

Universidad de Chile
Facultad de Ciencias Sociales (FACSO)
Departamento de Sociología



UNIVERSIDAD DE CHILE

Guía introductoria al uso de software en Estadística

Ayudantía Estadística I 2014

Equipo de ayudantes:

Ignacio Díaz

Carolina García

Magdalena León

Felipe Ruiz

Francisca Torres

Docentes: Paulina Lizama, Giorgio Boccardo,

Agosto de 2014

Contenido

Introducción.....	3
SPSS: uno de los programas más utilizados en ciencias sociales	3
Descarga de una base de datos para el ejemplo	4
Ejemplo de uso SPSS: modo botones y modo sintaxis:	11
Excel: uno de los olvidados.....	17
Vista de datos en Excel (la base de datos fue exportada, más adelante se ahondará en esto)	18
Obtención de tablas de frecuencias en Excel	20
Fuentes de datos secundarios	23
Elementos descargables diferentes: utilidades diferentes	25
Otros programas de análisis de datos: ¿qué alternativas existen?	26
Stata	27
R: un incomprendido	28

Introducción

Siguiendo a Paula Elousa, podemos afirmar que aquellas tareas “relacionadas con el tratamiento y análisis de datos en el campo de las ciencias sociales y de la salud pueden ejecutarse utilizando varias herramientas informáticas, de entre las cuales, sobresalen, por su extendido uso en nuestro entorno, SPSS, SAS, STATISTICA, Systat, Stata o GenStat. Las diferencias entre ellas residen básicamente en su mayor o menor grado de generalidad, en el costo o en la facilidad de uso. Este **triple criterio (generalidad, costo, facilidad de uso)** además de la tradición de uso, guían la elección de uno u otro software para la investigación y para la docencia. Sin duda, se trata, en todos los casos, de aplicaciones eficientes basadas en el uso de potentes interfaces gráficas (GUI Graphical User Interface) con sistemas de menús y ventanas desplegadas que tienden un puente al usuario en el proceso de modelización estadística. **Las GUIs ejercen de intermediarios entre los modelos formales y el usuario, que sin entrenamiento excesivamente intensivo es capaz de explorar, analizar y modelizar sólo con un simple click de ratón.** Los entornos de trabajo amigables han simplificado enormemente los procesos relacionados con el tratamiento de datos y, como consecuencia, el análisis de datos se ha socializado. Aunque los efectos positivos de la generalización puedan a veces verse ensombrecidos con una aplicación incorrecta de los modelos o una torpe interpretación de los resultados, los beneficios superan sobradamente los estropicios con los que en ocasiones nos encontramos o incluso cometemos.” (Elousa, 2009, pág. 652) (Los énfasis son nuestros)

La manera más fácil de familiarizarse con estos programas es haciendo uso de ellos. Sin embargo, su aplicación debe hacerse sin perder de vista su cualidad de **herramientas**. En todo momento somos nosotros quienes damos instrucciones a los programas, teniendo en cuenta el problema que se está investigando, el tipo de información con que se trabaja, los objetivos del estudio, etc. Así, los resultados que se obtengan dependen exclusivamente de lo se le “pida” al programa y, si no se cumplen las condiciones necesarias, éstos pueden no ser interpretables desde el punto de vista de la investigación. El programa no piensa por quien lo está utilizando, no señala los errores ni las incoherencias, por eso su uso debe hacerse en relación a nuestras decisiones previas en el ámbito de la investigación.

SPSS: uno de los programas más utilizados en ciencias sociales

SPSS es una de las herramientas más extendidas para uso en ciencias sociales, de hecho, muchos de los manuales para realizar distintos tipos de análisis estadísticos traen ejemplos e instrucciones obtenidos

con este programa. SPSS permite tanto crear bases de datos que contengan la información de una encuesta, como realizar distintas operaciones utilizando aquellos datos: procedimientos de estadística descriptiva, inferencial o modelos predictivos, en sus variantes univariada, bivariada y multivariada, entre otras aplicaciones.

Este paquete estadístico requiere, para **usos oficiales o institucionales**, ser usado con una **licencia**. De esta forma, puede adquirirse el **programa básico** e incorporar **otros paquetes de análisis de datos**, según lo que se requiera. También existen actualizaciones periódicas del programa, las cuales incluyen tanto mejoras en la interfaz de usuario como nuevas aplicaciones para el análisis de datos. Para efectos de **uso personal** es posible crackear¹ la última versión del programa e ir más allá del uso que se le da en las tareas universitarias.

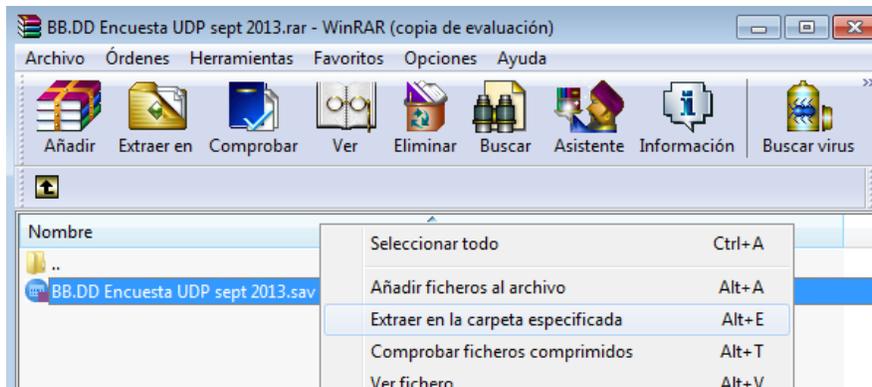
Descarga de una base de datos para el ejemplo

Para ir introduciéndonos en el uso del programa procederemos a descargar y ejecutar una base de datos, que luego abriremos. Para ello, accederemos al sitio: <http://www.encuesta.udp.cl/>

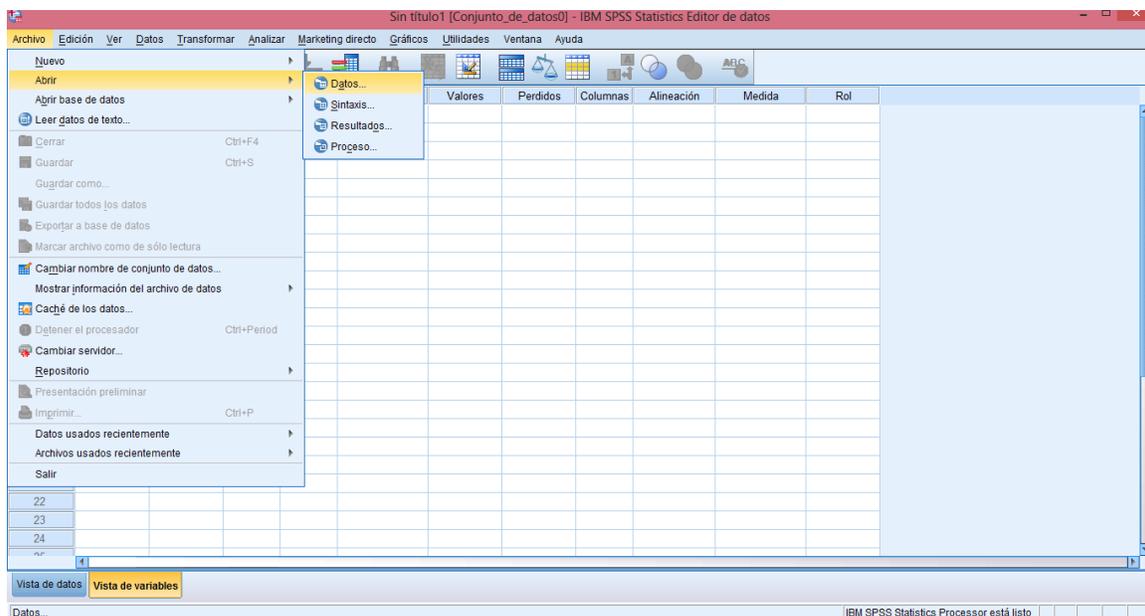


Guardamos el archivo comprimido (.rar) en un lugar fácil de encontrar (en el Escritorio, por ejemplo), y lo abrimos para extraer (también puede ser en el escritorio) la base de datos en un formato que podamos manejar.

¹ Cabe recordar que si nos estamos desempeñando en un contexto institucional universitario, por ejemplo, o en un ámbito que esté regido por las normativas públicas, esto no es posible. Emplear un software sin licencia constituye una infracción legal que, en tales contextos, puede ser fuertemente penalizada, ya que implica una malversación de fondos públicos.



Ahora, para abrir la base de datos, realizamos la siguiente operación luego de abrir el SPSS. Debes hacer clic en el menú 'Archivo', luego en el submenú 'Abrir' y luego en la opción 'Datos'. Se abrirá un cuadro de diálogo en el cual debes buscar y especificar la ubicación de la base de datos (carpeta donde la guardamos), haciendo doble clic sobre el archivo que contiene los datos.



Una vez efectuado el comando *abrir*, SPSS desplegará varias ventanas que pasaremos a definir.

Vista de variables: Contiene las preguntas de la encuesta [variables] y las características de los datos; podemos señalar rápidamente como importantes las columnas: *nombre*, *tipo*, *decimales*, *etiqueta*, *valores*, *perdidos* y *medida*.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	folio	Númerico	8	0		Ninguna	Ninguna	3	Derecha	Escala	Entrada
2	PROYECCI...	Númerico	8	5	PONDERADO...	Ninguna	Ninguna	5	Derecha	Escala	Entrada
3	POND_MU...	Númerico	8	5	PONDERADO...	Ninguna	Ninguna	7	Derecha	Escala	Entrada
4	P1	Númerico	8	0	P1.- ¿Cómo ev...	{1, Muy Bue...	Ninguna	5	Derecha	Escala	Entrada
5	P2	Númerico	8	0	P2.- En general...	{1, Muy Bue...	Ninguna	6	Derecha	Escala	Entrada
6	P3	Númerico	8	0	P3.- A su juicio...	{1, Ahora E...	Ninguna	10	Derecha	Nominal	Entrada
7	P4	Númerico	8	0	P4.- ¿Y en el fu...	{1, En El Fu...	Ninguna	10	Derecha	Nominal	Entrada
8	P5	Númerico	8	0	P5.- ¿Cómo ev...	{1, Muy Bue...	Ninguna	10	Derecha	Escala	Entrada
9	P6	Númerico	8	0	P6.- ¿Y la situa...	{1, En El Fu...	Ninguna	10	Derecha	Nominal	Entrada
10	P7	Númerico	8	0	P7.- Pensando ...	{1, No Alca...	Ninguna	10	Derecha	Nominal	Entrada
11	P8	Númerico	8	0	P8.- Si tuviera ...	{1, Alto}...	Ninguna	10	Derecha	Escala	Entrada
12	P9	Númerico	8	0	P9.- Considera ...	{1, Ha Aum...	Ninguna	10	Derecha	Nominal	Entrada
13	P10	Númerico	8	0	P10.- ¿Cree us...	{1, Ha Aum...	Ninguna	10	Derecha	Nominal	Entrada
14	P11	Númerico	8	0	P11.- Del sigui...	{1, La Educ...	Ninguna	10	Derecha	Nominal	Entrada
15	P12	Númerico	8	0	P12.- ¿Qué tan...	{1, Muy Inte...	Ninguna	10	Derecha	Nominal	Entrada
16	P13	Númerico	8	0	P13.- Hablando ...	{1, Aprueba}	Ninguna	10	Derecha	Nominal	Entrada
17	P14	Númerico	8	0	P14.- ¿Ud. apr...	{1, Aprueba}	Ninguna	10	Derecha	Nominal	Entrada
18	P15	Númerico	8	0	P15.- ¿Ud. apr...	{1, Aprueba}	Ninguna	10	Derecha	Nominal	Entrada
19	P16	Númerico	8	0	P16.- ¿Ud. apr...	{1, Aprueba}	Ninguna	10	Derecha	Nominal	Entrada
20	P17	Númerico	8	0	P17.- Pasando ...	{1, Mucho ...	Ninguna	10	Derecha	Escala	Entrada
21	P18	Númerico	8	0	P18.- ¿Cuál de ...	{1, Partido ...	Ninguna	10	Derecha	Nominal	Entrada
22	P19	Númerico	8	0	P19.- ¿Cuál de ...	{1, Alianza ...	Ninguna	10	Derecha	Escala	Entrada
23	P20	Númerico	8	0	P20.- ¿Con cuál...	{1, Derecha}	Ninguna	4	Derecha	Nominal	Entrada
24	P21_1	Númerico	8	0	P21.- Utilizand...	{1, Nada}...	Ninguna	10	Derecha	Escala	Entrada

Vista de datos: Es similar a una hoja de cálculo en Excel. Es una cuadrícula de caso [fila] x variable [columna]. De esta forma, podemos identificar rápidamente a un caso particular dentro de nuestro conjunto de casos; por ejemplo, podemos querer analizar el comportamiento del caso 127 en la variable P10. No tenemos más que buscar tal caso por su numeración en las filas y revisar los valores que presenta en cada una de las variables (columnas).

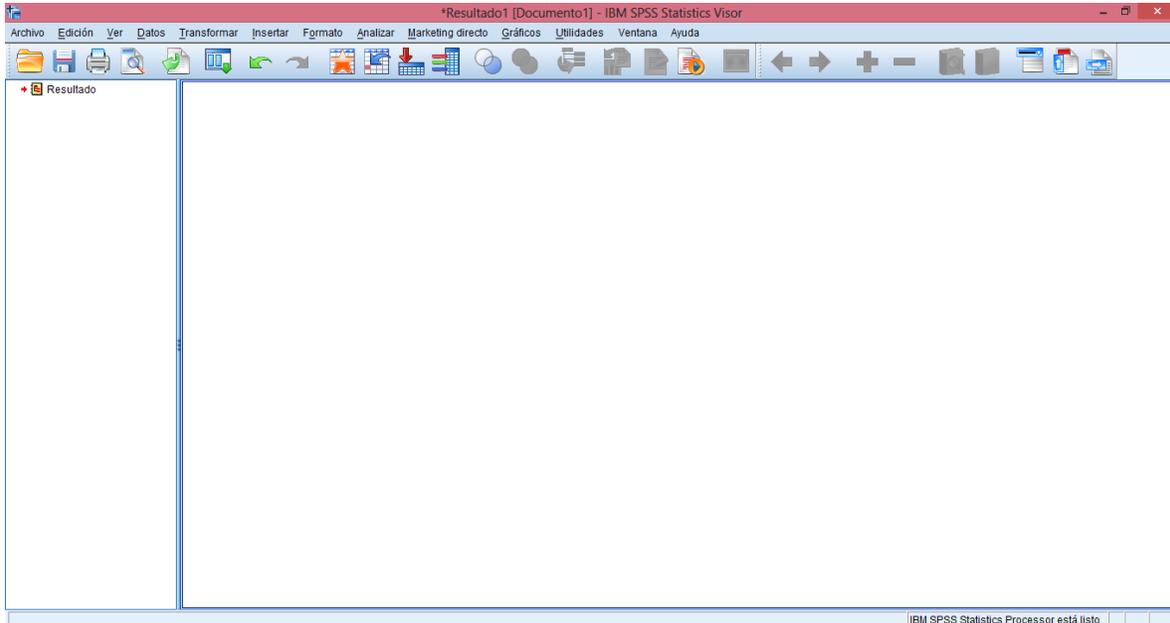
Visible: 194 de 194 variables

	folio	PROYECCION	POND_MU ESTRAL	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
1	1	5925,85	,80803	3	2	2	1	4	2	3	3	2	2	2	2
2	2	8632,25	1,17706	3	3	2	2	3	2	2	3	2	2	2	2
3	3	4655,96	,63487	4	4	3	2	2	1	3	3	1	1	1	1
4	4	6254,00	,85277	2	2	2	2	3	2	2	2	2	1	1	1
5	5	7354,81	1,00287	2	2	2	1	2	1	2	3	1	3	1	1
6	6	9794,68	1,33557	2	2	2	1	3	1	3	3	2	1	2	2
7	8	8632,25	1,17706	3	3	2	2	3	2	3	3	2	3	2	2
8	10	8432,03	1,14976	3	3	2	2	2	2	3	2	1	2	3	3
9	14	9794,68	1,33557	1	3	1	1	1	2	3	2	1	3	1	1
10	15	10608,0	1,44646	1	2	2	1	2	1	3	2	1	1	1	1
11	16	9794,68	1,33557	4	3	2	2	4	1	4	1	1	2	12	12
12	19	5248,96	,71573	3	2	2	2	1	1	4	1	88	88	1	1
13	20	5202,43	,70938	2	2	1	1	2	1	3	2	3	2	3	3
14	21	9794,68	1,33557	3	2	2	2	2	1	4	2	3	1	3	3
15	22	8578,88	1,16978	2	2	1	1	1	1	4	1	2	3	3	3
16	27	3928,14	,53563	3	4	2	2	2	2	4	2	1	1	1	1
17	28	7354,81	1,00287	88	2	2	1	2	1	4	2	88	88	3	3
18	29	5202,43	,70938	2	2	1	1	1	1	4	2	3	3	3	3
19	30	5248,96	,71573	2	3	1	1	2	1	4	2	2	2	2	2
20	31	8003,39	1,09131	3	3	2	2	2	2	3	2	1	1	2	2
21	32	5202,43	,70938	1	88	2	88	3	3	3	3	2	3	2	2
22	33	4655,96	,63487	2	3	2	2	2	1	4	2	2	2	1	1

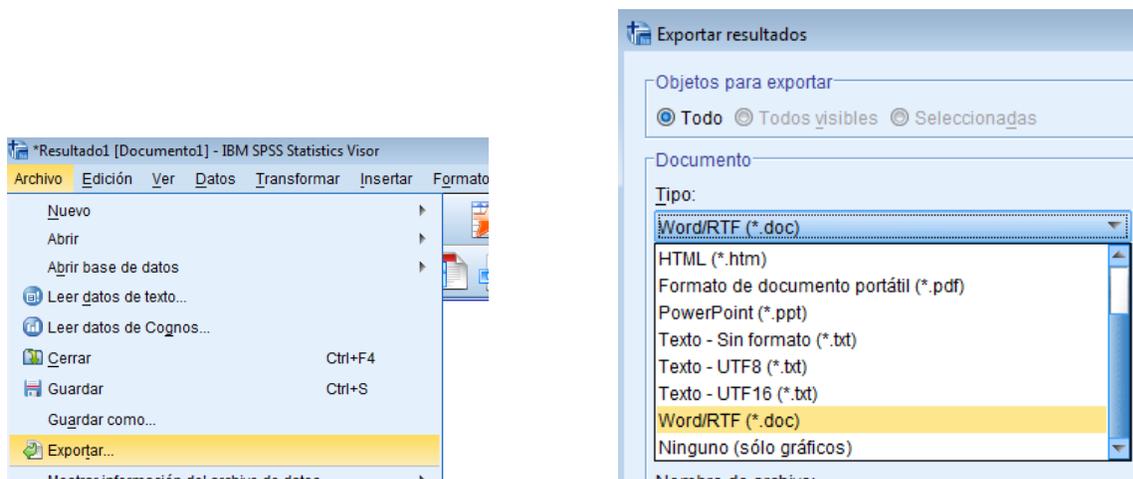
Vista de datos | Vista de variables

IBM SPSS Statistics Processor está listo | Ponderación activada

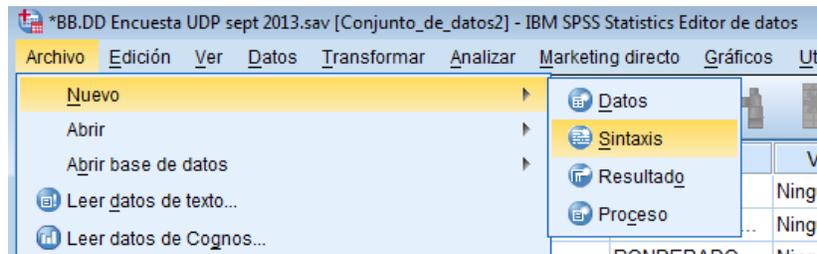
Ventana de Resultados: Esta ventana, que ahora nos aparecerá en blanco, será aquella donde el programa irá mostrando los resultados de los diferentes análisis que ejecutemos. La visualización de tablas, estadísticos y gráficos se da en esta ventana.



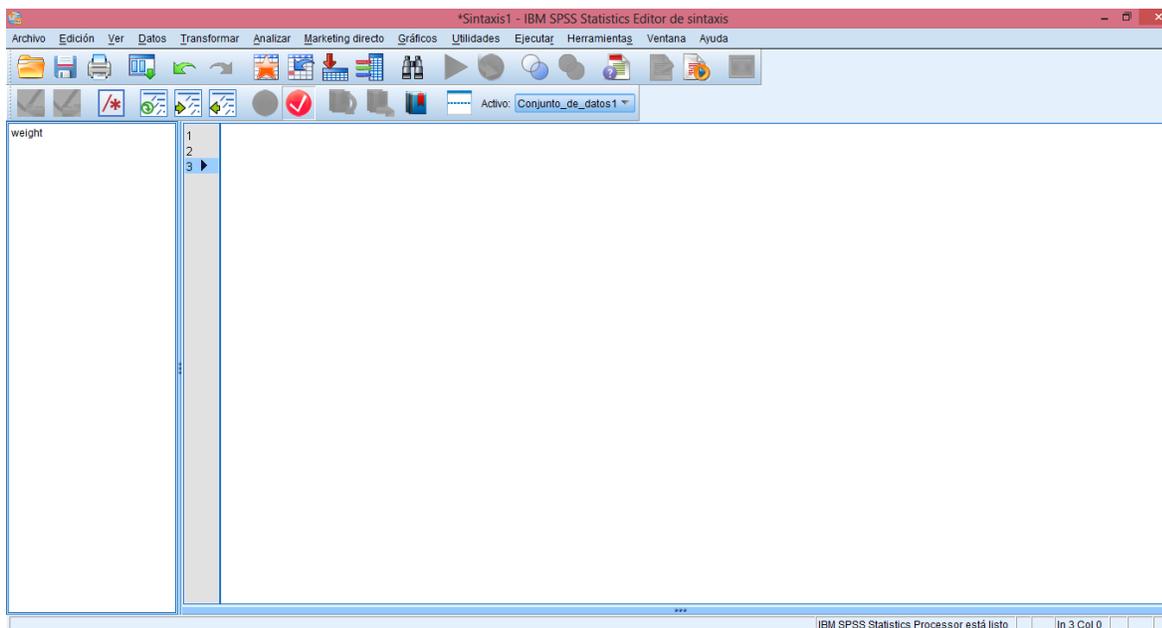
Cabe señalar que mediante la opción *exportar*, podemos guardar este archivo en diferentes formatos, haciendo más fácil su edición y manejo para la presentación de resultados en reuniones, informes etc.



Ventana de sintaxis: Para visualizar la ventana de sintaxis, una vez abierta una base de datos, hacemos la siguiente operación:



Se presenta en blanco cuando la abrimos por primera vez. En esta pestaña podemos escribir e intervenir de manera directa los comandos con los que trabaja el software.



Como ya hemos visto, **SPSS presenta dos interfaces de uso**, una gráfica en la que nos desplegamos a través de ventanas y botones y el editor de sintaxis.

La primera modalidad es similar al uso que damos a la mayoría de los programas: desde un menú con botones desplegable se seleccionan las acciones que se quieren realizar y se señalan distintas instrucciones o especificaciones al programa según lo que se necesite.

La segunda forma corresponde al editor de sintaxis, el cual consiste en escribir instrucciones para el programa en forma de comandos computacionales, en el *visor* de sintaxis. Este modo permite realizar

las mismas funciones que el uso mediante botones, pero tiene mayores beneficios ya que la sintaxis puede ser guardada como texto, copiada, enviada, etc., ahorrando el trabajo de señalar las instrucciones desde el menú repetidas veces (esta ventaja seguramente será difícil de apreciar en un ramo introductorio a la estadística, pero es preciso desde ya familiarizarse con tal uso); es sobre todo cuando requiramos realizar **múltiples operaciones estadísticas simultáneamente**, con la posibilidad de replicar los análisis rápidamente para ejecutarlos en distintas bases de datos, que esta modalidad adquiere mayor utilidad. En este sentido, la utilidad principal de manejar el modo sintaxis es que es posible modificar el texto de la sintaxis y usarla para distintas bases de datos en las que quiera hacerse análisis similares o también para cambiar las especificaciones del análisis que se está realizando sin entregar las instrucciones desde cero (como sí requiere el uso del menú). Manejar el análisis estadístico a nivel de sintaxis nos permitirá **agregar, eliminar o modificar rápidamente, ciertas partes del comando a ejecutar, sin perder el grueso del trabajo** ya hecho.

Ejemplo de uso SPSS: modo botones y modo sintaxis:

Para ejemplificar el uso de las distintas modalidades de SPSS, elaboraremos una tabla de frecuencias para la variable P12: *¿Qué tan interesado está usted en la política?*, que cuenta con las siguientes categorías de respuesta:

1 = Muy interesado

2 = Algo interesado

3 = Poco interesado

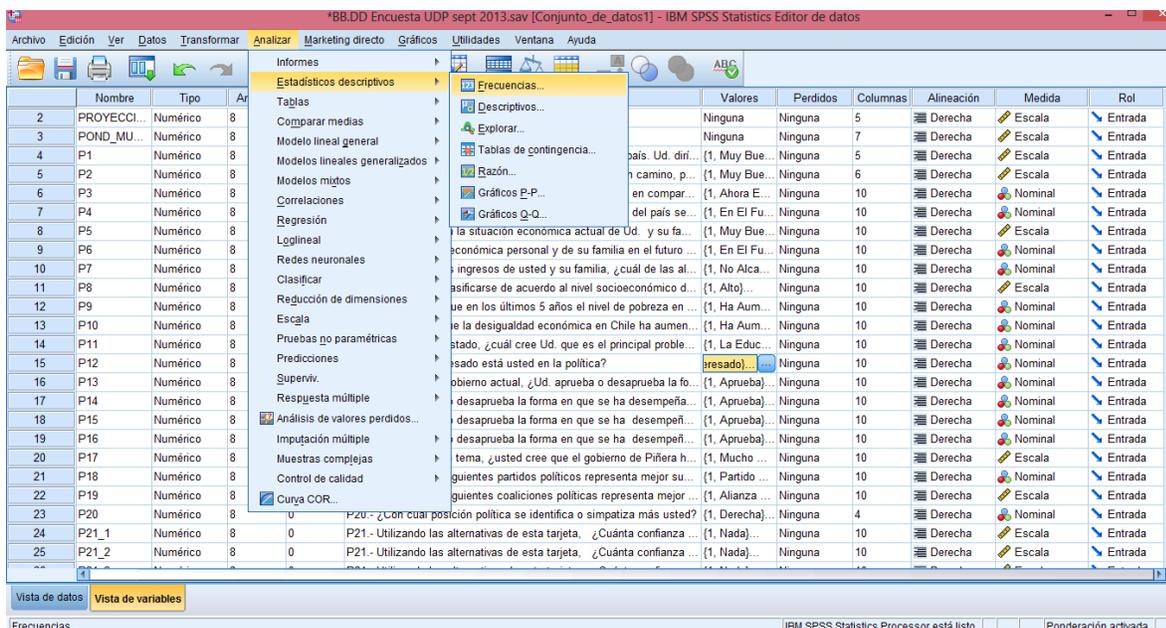
4 = Nada interesado

88 = No sabe

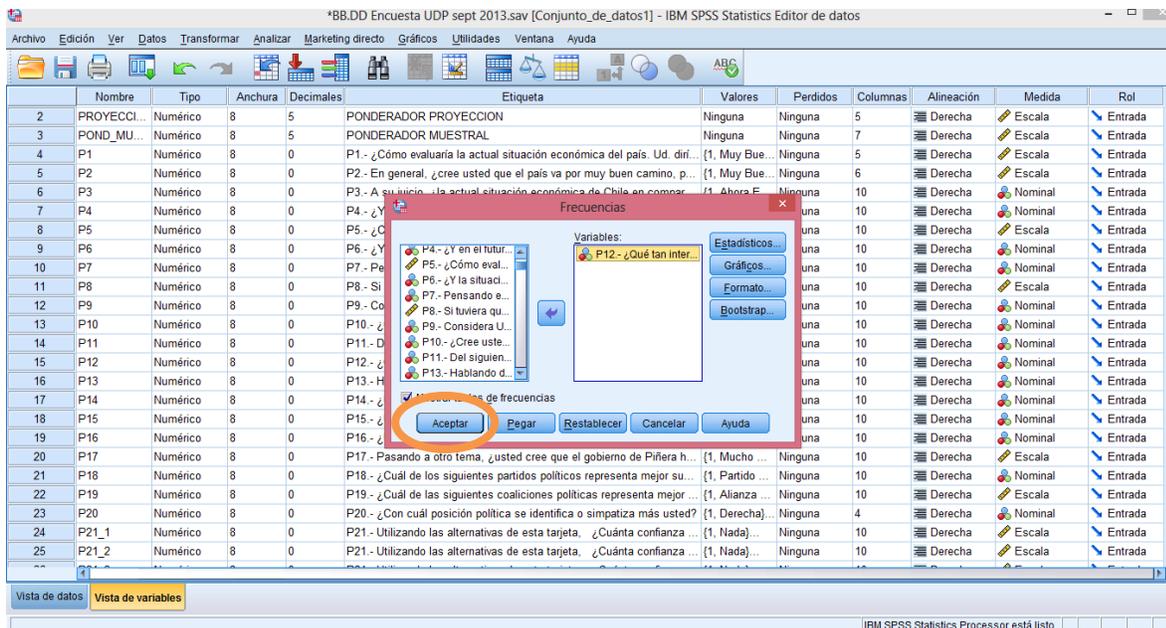
99 = No responde

Primero lo realizaremos desde la interfaz de botones:

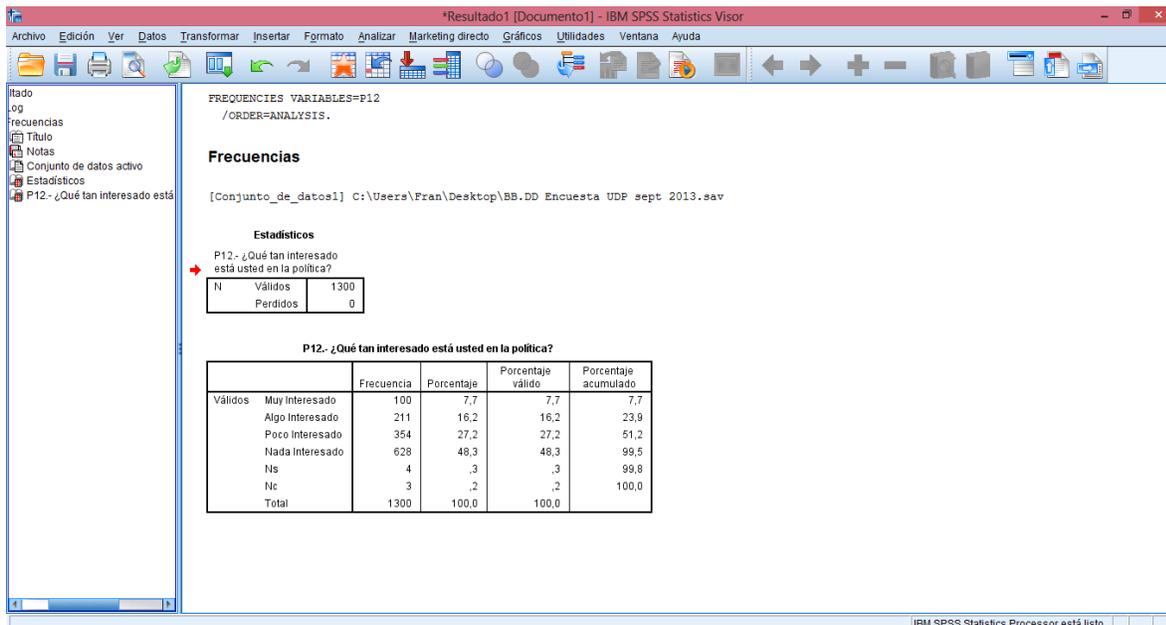
Desde la opción *analizar*, a través de la opción *estadísticos descriptivos*, le damos al botón que dice *frecuencias*:



Luego, en el cuadro de diálogo *Frecuencias* buscamos la variable que queremos analizar y la llevamos hacia el espacio en blanco titulado *Variables*, donde deben incluirse la(s) variable(s) a analizar. Una vez hecho tal procedimiento presionamos el botón *Aceptar*:



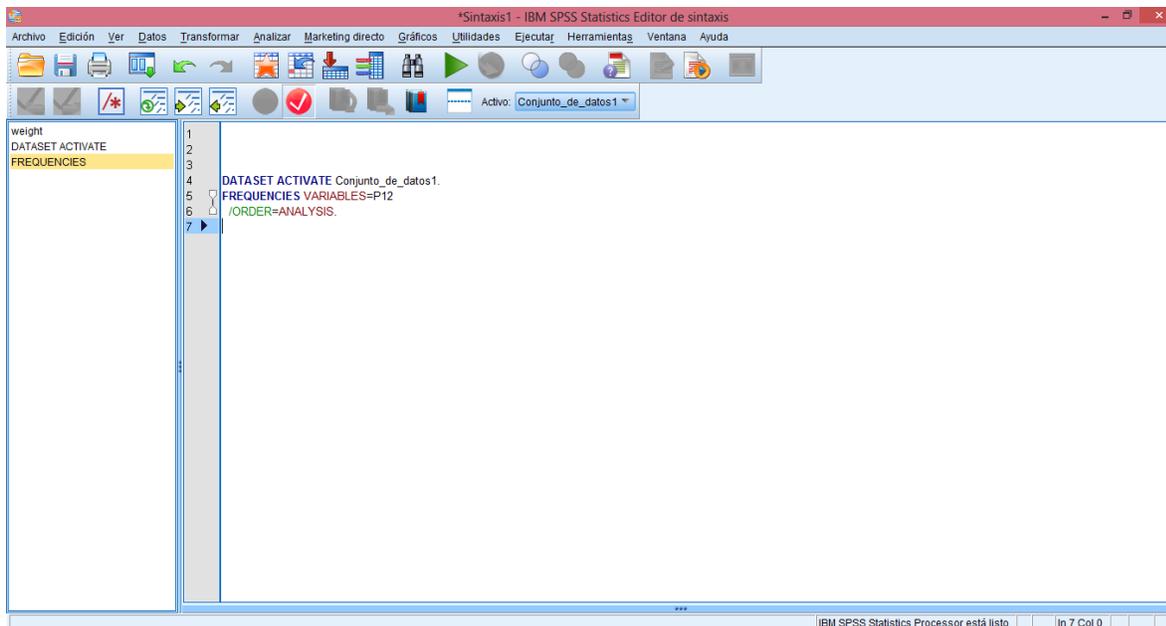
En la ventana *Resultados* podremos observar los resultados de la operación efectuada, obteniendo una tabla de frecuencias que nos indica la cantidad de casos por cada categoría de la variable analizada:



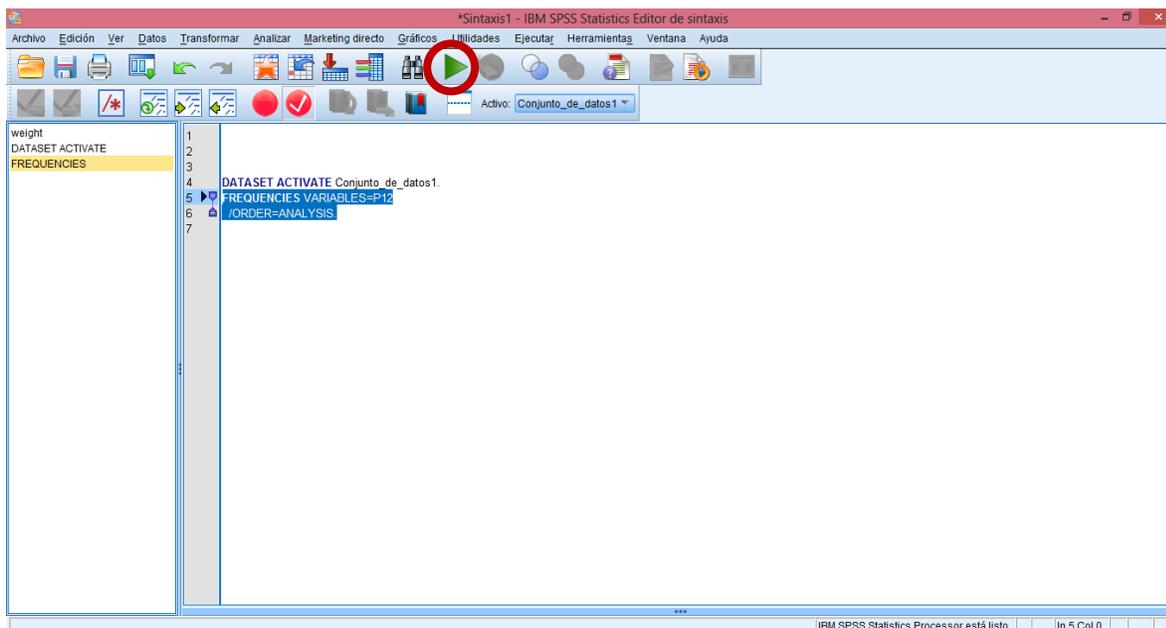
Ahora realizaremos el mismo ejercicio, pero empleando el editor de sintaxis. Para ello, escribimos el siguiente comando en el cuadro de diálogo:

FRECUENCIAS VARIABLES=P12

/ORDER=ANALYSIS.



Seleccionamos el texto recién escrito y presionamos el botón con la flecha (play) verde 'Ejecutar selección'.



Como podrán ver, la ventana de resultados que se obtiene es la misma que la obtenida mediante el procedimiento anterior.

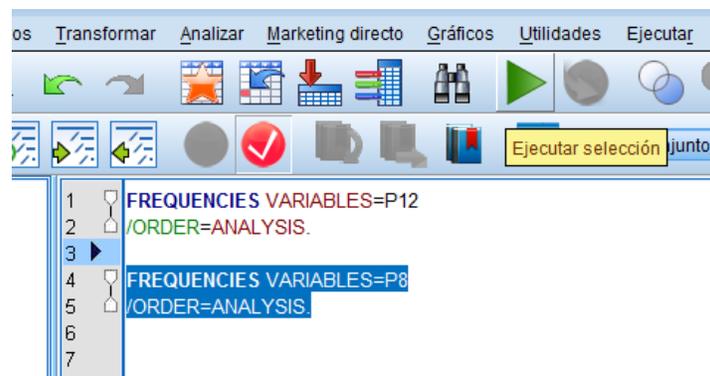
También puede cambiarse la variable que se quiere describir modificando la sintaxis de manera directa, por ejemplo:

Tomamos el comando inicial y cambiamos el nombre de la variable:

```
FRECUENCIAS VARIABLES=P12  
/ORDER=ANALYSIS.
```

```
FRECUENCIAS VARIABLES=P8  
/ORDER=ANALYSIS.
```

Y ejecutamos el comando recién modificado directamente desde el editor de sintaxis (no olvidar seleccionar el comando a ejecutar):



Se ha modificado la variable P12 e introducido la variable P8: *Si tuviera que clasificarse de acuerdo al nivel socioeconómico de su familia, ¿en qué nivel se clasificaría Ud.?*

1 = Alto

2 = Medio-alto

3 = Medio

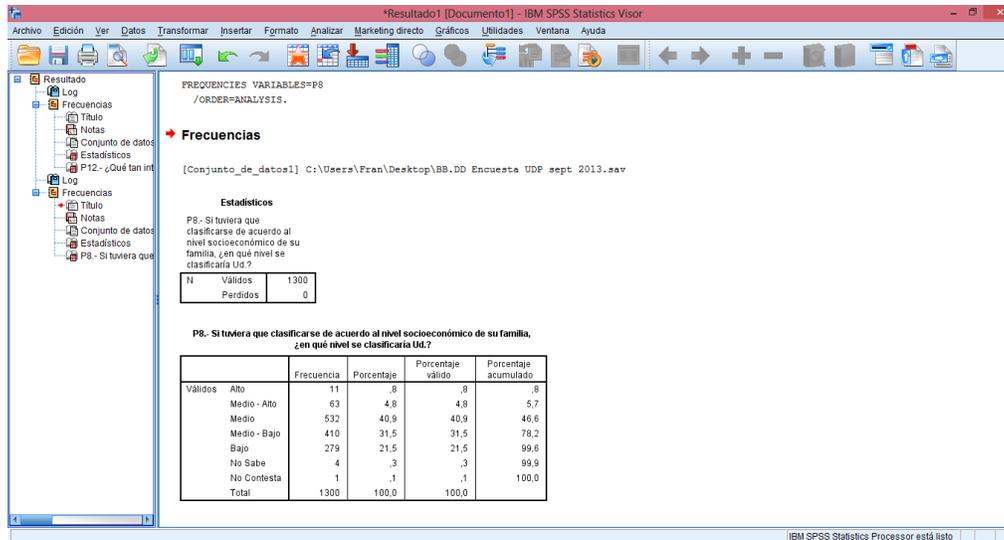
4 = Medio-bajo

5 = Bajo

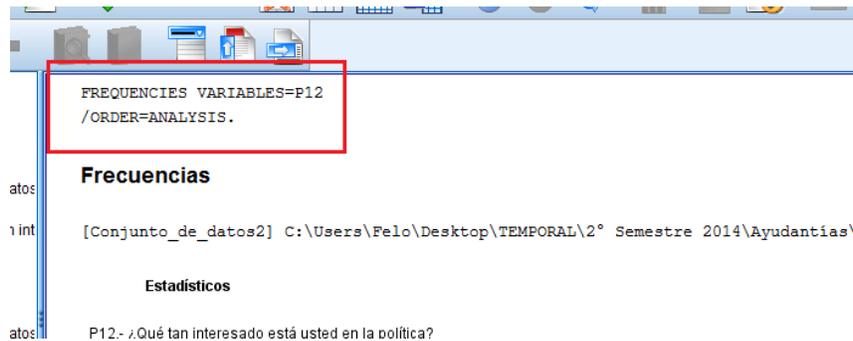
88 = No sabe

99 = No contesta

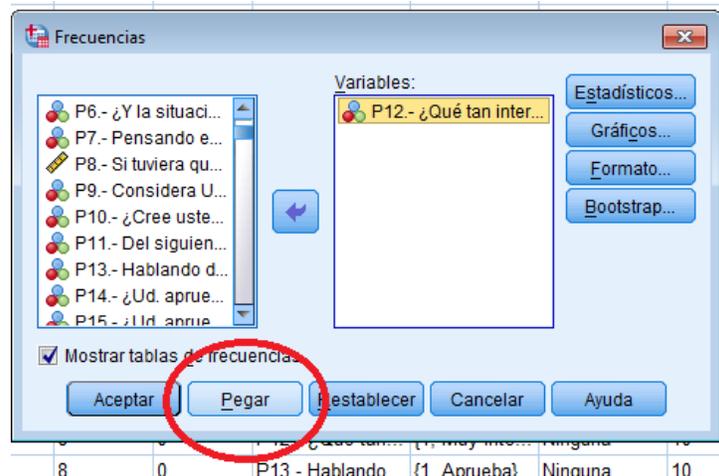
De esta forma, podemos ejecutar análisis directamente desde el editor de sintaxis de forma más rápida, sin realizar la operación paso a paso como es requerido en la interfaz de botones. En la ventana de resultados se mostrarán los resultados del segundo análisis ejecutado, a continuación de los resultados del primero.



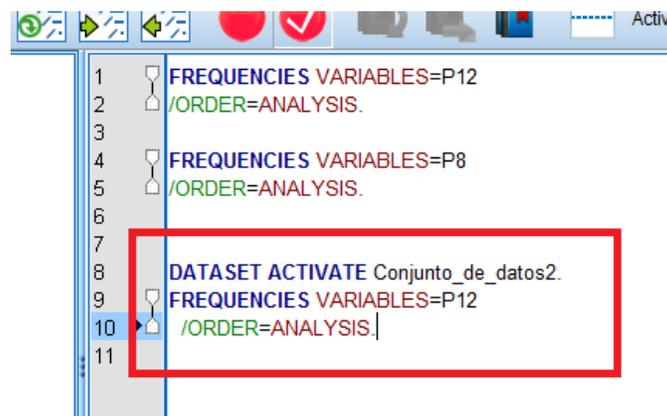
Ahora bien, mientras hacemos uso de los botones, también estamos haciendo uso de la sintaxis sólo que sin verla de manera directa, pero si nos fijamos en el visor de resultados, podremos ver el listado de comandos que fueron ejecutados mediante los botones:



Una opción para ir dejando registro de nuestra configuración mediante botones en una sintaxis que sea susceptible de posterior modificaciones, es acostumbrarnos a usar el botón pegar, antes de echar a correr cualquier análisis desde el visor de botones:



Al emplear este botón no echaremos a correr inmediatamente el análisis, sino que activaremos un paso intermedio: nuestra configuración mediante botones quedará registrada – una vez apretado el botón *Pegar* – en el editor de sintaxis. Al apretar tal botón desaparecerá el cuadro de diálogo sin dar inicio al análisis. Para comenzarlo debemos ir al editor de sintaxis, seleccionar el comando recién creado y ejecutarlo de la manera recién vista.



Esta posibilidad que da SPSS puede llegar a ser bastante útil para mezclar el uso de botones (que no requiere memorizar códigos dado su uso intuitivo) a la vez que se cumple el objetivo de contar con un registro modificable de los análisis solicitados al programa.

Excel: uno de los olvidados

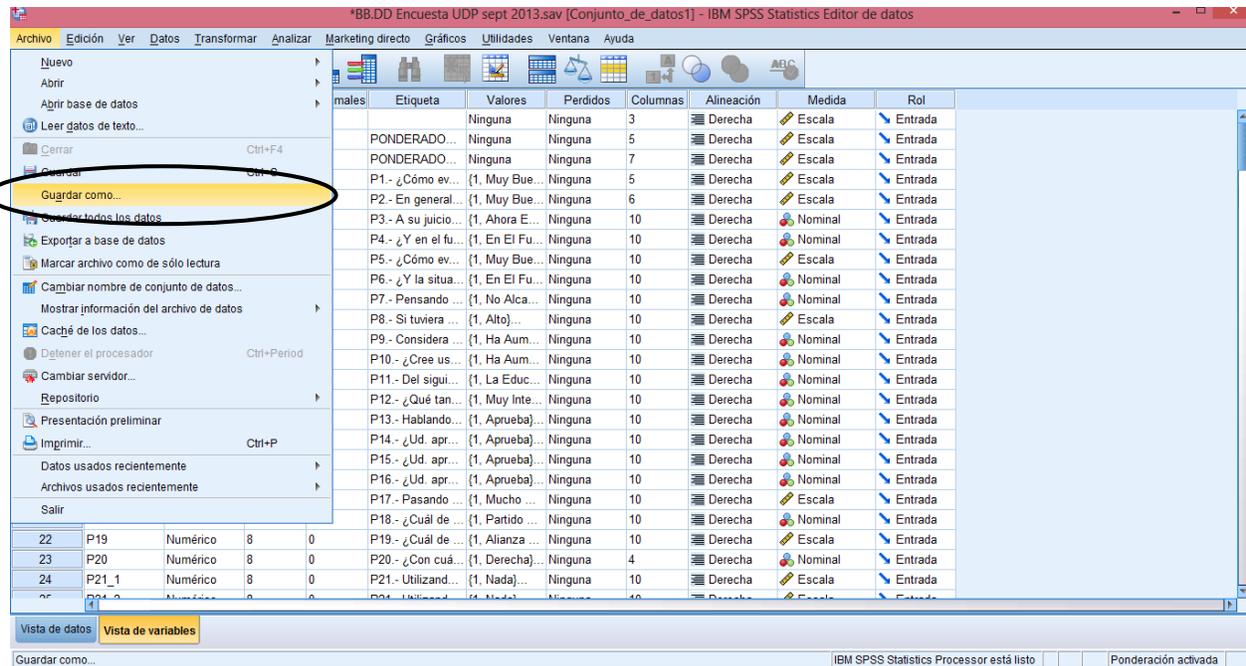
Por otra parte, Excel “es un software que permite crear tablas, y calcular y analizar datos. Este tipo de software se denomina software de hoja de cálculo. Excel permite crear tablas que calculan de forma automática los totales de los valores numéricos que especifica, imprimir tablas con diseños cuidados, y crear gráficos simples. Excel forma parte de “Office”, un conjunto de productos que combina varios tipos de software para crear documentos, hojas de cálculo y presentaciones, y para administrar el correo electrónico.” También corresponde a un software pagado (requiere comprar una licencia de Microsoft Office), sin embargo, es usual contar ya con una licencia básica al comprar un computador o conseguir una clave de licencia.

Muchas de las bases de datos disponibles, ya sea que provengan de instituciones del gobierno o de otro tipo, se encuentran en formato .xls, lo que implica que **como investigadoras e investigadores debemos familiarizarnos con esta herramienta** y ser capaces de trabajar con ambas (y otras si es posible) en comunicación (nociones claves en este sentido son ‘extensión de archivo’ y ‘exportar’). Además, Excel permite realizar operaciones relativas a estadística descriptiva y multivariada, y cuenta con un interfaz más cómoda para la digitación de datos.

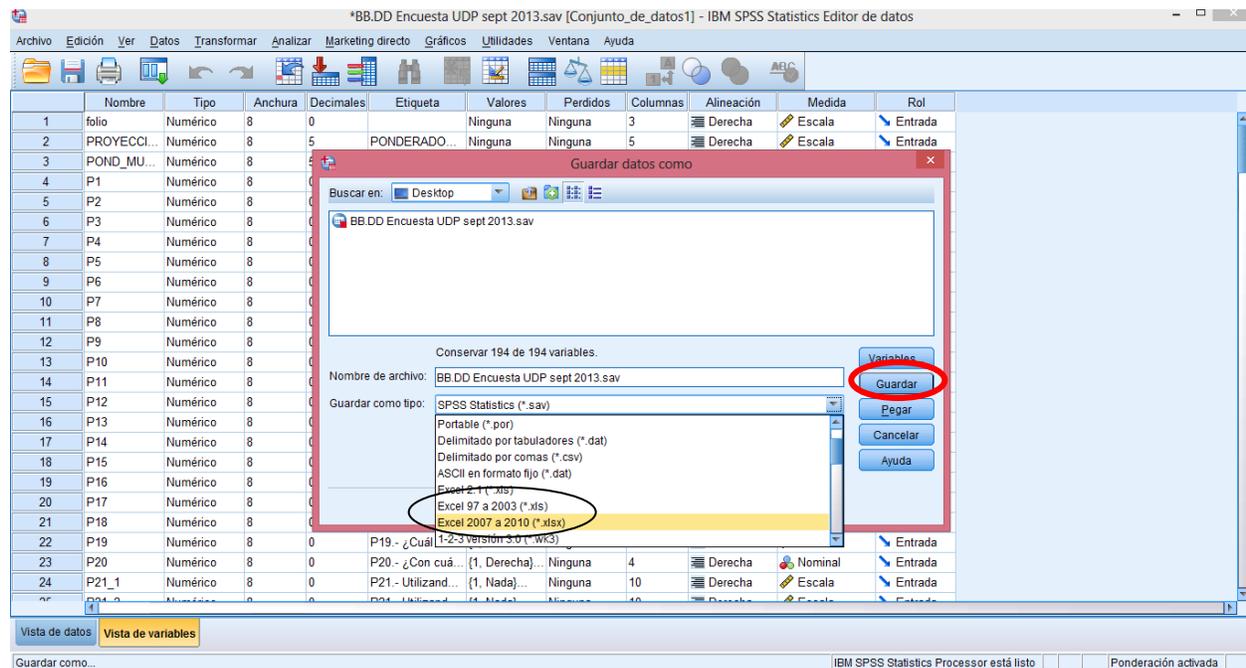
La interfaz de este software se presenta como una hoja de cálculo compuesta por filas numeradas y columnas ordenadas alfabéticamente; la intersección entre ellas se denomina ‘celda’. En cada celda es posible ingresar datos en distintos formatos, tales como texto, números, hora, fecha, etc. Además, es posible operar con estos datos a partir de funciones (que también se escriben en una celda). Una función opera de la misma manera en que lo aprendimos en matemática. Se le da la instrucción al programa para que efectúe una determinada operación matemática con la **información disponible en un conjunto de celdas**. Esto, según el tipo de datos, puede ir desde sumar, restar, calcular medias u otros estadísticos más sofisticados. Incluso, puede realizar operaciones de tremenda utilidad cuando estamos manejando conjuntos de datos extensos, tales como ordenar, contar, filtrar, juntar, dividir o borrar casos, etc.

Vista de datos en Excel

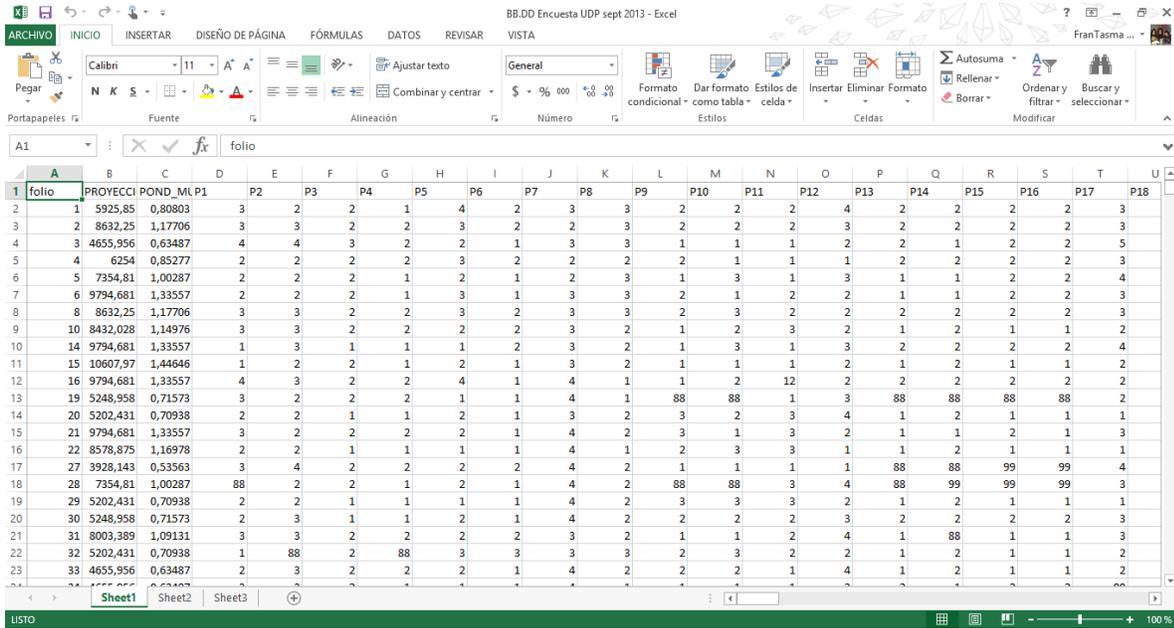
Con la base de datos abierta ir al menú 'Archivo', luego 'Guardar como'.



Luego en la barra 'Guardar como tipo': seleccionar el formato de Excel (.xls o .xlsx según la versión), y hacer clic en 'Guardar'.



Al abrir la base de datos con Excel nos encontramos con que la interfaz se corresponde con la interfaz de datos de SPSS (casos por fila, variables por columna) sin tener un modo de visualización por variables, por ejemplo.



Obtención de tablas de frecuencias en Excel

BB.DD Encuesta UDP sept 2013 - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	folio	PROYECCI	POND_ML	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18
2	1	5925,85	0,80803	3	2	2	1	4	2	3	3	2	2	2	4	2	2	2	2	2	3
3	2	8632,25	1,17706	3	3	2	2	3	2	2	3	2	2	2	3	2	2	2	2	2	3
4	3	4655,956	0,63487	4	4	3	2	2	1	3	3	1	1	1	2	2	1	2	2	5	
5	4	6254	0,85277	2	2	2	2	3	2	2	2	1	1	1	1	2	2	2	2	3	
6	5	7354,81	1,00287	2	2	2	1	2	1	2	3	1	3	1	3	1	1	2	2	4	
7	6	9794,681	1,33557	2	2	2	1	3	1	3	3	2	1	2	2	1	1	2	2	3	
8	8	8632,25	1,17706	3	3	2	2	3	2	3	3	2	3	2	2	2	2	2	2	3	
9	10	8432,028	1,14976	3	3	2	2	2	2	3	2	1	2	3	2	1	2	1	1	2	
10	14	9794,681	1,33557	1	3	1	1	1	2	3	2	1	3	1	3	2	2	2	2	4	
11	15	10607,97	1,44646	1	2	2	1	2	1	3	2	1	1	1	2	1	2	1	1	2	
12	16	9794,681	1,33557	4	3	2	2	4	1	4	1	1	2	12	2	2	2	2	2	2	
13	19	5248,958	0,71573	3	2	2	2	1	1	4	1	88	88	1	3	88	88	88	88	2	
14	20	5202,431	0,70938	2	2	1	1	2	1	3	2	3	2	3	4	1	2	1	1	1	
15	21	9794,681	1,33557	3	2	2	2	2	1	4	2	3	1	3	2	1	1	2	1	3	
16	22	8578,875	1,16978	2	2	1	1	1	1	4	1	2	3	3	1	1	2	1	1	1	
17	27	3928,143	0,53563	3	4	2	2	2	2	4	2	1	1	1	1	88	88	99	99	4	
18	28	7354,81	1,00287	88	2	2	1	2	1	4	2	88	88	3	4	88	99	99	99	3	
19	29	5202,431	0,70938	2	2	1	1	1	1	4	2	3	3	3	2	1	2	1	1	1	
20	30	5248,958	0,71573	2	3	1	1	2	1	4	2	2	2	2	3	2	2	2	2	3	
21	31	8003,389	1,09131	3	3	2	2	2	2	3	2	1	1	2	4	1	88	1	1	3	
22	32	5202,431	0,70938	1	88	2	88	3	3	3	2	3	2	2	2	1	2	1	1	2	
23	33	4655,956	0,63487	2	3	2	2	2	1	4	2	2	2	1	4	1	2	1	1	2	

Utilizaremos dos hojas de cálculo: una que contiene todos los datos y otra para realizar las tablas de frecuencias (Hoja 2 que nombraremos como *Tab_freq*). En esta hoja renombrada, creamos las tablas para las variables que queremos describir: P12 (Interés en la política) y P8 (Nivel socioeconómico), con sus respectivas categorías de respuesta, una celda para el total, y columnas para expresar las frecuencias y los porcentajes equivalentes.

BB.DD Encuesta UDP sept 2013 - Excel

P12	Frecuencia	Porcentaje
Int_pol		
Muy interesado		
Algo interesado		
Poco interesado		
Nada interesado		
No sabe		
No responde		
Total		

P8	Frecuencia	Porcentaje
NSE		
Alto		
Medio-alto		
Medio		
Medio-bajo		
Bajo		
No sabe		
No responde		
Total		

Usando la fórmula CONTAR.SI podemos calcular la frecuencia de cada categoría para las variables P12 y P8. La tabla azul muestra cómo han sido utilizadas las fórmulas (están escritas entre comillas para que Excel no efectúe la operación). El círculo rojo encierra dos funciones 'Estilo porcentual' (%) para mostrar los decimales como porcentajes y 'Estilo millares' para pasar a decimales en este caso. Sobre el círculo se encuentra una barra que dice 'General', ésta sirve para cambiar el formato de la celda en caso de trabajar con fechas, texto, divisas, etc.

Para calcular las frecuencias empleamos la función CONTAR.SI; de manera general, cuando escribimos una función en una celda de Excel, se le debe anteponer un signo *igual* ("=") para que el programa reconozca que se trata de una función. Al momento de ingresarla, el mismo software nos indica el orden de los elementos que componen la función.

La especificación de la función CONTAR.SI es la siguiente: **=Función(Rango;Criterio)**

El rango indica dónde – en qué cuadrante de celdas - buscar los datos (en el caso que ejemplificaremos caso es en la Hoja1 y la columna O completa). El criterio es el valor que se está buscando (en nuestro ejemplo sería el valor 1) si es texto se pone entre comillas.

=CONTAR.SI(Hoja1!O:O;1)

Para realizar el cálculo de % (frecuencia/total) a partir del conjunto de datos, le indicamos al programa que haga el mismo cálculo para contar las frecuencias, pero que a cada valor lo divida por el total de casos. Para ello se ocupa el signo \$ para indicar que la celda de referencia (el divisor de todas las celdas) debe quedar fija. Para ingresarlo rápidamente es posible apretar la tecla F4 justo después del nombre de la celda escrita: 'J16'+F4 → '\$J\$16'

P12

Int_pol	Frecuencia	Porcentaje
Muy interesado	100	8%
Algo interesado	211	16%
Poco interesado	354	27%
Nada interesado	628	48%
No sabe	4	0%
No responde	3	0%
Total	1300	100%

P8

NSE	Frecuencia	Porcentaje
Alto	11	1%
Medio-alto	63	5%
Medio	532	41%
Medio-bajo	410	32%
Bajo	279	21%
No sabe	4	0%
No responde	1	0%
Total	1300	100%

Formulas:

Int_pol	Frecuencia	Porcentaje
Muy interesado	"=CONTAR.SI(Hoja1!O:O;1)"	"=F4/\$F\$10"
Algo interesado	"=CONTAR.SI(Hoja1!O:O;2)"	"=F5/\$F\$10"
Poco interesado	"=CONTAR.SI(Hoja1!O:O;3)"	"=F6/\$F\$10"
Nada interesado	"=CONTAR.SI(Hoja1!O:O;4)"	"=F7/\$F\$10"
No sabe	"=CONTAR.SI(Hoja1!O:O;88)"	"=F8/\$F\$10"
No responde	"=CONTAR.SI(Hoja1!O:O;99)"	"=F9/\$F\$10"
Total	"=CONTAR(Hoja1!O:O)"	"=F10/\$F\$10"

Explicación de fórmulas:

=CONTAR.SI(Hoja1!O:O;1)
=Función(Rango;Criterio)

El rango indica dónde buscar los datos (en este caso es en la Hoja1 y la columna O completa. El criterio es el valor que se está buscando, si es texto se pone entre comillas.

Cálculo de % (frecuencia/total)
 El signo \$ es para indicar que la celda de referencia debe quedar fija.

Fuentes de datos secundarios

Uno de los principales conocimientos que se requieren para llevar a cabo análisis estadísticos es conocer desde donde podemos obtener fuentes de datos secundarios; si bien parece algo trivial, no siempre lo es, pues hay que tener claro en primer lugar **qué** es lo que necesitamos, para saber dónde buscar.

Existe una variada gama de encuestas y estudios, que son realizados de manera periódica en nuestro país, dentro de ellas podemos destacar, dentro de las más empleadas:

Institución	Encuesta	Enlace
Centro de Estudios Públicos	Encuesta Nacional de Opinión Pública	http://www.cepchile.cl/dms/lang_1/encuestasCEP.htm
Universidad Diego Portales (ICSO)	Encuesta Nacional UDP	http://www.encuesta.udp.cl/
Programa de las Naciones Unidas para el Desarrollo (PNUD)	Desarrollo Humano en Chile	http://www.desarrollohumano.cl/base-datos/base-datos.htm
Ministerio de Salud	Encuesta Nacional de Salud	http://epi.minsal.cl/vigilancia-epidemiologica/sistema-de-vigilancia-de-enfermedades-no-transmisibles-vent/solicita-base/ (Se debe solicitar creando un usuario)

Sin embargo, es preciso tener en cuenta que cada uno de estos estudios presenta sus objetivos particulares, por lo que la información que queramos obtener desde ellos estará delimitada por tales orientaciones. En tal sentido, una recomendación importante es la importancia de contar con una relativa **orientación** al comenzar cualquier trabajo de investigación (objetivos, hipótesis, áreas temáticas de interés, etc.), pues sólo tal condición nos ahorrará tiempo valioso al momento de buscar posibles

fuentes de información (sin estar recorriendo cuestionarios incansablemente, hasta que por arte de magia se nos ocurra sobre qué hacer nuestro informe.

El siguiente cuadro resume las principales orientaciones de los estudios ya mencionados.

Institución	Características
Centro de Estudios Públicos	“La Encuesta Nacional de Opinión Pública CEP representa un estudio académico y objetivo de las actitudes y predisposiciones políticas, económicas y sociales de todos los habitantes del país . A través de este tipo de estudios se desea conocer las preocupaciones, preferencias y necesidades de la población.” (CEP, 2013)
Universidad Diego Portales (ICSO)	“ Conocer la opinión pública a nivel nacional y regional de los siguientes temas: Políticas públicas y sociales, opiniones sobre la coyuntura política y económica, confianza en instituciones, democracia y representación, familia e intimidad, percepción sobre los empresarios, participación ciudadana y medio ambiente, y discriminación. ” (UDP, 2013)
Programa de las Naciones Unidas para el Desarrollo (PNUD)	<p><i>Las características dependen del área de estudio específica definida para el año en cuestión.</i></p> <p>“Lo que aquí se pretende mostrar es que, para resignificar el horizonte del desarrollo poniendo al centro los proyectos de vida de las personas, es necesario ampliar la mirada y asumir en toda su complejidad una dimensión hasta ahora menos atendida del desarrollo, y de la cual el fenómeno de la felicidad forma parte: la subjetividad. Se entiende por subjetividad el espacio y el proceso en que los individuos construyen una imagen de sí, de los otros y del mundo en el contexto de sus experiencias sociales. Este ámbito está formado por sus emociones, imágenes, percepciones, deseos, motivaciones y evaluaciones, entre otros elementos.” (PNUD, 2012)</p>
Ministerio de Salud	-----

Elementos descargables diferentes: utilidades diferentes

Cuando accedamos a alguno de los links que ya hemos anotado, nos encontraremos con la posibilidad de descargar diferentes elementos, que es preciso diferenciar para saber cómo ocupar.

En primer lugar, muchas veces es bastante visible la **presentación, informe o sinopsis** (o algún nombre afín). Este archivo por lo general no nos será tan útil, ya que se trata de la presentación que hace la institución de los principales resultados que arrojó el estudio. A veces se trata de una presentación powerpoint que incluye frecuencias, porcentajes y gráficos. Evidentemente, un documento así no sirve para nuestras intenciones: a nosotros nos interesa analizar datos, no que nos informen sobre el análisis que hizo otro; sin embargo, en algunos casos (como los informes del PNUD, en este documento se suelen incluir aspectos teóricos relevantes que pueden sernos de utilidad a la hora de encara el análisis de datos). No suelen incluir precisiones metodológicas.

En segundo lugar podemos mencionar la **base de datos**. Usualmente es posible encontrar bases de datos en formato *.xls* (planilla Excel) o *.sav* (formato SPSS) susceptibles de abrir con los programas. Cualquiera sea el formato, una base de datos es – en formato virtual – un pliego de papel blanco cuadriculado donde podemos ir poniendo números. Evidentemente éste archivo nos interesa, ya que si lo tenemos en nuestro poder podemos iniciar nuestros propios análisis sobre la información producida por otras instituciones. Es el archivo que abriremos con nuestro software de análisis estadístico.

En tercer lugar, tenemos el **cuestionario**. En algunos casos (Encuesta CEP, por ejemplo), este viene incluido en el manual de usuario que describiremos en el siguiente párrafo. Más allá de que la **base de datos** sea el archivo clave, por incluir la información producida, gracias al **cuestionario** será posible entender el estudio y decidir sobre qué variables (preguntas) emplearemos en nuestro análisis. Muchas veces al abrir una base de datos se hará dificultosa la búsqueda de variables de interés, procedimiento que se realiza de manera expedita al inspeccionar el cuestionario.

En último lugar, es preciso mencionar el **manual de usuario**. En este documento encontraremos la **información técnica** necesaria para hacernos una idea de las principales características y alcances del estudio. Datos como tamaño de la muestra, población representada, nivel de confianza, margen de error, error de cobertura, tasas de reemplazo y factores de ponderación de los datos, estarán indicados con detalle en este documento.

Otros programas de análisis de datos: ¿qué alternativas existen?



Ya hemos revisado dos de los programas computacionales más empleados para análisis estadístico de datos. Sin embargo, siguiendo a Elousa (Elousa, 2009), y adoptando como criterio evaluador para estos programas el triple criterio de (1) generalidad, (2) costo y (3) facilidad de uso, ¿podemos decir que estos software sean los más indicados? ¿existen otras alternativas para su uso?

En nuestra opinión, tanto SPSS como Excel cumplen con el criterio (3), **facilidad de uso**. Son softwares amigables, que ponen a disposición del usuario una amplia variedad de herramientas de uso intuitivo, lo que ayuda mucho en la introducción al análisis de datos empleando softwares computacionales. Sin embargo, podemos afirmar que se quedan cortos en cuanto a los criterios **(1) generalidad y (2) costo**. Si bien para los análisis que llevaremos a cabo en esta asignatura, ambos programas cuentan con las herramientas suficientes, para análisis más sofisticados, se quedan cortos en las herramientas que disponen, así como en la posibilidad de realizar análisis más personalizados y no por defecto.

En estricto rigor, el punto del **costo** no deja de ser menor; en cualquier caso, tendremos que contar con una licencia para poder ocuparlos correctamente en nuestros computadores personales, y para ello, debemos comprar una (son bastante costosas) o si no, deberemos piratearlo (situación que a veces puede resultar algo engorrosa, sobre todo para el usuario que no es tan experimentado en tal tipo de procedimientos computacionales).

Así y todo en el caso de SPSS, por ejemplo, la versión básica no viene con todos los paquetes de análisis (el paquete AMOS, para ecuaciones estructurales por ejemplo, se vende por separado) lo que entrapa aún más la utilidad de estos programas en la medida que nuestros requerimientos aumentan.

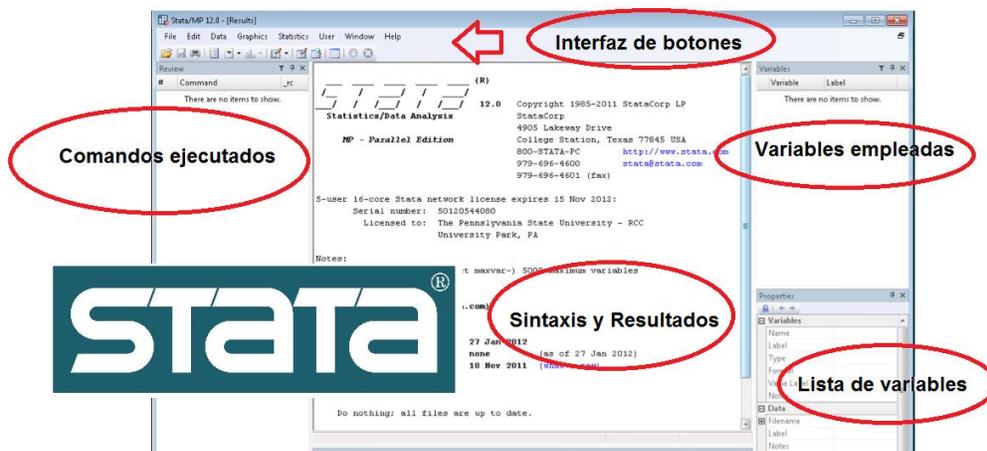
La experiencia dice, además, que son programas algo pesados, que ocupan bastante memoria para poder correr los análisis. Contar con una interfaz llena de colores y botones tiene el costo de que consumirá más memoria de nuestros computadores, haciendo difícil la realización de varias tareas de manera simultánea, sobre todo cuando analicemos bases de datos extensas (como en el caso de censos, por ejemplo) o corramos análisis que requieran de muchos procedimientos complejos.

Stata

Un programa de análisis estadístico de amplio uso y que creemos cumple de mejor manera con el requerimiento (1) de **generalidad** es Stata. Si bien también tiene la problemática del costo, es fácilmente crackeable y presenta una serie de ventajas ante las limitantes que presentan Excel y SPSS.

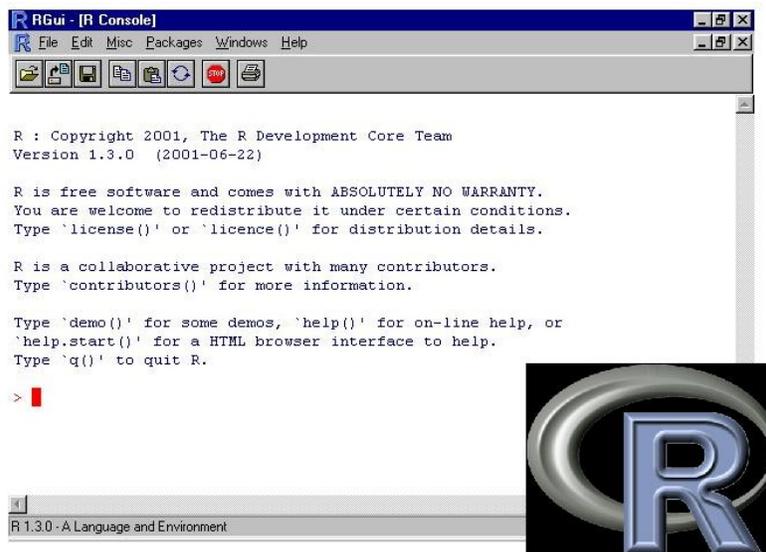
En primer lugar, dispone de más herramientas de análisis en su modalidad básica y permite un uso bastante fino de ellas. En segundo lugar, su interfaz básica articula de manera simultánea una interfaz de botones, con un visor de resultados y editor de sintaxis, un visor que despliega los comandos ejecutados recientemente, un visor que despliega las variables empleadas recientemente en los análisis y una ventana que despliega el listado de variables de la base de datos.

De esta forma, se constituye como un programa amigable, al contar con un modo botones, pero que al mismo tiempo combina en una sola pantalla, distintas herramientas que hacen más expedita la labor de análisis. A la vez, es un programa más liviano respecto a los dos anteriores, y permite modificar la cantidad de memoria en uso por el software con el fin de manejar la potencia de los análisis estadísticos a criterio del usuario.



R: un incomprendido

Finalmente tenemos a R, un freeware (programa de distribución gratuita). La modalidad básica de este programa de análisis estadístico dista mucho de cumplir el requerimiento **(3) facilidad de uso**. Como vemos en la imagen a continuación, la interfaz no dista mucho de la sofisticación de un bloc de notas ya que sólo presenta un espacio en blanco donde escribir los comandos, mientras que ninguno de los botones sirven para realizar análisis estadísticos. Básicamente es un editor de sintaxis, lo que en principio requeriría la habilidad de manejar al dedillo todos y cada uno de los comandos necesarios para ejecutar algún tipo de análisis mediante el particular lenguaje de programación de este freeware. Es fundamentalmente esta característica, la que inhibe al usuario en su uso, aunque presenta atributos que lo sitúan por encima de cualquier otro software que hayamos mencionado.



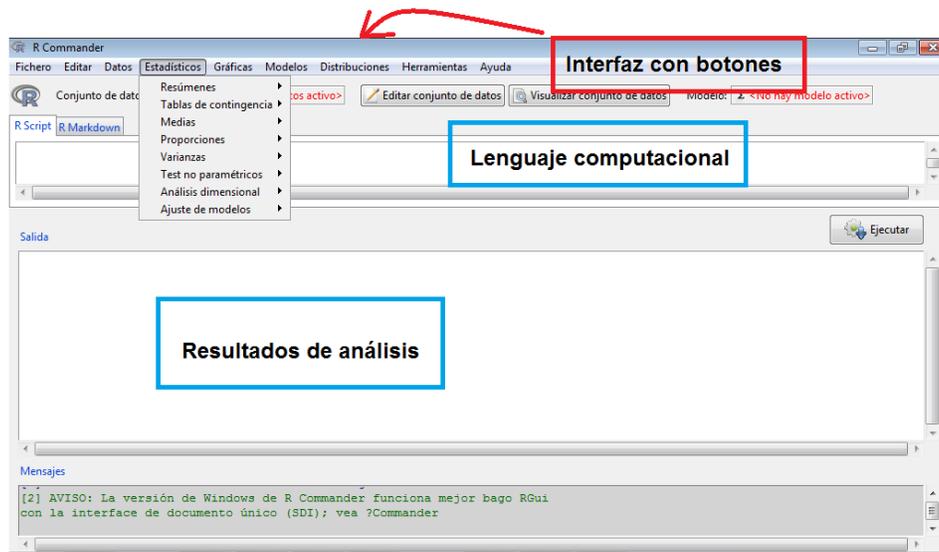
Respecto a nuestra primera exigencia de **generalidad**, es quizá uno de los softwares con una mayor variedad de herramientas de análisis estadístico, no sólo para las ciencias sociales sino también para otras disciplinas². Al ser un programa de contribución libre, vemos cumplida la exigencia número (2) relativa al **costo**, pues existe una amplia variedad de paquetes descargables de modo gratuito

² Como lo plantea Paula Elousa, R "está constituido por más de 1.400 paquetes integrados con los que es posible ejecutar simples análisis descriptivos o aplicar los más complejos y novedosos modelos formales. Además, la incorporación a R de interfaces gráficas como Rcommander que crean entornos de trabajo amigables muy similares al entorno del SPSS permiten saltar la barrera de la accesibilidad, y utilizarlo sin ningún tipo de reparo en la docencia. ¿Existe algo mejor? Libre, gratuito, asequible, accesible y siempre a la vanguardia." (Elousa, 2009, pág. 652)

desarrollados por usuarios a lo largo de todo el mundo, lo que nos permite estar al día en cuanto a exigencias de nuevas y más sofisticadas herramientas de un modo gratuito y libre.³ Si a esto sumamos que se trata de un programa bastante liviano, que gasta poca memoria en nuestros computadores, estamos ante una alternativa bastante interesante.

Ahora bien, respecto a la exigencia **(3) facilidad de uso**, si bien la versión básica del programa dista mucho de ser amigable, no hay que desesperar pues existen soluciones (gratuitas) para este asunto. Es posible descargar e instalar paquetes adicionales que nos permitan contar con una máscara de botones que actúa como una interfaz amigable entre el usuario y el entorno computacional que opera con códigos; entre las diferentes opciones que existen, aquí destacamos aquella denominada *Rcommander*, pues “incluye funciones para ejecutar análisis estadísticos y generar gráficos. Los menús desplegables permiten: importar/exportar datos, manipular variables (recodificar, calcular...), seleccionar casos, describir variables, generar gráficos, obtener estadísticos básicos, coeficientes de fiabilidad, contrastar hipótesis, aplicar el modelo lineal general o modelos de análisis multivariados como el análisis factorial, el análisis de correspondencias o el análisis de componentes principales. Las opciones que ofrece *Rcommander* cubrirían los contenidos de los cursos básicos de análisis de datos en las ciencias sociales y de la salud.” (Elousa, 2009, pág. 654)

³ “R, en tanto en cuanto software libre, se inscribe dentro del proyecto GNU *General Public Licence* (Licencia Pública General, GNU). Se trata de una licencia creada por *Free Software Foundation* (Fundación para el software libre), organización fundada por Richard Matthew Stallman en el año 1985. El principal propósito de la licencia GNU es declarar la libertad del uso, modificación y distribución del software y protegerlo de intentos de privatización que puedan de algún modo restringir su uso (...) Parte de la vasta información disponible sobre R es accesible a través de la web CRAN (*Comprehensive R Archive Network*; <http://cran.r-project.org/>), sitio oficial de R. Es la página base del proyecto R, desde la cual se puede descargar la última versión del programa (un equipo formado por unas doce personas, *R Development Core Team*; asumió en 1997 las labores de actualización semestral del código de R), consultar manuales sobre R, obtener ayuda sobre su funcionamiento a través de un sistema de ayuda on line, y, en definitiva, estar al corriente de los movimientos en este entorno de trabajo.” (Elousa, 2009, pág. 653)



En definitiva quizá sea R la alternativa una vez que se hayan aprendido las nociones básicas de Excel y SPSS.

Hemos expuesto las características básicas de las diferentes alternativas de software de análisis de datos que existen, sin embargo, en este ramo, trabajaremos fundamentalmente con Excel y SPSS; la idea es, sin embargo, que ustedes conozcan las alternativas que existen y se atrevan a explorar. Material bibliográfico, archivos de instalación, así como este material, quedarán colgados en la plataforma U-Cursos.

Trabajos citados

Elousa, P. (2009). ¿Existe vida más allá del SPSS? Descubre R. *Psicothema*, 21, 652-655.