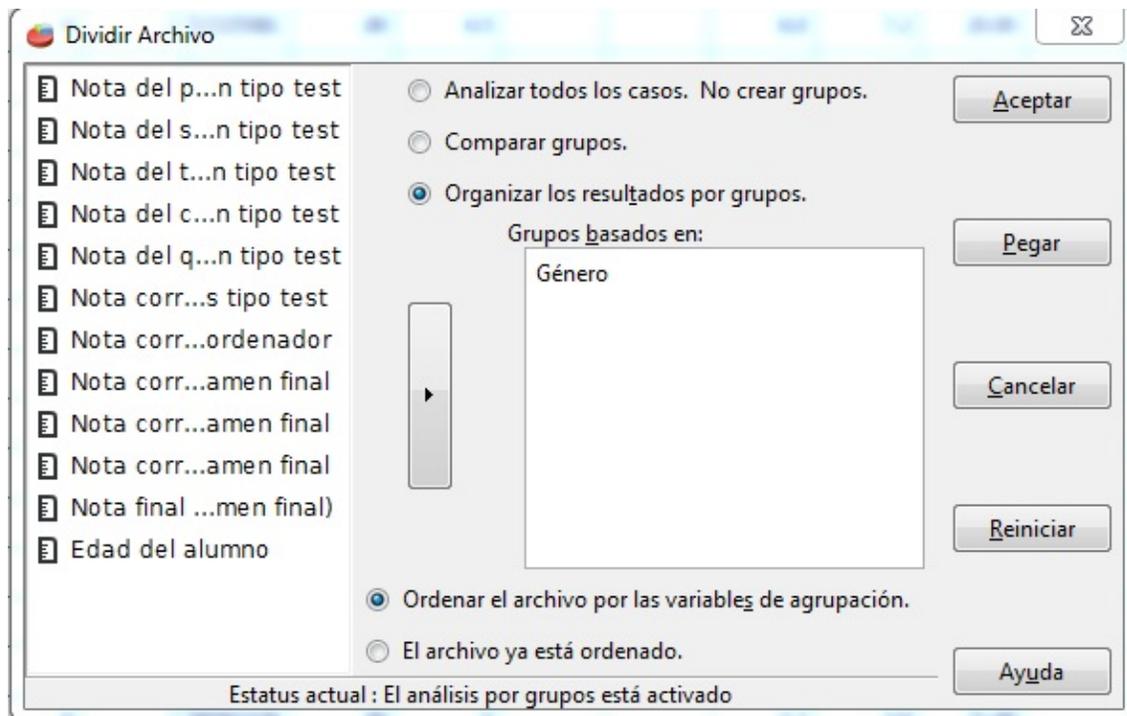


Figura 20: Procedimiento *Dividir Archivo...* de *Datos*

Objetivo — PSPPIRE Visor de resultados

Archivo Editar Ventanas Ayuda

NPAR TESTS

NPAR TEST
/KOLMOGOROV-SMIRNOV (NORMAL) = Calificación.

Prueba Kolmogorov_Smirnov para una muestra

		Nota final de toda la asignatura (engloba evaluación continua y examen final)
N		34
Parámetros Normal	Media	3.01
	Desviación Estándar	2.25
Diferencias Más Extremas	Absoluto	.13
	Positivo	.13
	Negativo	-.09
Z de Kolmogorov-Smirnov		.77
Sig. Asint. (2-colas)		.586

Prueba Kolmogorov_Smirnov para una muestra

		Nota final de toda la asignatura (engloba evaluación continua y examen final)
N		55
Parámetros Normal	Media	5.85
	Desviación Estándar	2.36
Diferencias Más Extremas	Absoluto	.09
	Positivo	.07
	Negativo	-.09
Z de Kolmogorov-Smirnov		.65
Sig. Asint. (2-colas)		.790

Figura 21: Resultados sobre la normalidad de *Calificaciones* distinguiendo en función de *Género*