

VICERRECTORADO DE DOCENCIA AYUDAS DE LA
UNIVERSIDAD DE SALAMANCA A LA INNOVACIÓN DOCENTE

Curso: 2013/2014

Referencia: **ID2013/263**

MEMORIA DE RESULTADOS

**INCORPORACIÓN DEL SOFTWARE LIBRE OPENSTAT PARA LA REALIZACIÓN DE PRÁCTICAS DE
ANÁLISIS DE DATOS EN LA ASIGNATURA “METODOLOGÍA DE LA INVESTIGACIÓN SOCIO-
EDUCATIVA”**

Referencia: ID 2013/263

Responsable del Proyecto de Innovación

Fernando Martínez Abad

Departamento de Didáctica, Organización y Métodos de Investigación.

Facultad de Educación.

Instituto Universitario de Ciencias de la Educación.

Universidad de Salamanca

Miembros del equipo

Susana Olmos Migueláñez

M^a José Rodríguez Conde

M^a Esperanza Herrera García

Juan Francisco Martín Izard

Eva María Torrecilla Sánchez

Patricia Torrijos Fincias

Salamanca, 24 de Junio de 2014



ÍNDICE GENERAL

| | |
|---|----|
| ÍNDICE GENERAL..... | 2 |
| 0. Justificación y contextualización | 3 |
| 1. Introducción | 4 |
| 2. Objetivos | 6 |
| 3. Método: Procedimiento de implementación..... | 7 |
| 3. Resultados | 9 |
| 4. Conclusiones y discusión..... | 13 |
| 5. Bibliografía | 14 |
| ANEXO I | 16 |
| ANEXO II | 52 |
| ANEXO III | 56 |

0. Justificación y contextualización

Durante los cursos 2011/12 y 2012/13, en la asignatura ‘Metodología de la Investigación Socio-educativa’ se han venido incluyendo como parte de la formación prácticas en el aula de informática consistentes en el empleo del software propietario IBM SPSS para el análisis cuantitativo de datos. Durante el curso 2012/13, como tarea añadida a estas prácticas de aula, se planteó el desarrollo de un proyecto de investigación, consistente en el planteamiento y desarrollo de unas hipótesis de investigación partiendo de una base de datos proveniente del Instituto Nacional de Estadística.

Es a partir de esta experiencia, en la que los alumnos muestran su disconformidad por tener que emplear una herramienta informática que no está a su alcance fuera de la institución universitaria, cuando el equipo de profesores de la asignatura, junto con otros miembros del IUCE, nos planteamos la posibilidad de localizar y emplear una herramienta de software libre para estas actividades formativas durante el curso 2013-14. Se pretende con ello integrar en la medida de lo posible el software libre en la vida tanto de las instituciones de educación superior como de los estudiantes que acogen. Las razones que nos llevan a esta propuesta se basan, más allá del mero beneficio económico para la Universidad, en razones tanto de mejora de la educación como morales:

En lo que respecta a la **mejora de la educación** en la Universidad, el empleo de software libre como recurso docente permite a los estudiantes disponer de este recurso y todo su potencial fuera de la propia institución, ya durante su vida académica y más adelante, en eventuales necesidades surgidas en el ámbito laboral. Esto conlleva necesariamente un impacto positivo sobre la calidad de la docencia ofertada.

En cuanto a los **aspectos morales**, el empleo de software libre implica la transmisión de unos valores innegables e innegociables, que toda universidad debe transmitir a sus estudiantes como formación transversal. Así, la naturaleza del software libre transmite valores relacionados con la colaboración y trabajo en equipo, independencia y libertad o democracia y razonamiento crítico.

1. Introducción

“La formación en este siglo XXI se está orientando hacia nuevos modelos de enseñanza masivos en abierto y gratuitos. Estos modelos interactivos, colaborativos y online aumentan y posibilitan el acceso a la formación superior de manera universal” (Vázquez Cano, López Meneses, & Sarasola Sánchez-Serrano, 2013, p. 13)

Nos encontramos en una realidad social en la que, debido al avance tecnológico, el mundo está interconectado y todo se puede localizar, exponer, intercambiar, transferir, recibir, vender o comprar sin importar el lugar en que nos encontremos (De Pablos, 2010); por ello, es necesario que las instituciones de educación superior se transformen y adapten a los cambios, para hacer frente a las nuevas demandas educativas de la sociedad de la información. La universidad y particularmente los profesores universitarios tienen la responsabilidad de contribuir, con una práctica educativa innovadora, en la formación integral de individuos con competencias suficientes para incorporarse y desenvolverse fácilmente en la realidad que tengan que vivir (González Mariño, 2008).

En concreto, como una reacción al movimiento en abierto global¹ y de promoción general del Conocimiento en Abierto, surge en el campo de la educación superior la iniciativa OpenCourseWare (OCW), propuesta en sus inicios desde el Instituto Tecnológico de Massachusetts (MIT). Dado el éxito y potencial del movimiento OCW en el ámbito de la educación superior, la propia UNESCO resalta su valor en el informe final del “Foro sobre las Incidencias de los Programas Educativos Informáticos Abiertos” (2002), acuñando el término “Recursos Educativos Abiertos” (REA) para todo este tipo de materiales educativos de acceso libre. Finalmente, este término se consolida en la Declaración de París sobre los recursos educativos abiertos (UNESCO, 2012). Es en este contexto y bajo esta filosofía dónde el empleo de software libre en educación alcanza la importancia que aquí se le concede. De hecho, en los últimos años, debido a esta vigencia y fuerza del movimiento en abierto general, ha surgido en el campo educativo una potente rama (Antonenko, Toy, & Niederhauser, 2004; Aydin & Tirkes, 2010; MIT news, 2001), sobre todo en lo que tiene que ver con la publicación y empleo de materiales y recursos libres y/o en abierto.

Por otro lado, en base a la Teoría de Acción Razonada (Ajzen & Fishbein, 1980), las creencias de usuarios hacia las tecnologías influirán en su actitud hacia las mismas, y el propio empleo de las TIC se basará en las actitudes individuales hacia éstas. Esto nos enseña que en la

¹ Se denomina así a todos los movimientos sociales surgidos con la intención de facilitar el conocimiento abierto en cualquier campo: informática (software libre y código abierto), biología (Proyecto Genoma Humano), educación (OpenCourseWare y MOOCs), etc.



incorporación de las TIC no se puede dejar de lado el componente actitudinal. Cuando el usuario muestra actitudes positivas hacia el empleo de herramientas informáticas, su integración efectiva tiene una probabilidad mucho mayor de ser exitosa. Por ello, la primera medida que debemos tomar si deseamos que las TIC se integren en las prácticas personales y profesionales de nuestros estudiantes, es fomentar en los mismos una actitud adecuada hacia las mismas.

A partir de la Teoría de la Acción Razonada, Fred D. Davis propone el Modelo de Aceptación Tecnológica (Davis, Bagozzi, & Warshaw, 1989; Davis, 1989, 1993), que explica el empleo actual de tecnologías del usuario en base a sus actitudes hacia el propio uso de las tecnologías. Las actitudes hacia las tecnologías se forman en función de la utilidad percibida y la percepción sobre su facilidad de empleo.

En suma, el comportamiento tecnológico mostrado por una persona (Ajzen & Fishbein, 1980) y más en concreto el empleo de un recurso tecnológico (Davis et al., 1989; Davis, 1989, 1993), está determinado por la utilidad y facilidad de uso percibidas y por las actitudes hacia el mismo.

Por tanto, a partir de lo dicho hasta ahora, se desarrolla el presente proyecto de investigación. Se espera que las actitudes de los estudiantes hacia el empleo de software educativo libre sean más positivas que hacia el empleo de software educativo propietario, y con ello la intención de uso y el uso final muestren mejores índices. Esta mejora de los índices de intención de uso y empleo final es la que debe promocionar una más efectiva integración de estos recursos informáticos valiosos en el desempeño personal y profesional de los estudiantes.



2. Objetivos

Justificado a partir de las razones educativas y morales planteadas más arriba, se plantea como **objetivo general** la integración del software libre, en concreto del software de análisis estadístico de datos PSPP, en la vida tanto de las instituciones de educación superior como de sus estudiantes.

Como **objetivos específicos**, se pueden plantear los siguientes:

Promover el empleo de software libre en la universidad y en el entorno educativo.

Facilitar al alumnado el acceso a recursos informáticos abiertos de análisis de datos.

Valorar el impacto del empleo del software libre PSPP en el aprendizaje de los estudiantes.

Promover la mejora de la calidad de la enseñanza universitaria a través de la incorporación de recursos informáticos educativos de acceso abierto.

Se trata de aprovechar la implementación inicial del proyecto durante el curso académico 2013-14 para observar el funcionamiento y utilidad educativa de la herramienta en una asignatura concreta, *Metodología de la Investigación Educativa*, de primer curso del Grado en Educación Social para, a partir de los resultados obtenidos, valorar su integración y generalización en otras asignaturas impartidas por el PDI relacionadas con el análisis estadístico de datos en Ciencias Sociales, como alternativa al empleo del software propietario IBM SPSS.

3. Método: Procedimiento de implementación

En primer lugar se debe señalar que, posteriormente a la presentación del proyecto, y previamente al comienzo en el desarrollo del mismo, debido a la implementación inicial de un procedimiento de búsqueda de un software libre de análisis estadístico más sencillo y adaptado a las necesidades concretas de las prácticas a desarrollar, se sustituye el software que estaba proyectado emplear inicialmente, **OpenStat**, por otro software igualmente libre bajo licencia GNU, **PSPP**. A pesar de esta modificación inicial en el software empleado, las fases y desarrollo del proyecto se mantienen intactas, desarrollándose todas y cada una tal y como estaban previstas.

En concreto, el trabajo se desarrolla en las siguientes fases:

1. Adaptación de las bases de datos empleadas en la asignatura al formato PSPP. El software PSPP trabaja con las bases de datos en formato .sav, el formato habitual de IBM SPSS, por lo que esta fase se cumplió perfectamente con una simple revisión del funcionamiento de las bases de datos en el nuevo software. Fruto de este procedimiento se pudieron realizar algunas correcciones en la base que se detectaron a lo largo de la revisión.

2. Elaboración de un Manual para el empleo de PSPP adaptado a las necesidades de los alumnos de la asignatura. En esta fase se diseña un manual en formato .pdf en el que se explican con detalle las rutas de navegación necesarias para aplicar las técnicas estadísticas explicadas en la asignatura, y la interpretación de las salidas de datos. El manual completo se adjunta en el anexo I. Por otro lado, de cara a mantener la filosofía del conocimiento abierto, se publica el manual en abierto en el siguiente enlace: <http://www.slideshare.net/Plastilino/manual-paquete-estadstico-pspp-espaol>

3. Formación de estudiantes e implementación de las prácticas en el aula con software PSPP. Se disponen cuatro sesiones formativas completas de una hora de duración cada una en las que se realizan dos prácticas de análisis de datos con PSPP, en las que se analizan varias variables de la base de datos publicada por el Centro de Investigaciones Sociológicas (CIS) a partir de la encuesta sobre Salud y Hábitos Sexuales, desarrollada y publicada en 2003. Las dos prácticas señaladas están disponibles en el anexo II.

3. Diseño de encuestas de satisfacción con el software empleado. Se diseña una encuesta que incorpora un primer bloque sobre cuestiones demográficas y tres dimensiones acerca del software libre:

- Actitudes personales acerca del software libre



- Actitudes hacia el software libre en educación
- Actitudes hacia el empleo del software libre PSPP en la asignatura

Todos los ítems del cuestionario diseñado están valorados en una escala tipo Likert con puntuaciones que van desde el 1 (totalmente en desacuerdo) hasta el 5 (totalmente de acuerdo) Las variables o ítems incluidos en la encuesta se pueden consultar en el Anexo III de este informe.

4. Implementación de las encuestas en GoogleDocs y aplicación a los estudiantes. Para mantener la filosofía ligada al software libre, se emplea la herramienta gratuita y de acceso abierto de diseño y aplicación de encuestas en línea GoogleDocs, que los alumnos contestaron a través de un enlace incluido en la plataforma Studium. El enlace del cuestionario es el que se muestra a continuación: https://docs.google.com/forms/d/1gV3zfl2SK2y_KQTgwa8JXTIsMIQErNnBa42hDHGNrbg/viewform

4. Análisis e interpretación de los datos recogidos en las encuestas. Informatizadas las respuestas automáticamente gracias al sistema de GoogleDocs empleado, se extrae la información a una hoja de cálculo y posteriormente se importa al software PSPP para su análisis estadístico.

3. Resultados

Además de los propios recursos diseñados, señalados en el apartado anterior, y mostrados en los anexos incluidos en este informe, se presentan aquí los resultados obtenidos en el cuestionario, que fue aplicado a 25 de los estudiantes que participaron en la asignatura asiduamente durante el curso 2013-14.

En cuanto a la distribución por sexo, se observa en el gráfico I cómo casi un 80% de la muestra está conformado por mujeres.

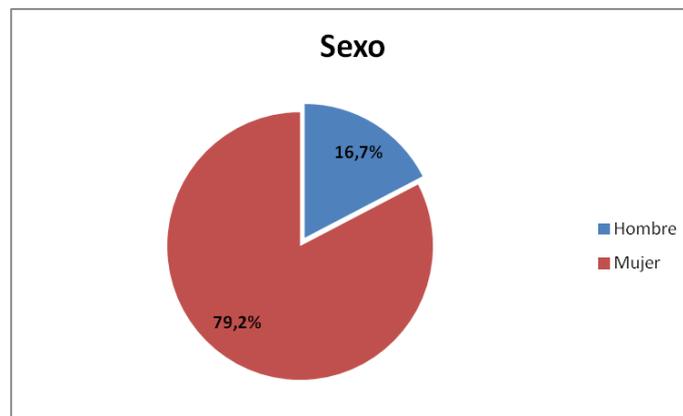


Gráfico I. Distribución de la muestra por sexo

Estas diferencias por sexo se deben a que nos encontramos en una titulación del área de Ciencias de la Educación, y son las habituales en estos casos.

En cuanto a la distribución por edad, la mayor parte de los casos se encuentra entre los 18 y los 20 años (70.8%), como era de esperar.

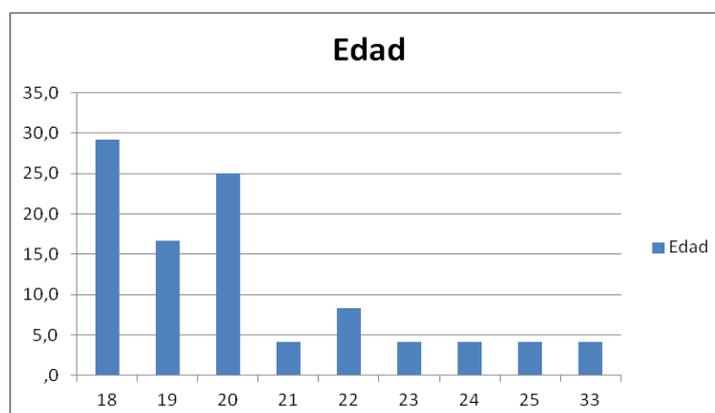


Gráfico II. Distribución de la muestra por edad

Por otro lado, en cuanto a las respuestas a los ítems de la escala, en la tabla I se muestran los resultados obtenidos en la media y la desviación típica en los ítems de la escala.

Tabla I. Media y desviación típica obtenida en los ítems de la escala

| | Media | Desv. típ. |
|---|-------|------------|
| 19. La Universidad de Salamanca debería fomentar el empleo de software libre como PSPP en las prácticas de análisis de datos de las titulaciones frente a otro software de pago | 4,58 | 0,584 |
| 14. Las instituciones públicas deberían fomentar el desarrollo de software libre educativo | 4,50 | 0,659 |
| 12. Entre el software libre y de pago, prefiero que mis profesores me presenten la alternativa libre | 4,46 | 0,833 |
| 9. El empleo de software libre fomenta la igualdad de oportunidades en educación | 4,43 | 0,662 |
| 8. En definitiva, creo que el software libre es beneficioso para la sociedad en su conjunto | 4,30 | 1,063 |
| 15. En definitiva, el empleo de software libre en la educación mejora la calidad de la formación recibida por los estudiantes | 4,25 | 0,737 |
| 13. El empleo de software libre en las instituciones educativas transmite a los estudiantes valores esenciales para la educación | 4,25 | 0,847 |
| 17. En el caso de necesitar realizar cualquier análisis de datos, el conocimiento de la herramienta PSPP me hace más autónomo e independiente | 4,21 | 0,932 |
| 1. El software libre fomenta la igualdad de oportunidades en la sociedad | 4,21 | 0,884 |
| 11. Creo que el conocimiento de herramientas de software libre frente al software de pago va a ser ventajoso para mi futuro laboral | 4,21 | 0,884 |
| 10. El uso de software libre nos hace a los estudiantes más autónomos e independientes | 4,17 | 0,816 |
| 6. Creo que todas las instituciones públicas deberían emplear software libre en sus procedimientos administrativos | 4,13 | 1,058 |
| 16. La enseñanza de PSPP para el análisis de datos, frente a la enseñanza de otra alternativa de pago, fomenta la igualdad de oportunidades en la sociedad | 4,13 | 1,076 |
| 18. Creo que el conocimiento de PSPP frente a otro software de análisis de datos de pago puede aportarme ventajas en mi futuro laboral | 4,08 | 0,776 |
| 3. El software libre fomenta la independencia de las personas e instituciones en la sociedad | 4,04 | 0,690 |
| 20. En definitiva, creo que el empleo de PSPP (frente al empleo de software de pago) mejora la calidad de la formación que recibo en la universidad | 4,00 | 0,933 |
| 4. El empleo de software libre fomenta el funcionamiento democrático de las instituciones | 3,96 | 0,751 |
| 2. Los planteamientos del software libre fomentan el pensamiento crítico de las personas | 3,79 | 0,884 |
| 5. El empleo de software libre por instituciones y empresas les beneficia no sólo en el aspecto económico | 3,67 | 1,090 |
| 7. Si existe una alternativa de software libre frente a otro de pago, la utilizo en mi ordenador personal por cuestiones principalmente morales | 3,58 | 1,412 |

Se observa a nivel general cómo los ítems de la primera dimensión, sobre las actitudes personales acerca del software libre, son los que tienen una valoración más baja. Por otro lado los ítems de las dos siguientes dimensiones comparten posiciones elevadas.

El gráfico III nos ayuda a entender mejor esta distribución. La mayor parte de los ítems de los bloques II y III se encuentran en puntuaciones entre 4,5 y 4, mientras que los del primer bloque se acercan más a puntuaciones entre 4 y 3,5. Esto nos indica que los estudiantes valoran más positivamente el empleo educativo de estas herramientas de software libre frente al empleo personal de las mismas.

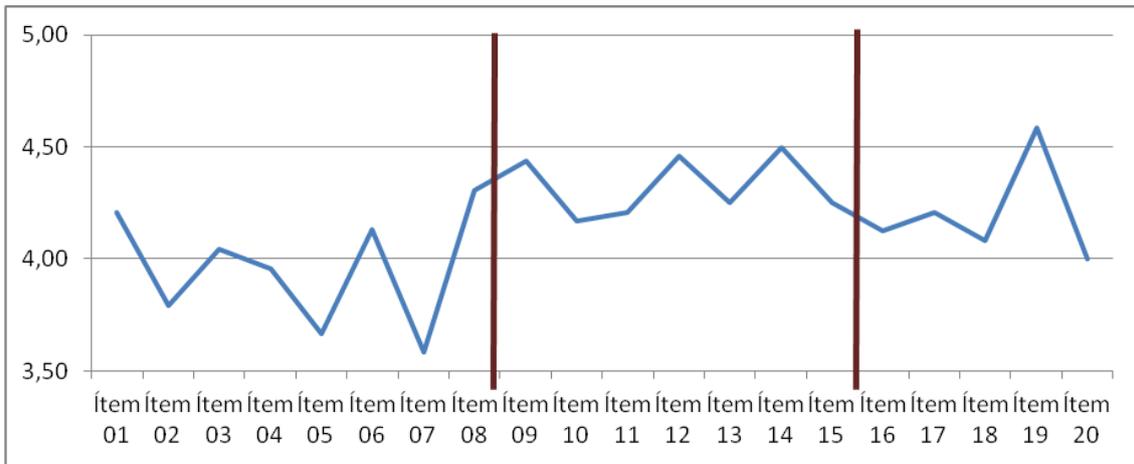


Gráfico III. Tendencia de las medias de los ítems de la escala

Finalmente, si realizamos el análisis por dimensiones, parece que la dimensión 2 '**Actitudes hacia el software libre en educación**' está mejor valorada que el resto, con una diferencia de casi medio punto.

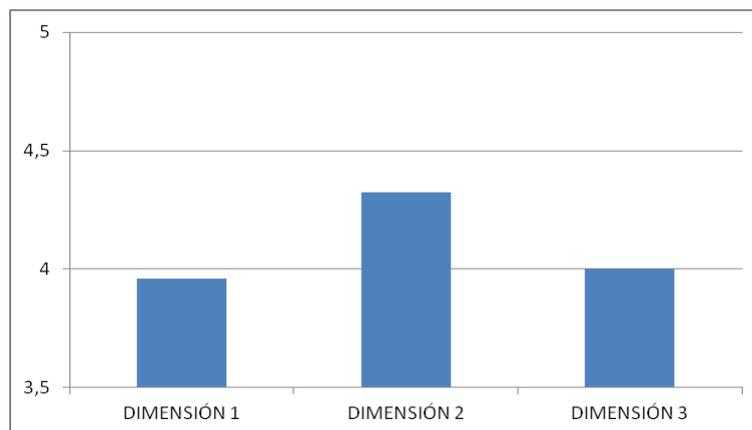


Gráfico IV. Tendencia de las medias de los ítems de la escala

Sin embargo, los resultados de la prueba no paramétrica de los signos de Wilcoxon nos muestran que sólo existen diferencias significativas entre las dimensiones 1 y 2 del

cuestionario, no teniendo la dimensión 3 diferencias significativas con ninguna de las otras dos medidas.

Tabla II. Contraste de hipótesis. Prueba de Wilcoxon dimensiones

| | D1 - D2 | D3 - D2 | D3 - D1 |
|---------------------------------|----------------|----------------|----------------|
| Estadístico de contraste | -3,246 | -0,237 | -1,651 |
| p-valor | ,001 | ,813 | ,099 |

Estos resultados están indicando que los estudiantes valoran de manera significativamente superior la integración del software libre a nivel general en los procesos educativos en la educación, mientras que prácticamente del mismo modo el empleo personal del software libre y su utilización en concreto en las tareas de análisis de datos estadístico.

En conclusión, a pesar de que la valoración es muy positiva en las tres escalas, parece que los estudiantes ven con mejores ojos la integración del software libre en los procesos educativos generales que en su ámbito personal y/o profesional futuro, cuestión que se debería trabajar en futuro.

4. Conclusiones y discusión

A nivel general, parece que el movimiento en abierto está teniendo un impacto significativo en el mundo de la educación (UNESCO, 2002, 2012), y que por tanto las universidades deben adaptarse a este cambio, como añadido al esfuerzo realizado con el OCW (MIT news, 2001), integrando software libre en su funcionamiento administrativo, docente y pedagógico.

El esfuerzo realizado en el proyecto aquí presentado ha ido en esa dirección. En concreto, se ha tratado de integrar herramientas de software libre en las prácticas de un campo en el que tradicional y masivamente se ha empleado el software propietario IBM SPSS, sin proponerse una alternativa clara al respecto.

Los resultados aquí obtenidos, más allá de los beneficios intangibles vinculados a principios éticos y educativos inquebrantables, indican que los estudiantes están comprometidos con la integración del software libre en los procesos de enseñanza-aprendizaje, y que valoran positivamente su empleo en el ámbito educativo en general.

Por otro lado, el presente proyecto ha posibilitado la integración de otros

Sin embargo, los resultados también sugieren que parte de los objetivos que se plantean en este proyecto que acaba de nacer este curso, y que se espera tenga continuidad, están insatisfechos al 100%, sobre todo en lo que respecta a alcanzar un mayor nivel de compromiso personal de los estudiantes sobre el software libre. Quizás este objetivo supere el campo de acción planteado marginalmente desde una asignatura de un grado, y sería esencial diseñar un plan de acción coordinado y conjunto entre el profesorado de todo el Grado para alcanzar esta movilización de la conciencia y actitudes de los estudiantes.

Fruto de esta experiencia, por otro lado, dado el éxito de la misma en todos los sentidos, se plantea como una línea futura que probablemente se integre a lo largo del curso 2014-15, la ampliación del proyecto a otras materias del área de Métodos de Investigación y Diagnóstico en Educación que impartan prácticas con IBM SPSS. Así, probablemente consista en la migración del software de estas prácticas hacia el PSPP.

5. Bibliografía

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, N.J.: Prentice-Hall.

Antonenko, P., Toy, S., & Niederhauser, D. (2004). Modular Object-Oriented Dynamic Learning Environment: What Open Source Has to Offer. En *Association for Educational Communications and Technology*. Recuperado a partir de <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED485088>

Aydin, C. C., & Tirkes, G. (2010). Open Source Learning Management Systems in Distance Learning. *Turkish Online Journal of Educational Technology*, 9(2), 175-184.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319. doi:10.2307/249008

Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38(3), 475–487. doi:10.1006/imms.1993.1022

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, 35(8), 982-1003. doi:10.1287/mnsc.35.8.982

De Pablos, J. de. (2010). Universidad y sociedad del conocimiento. Las competencias informacionales y digitales. *Revista de Universidad y Sociedad de Conocimiento*, 7(2), 6-16.

González Mariño, J. C. (2008). TIC y la transformación de la práctica educativa en el contexto de las sociedades del conocimiento. *Revista de Universidad y Sociedad del*



Conocimiento, 5(2). Recuperado a partir de
<http://www.uoc.edu/rusc/5/2/dt/esp/gonzalez.html>

MIT news. (2001, abril 1). *MIT OpenCourseWare -- Fact Sheet. MIT's News Office*. Recuperado
13 de enero de 2014, a partir de <http://web.mit.edu/newsoffice/2001/ocw-facts.html>

UNESCO. (2002). *Forum on the impact of open courseware for higher education in developing
countries: Final report*. París: UNESCO. Recuperado a partir de
unesdoc.unesco.org/images/0012/001285/128515e.pdf

UNESCO. (2012). *Declaración de París de 2012 sobre los REA*. París: UNESCO.

Vázquez Cano, E., López Meneses, E., & Sarasola Sánchez-Serrano, J. L. (2013). *La expansión del
conocimiento en abierto: los MOOC*. Barcelona: Octaedro ICE.

ANEXO I

MANUAL DE EMPLEO DE PSPP

AUTORES:

Fernando Martínez Abad

Susana Olmos Migueláñez

Metodología de la Investigación Socio-educativa

Grado en Educación Social

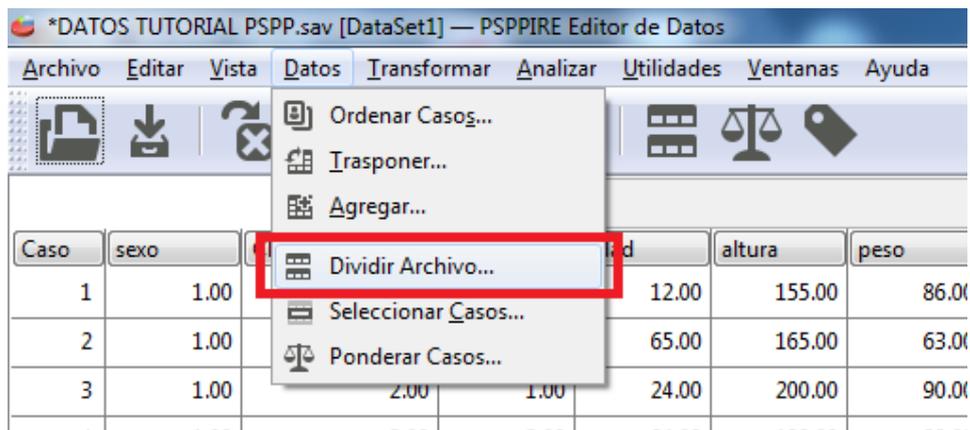
Universidad de Salamanca

ÍNDICE

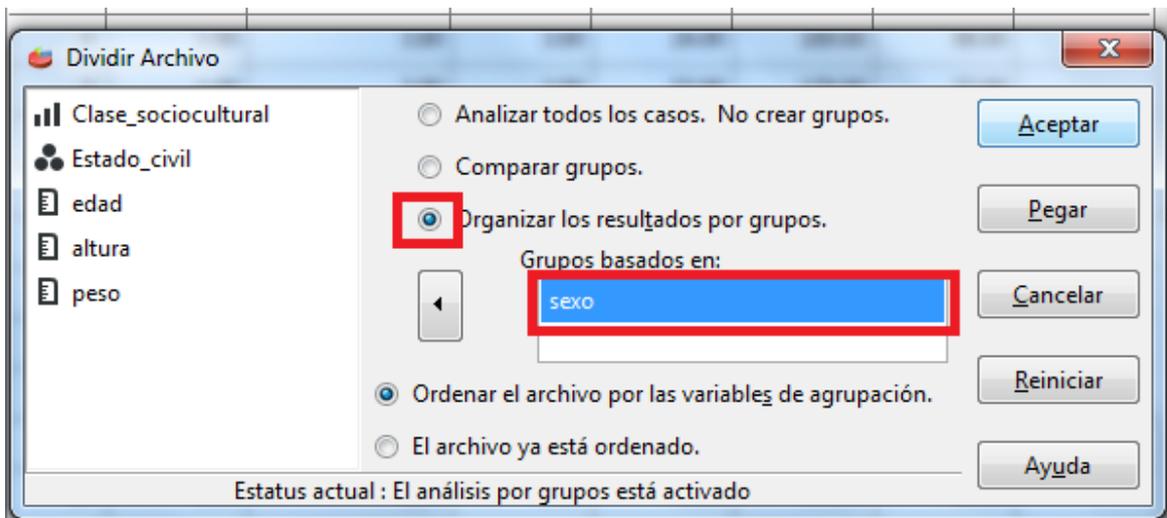
| | |
|--|----|
| DIVIDIR ARCHIVO | 3 |
| CALCULAR UNA NUEVA VARIABLE | 5 |
| RECODIFICAR VARIABLE | 7 |
| TABLA DE FRECUENCIAS y ESTADÍSTICOS DESCRIPTIVOS | 9 |
| COEFICIENTE DE CORRELACIÓN DE PEARSON | 11 |
| OTROS COEFICIENTES DE CORRELACIÓN | 12 |
| REGRESIÓN | 14 |
| COMPARAR MEDIAS ENTRE DOS GRUPOS DIFERENTES | 16 |
| COMPARAR MEDIAS EN UN MISMO GRUPO | 18 |
| ANÁLISIS DE VARIANZA | 19 |
| GENERAR GRÁFICOS EN PSPP | 21 |
| EXPORTAR TABLAS Y GRÁFICOS DE PSPP | 27 |

DIVIDIR ARCHIVO

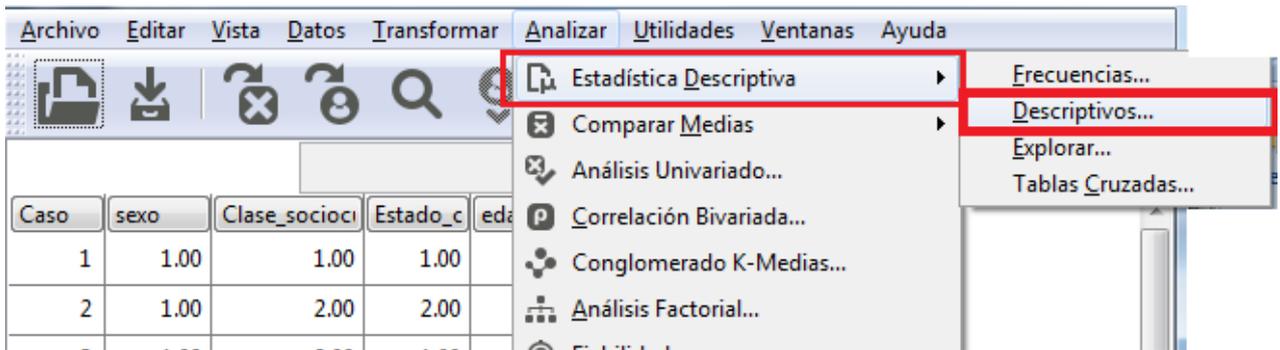
La opción 'dividir archivo' nos servirá para dividir la muestra completa en varias submuestras, de manera que podamos realizar cualquier tipo de análisis para cada submuestra, en lugar de para toda la muestra completa, como hace habitualmente el PSPP. Así, debemos ir a **Datos** → **dividir archivo**:



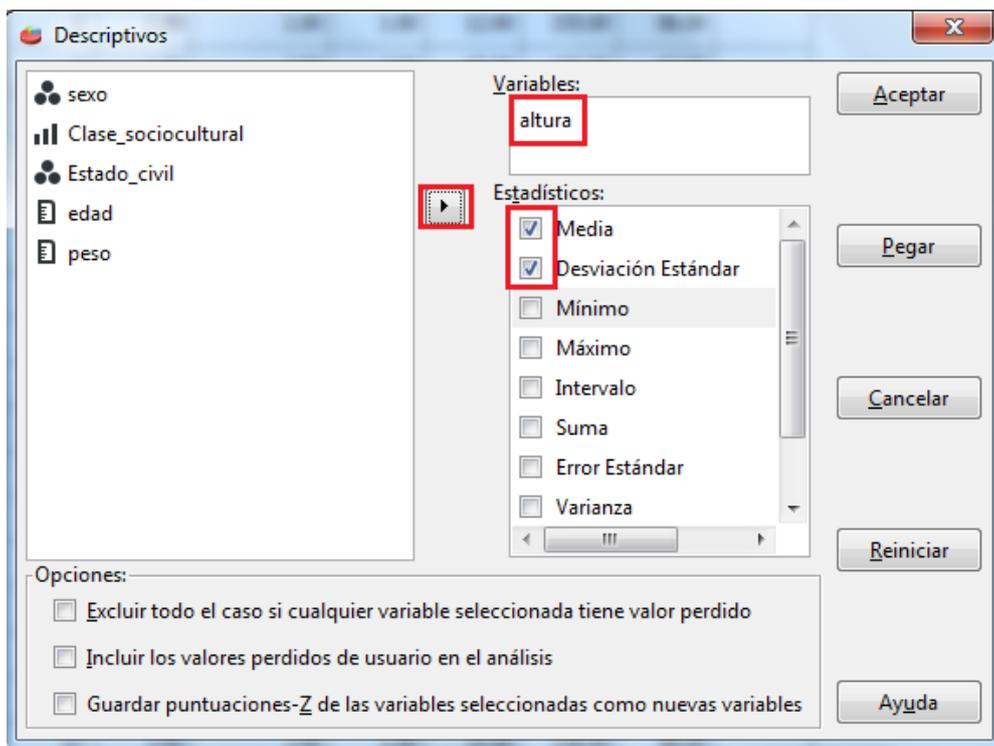
En la ventana emergente se selecciona 'organizar los resultados por grupos' y se añade la variable a partir de la que queremos que se segmente el archivo:



Una vez hecho esto, si solicitamos el cálculo de cualquier estadístico descriptivo, la hoja de resultados nos devolverá por separado los resultados para hombres y para mujeres:



Veamos el ejemplo para la media y desviación típica de la altura. Una vez introducimos la variable y seleccionamos los estadísticos, pulsamos 'aceptar' y en la ventana emergente se mostrarán los estadísticos:



DESCRIPTIVES

DESCRIPTIVES

/VARIABLES= altura
/STATISTICS=MEAN STDDEV.

Variable Valor Etiqueta
sexo 1.00 Hombre

Casos válidos = 13; casos con valor(es) perdido(s) = 0.

| Variable | N | Media | Desv Std |
|----------|----|--------|----------|
| altura | 13 | 175.38 | 14.02 |

Variable Valor Etiqueta
sexo 2.00 Mujer

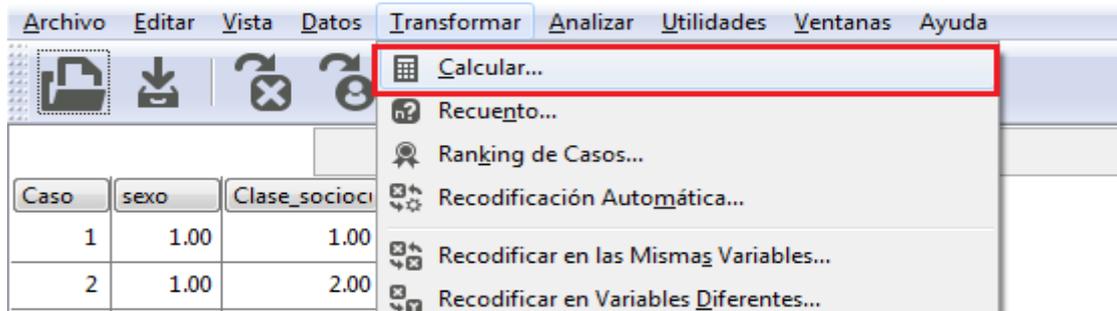
Casos válidos = 12; casos con valor(es) perdido(s) = 0.

| Variable | N | Media | Desv Std |
|----------|----|--------|----------|
| altura | 12 | 160.42 | 10.39 |

OJO!!! no se nos olvide quitar la segmentación una vez terminado, para volver a pedir estadísticos para la muestra completa **Datos → dividir archivo → Analizar todos los casos. No crear grupos**

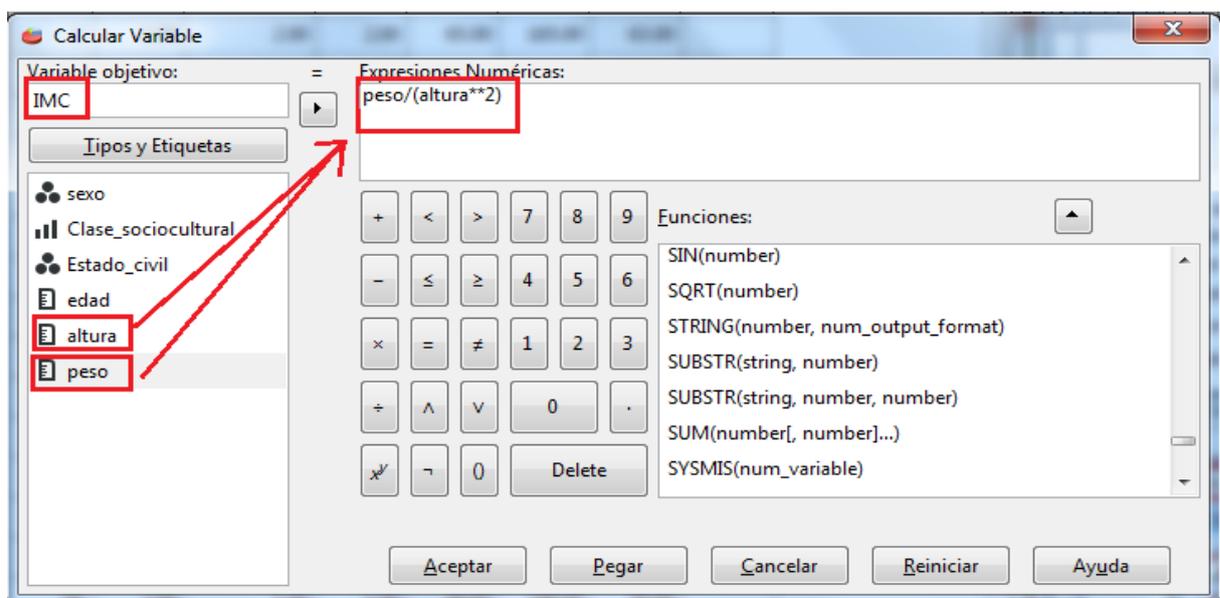
CALCULAR UNA NUEVA VARIABLE

Puede que necesitemos calcular una nueva variable haciendo, por ejemplo, la suma de varias variables (o la combinación de operaciones que nos interese). Para ello, **Transformar** → **Calcular...**:



Nos aparece en la ventana emergente una especie de calculadora, ahí iremos poniendo la combinación que nos interese. En '**variable objetivo**' introduciremos el nombre de la variable que vamos a crear, y en '**expresiones numéricas**' introduciremos la operación a realizar. En nuestro ejemplo, podemos calcular una nueva variable que sea el índice de masa corporal

(IMC) de las personas de la muestra, que se calcula con la fórmula $IMC = \frac{PESO}{ALTURA^2}$. Así, llamaremos a la nueva variable '**IMC**':



Como se puede observar en la anterior imagen, para poner una potencia, se deben añadir dos asteriscos antes del valor o variable que forma la potencia.

Una vez hecho esto y pulsado aceptar, debemos comprobar en la ventana de resultados que las operaciones se han hecho correctamente:

```

COMPUTE
COMPUTE IMC = peso/(altura**2).

EXECUTE
EXECUTE.
    
```

En el caso de que no se haya escrito correctamente la operación nos aparecerá un mensaje de error similar al siguiente:

```

COMPUTE
.21-26: error: Error de sintaxis en `^`: Caracter erróneo `^` en la entrada.
.28: error: Error de sintaxis en `2`: esperando `)`.
COMPUTE IMC = peso/(altura^2).
    
```

Si todo se ha hecho correctamente, nos aparecerá en la última columna de la **Vista de datos** una nueva variable llamada **IMC** (en el caso de que pongamos como nombre de la nueva variable alguno ya existente se sustituirán los valores de dicha variable existente):

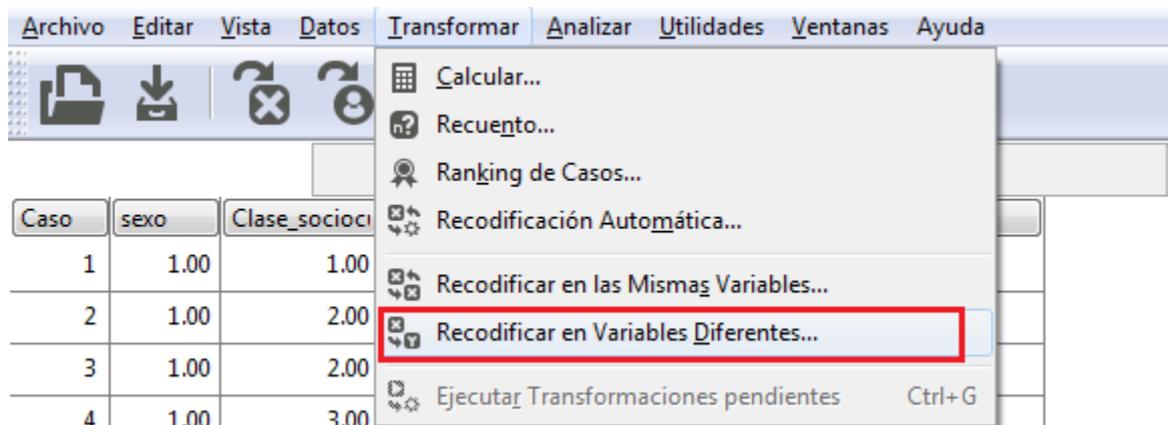
| Caso | sexo | Clase_socioci | Estado_c | edad | altura | peso | IMC |
|------|------|---------------|----------|-------|--------|--------|-------|
| 1 | 1.00 | 1.00 | 1.00 | 12.00 | 1.55 | 86.00 | 35.80 |
| 2 | 1.00 | 2.00 | 2.00 | 65.00 | 1.65 | 63.00 | 23.14 |
| 3 | 1.00 | 2.00 | 1.00 | 24.00 | 2.00 | 90.00 | 22.50 |
| 4 | 1.00 | 3.00 | 3.00 | 24.00 | 1.89 | 98.00 | 27.43 |
| 5 | 1.00 | 3.00 | 2.00 | 53.00 | 1.79 | 72.00 | 22.47 |
| 6 | 1.00 | 2.00 | 3.00 | 53.00 | 1.70 | 80.00 | 27.68 |
| 7 | 1.00 | 2.00 | 2.00 | 22.00 | 1.95 | 109.00 | 28.67 |
| 8 | 1.00 | 2.00 | 1.00 | 12.00 | 1.55 | 80.00 | 33.30 |
| 9 | 1.00 | 2.00 | 3.00 | 45.00 | 1.78 | 80.00 | 25.25 |
| 10 | 1.00 | 3.00 | 3.00 | 54.00 | 1.70 | 73.00 | 25.26 |
| 11 | 1.00 | 2.00 | 2.00 | 32.00 | 1.85 | 93.00 | 27.17 |
| 12 | 1.00 | 3.00 | 2.00 | 21.00 | 1.68 | 62.00 | 21.07 |

RECODIFICAR VARIABLE

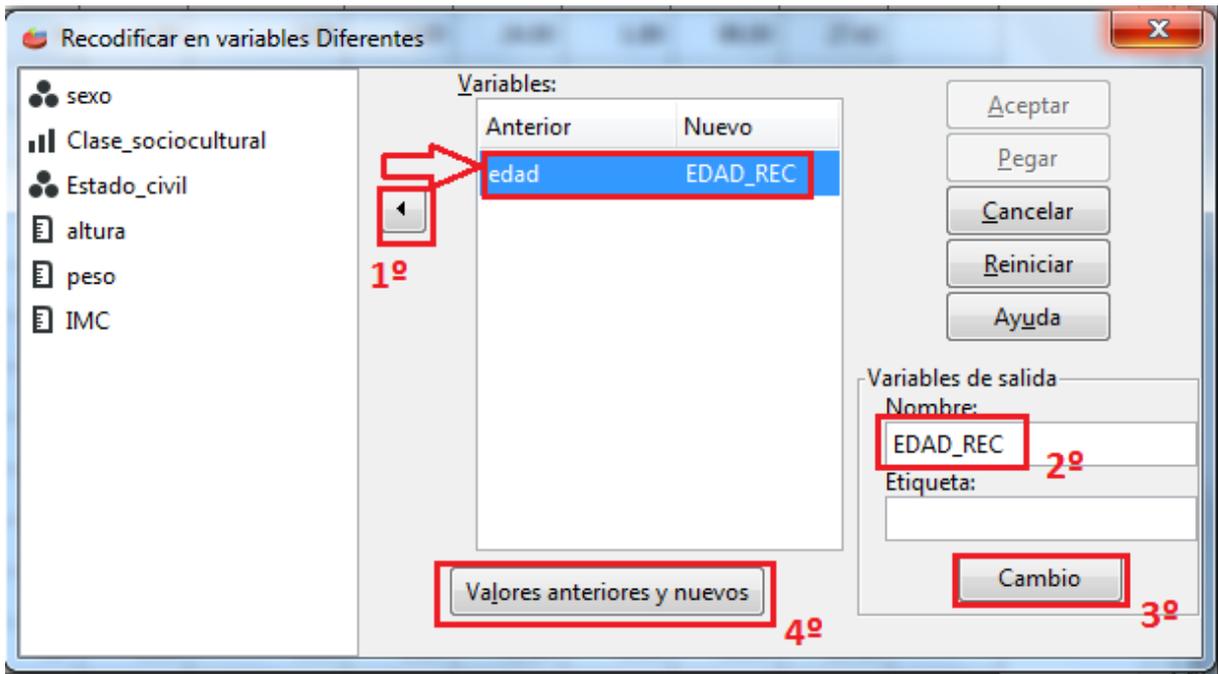
En algunos casos, puede interesarnos recodificar o categorizar una variable cuantitativa para hacerla cualitativa. Por ejemplo, puede que nos interese categorizar la variable '**edad**' en una nueva variable en la que agrupemos a **jóvenes**, **adultos** y **ancianos**. Si tenemos edades comprendidas entre 12 y 74 años, podemos hacer la siguiente división:

- JOVEN → Personas de entre 12 y 30 años
- ADULTO → Personas de entre 31 y 60 años
- ANCIANO → Personas de entre 61 y 74 años

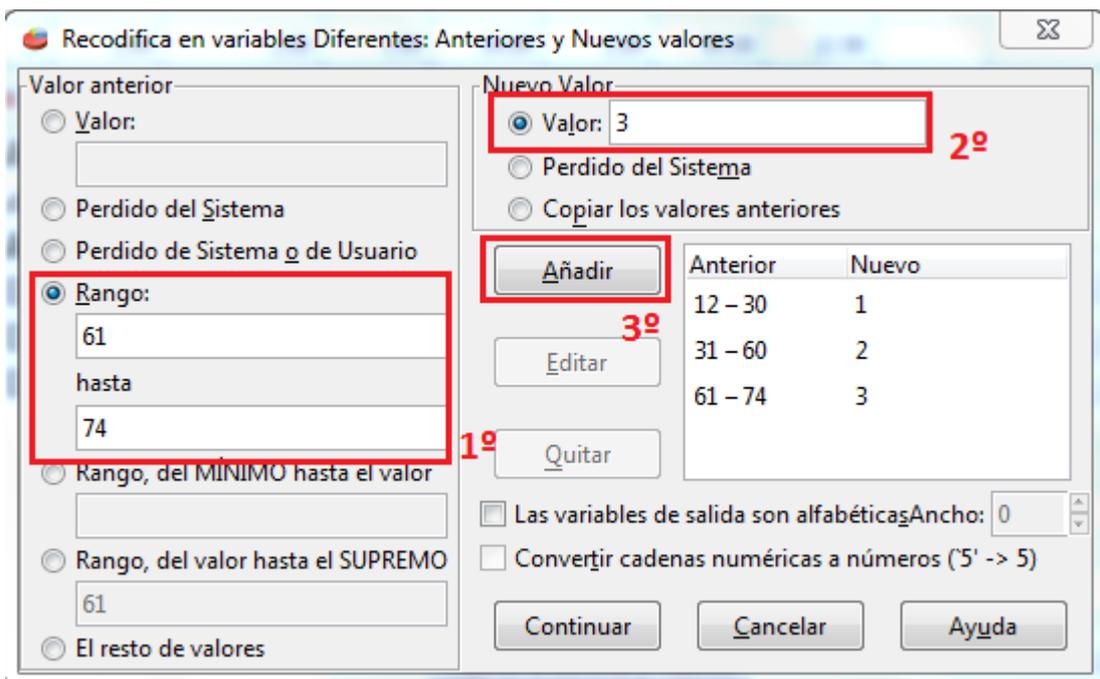
Entonces pues, **Transformar** → **Recodificar en variables diferentes**:



En la ventana emergente debemos, en primer lugar, introducir la variable **edad** en la ventana central, hacer un click sobre la variable y en la parte derecha, en '**variables de salida**', poner el nombre de la nueva variable, por ejemplo **EDAD_REC**. Finalmente, hacer click en '**cambio**'. En la ventana central debería aparecer **edad** (anterior) **EDAD_REC** (nuevo). Una vez hecho esto, hacer click en '**valores anteriores y nuevos**':



Aparecerá una nueva ventana en la que seleccionaremos los valores de la nueva variable de uno en uno. En la columna de la izquierda de **valor anterior**, seleccionaremos rango y pondremos el primero 12-30. Después en la columna de la derecha **nuevo valor**, pondremos el nuevo valor de los sujetos que tienen entre 12 y 30 años, en este caso 1. Finalmente hacemos click en 'añadir' y nos debe aparecer la transformación en la ventana de la derecha. Así lo haremos con todas las categorías hasta el final, que haremos click en **continuar**:

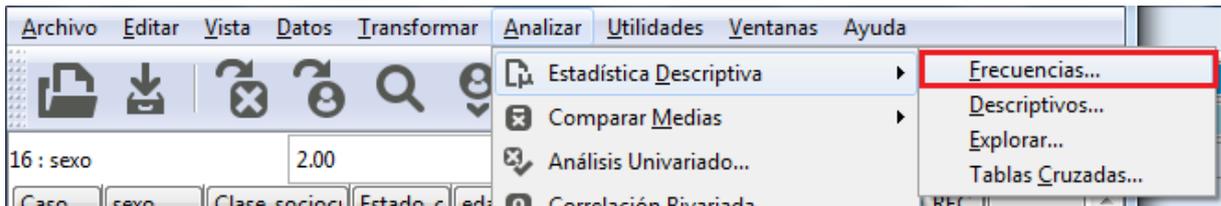


Nos volverá a la ventana anterior, donde haremos click en **aceptar**. Finalizado el proceso debe habernos aparecido la nueva variable:

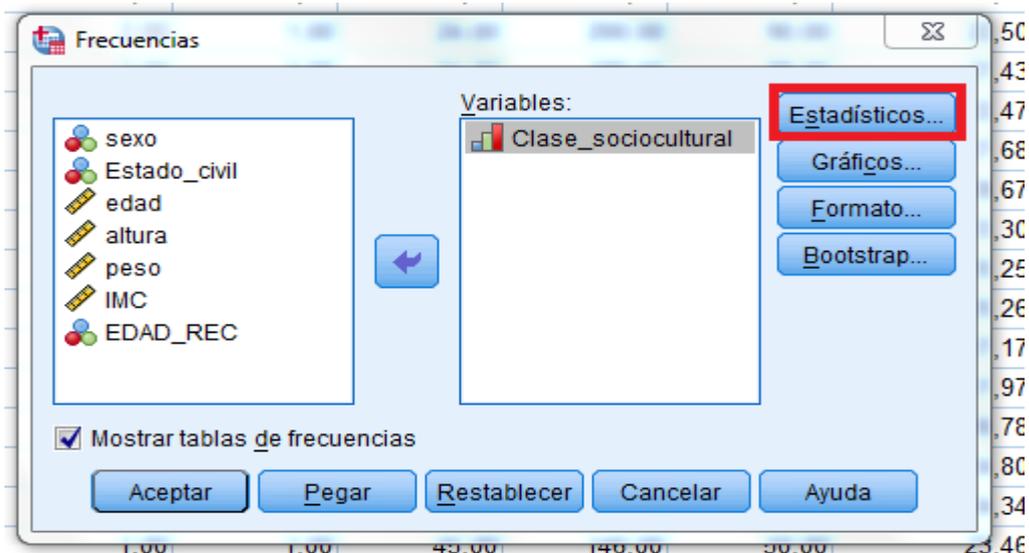
| Caso | sexo | Clase_socioci | Estado_c | edad | altura | peso | IMC | EDAD_REC |
|------|------|---------------|----------|-------|--------|--------|-------|----------|
| 1 | 1.00 | 1.00 | 1.00 | 12.00 | 1.55 | 86.00 | 35.80 | 1.00 |
| 2 | 1.00 | 2.00 | 2.00 | 65.00 | 1.65 | 63.00 | 23.14 | 3.00 |
| 3 | 1.00 | 2.00 | 1.00 | 24.00 | 2.00 | 90.00 | 22.50 | 1.00 |
| 4 | 1.00 | 3.00 | 3.00 | 24.00 | 1.89 | 98.00 | 27.43 | 1.00 |
| 5 | 1.00 | 3.00 | 2.00 | 53.00 | 1.79 | 72.00 | 22.47 | 2.00 |
| 6 | 1.00 | 2.00 | 3.00 | 53.00 | 1.70 | 80.00 | 27.68 | 2.00 |
| 7 | 1.00 | 2.00 | 2.00 | 22.00 | 1.95 | 109.00 | 28.67 | 1.00 |
| 8 | 1.00 | 2.00 | 1.00 | 12.00 | 1.55 | 80.00 | 33.30 | 1.00 |
| 9 | 1.00 | 2.00 | 3.00 | 45.00 | 1.78 | 80.00 | 25.25 | 2.00 |
| 10 | 1.00 | 3.00 | 3.00 | 54.00 | 1.70 | 73.00 | 25.26 | 2.00 |
| 11 | 1.00 | 2.00 | 2.00 | 32.00 | 1.85 | 93.00 | 27.17 | 2.00 |
| 12 | 1.00 | 3.00 | 2.00 | 21.00 | 1.68 | 62.00 | 21.97 | 1.00 |

TABLA DE FRECUENCIAS y ESTADÍSTICOS DESCRIPTIVOS

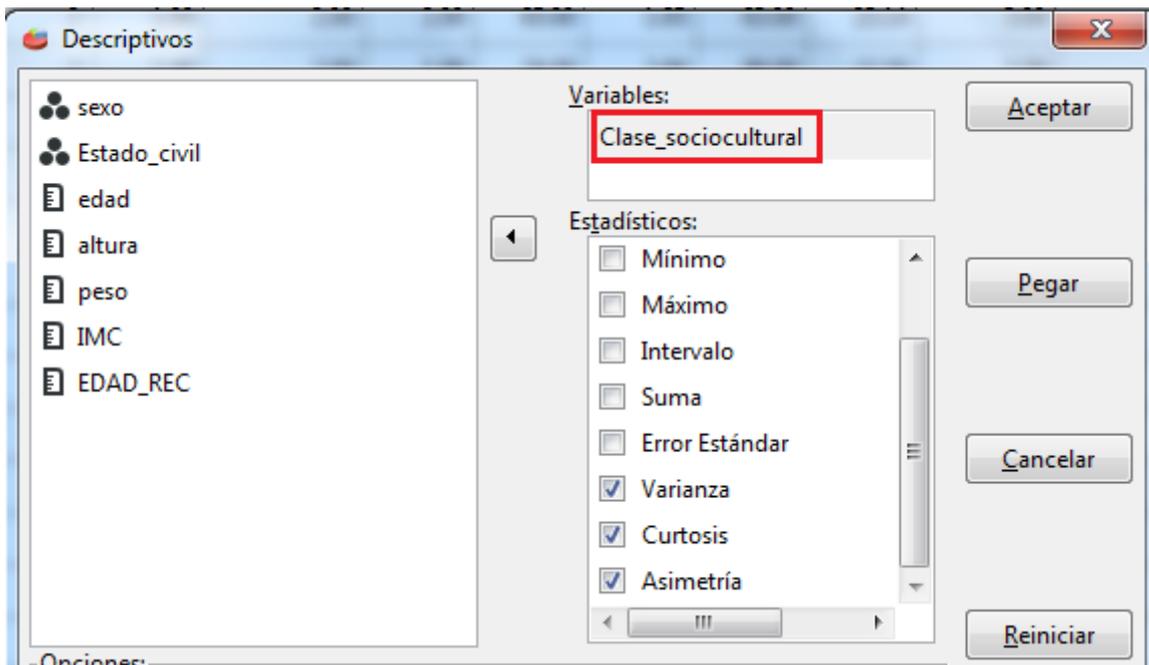
En PSPP podemos pedir la tabla de frecuencias y los descriptivos básicos de cualquier variable, por ejemplo del nivel socio-económico. En **analizar**→**estadística descriptiva**→**descriptivos**:



Se abrirá una ventana emergente. Ahí introduciremos la variable o variables que queremos analizar:



Para solicitar los estadísticos descriptivos deseados, vamos a 'estadísticos' y seleccionamos los que deseemos, por ejemplo, la media, desviación típica, varianza, asimetría y curtosis.



Cuando hacemos click en **continuar** y **aceptar**, en la ventana de resultados saldrán los estadísticos solicitados en una tabla:

Casos válidos = 25; casos con valor(es) perdido(s) = 0.

| Variable | N | Media | Desv Std | Varianza | Curtosis | Err.Est.Curt. | Asimetría | Err.Est.Asim. |
|---------------------|----|-------|----------|----------|----------|---------------|-----------|---------------|
| Clase_sociocultural | 25 | 2.12 | .73 | .53 | -.97 | .90 | -.19 | .46 |

Para solicitar la tabla de frecuencias deberemos ir a **analizar**→**estadística descriptiva**→**frecuencias**. La opción '**frecuencias**' también permite solicitar los estadísticos deseados, incluida la moda y mediana, que no se permiten en la opción '**descriptivos**'. Por otro lado,

Por ejemplo, podemos pedir las frecuencias para la variable '**peso**', junto con algunos descriptivos básicos y un histograma (en '**gráficos**')

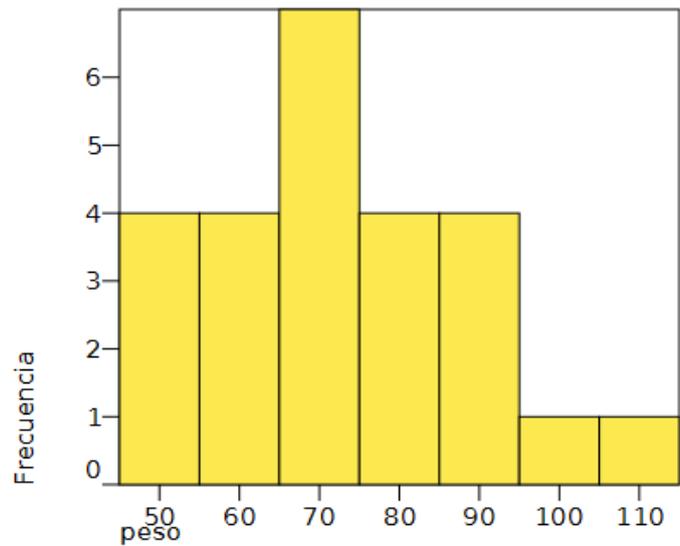
peso

| Etiqueta de Valor | Valor | Frecuencia | Porcentaje | Porcentaje Válido | Porcentaje Acumulado |
|-------------------|--------------|------------|--------------|-------------------|----------------------|
| | 48.00 | 1 | 4.00 | 4.00 | 4.00 |
| | 49.00 | 1 | 4.00 | 4.00 | 8.00 |
| | 50.00 | 2 | 8.00 | 8.00 | 16.00 |
| | 59.00 | 1 | 4.00 | 4.00 | 20.00 |
| | 62.00 | 1 | 4.00 | 4.00 | 24.00 |
| | 63.00 | 2 | 8.00 | 8.00 | 32.00 |
| | 65.00 | 1 | 4.00 | 4.00 | 36.00 |
| | 70.00 | 2 | 8.00 | 8.00 | 44.00 |
| | 72.00 | 2 | 8.00 | 8.00 | 52.00 |
| | 73.00 | 2 | 8.00 | 8.00 | 60.00 |
| | 75.00 | 1 | 4.00 | 4.00 | 64.00 |
| | 80.00 | 3 | 12.00 | 12.00 | 76.00 |
| | 86.00 | 1 | 4.00 | 4.00 | 80.00 |
| | 90.00 | 2 | 8.00 | 8.00 | 88.00 |
| | 93.00 | 1 | 4.00 | 4.00 | 92.00 |
| | 98.00 | 1 | 4.00 | 4.00 | 96.00 |
| | 109.00 | 1 | 4.00 | 4.00 | 100.00 |
| | Total | 25 | 100.0 | 100.0 | |

peso

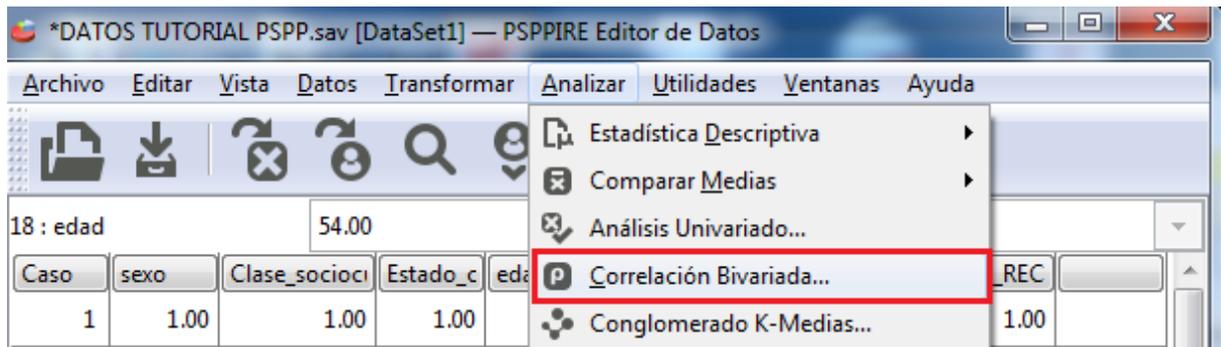
| | | |
|----------------------|--------------|--------|
| <i>N</i> | Válido | 25 |
| | Perdidos | 0 |
| <i>Media</i> | | 72.80 |
| <i>Modo</i> | | 80.00 |
| <i>Desv Std</i> | | 15.91 |
| <i>Varianza</i> | | 253.25 |
| <i>Curtosis</i> | | -.23 |
| <i>Err.Est.Curt.</i> | | .90 |
| <i>Asimetría</i> | | .30 |
| <i>Err.Est.Asim.</i> | | .46 |
| <i>Intervalo</i> | | 61.00 |
| <i>Mínimo</i> | | 48.00 |
| <i>Máximo</i> | | 109.00 |
| | 50 (Mediana) | 72.00 |

HISTOGRAM

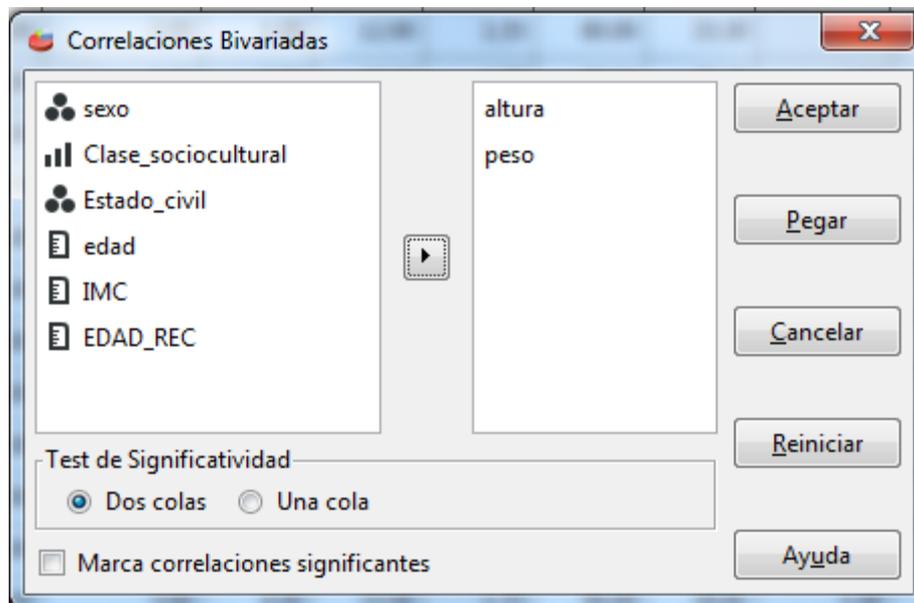


COEFICIENTE DE CORRELACIÓN DE PEARSON

Puede que nos interese ver la relación entre dos variables **CUANTITATIVAS**. Para ello seleccionaremos **analizar** → **correlación bivariadas**



En la ventana emergente se seleccionarán las variables entre las que se quiere obtener el coeficiente de correlación de Pearson, por ejemplo **altura** y **peso**:



Una vez hagamos click en **aceptar**, nos aparecerá en la ventana de resultados la tabla con el coeficiente de correlación de PEARSON entre las variables:

Correlaciones

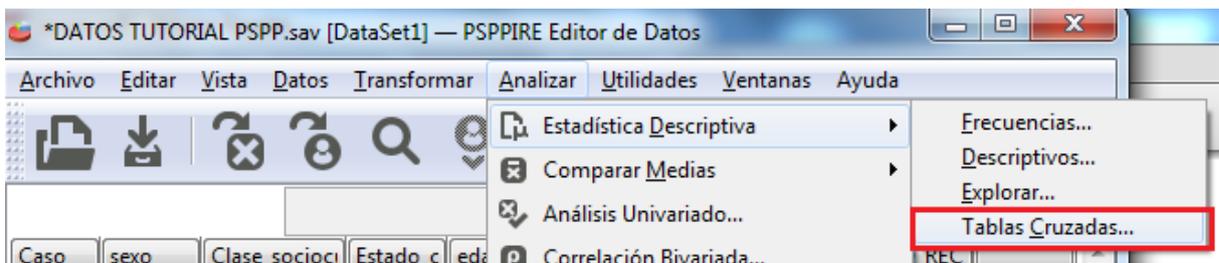
| | | altura | peso |
|--------|------------------------|--------|------|
| altura | Correlación de Pearson | 1.00 | .77 |
| | Sign. (2-colas) | | .00 |
| | N | 25 | 25 |
| peso | Correlación de Pearson | .77 | 1.00 |
| | Sign. (2-colas) | .00 | |
| | N | 25 | 25 |

En este caso la correlación es de 0.77, correlación directa alta.

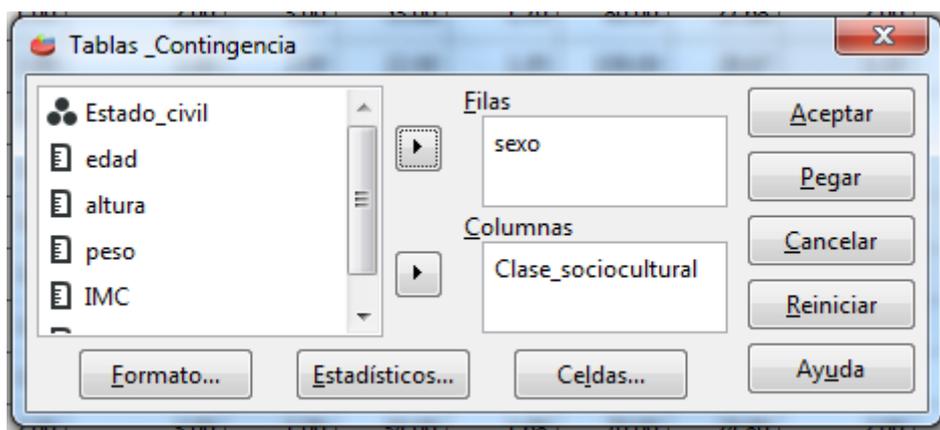
NOTA: La **correlación biserial puntual** también se puede calcular con esta opción.

OTROS COEFICIENTES DE CORRELACIÓN

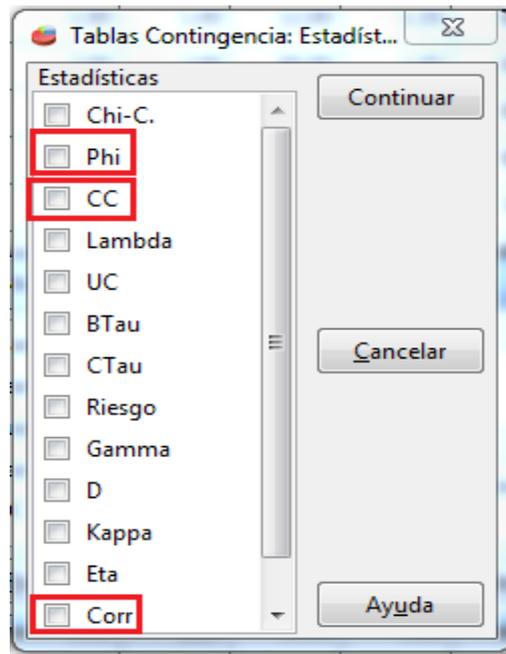
Puede que queramos medir la correlación entre dos variables cualitativas o ordinales, para ello debemos ir a **analizar**→**estadística descriptiva**→**tablas cruzadas**.



En la ventana emergente seleccionaremos las variables que queremos correlar. Por ejemplo el nivel socio-económico con el sexo:



Ahora, en **estadísticos**, podemos solicitar que calcule el **coeficiente de contingencia (CC)** (como es una variable dicotómica y otra politómica es lo más correcto). **También nos permite el cálculo del coeficiente ϕ (Phi), de la correlación de Spearman (Corr.)**



Así, en los resultados puedo observar tanto la tabla de contingencia, que indica las frecuencias por sexo y nivel socio-económico, como el valor del coeficiente de contingencia, de .33 en este caso:

sexo * Clase_sociocultural [recuento].

| sexo | Clase_sociocultural | | | Total |
|--------|---------------------|-------|------|-------|
| | alta | media | baja | |
| Hombre | 1.00 | 8.00 | 4.00 | 13.00 |
| Mujer | 4.00 | 4.00 | 4.00 | 12.00 |
| Total | 5.00 | 12.00 | 8.00 | 25.00 |

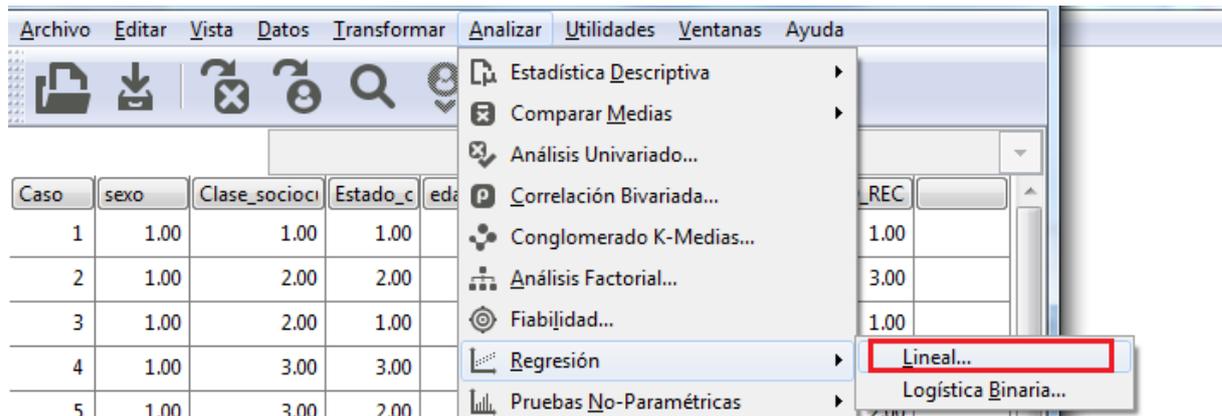
Medidas simétricas.

| Categoría | Estadístico | Valor | Err. Est. Asint. | T Aproxim. | Sign. Aproxim. |
|----------------------------------|-----------------------------|-------|------------------|------------|----------------|
| Nominal según N de casos válidos | Coeficiente de Contingencia | .33 | | | |
| | | 25 | | | |

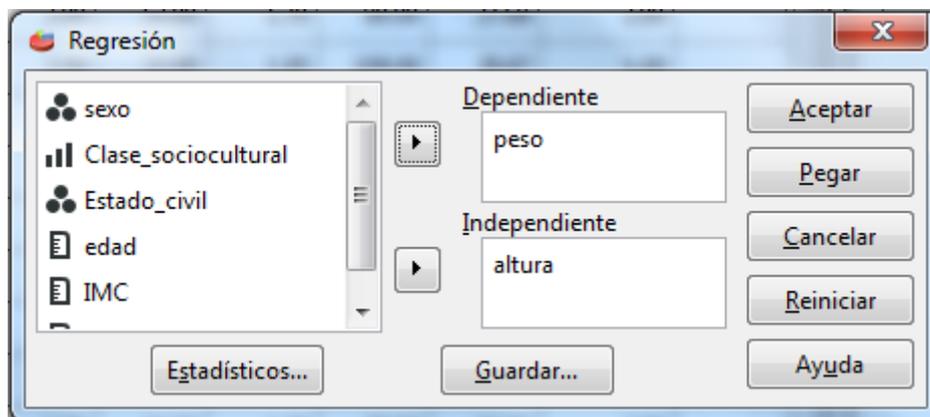
Un valor de .33 nos indica, en este caso, que el hecho de ser mujer se asocia positiva, aunque moderadamente, con tener un nivel socio-económico más alto.

REGRESIÓN

Para obtener la ecuación de regresión lineal simple o múltiple deberemos ir a **analizar**→**regresión** → **lineal**.



Si, por ejemplo, queremos predecir el peso a partir de la altura, pondremos como **variable dependiente** el PESO y como **variable independiente** la ALTURA:



En cuanto hagamos click en **aceptar** nos aparecerá la tabla de datos. Nos interesa la tabla **resumen del modelo** y la tabla **coeficientes**:

Resumen del modelo (peso)

| R | R Cuadrada | R Cuadrada Ajustada | Error estándar del Estimador |
|-----|------------|---------------------|------------------------------|
| .77 | .59 | .57 | 10.38 |

ANOVA (peso)

| | Suma de Cuadrados | df | Cuadrado medio | F | Sign. |
|-----------|-------------------|----|----------------|-------|-------|
| Regresión | 3598.61 | 1 | 3598.61 | 33.38 | .00 |
| Residual | 2479.39 | 23 | 107.80 | | |
| Total | 6078.00 | 24 | | | |

Coefficientes (peso)

| | Unstandardized Coefficients | | Standardized Coefficients | t | Sign. |
|------------|-----------------------------|----------------|---------------------------|-------|-------|
| | B | Error Estándar | Beta | | |
| (Constant) | -70.70 | 24.92 | .00 | -2.84 | .01 |
| altura | 85.32 | 14.77 | .77 | 5.78 | .00 |

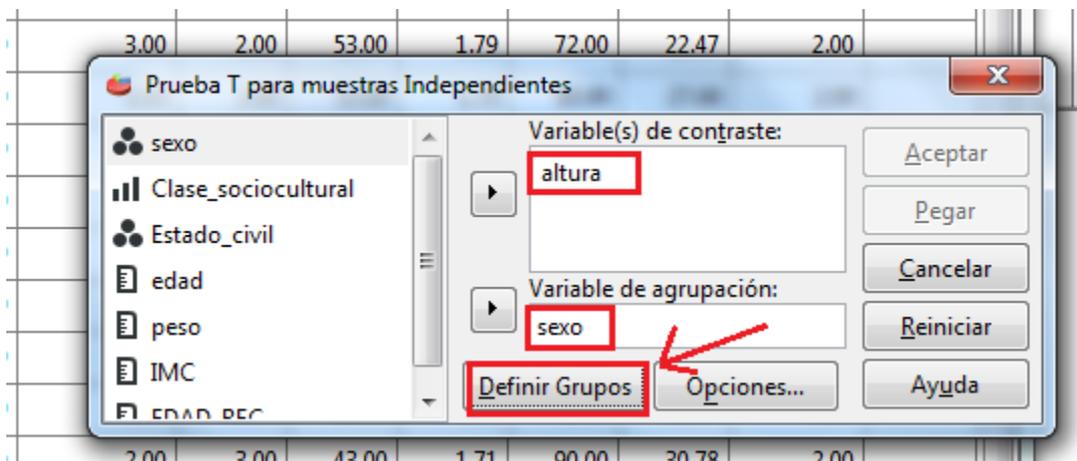
Para construir el modelo $Y=a+bx$, cogeremos de la tabla **coeficientes** los dos valores de la primera columna. En este caso tenemos el siguiente modelo:

$$y = -70.701 + 85.32x$$

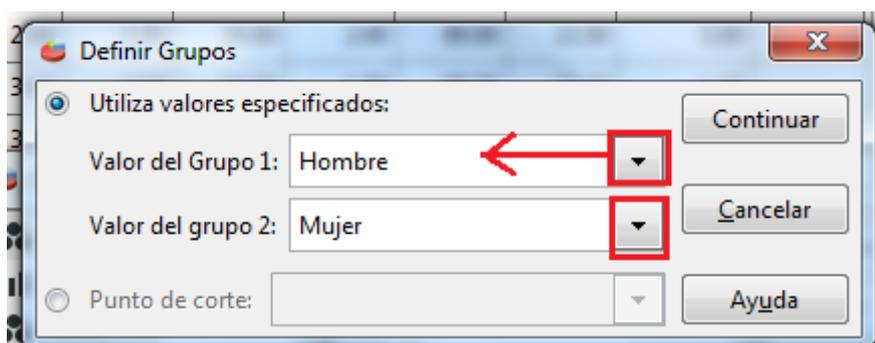
En la tabla **resumen del modelo**, el R se refiere al coeficiente de correlación de pearson (0.77) y el R cuadrado al coeficiente de determinación o ajuste global del modelo ($R^2=0.59$).

COMPARAR MEDIAS ENTRE DOS GRUPOS DIFERENTES

Si queremos comparar si la diferencia entre dos medias para una misma variable en dos grupos diferentes es estadísticamente significativa, deberemos ir a **analizar** → **comparar medias** → **prueba T para muestras independientes**. En la ventana emergente, debemos introducir la variable de la que se extraerán las medias (Variable dependiente) en **'variable(s) de contraste'** y la variable de agrupación (variable independiente) en **'variable de agrupación'**:



Para definir sobre qué dos grupos se debe realizar el contraste, debemos hacer click en **'definir grupos'**, y señalar las dos categorías que definirán los grupos de contraste, en este caso es 'hombre' y 'mujer':



Después presionamos **continuar** y **aceptar** y nos aparecerán en la ventana de resultados los datos del contraste.

Los datos que nos interesan son los que aparecen en la tabla **'prueba para muestras independientes'**. En primer lugar deberemos observar si se **asumen varianzas iguales**. Eso se contrasta en las dos primeras columnas de datos, en la **prueba de Levene para la igualdad de varianzas**.

- Si se consideran iguales las varianzas (**Sign.>0.05**), se cogen los datos que se muestran en la primera fila.
- Si se consideran las varianzas diferentes (**Sign.<0.05**), se cogen los datos que se muestran en la segunda fila.

En el ejemplo, se consideran las varianzas iguales:

Estadísticas de grupo

| | sexo | N | Media | Desviación Estándar | Err.Est.Media |
|--------|--------|----|-------|---------------------|---------------|
| altura | Hombre | 13 | 1.75 | .14 | .04 |
| | Mujer | 12 | 1.60 | .10 | .03 |

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 |
| altura | Se asume igualdad de varianzas | 1.44 | .24 | 3.01 | 23.00 | .01 | .15 | .05 | .05 | .25 |
| | Igualdad de varianzas no asumida | | | 3.05 | 22.02 | .01 | .15 | .05 | .05 | .25 |

Se interpretan estos datos, el estadístico de contraste $t=3.01$ tiene una probabilidad asociada inferior a 0.05 (0.01), por lo que se rechaza la hipótesis nula y existen diferencias significativas entre las medias

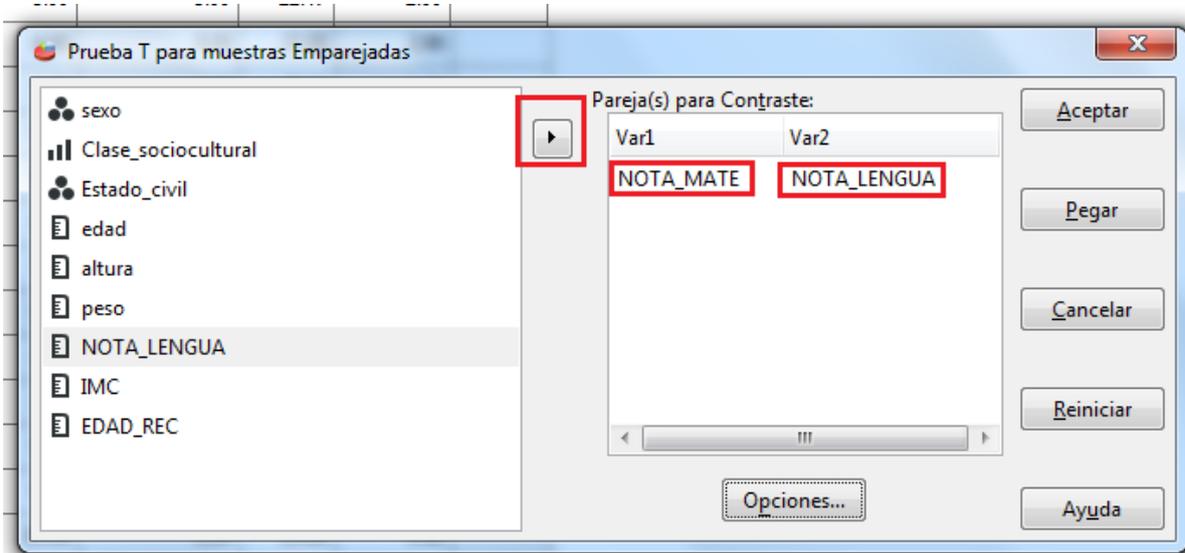
Se observa cómo los hombres tienen una mayor altura media (1.75 metros los hombres por 1.60 las mujeres). Además, existen diferencias significativas ($\text{Sign.}<0.05$) en el contraste. Por lo tanto, se puede afirmar que las diferencias en la altura media de hombres y mujeres es significativa, y que las diferencias obtenidas en la muestra no son debidas al azar (se pueden generalizar a nivel poblacional).

COMPARAR MEDIAS EN UN MISMO GRUPO

En algunas ocasiones lo que interesa es comparar las medias de variables diferentes en todos los sujetos de la muestra. Para poder realizar este contraste es necesario que las variables que se emparejan tengan una escala igual.

Para pedir el contraste **analizar** → **comparar medias** → **prueba T para muestras emparejadas**.

En la tabla deberemos seleccionar las dos variables sobre las que queremos hacer el contraste y emparejarlas:



En este caso, una vez presionemos **aceptar**, de los resultados nos interesa la tabla **prueba de muestras emparejadas**:

Prueba de muestras emparejadas

| | Diferencias emparejadas | | | | | t | df | Sign. (2-colas) |
|----------------------------------|-------------------------|---------------------|------------------|---|----------|------|----|-----------------|
| | Media | Desviación Estándar | Error Est. Media | Intervalo de confianza 95% de la Diferencia | | | | |
| | | | | Inferior | Superior | | | |
| Pareja 1 NOTA_MATE - NOTA LENGUA | .08 | 3.20 | .64 | -1.24 | 1.40 | .125 | 24 | .902 |

En este caso, el valor del estadístico de contraste es muy pequeño, $T=0.125$, y su probabilidad asociada es de $0.902 (>0.05)$, por lo que no rechazo la Hipótesis Nula y puedo señalar que las diferencias entre la nota de matemáticas y la de lengua no son significativas

ANÁLISIS DE VARIANZA

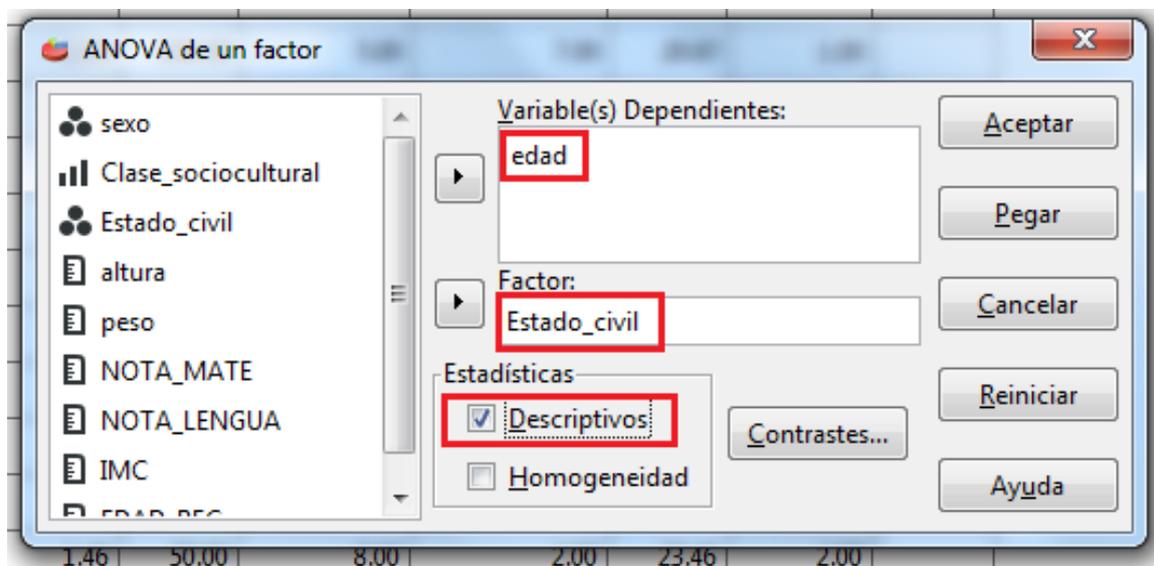
COMPARAR MEDIAS ENTRE MÁS DE DOS GRUPOS INDEPENDIENTES

Puede ocurrir que en vez de querer comparar sólo dos grupos (como el sexo, hombre y mujer), queramos comparar más de dos grupos (como el estado civil: casado, soltero, divorciado, viudo). En este caso debemos hacer otro contraste llamado **Análisis de Varianza**. Para realizar el contraste:

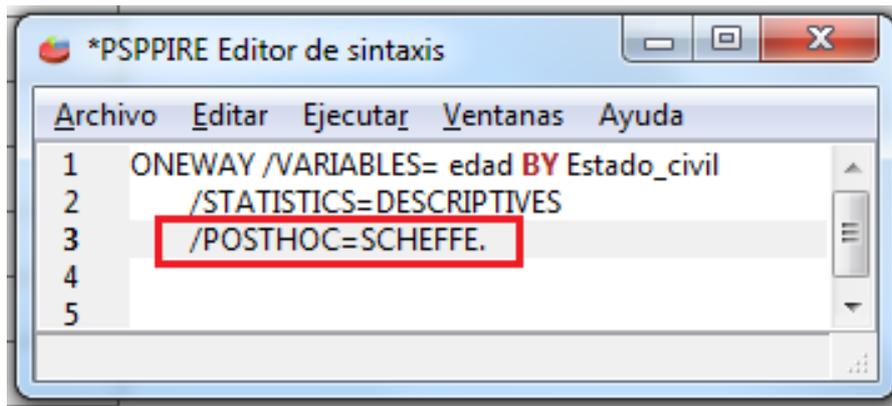
analizar → **comparar medias** → **ANOVA de un factor**.

Imaginemos que queremos contrastar si el estado civil tiene influencia sobre la variable edad.

En la ventana introduciremos la variable de agrupación (Estado civil) en **factor** y la variable dependiente en **variable(s) dependientes** (mediremos la edad). En el cuadro '**estadísticas**' también he seleccionado la opción '**descriptivos**'. Si se deja así el análisis, solamente se sabrá si existen diferencias significativas entre algún grupo, sin saber entre qué grupos en concreto existe.



Para saber entre qué grupos concretamente se encuentran las diferencias, debo pedirlo. Eso se hace, en lugar de presionando aceptar, presionando '**pegar**'. Se abrirá una nueva ventana a la que tengo que quitar el punto final que aparece después de 'DESCRIPTIVES' y añadir más abajo el código '/POSTHOC=SCHEFFE'.



Posteriormente, hago click en **ejecutar** → **todos**, y en la ventana de resultados se habrá reproducido:

Descriptivos

| | | N | Media | Desviación Estándar | Error Estándar | Intervalo de Confianza 95% para la Media | | Mínimo | Máximo |
|------|------------|----|-------|---------------------|----------------|--|-----------------|--------|--------|
| | | | | | | Límite Inferior | Límite Superior | | |
| edad | soltero | 8 | 20.00 | 6.57 | 2.32 | 14.51 | 25.49 | 12.00 | 32.00 |
| | casado | 8 | 34.38 | 8.73 | 3.09 | 27.07 | 41.68 | 22.00 | 45.00 |
| | divorciado | 6 | 48.33 | 5.57 | 2.28 | 42.48 | 54.18 | 42.00 | 54.00 |
| | Viudo | 3 | 64.33 | 10.02 | 5.78 | 39.45 | 89.22 | 54.00 | 74.00 |
| | Total | 25 | 36.72 | 16.54 | 3.31 | 29.89 | 43.55 | 12.00 | 74.00 |

ANOVA

| | | Suma de Cuadrados | df | Cuadrado medio | F | Sign. |
|------|--------------|-------------------|----|----------------|-------|-------|
| edad | Entre Grupos | 5377.16 | 3 | 1792.39 | 31.58 | .00 |
| | Intra Grupos | 1191.87 | 21 | 56.76 | | |
| | Total | 6569.04 | 24 | | | |

Existen diferencias significativas entre alguno de los grupos

Múltiples Comparaciones (edad)

| (I) Estado_civil | | (J) Estado_civil | Diferencia Media (I - J) | Error Estándar | Sign. | Intervalo de Confianza del 95% | |
|------------------|------------|------------------|--------------------------|----------------|-------|--------------------------------|-----------------|
| | | | | | | Límite Inferior | Límite Superior |
| Scheffé soltero | | casado | -14.38 | 3.77 | .01 | -25.81 | -2.94 |
| | | divorciado | -28.33 | 4.07 | .00 | -40.69 | -15.98 |
| | | Viudo | -44.33 | 5.10 | .00 | -59.82 | -28.85 |
| casado | soltero | | 14.38 | 3.77 | .01 | 2.94 | 25.81 |
| | divorciado | | -13.96 | 4.07 | .02 | -26.31 | -1.61 |
| | Viudo | | -29.96 | 5.10 | .00 | -45.44 | -14.47 |
| divorciado | soltero | | 28.33 | 4.07 | .00 | 15.98 | 40.69 |
| | casado | | 13.96 | 4.07 | .02 | 1.61 | 26.31 |
| | Viudo | | -16.00 | 5.33 | .05 | -32.17 | .17 |
| Viudo | soltero | | 44.33 | 5.10 | .00 | 28.85 | 59.82 |
| | casado | | 29.96 | 5.10 | .00 | 14.47 | 45.44 |
| | divorciado | | 16.00 | 5.33 | .05 | -1.17 | 32.17 |

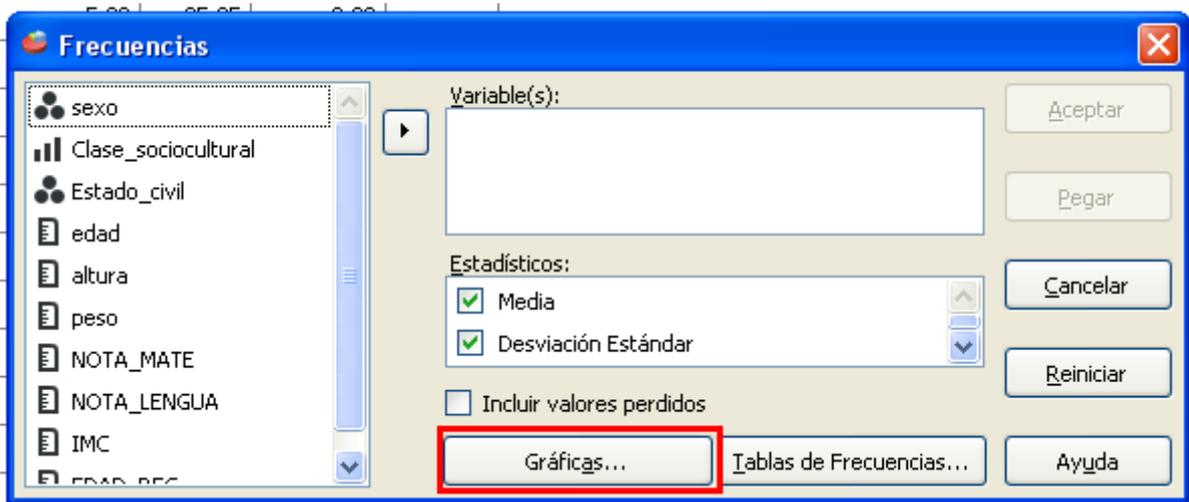
Se observan diferencias entre todos los grupos excepto entre viudos y divorciados

Se observan diferencias significativas entre los grupos ($F=31.58$; $Sig<0.05$), y si se analizan las **pruebas post hoc**, se observa una $sig<0.05$ entre solteros-casados; solteros-divorciados; solteros-viudos; casados-divorciados y casados-viudos. Si analizamos las medias de ambos grupos, esto es, la primera tabla de **descriptivos**, podemos ver cómo los solteros tienen una

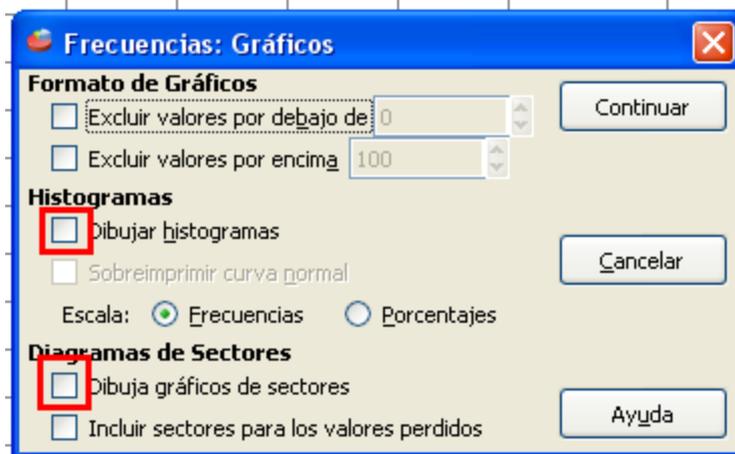
edad media inferior a los casados, estos inferior a los divorciados y los divorciados a los viudos. Podemos afirmar pues, que los solteros tienen una edad significativamente inferior al resto de sujetos de otros estados civiles y que los casados tienen a su vez una edad significativamente inferior a los divorciados y los viudos.

GENERAR GRÁFICOS EN PSPP

Para generar un sectorgrama (variables cualitativas) o un histograma (variables cuantitativas continuas), debemos ir a **analizar** → **estadística descriptiva** → **frecuencias**. Haciendo click en el botón '**gráficos**' podremos acceder a una ventana que nos permite introducir ambos gráficos para las variables seleccionadas en la ventana inicial:

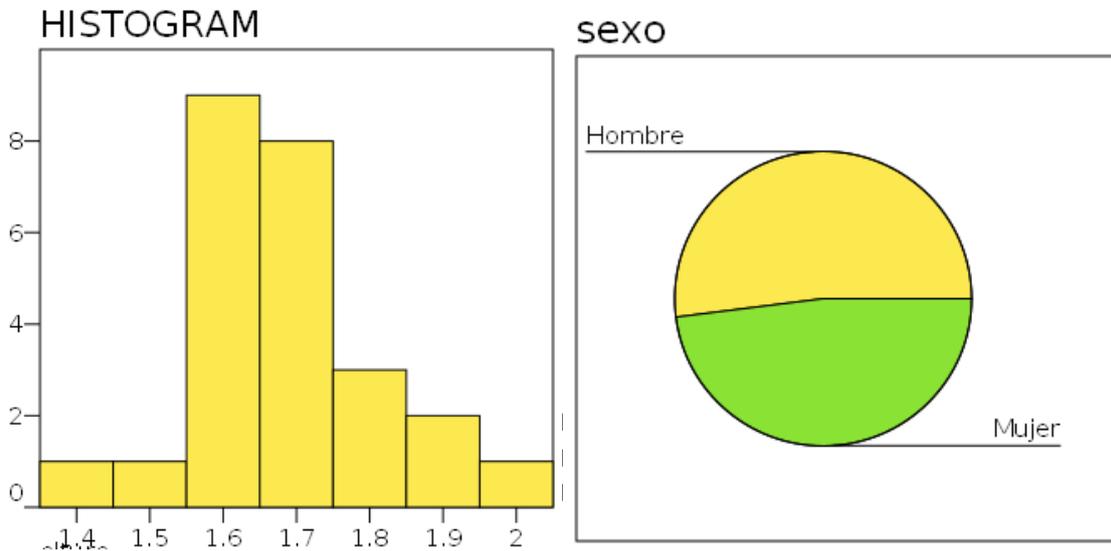


Veamos la ventana emergente que surge:



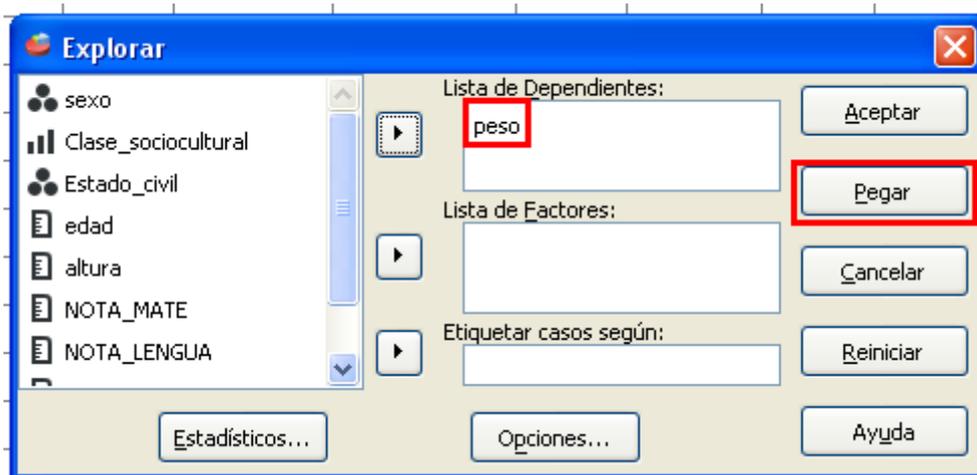
El PSPP no diferencia entre el tipo de variables a la hora de devolver los gráficos. Por ello, podemos pedir un sectorgrama para una variable cuantitativa o un histograma para una variable cualitativa y que en la ventana de resultados nos aparezcan dichos gráficos. Sin embargo, esa operación está mal realizada y debemos desechar inmediatamente el gráfico generado que no sea correcto.

Por ejemplo, si pedimos un histograma para la variable 'altura' de nuestra base de datos y un gráfico de sectores o sectograma para la variable 'sexo', los gráficos obtenidos son los que se muestran a continuación:



El programa PSPP no dispone de la opción de generar un **gráfico de caja y bigotes** directamente en el menú. Sin embargo, es posible hacerlo mediante el uso de sintaxis².

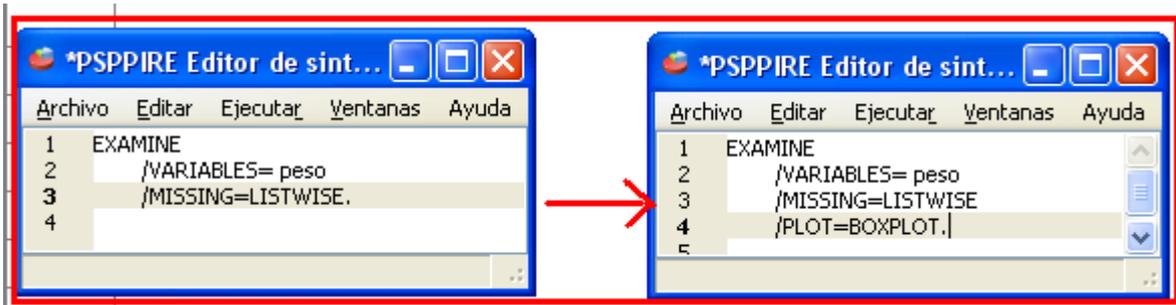
Para ello debemos ir a la opción 'explorar' en **analizar** → **estadística descriptiva** → **explorar**. Una vez ahí, introducimos la variable de la que queremos sacar el gráfico en la ventana 'lista de dependientes'. Si por ejemplo queremos ver el diagrama de cajas de la variable peso:



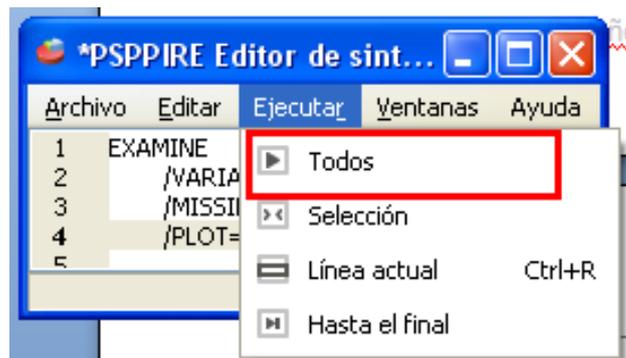
² La sintaxis es el código que el programa necesita para dar la orden de llevar a cabo los análisis. A nivel general, cuando hacemos click en una ventana el código se genera de modo automático, pero existen algunas opciones que no existen en las ventanas y menús. Para activarlas hay que ir directamente a la sintaxis y escribir el código correspondiente.

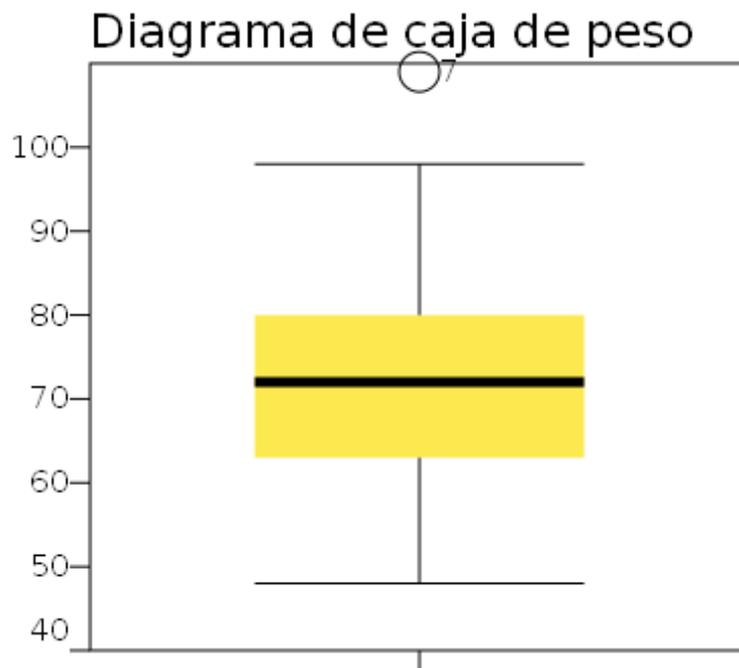
Una vez hecho esto, debemos hacer click en el botón 'pegar', para que nos aparezca una ventana emergente con el código o sintaxis que estamos pidiendo.

En esa ventana de código deberemos añadir una línea nueva que diga `'/PLOT= BOXPLOT.'`, teniendo cuidado con el punto final, que debe aparecer justo al final de esta línea y no de la anterior como estaba puesto previamente:



Una vez escrito el código como se muestra en la ventana de la derecha, debemos ir a ejecutar → todos, y en la ventana emergente nos aparecerá el gráfico solicitado:



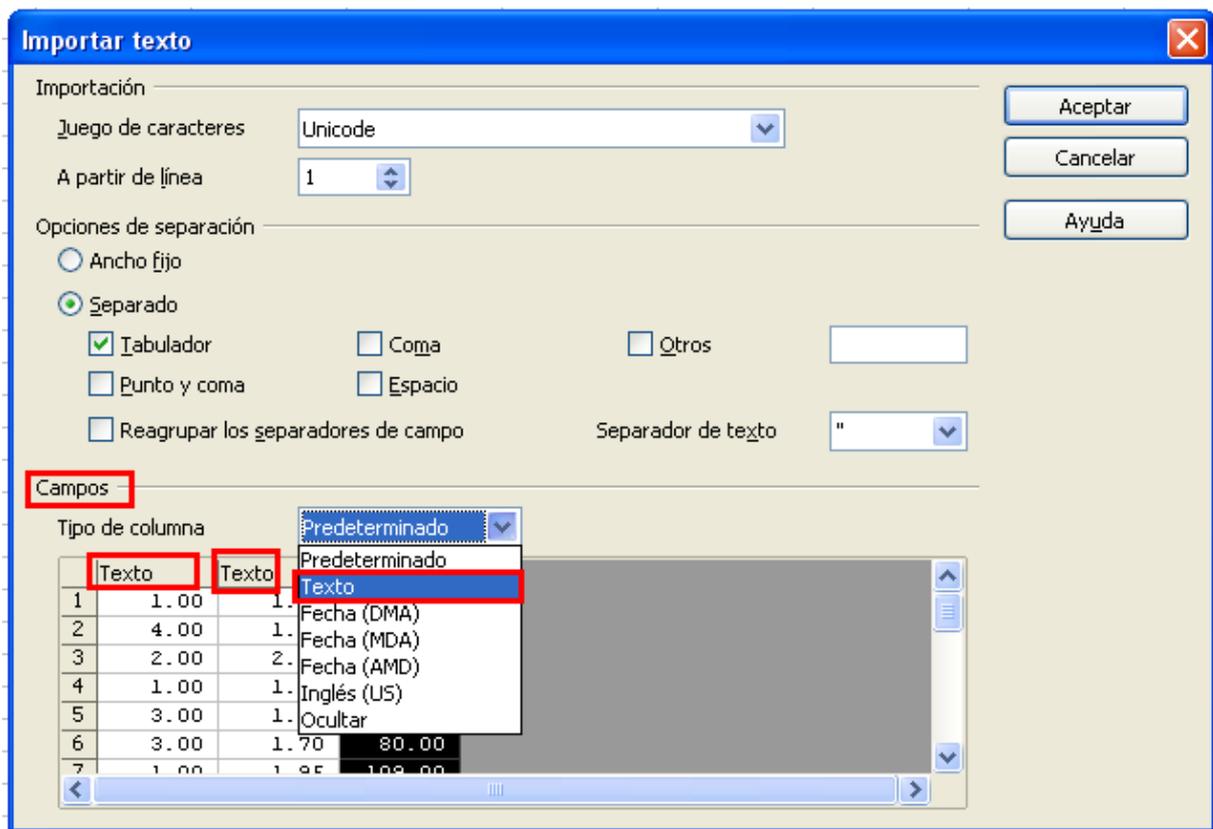


Por desgracia, en PSPP no se pueden realizar los gráficos de dispersión. Sin embargo, esta operación se puede solicitar fácilmente en una hoja de cálculo (ya sea de Microsoft Office o de OpenOffice).

Para ello, primero debemos transportar los datos a la hoja de cálculo (se explicará aquí en OpenOffice, pero la operación es igual, o más sencilla para Excel). Lo primero es copiar las variables deseadas desde el PSPP. Seleccionamos las variables haciendo click en sus nombres (por ejemplo 'altura' y 'peso'), y se copia:

| edad | altura | peso | NOTA_MATE | NOTA LENGUA | IMC | EDAD_REC |
|-------|--------|--------|-----------|-------------|-------|----------|
| 12.00 | 1.55 | 86.00 | 6.00 | 7.00 | 35.80 | 1.0 |
| 65.00 | 1.65 | 63.00 | 9.00 | 8.00 | 23.14 | 3.0 |
| 24.00 | 2.00 | 90.00 | 8.00 | 5.00 | 22.50 | 1.0 |
| 24.00 | 1.89 | 98.00 | 2.00 | 4.00 | 27.43 | 1.0 |
| 53.00 | 1.79 | 72.00 | 8.00 | 8.00 | 22.47 | 2.0 |
| 53.00 | 1.70 | 80.00 | 6.00 | 9.00 | 27.68 | 2.0 |
| 22.00 | 1.95 | 109.00 | 5.00 | 7.00 | 28.67 | 1.0 |
| 12.00 | 1.55 | 80.00 | 9.00 | 7.00 | 33.30 | 1.0 |
| 45.00 | 1.78 | 80.00 | 6.00 | 5.00 | 25.25 | 2.0 |

Vamos a la hoja de cálculo y pegamos. Puede que nos aparezca una ventana emergente pidiéndonos más información antes del pegado. Debemos dejar todo igual, salvo en los campos, donde tenemos que seleccionar cada variable e indicar para cada caso en el desplegable que es de tipo 'texto'. Si no hacemos esto, podemos tener problemas con la lectura del tipo de datos:



Por último, una vez pegados los datos, debemos ir al menú **editar** → **buscar y reemplazar**, y añadir en la ventana 'buscar' un punto y en la ventana 'reemplazar por' una coma, como se observa en la imagen siguiente:



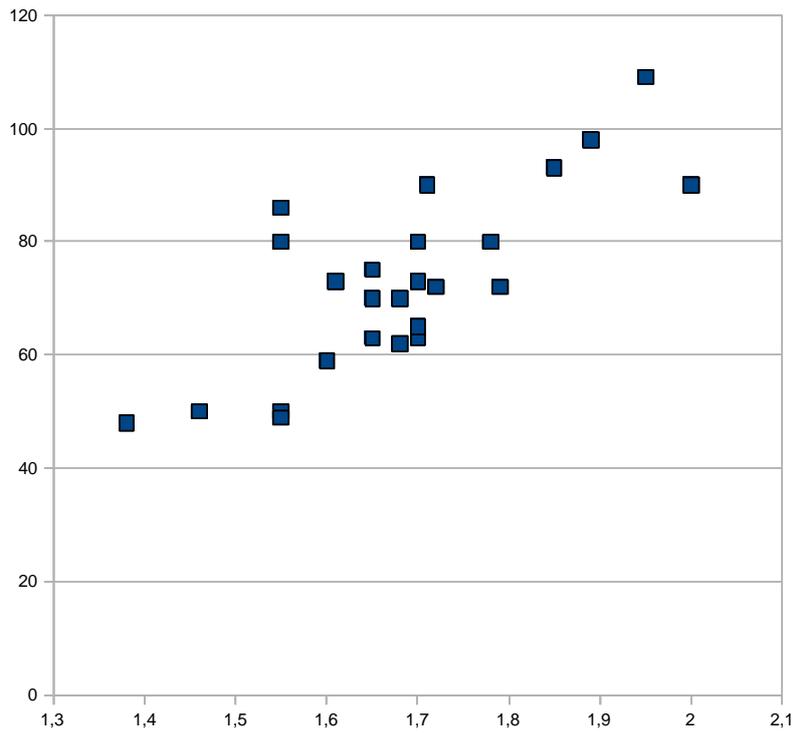
Finalmente hacemos click en '**reemplazar todo**' y ya nos habrá identificado la hoja de cálculo las variables correctamente.

Ahora ya sólo tenemos que pedirle el gráfico en la opción **insertar** → **gráfico** → **XY (dispersión)** teniendo seleccionadas las variables deseadas:

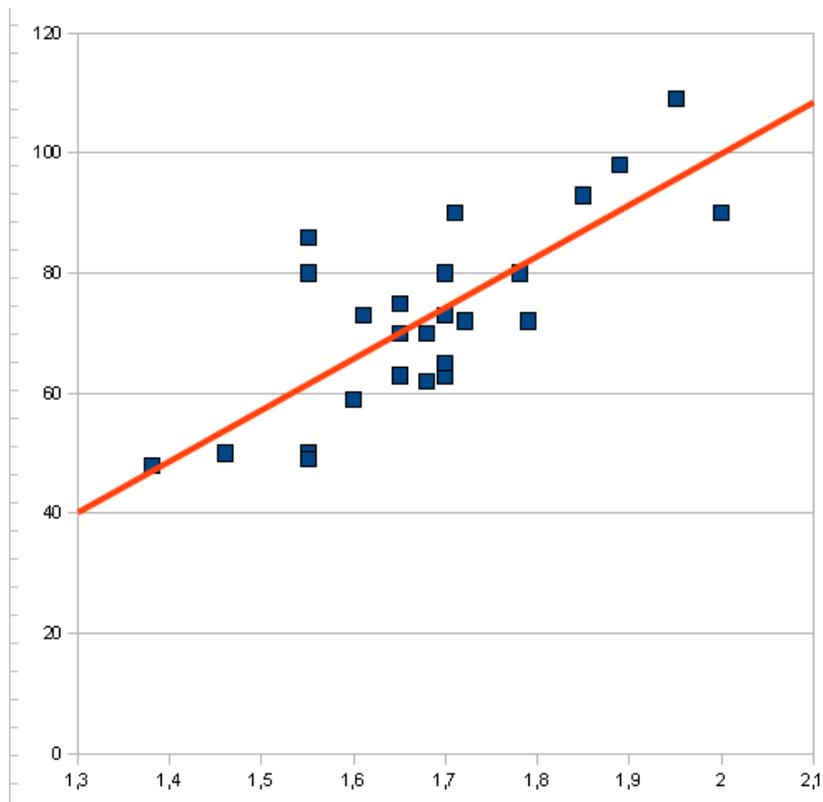
The screenshot shows the 'Asistente de gráficos' (Chart Wizard) dialog box in a spreadsheet application. The spreadsheet has two columns: 'PESO' (Weight) and 'ALTURA' (Height). The 'Asistente de gráficos' dialog box is open, and the 'Selección de un tipo de gráfico' (Select a chart type) section is active. The 'XY (dispersión)' (XY (Scatter)) option is selected and highlighted with a red box. The 'Sólo puntos' (Only points) sub-option is also highlighted with a red box. The 'Finalizar' (Finish) button at the bottom right is also highlighted with a red box. The 'Pasos' (Steps) list on the left shows the current step is '1. Tipo de gráfico'.

| A | B |
|------|--------|
| PESO | ALTURA |
| 1,55 | 86 |
| 1,65 | 63 |
| 2 | 90 |
| 1,89 | 98 |
| 1,79 | 72 |
| 1,7 | 80 |
| 1,95 | 109 |
| 1,55 | 80 |
| 1,78 | 80 |
| 1,7 | 73 |
| 1,85 | 93 |
| 1,68 | 62 |
| 1,71 | 90 |
| 1,68 | 70 |
| 1,72 | 72 |
| 1,46 | 50 |
| 1,38 | 48 |
| 1,65 | 75 |
| 1,55 | 50 |
| 1,7 | 63 |
| 1,6 | 59 |
| 1,55 | 49 |
| 1,61 | 73 |
| 1,7 | 65 |
| 1,65 | 70 |

Y ya tendremos nuestro gráfico de dispersión que podemos editar al gusto:



También podemos insertar la recta de regresión en el menú **insertar** → **líneas de tendencia...** → **lineal**:

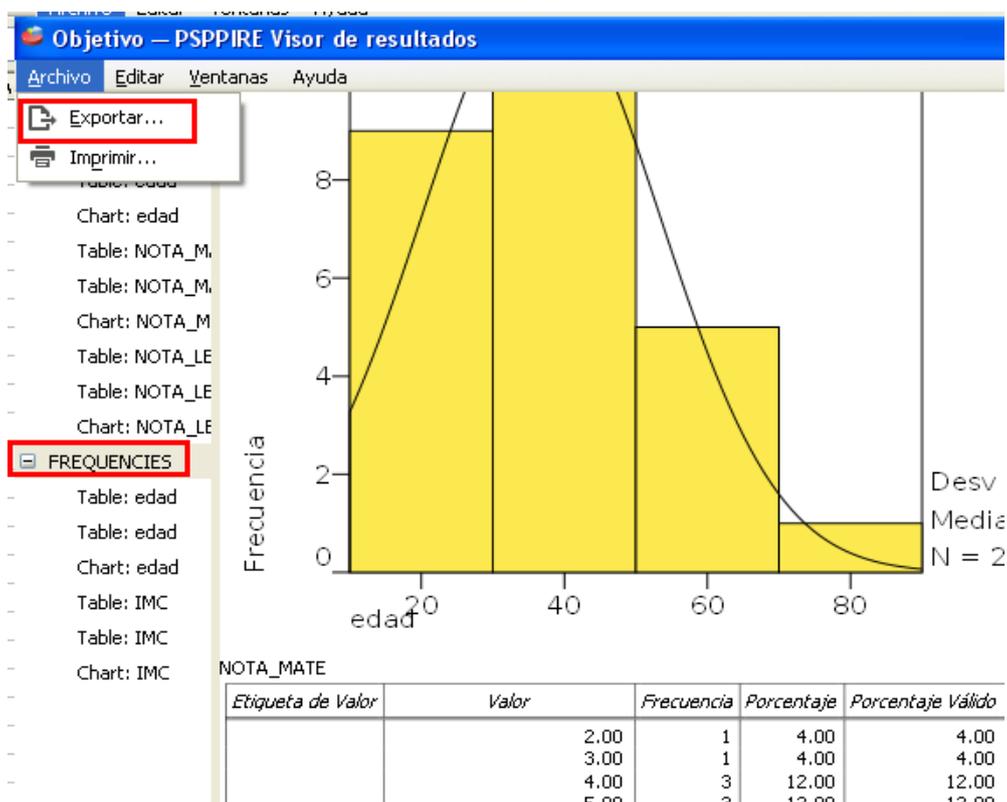


EXPORTAR TABLAS Y GRÁFICOS DE PSPP

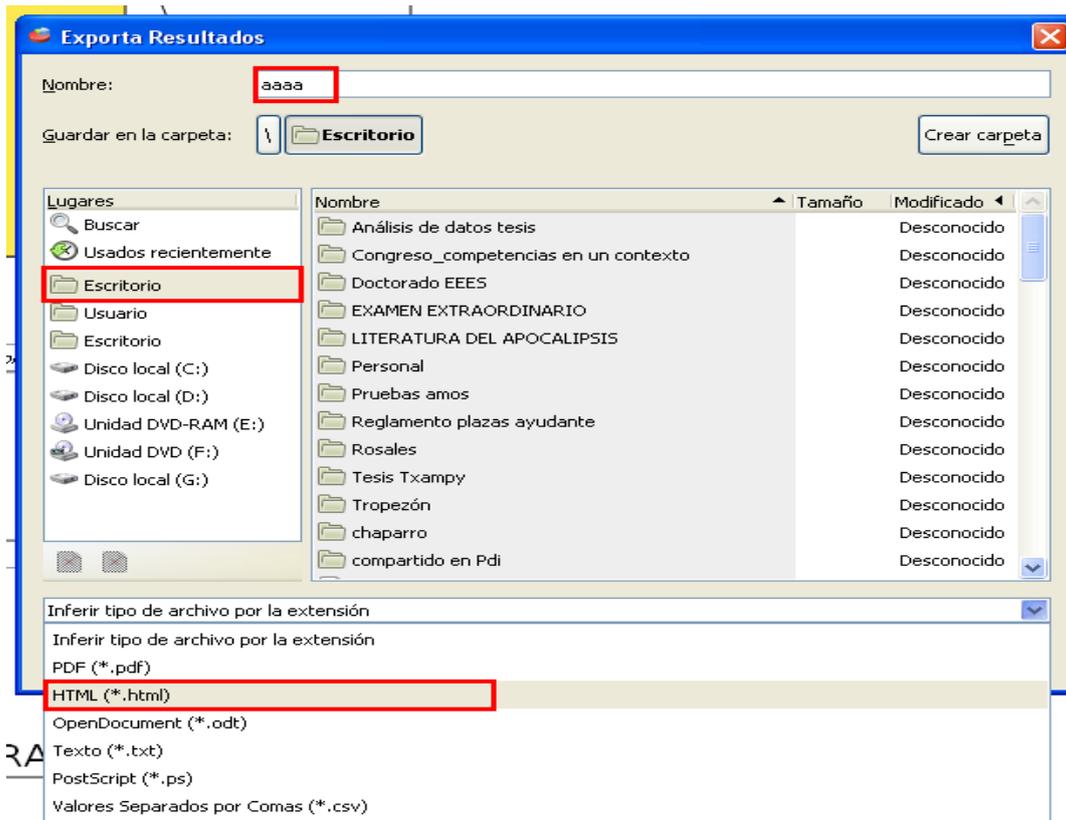
En PSPP, la ventana de resultados se muestra como un documento plano y sobre el que no se puede interactuar. Sin embargo, con la opción **archivo → exportar**, podemos copiar fácilmente las tablas a un formato que nuestro procesador de textos (Word o Swriter) puede leer fácilmente.

Simplemente debemos guardar el archivo como html, y después copiar y pegar las tablas deseadas en el procesador de textos.

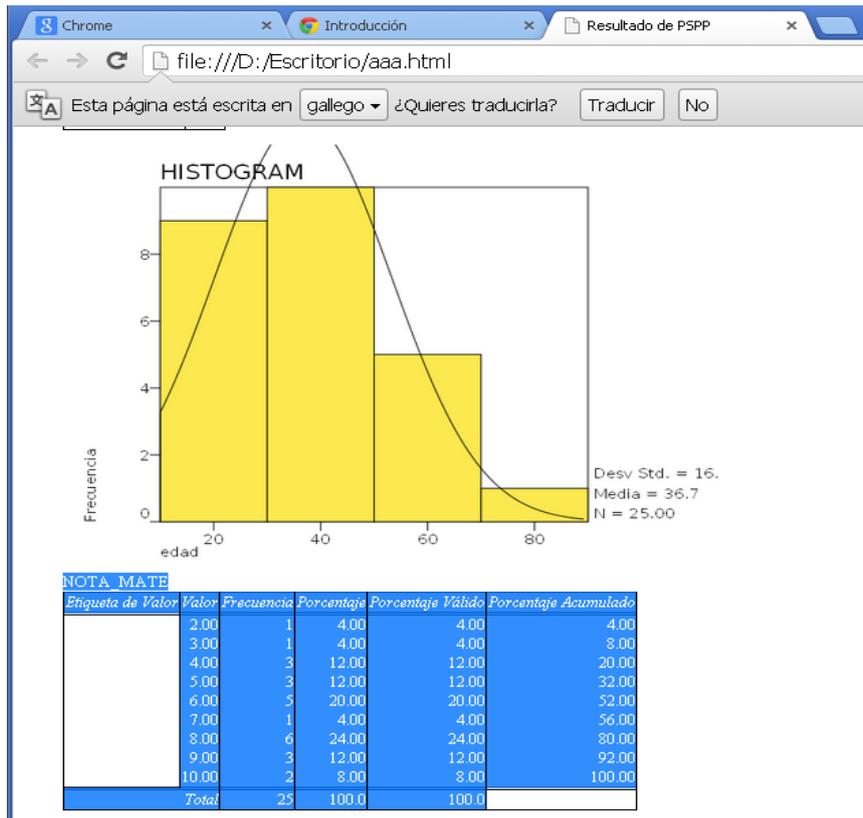
Así, en primer lugar, seleccionamos en la columna izquierda de la ventana de resultados el resultado que queremos guardar, y hacemos click en **archivo → exportar**:



En la ventana emergente, seleccionamos el lugar donde queremos que se guarde, el nombre del archivo y el formato html. Es importante seleccionar bien el lugar dónde queremos guardar el archivo, ya que en PSPP funciona la selección un poco diferente que con otros programas de Windows:



Si abrimos el archivo creado, se verá en el explorador de internet, y las tablas se podrán copiar y pegar sin ningún problema en nuestro procesador de textos (para poder ver correctamente los gráficos se recomienda emplear Internet Explorer o Google Chrome):



ANEXO II

NOMBRE ALUMNO 1 _____

NOMBRE ALUMNO 2 _____

A partir de la base de datos de hábitos sobre salud y educación sexual obtener:

- 1- Tabla de frecuencias para las variables 'estado civil' y 'frecuencia de salidas nocturnas'

Tabla 1. Distribución de la muestra por estado civil

| | n_i |
|-------------------|-------|
| Soltero | |
| Casado | |
| Separado | |
| Divorciado | |
| Viudo | |

n=

Tabla 2. Distribución de la muestra por frecuencia de salidas nocturnas

| | n_i | n_a |
|-----------|-------|-------|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| n= | | |

¿Qué se puede decir de estas tablas de frecuencias?

2- Obtener los estadísticos descriptivos principales para la variable 'edad'

| \bar{x} | <i>Mdn</i> | <i>Mo</i> | S_x^2 | S_x | <i>As</i> | <i>Curt</i> |
|-----------|------------|-----------|---------|-------|-----------|-------------|
| | | | | | | |

¿Cuál es la edad más repetida en la muestra? ¿Qué frecuencia tiene la edad más repetida de la muestra?

Calcula el coeficiente de variación. ¿Qué se puede decir acerca de la dispersión de la variable edad? ¿Es la media un buen estadístico de tendencia central?

3- Dividir el archivo por la variable 'sexo' y calcular los estadísticos descriptivos de tendencia central por sexo

| | | | |
|---------------|--|--|--|
| | | | |
| Hombre | | | |
| Mujer | | | |

¿Qué se puede decir en cuanto a las diferencias entre hombres y mujeres de la muestra en lo que respecta al consumo de alcohol?

NOMBRE ALUMNO 1 _____

NOMBRE ALUMNO 2 _____

A partir de la base de datos de hábitos sobre salud y educación sexual obtener:

- 1- Coeficientes de correlación entre:
 - a. Número de parejas sexuales, edad y edad de inicio de las relaciones sexuales (cuantitativas)
 - b. Sexo y edad de inicio de las relaciones sexuales (cualitativa dicotómica y cuantitativa)
 - c. Estudios del padre y estudios de la madre (ordinales)
 - d. Sexo y consumo de alcohol en los últimos 12 meses (cualitativas dicotómicas)
- 2- Estima el modelo de regresión que prediga el número de parejas sexuales a partir de la edad de inicio de las relaciones sexuales.
 - a. Escribe la ecuación de regresión, el valor del coeficiente de correlación de Pearson y el porcentaje de varianza explicada
 - b. Vuelve a hacer un modelo en el que la VD siga siendo el número de parejas sexuales y las VI la edad de inicio de las relaciones sexuales, el sexo y la frecuencia de salidas nocturnas. Escribe la ecuación de regresión, r_{xy} y R^2

3- Entonces, ¿existen diferencias significativas entre hombres y mujeres en lo que respecta a la edad de inicio de las relaciones sexuales? Comprueballo con un contraste de hipótesis (inferencia estadística; Prueba de T para muestras independientes).

4- Comprueba si se puede afirmar que existan diferencias significativas entre el Tamaño del municipio (Considerando esta variable el factor) y la frecuencia de salidas nocturnas. (ANOVA de un factor).

ANEXO III

INCORPORACIÓN DE SOFTWARE LIBRE EN LA DOCENCIA UNIVERSITARIA

Uno de los objetivos básicos de la educación universitaria es proveer al alumno universitario de herramientas específicas que puedan servirle de ayuda en su futuro desempeño, tanto laboral, como personal y social. Así, el **software libre** facilita a cualquier persona el acceso a recursos que hasta hace pocos años eran prohibitivos para la mayoría de los ciudadanos.

Este cuestionario, integrado en el Proyecto de Innovación *Incorporación del software libre PSPP para la realización de prácticas de análisis de datos en la asignatura 'Metodología de la Investigación Socio-Educativa'* trata de recoger información sobre las actitudes y opiniones del alumnado de esta asignatura sobre la integración en la docencia de la herramienta **PSPP**.

CUESTIONES DEMOGRÁFICAS

- Sexo
- Edad
- Repetidor en la asignatura

OPINIÓN ACERCA DEL SOFTWARE LIBRE

(Indica tu grado de acuerdo con estas afirmaciones siendo 1 totalmente en desacuerdo y 5 totalmente de acuerdo)

- El software libre fomenta la igualdad de oportunidades en la sociedad
- Los planteamientos del software libre fomentan el pensamiento crítico de las personas
- El software libre fomenta la independencia de las personas e instituciones en la sociedad
- El empleo de software libre fomenta el funcionamiento democrático de las instituciones
- El empleo de software libre por instituciones y empresas les beneficia **no sólo** en el aspecto económico
- Creo que todas las instituciones públicas deberían emplear software libre en sus procedimientos administrativos
- Si existe una alternativa de software libre frente a otro de pago, la utilizo en mi ordenador personal por cuestiones principalmente morales
- En definitiva, creo que el software libre es beneficioso para la sociedad en su conjunto

VALOR DEL SOFTWARE LIBRE EN LA EDUCACIÓN

(Indica tu grado de acuerdo con estas afirmaciones siendo 1 totalmente en desacuerdo y 5 totalmente de acuerdo)

- El empleo de software libre fomenta la igualdad de oportunidades en **educación**
- El uso de software libre nos hace a los estudiantes más autónomos e independientes
- Creo que el conocimiento de herramientas de software libre frente al software de pago va a ser ventajoso para mi futuro laboral

- Entre el software libre y de pago, prefiero que mis profesores me presenten la alternativa libre
- El empleo de software libre en las instituciones educativas transmite a los estudiantes valores esenciales para la educación
- Las instituciones públicas deberían fomentar el desarrollo de software libre educativo
- En definitiva, el empleo de software libre en la educación mejora la calidad de la formación recibida por los estudiantes

VALORACIÓN DEL EMPLEO DEL SOFTWARE LIBRE PSPP EN LA ASIGNATURA

- La enseñanza de PSPP para el análisis de datos, frente a la enseñanza de otra alternativa de pago, fomenta la igualdad de oportunidades en la sociedad
- En el caso de necesitar realizar cualquier análisis de datos, el conocimiento de la herramienta PSPP me hace más autónomo e independiente
- Creo que el conocimiento de PSPP frente a otro software de análisis de datos de pago puede aportarme ventajas en mi futuro laboral
- La Universidad de Salamanca debería fomentar el empleo de software libre como PSPP en las prácticas de análisis de datos de las titulaciones frente a otro software de pago
- En definitiva, creo que el empleo de PSPP (frente al empleo de software de pago) mejora la calidad de la formación que recibo en la universidad

¡Muchas gracias por tu colaboración!