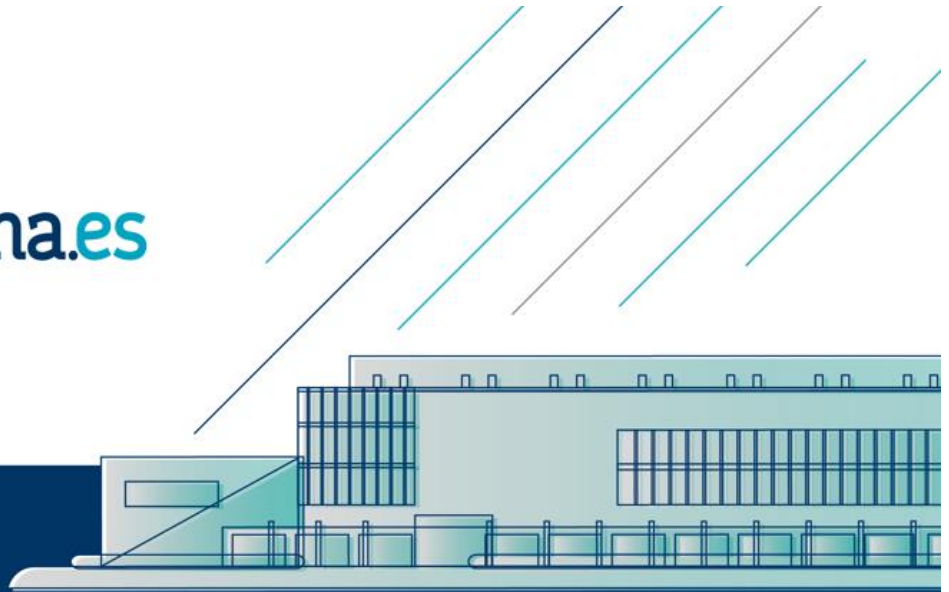




UNIVERSIDAD
DE MÁLAGA

| **uma.es**



GUÍA RÁPIDA DE SPSS

Ismael Pablo Soler García
Departamento de Economía Aplicada (Estadística y
Econometría)

Tel: 952 13 23 99 | ipsoler@uma.es

Facultad de Ciencias Económicas y Empresariales
Campus de El Ejido
Málaga.

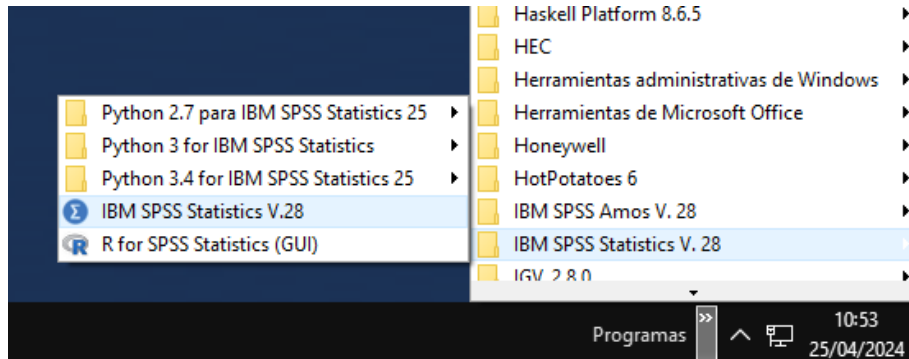
Metodología en
Ciencias
Jurídicas y
Sociales

Introducción

SPSS es posiblemente el programa más utilizado para el análisis de datos. Esto es debido fundamentalmente a su reducida curva de aprendizaje en comparación con otros programas, junto con la gran potencia para todo tipo de análisis.

Abrir SPSS

SPSS está disponible directamente en los ordenadores UMA mediante acceso directo a los programas (cerca del reloj):



o buscándolo en la ruta de Inicio.



Si estamos fuera de la UMA deberemos acceder a través de un PC Virtual disponible en el Campus Virtual. Se trata de la primera entrada del listado de asignaturas en las que estamos matriculados.

Másteres Oficiales de Posgrado

Aulas TIC | Programación Docente

<p>Mis asignaturas en Másteres Oficiales</p>	<p>Mis asignaturas en todo Campus Virtual</p>	<p>Mis a O</p>
--	---	--------------------

▼ Curso académico 2023-2024

PC virtuales para prácticas con software licenciado UMA

Seguimos las instrucciones y accedemos, salvo casos especiales, a la primera de las opciones.

 Novedades y anuncios

Acceso remoto al software para uso docente

El **software** utilizado en la docencia de las distintas Titulaciones que se imparten en el Centro se va a proporcionar mediante:

- Acceso remoto a **PC en Aulas TIC** del Centro con su software habitual.
- Acceso remoto a **PC virtual con software genérico**, común a todas las Aulas TIC.
- Acceso remoto a **PC virtual con software específico**.

IMPORTANTE:

- Cuando termines de trabajar con un PC remoto (PC en Aulas TIC o con un PC virtual) **no olvides**:
 1. **Copiar tus archivos desde el PC remoto**, por ejemplo a GDrive o MS OneDrive
 2. Revisar que no dejas información personal en el PC remoto.
 3. **Apagar o cerrar sesión**, para que el PC remoto quede disponible para su próximo uso lo antes posible.
- **Usa solo un acceso remoto a la vez**, si no lo haces, estarás quitando la posibilidad de conexión a otra persona.

Para **copiar archivos** desde tu ordenador personal al PC remoto (y viceversa) puedes **utilizar** los servicios proporcionados por las nubes de [Microsoft OneDrive](#) o [Google Drive](#).

Sobre las posibilidades que ofrecen estas nubes para los alumnos de la UMA, encontrarás más información en [software.uma.es > Alumnos > Microsoft / Google](#).

Pulsa el siguiente enlace para el:

 [Acceso remoto a PC virtual con software genérico](#)

Pulsa uno de los siguientes enlace para el acceso remoto a PC Virtuales con software específico (**solo si necesitas ese software**):

 [PC Virtual - Ubuntu](#)

 [PC Virtual Programas de CAD](#)

Instalamos el UDS Plugin download page para que podamos acceder al PC Virtual.

Acceso remoto a PC virtual con software genérico

PC Virtual con sistema operativo Windows 10 Pro con el [software genérico de Aulas TIC](#)

- IBM SPSS
- Wolfram Mathematica
- Matlab
- Eviews
- ESRI ArcGis
- ...



ar en el botón "Acceder al PC Virtual" puede aparecer un mensaje para descargar un programa necesario para acceder al PC Virtual, solo la primera vez. La dirección de descarga directa es [UDS Plugin download page](#)

Una vez se accede estaremos en virtualmente en un ordenador de la facultad. Este ordenador virtual se ejecuta como una pestaña del navegador.

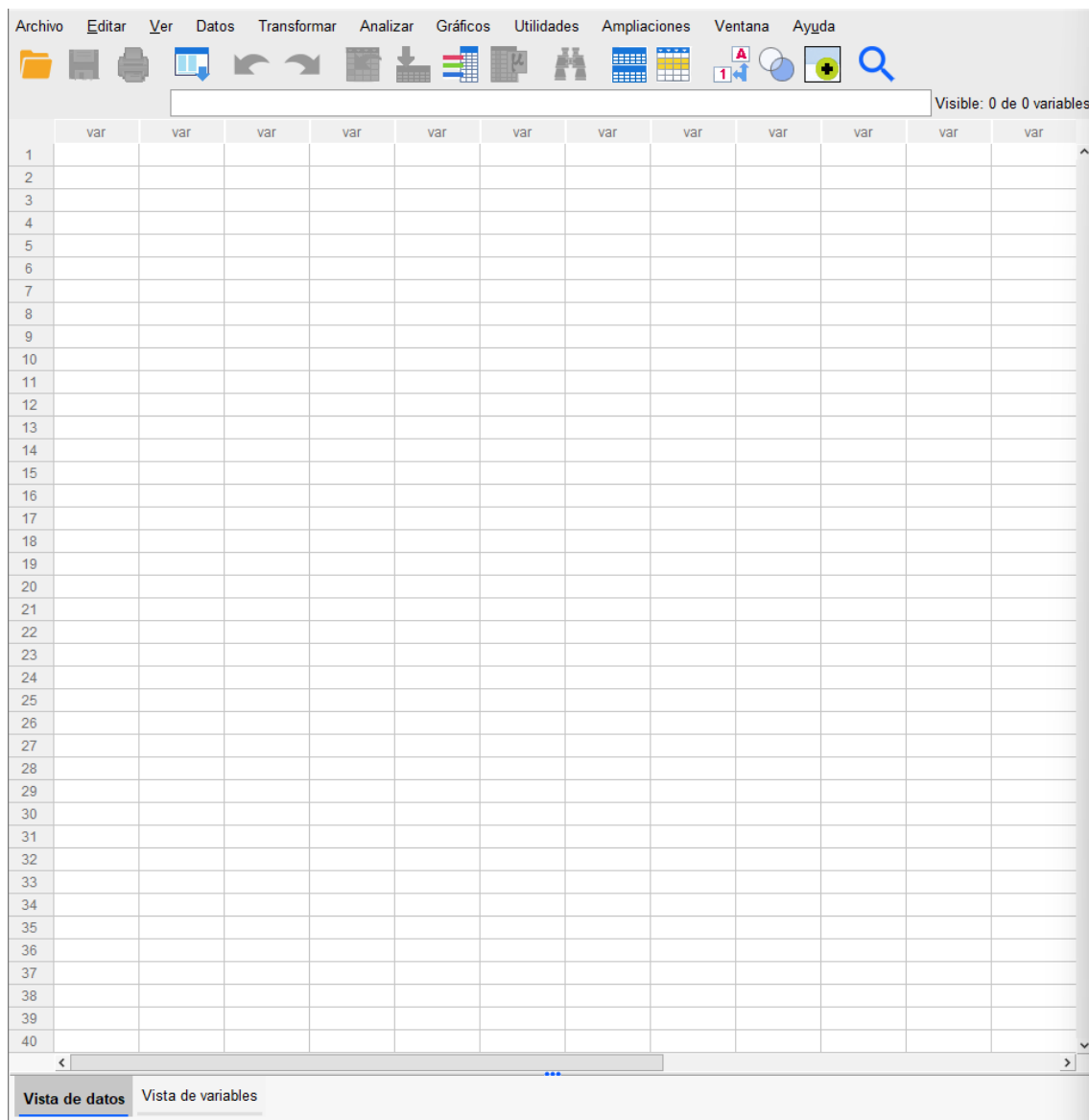
Recuerda copiar los datos desde el PC remoto, no dejar información personal y cerrar sesión cuando termines.

Interfaz

Una de las principales características del programa es que separa el trabajo en varias ventanas.

Por un lado, tenemos la ventana de edición de datos. En la que destacan principalmente dos pestañas:

La pestaña de vista de datos que aparece en la siguiente captura y cuyo diseño nos evoca a una hoja de cálculo de Excel.



En las filas, el programa coloca los casos, mientras que en las columnas incluirá las variables. Sin introducir directamente datos el programa le asignará un nombre y unas características a la variable por defecto.

Y la pestaña de vista de variables que sirve para configurar los datos que introducimos en el programa.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Ejemplo	Cadena	8	0		Ninguna	Ninguna	9	Izquierda	Nominal	Entrada
2											
3											
4											
5											
6											
7											
8											
9											

- Nombre: Nombre corto de la variable. No permite espacios, por ejemplo.
- Tipo: El tipo de dato que debe leer SPSS. Lo más común son numéricos, cadenas y fechas.
- Anchura: Define el número máximo de caracteres reservados para los registros.
- Decimales: Define el número de decimales visibles.
- Etiqueta: Describen o complementan en nombre de la variable.
- Valores: Permiten generar una leyenda o un cuadro de códigos.
- Perdidos: Le indica al programa como debe identificar los perdidos.
- Columnas: El tamaño asignado a la columna en el visor de datos.
- Alineación: Donde dentro de la casilla se muestran los datos.
- Medida: El tipo real de variable que estamos usando (Nominal, Ordinal, Escala).
- Rol: Las funciones de la variable en el análisis (por defecto entrada).

Abrir una base de datos

Podemos trabajar una base de datos nueva, abrir o importar datos ya generados anteriormente. Usa el icono de la carpeta o la pestaña archivo visible arriba a la izquierda del carrusel de herramientas.

También podemos acceder a una gran variedad de bases de datos integradas como muestra dentro del propio programa.

Bienvenido a IBM SPSS Statistics

Nuevos archivos:

- Nuevo conjunto de datos
- Nueva consulta de base de datos...

Archivos de muestra: Todo

- accidents.sav
- adl.sav
- advert.sav
- aflatoxin.sav
- aflatoxin20.sav
- anorectic.sav
- anticonvulsants.sav
- automaton.sps
- bankloan.sav
- bankloan_binning.sav
- bankloan_cs.sav
- bankloan_cs_noweights.sav
- behavior.sav
- behavior_ini.sav
- brakes.sav
- breakfast.sav

Novedades:

Sample Size Based on Confidence Interval

	N ^a	Actual CI Half Width	Assumed CI Half Width
Test for Mean Difference	64	4.996	5
	18	9.946	10
	7	18.497	20

a. Number of pairs.

La mayoría de los procedimientos de ANÁLISIS DE POTENCIA se han ampliado para permitir la estimación de márgenes de error asociados con varios intervalos de confianza, también conocidos como...

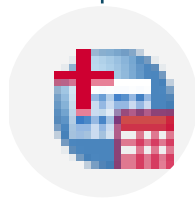
Ayuda y soporte | Guías de aprendizaje | Comunidad

Visite la [página de SPSS Statistics](#) para explorar los paquetes y ofertas disponibles.

Proporcionar con ... | Notificar problema

No mostrar este cuadro de diálogo en el futuro

Abrir | Cerrar



Ventana Editor de Datos



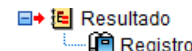
Ventana Visor de Resultados

Vista de variables

Pestaña vista de variables

Vista de datos

Pestaña vista de datos



Registro de Resultados

.frecuencias

[ConjuntoDatos1]

Estadísticos

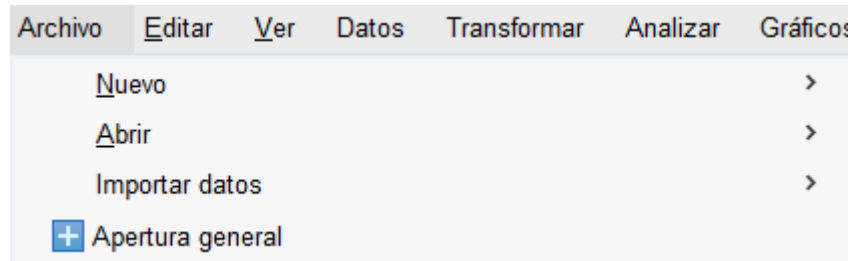
Ejemplo

N	Válido	27
	Perdidos	

Resultados

Importar datos

También es posible importar datos desde otros formatos, como por ejemplo Excel. Esto puede realizarse, utilizando el icono de abrir o la herramienta de importar disponible en la sección de Archivo.



Guardar datos de SPSS

SPSS permite guardar por un lado la base de datos (archivos .sav) y por otro lado los resultados (archivos.spv). Para guardar cualquiera de ellos, accede a la pestaña archivo o usa el icono del disquete de tres y medio¹.

¹ Felicidades a aquellos que recuerdan o, al menos conocen, el icónico disquete de tres y medio. Su figura como el símbolo de guardar ha resistido el paso del tiempo en nuestras interfaces digitales.

Pestañas clave

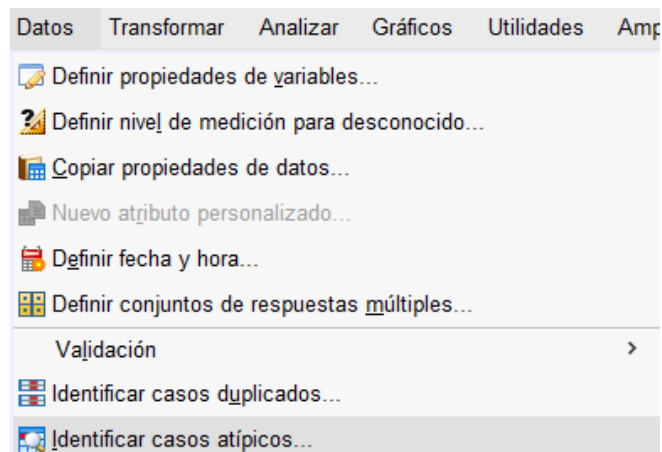
En esta sección se van a enumerar los procedimientos habituales disponibles en las principales pestañas de herramientas de SPSS

Datos

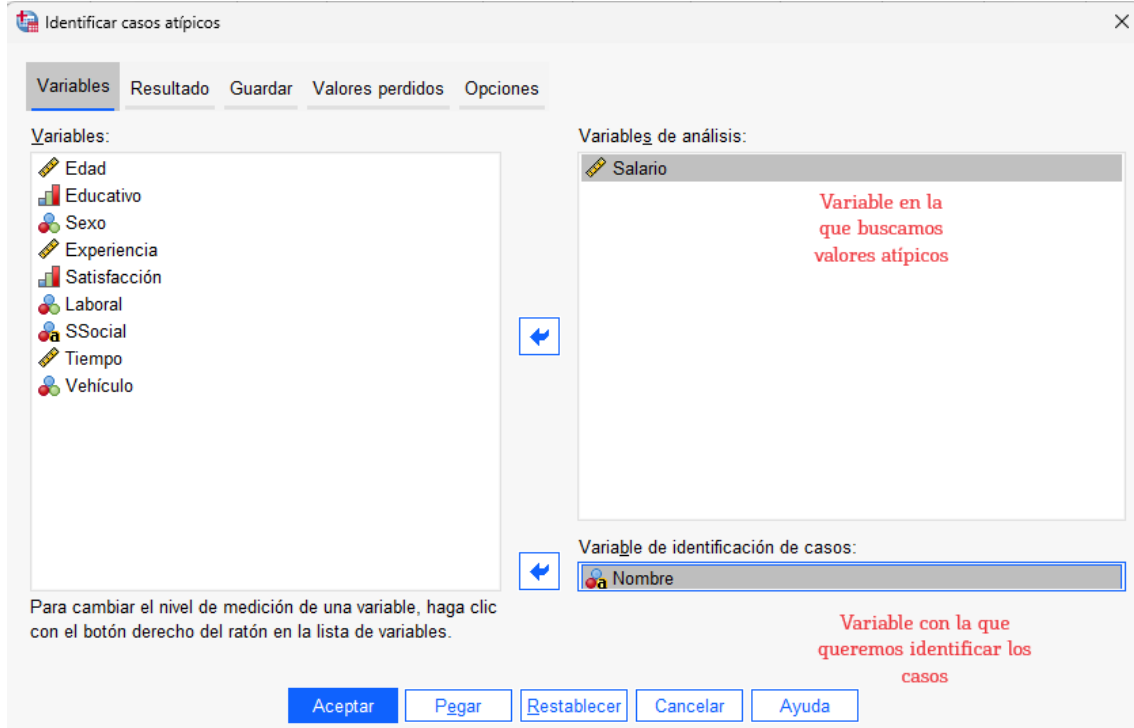
Esta pestaña incorpora procedimientos clave tales como: **Identificar casos atípicos** (mediante índice de anomalía), **Fusionar archivos**, **Segmentar archivos**, **Seleccionar casos**, **Ponderar casos**, entre otros.

- **Identificar casos atípicos:**

No es el método más común. Usa un procedimiento propio de SPSS basado en el índice de anomalía.



Se abrirá la siguiente ventana de trabajo.



El resultado será:

Detectar anomalía

Lista de índices de casos con anomalías

Caso	Nombre	Índice de anomalía
28	Raistlin Majere	13,500

Lista de ID de los homólogos de casos con anomalías

Caso	Nombre	ID de homólogo	Tamaño de homólogo	Porcentaje de tamaño de homólogo
28	Raistlin Majere	2	1	3,7%

Lista de motivos de casos con anomalías

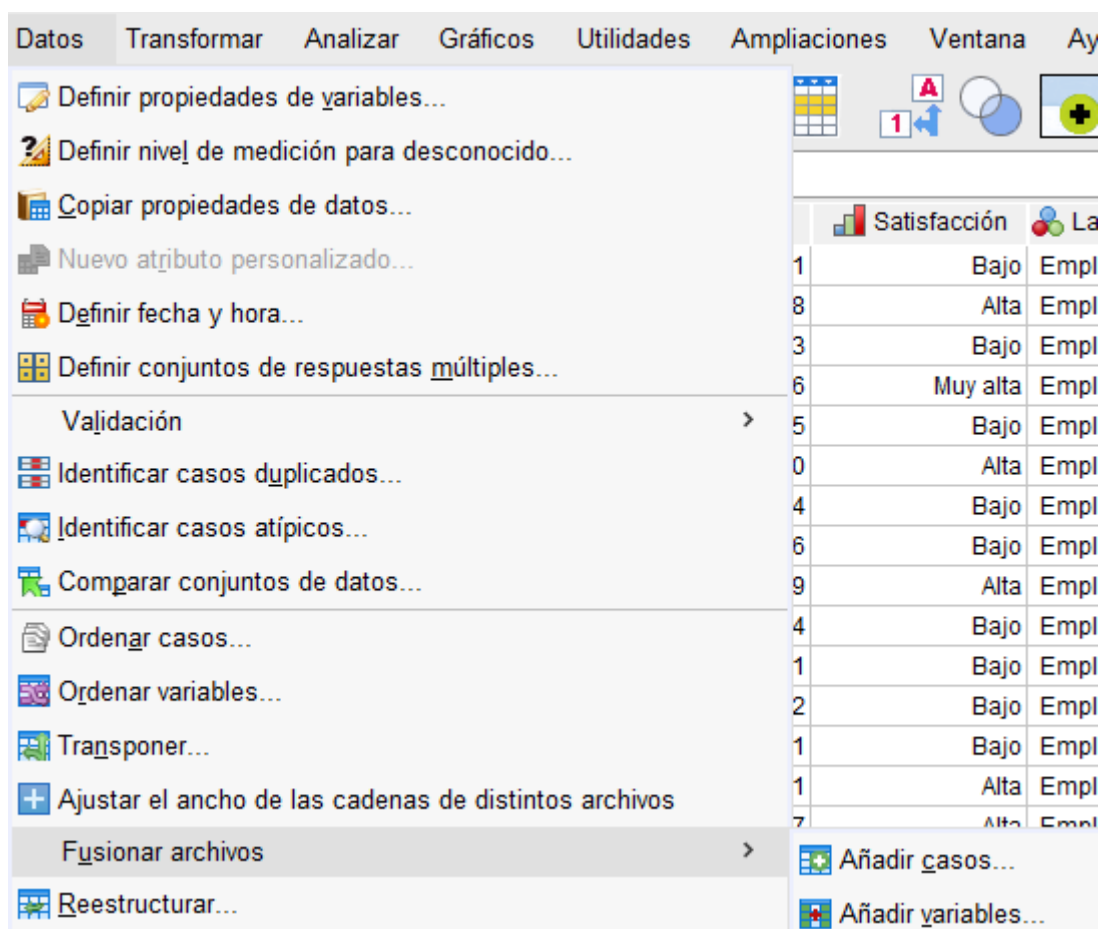
Razón: 1

Caso	Nombre	Variable de razón	Impacto de variable	Valor de variable	Norma de variable
28	Raistlin Majere	Salario	1,000	50000	50000,00

No indica que el salario del caso 28, es decir, el perteneciente a Raistlin Majere puede considerarse un valor atípico según el Índice de anomalía por defecto de SPSS.

- **Fusionar archivos:**

Esta herramienta tiene dos alternativas, ambas útiles para unir archivos de SPSS en un único archivo.



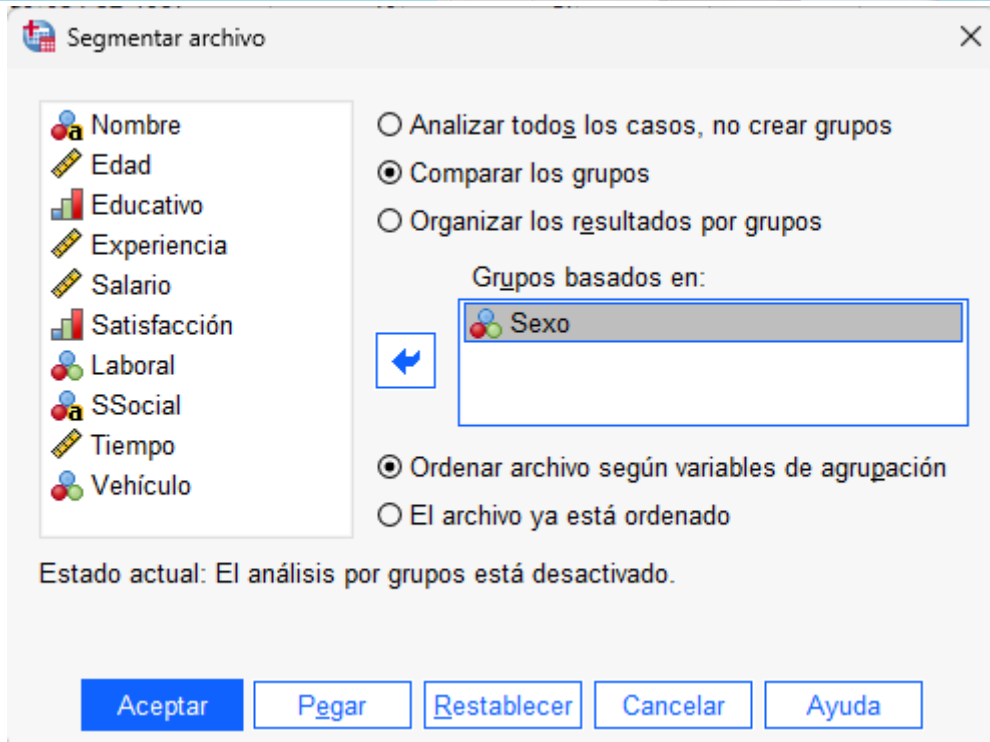
Si tuviésemos dos archivos, por ejemplo de dos centros de trabajo, y quisiéramos unir a los trabajadores de ambos centros en un mismo fichero, usaríamos **Añadir casos**. En cambio, si sobre los mismos trabajadores tuviésemos otra base de datos con información adicional sobre estos empleados acudiríamos a **Añadir variables**.

- **Segmentar archivos:**

Con esta herramienta lo que pretendemos es realizar un análisis separado siguiendo algún criterio.

	Nombre
1	Cheng Xin
2	Gohan
3	Hermione
4	Luffy
5	Mia Corvere
6	Naruto
7	Ichigo
8	Leia
9	Edward Elric
10	Katniss
11	Legolas
12	Malekith
13	Bruce Wayne
14	Peter Parker
15	Mei Ling
16	Cloud
17	Aloy
18	Kratos
19	Ellen Ripley
20	John Connor
21	Geralt of Rivia
22	Urza
23	Eren Jaeger
24	Shinji Ikari
25	Roberto Larcos
26	Lara Croft
27	Once
28	Raistlin Majere

Por ejemplo, si quisiéramos separar nuestro análisis por sexo:



A partir de que demos a aceptar nuestro análisis estará segmentado. En la siguiente captura se puede ver un análisis descriptivo simple con la segmentación aplicada.

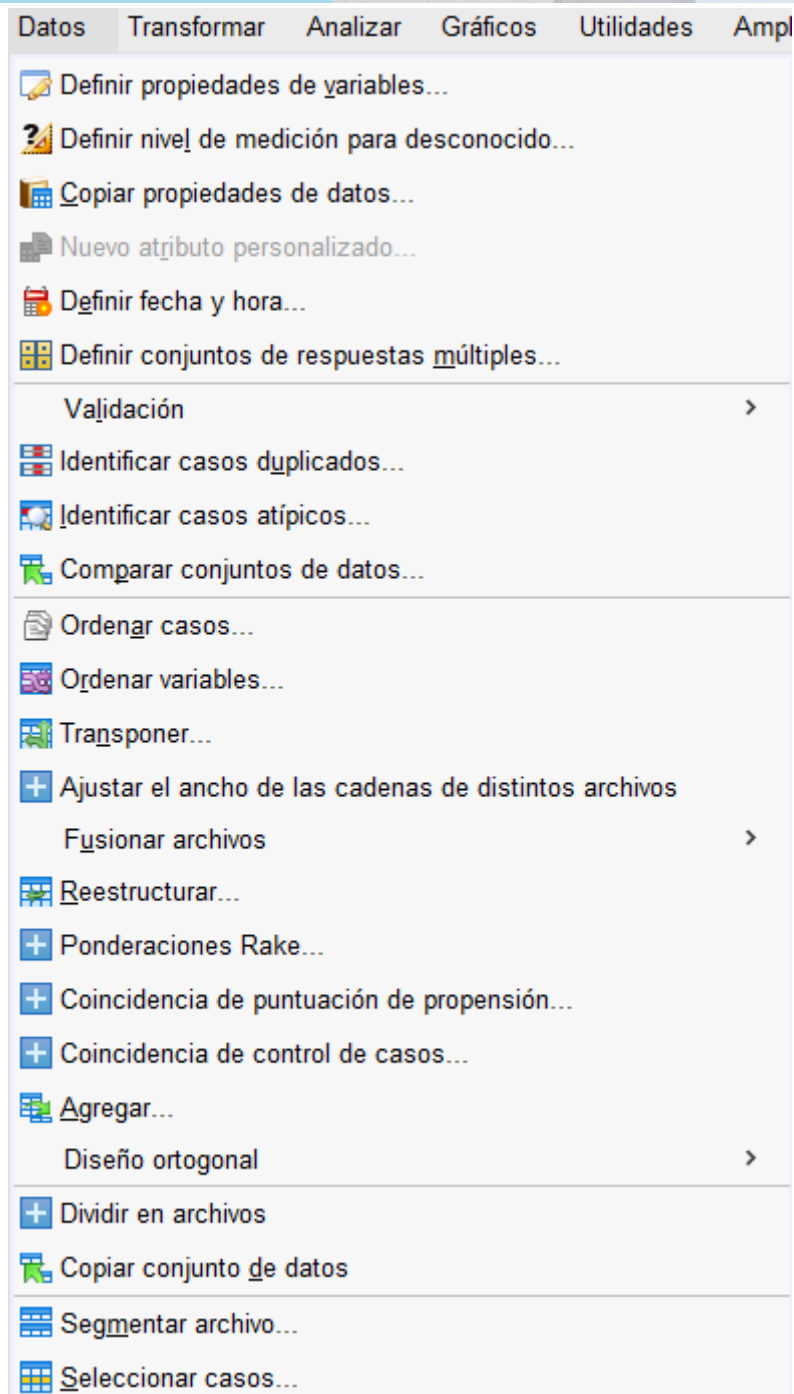
Estadísticos descriptivos

Sexo		N	Mínimo	Máximo	Media	Desv. estándar
Hombre	Salario	18	1	50000	2787,06	11782,799
	N válido (por lista)	18				
Mujer	Salario	9	7	20	14,11	3,822
	N válido (por lista)	9				

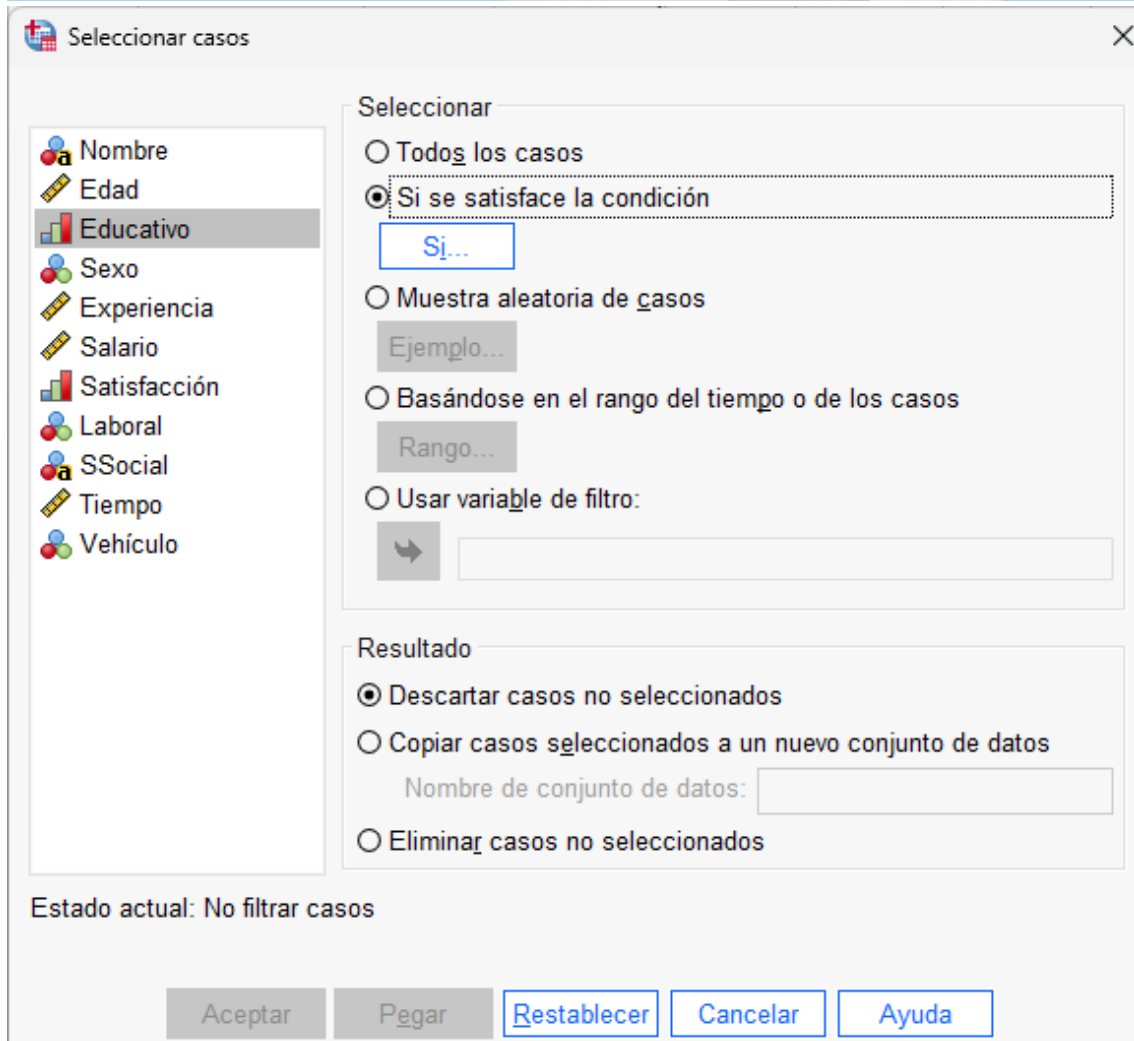
Para deshacer la segmentación y volver al análisis conjunto deberemos volver a **Analizar todos los casos, no crear grupos**.

- **Seleccionar casos:**

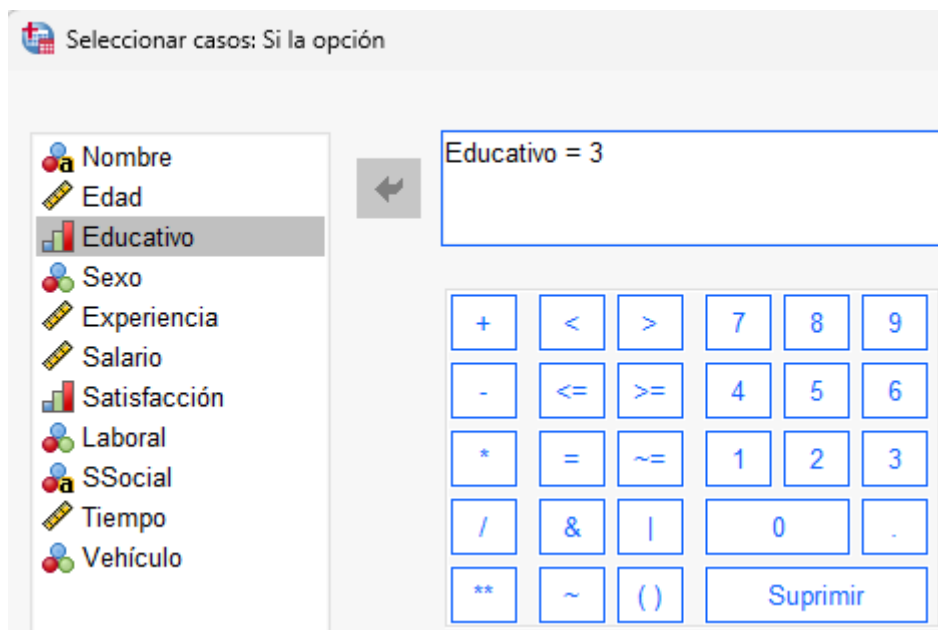
Esta herramienta sirve para filtrar los casos a analizar.



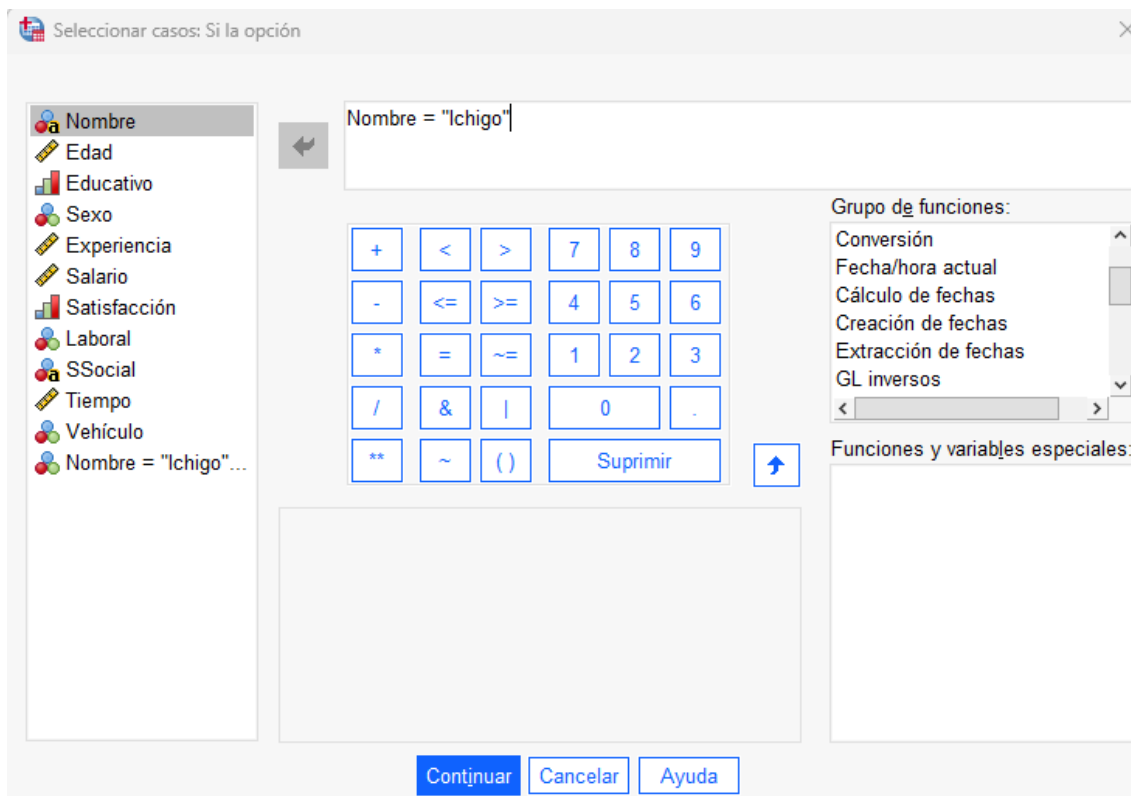
Por ejemplo, supongamos que queremos trabajar solo con aquellos trabajadores con un nivel educativo de Bachillerato.



Para que Aceptar aparezca como ejecutable tenemos que fijar la condición en Si. Como en mi lista de códigos para la variable Educación yo he establecido que el código para Bachillerato es 3. El condicional sería: Educativo = 3.



Si tuviésemos definidos las características como cadenas, introduciríamos el nombre entre comillas.



A partir de ese momento nuestra base de datos solo hará visible los casos que cumplan la condición requerida. En la siguiente captura se ve el filtrado de para el nivel educativo Bachillerato.

	Nombre	Edad	Educativo	Sexo	Experiencia	Salario
2	Luffy	19	Bachillerato	Hombre	11	6
3	Naruto	20	Bachillerato	Hombre	13	10
4	Ichigo	22	Bachillerato	Hombre	17	14
7	Malekith	5000	Bachillerato	Hombre	9	2
10	Cloud	24	Bachillerato	Hombre	13	12
11	Kratos	400	Bachillerato	Hombre	338	19
15	Eren Jaeger	21	Bachillerato	Hombre	12	9
16	Shinji Ikari	16	Bachillerato	Hombre	1	5
23	Katniss	26	Bachillerato	Mujer	3	14

Los casos que no cumple la condición no se han borrado, Para volver al análisis completo con todos los casos tenemos que deshacer el filtrado dentro de **Seleccionar Casos** recuperando la opción por defecto **Todos los casos**.

- **Ponderar casos**

Se utilizar para modificar en recuento de casos. Lo normal es que cada caso cuente como uno. Sin embargo, a veces es útil cambiar la ponderación de casos por ejemplo

para ahorrar tiempo en escribir casos recurrentes o que se repiten con mucha frecuencia.

Por ejemplo, se adjuntan a continuación los datos de plantilla del P.D.I. de la Universidad de Málaga a 31 de diciembre de 2022 extraídos de la [web](#) de la Universidad.

Plantilla de P.D.I. a 31/12/2022

FUNCIONARIOS DE CARRERA

CATEGORIA	MUJERES	HOMBRES	TOTAL
Catedrático de Universidad	74	275	349
Profesor Titular de Universidad	383	469	852
Catedrático de Escuela Universitaria	10	14	24
Profesor Titular de Escuela Universitaria	15	30	45

1.270

PROFESORADO CONTRATADO LABORAL INDEFINIDO

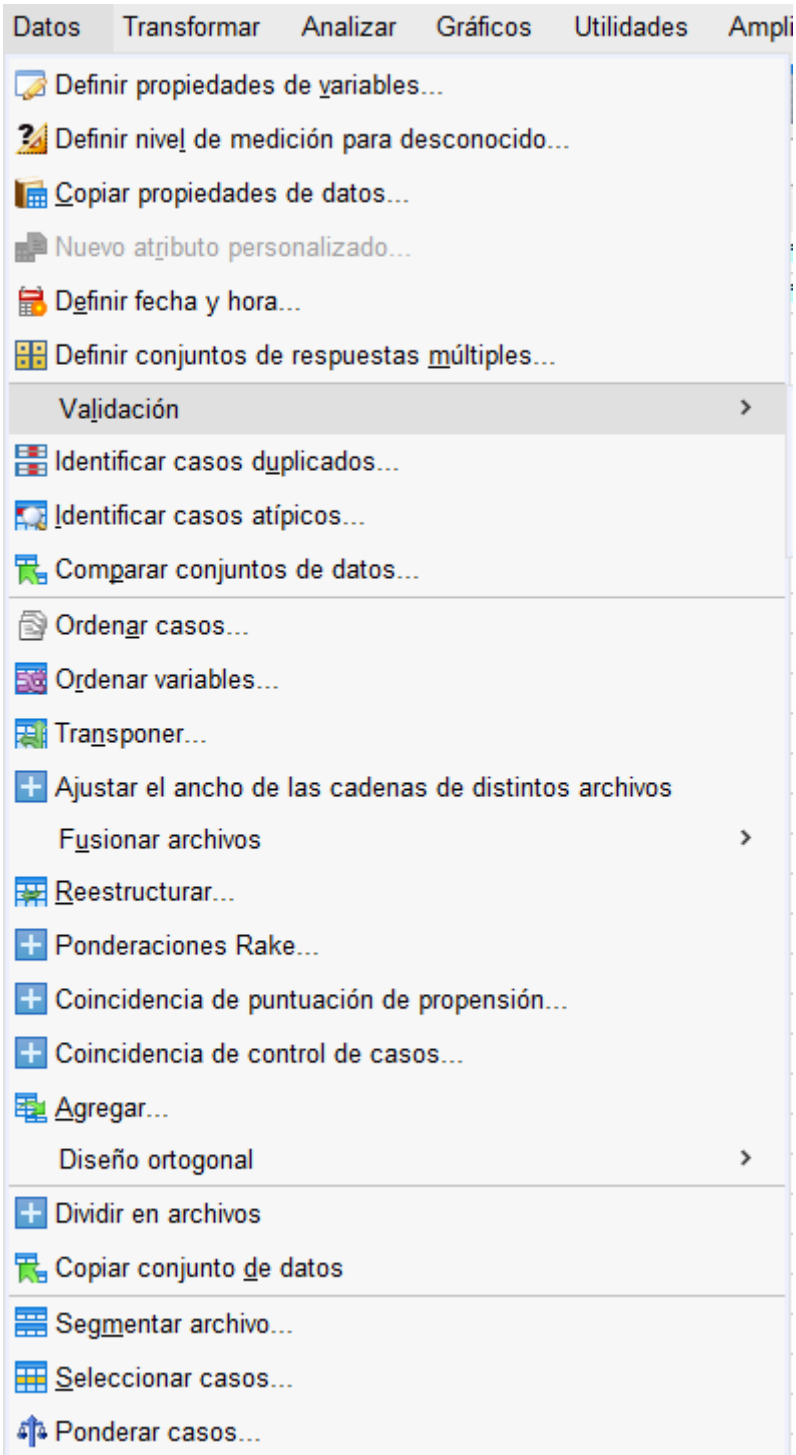
CATEGORIA	MUJERES	HOMBRES	TOTAL
Profesor Contratado Doctor	121	113	234
Profesor Colaborador	12	13	25

259

En lugar de tener que escribir un caso para cada uno de los docentes puede construir la base de datos como:

Categoría	Tipo	Sexo	Sueldo	Frecuencia
Profesor Colaborador	Profesorado Contratado Laboral Indefinido	Mujeres	31811,76	12
Profesor Contratado Doctor	Profesorado Contratado Laboral Indefinido	Mujeres	37425,60	121
Profesor Titular de Escuela Universitaria	Funcionarios de Carrera	Mujeres	32994,04	15
Catedrático de Escuela Universitaria	Funcionarios de Carrera	Mujeres	37425,60	10
Profesor Titular de Universidad	Funcionarios de Carrera	Mujeres	37425,60	383
Catedrático de Universidad	Funcionarios de Carrera	Mujeres	47114,30	74
Profesor Colaborador	Profesorado Contratado Laboral Indefinido	Hombres	31811,76	13
Profesor Contratado Doctor	Profesorado Contratado Laboral Indefinido	Hombres	37425,60	113
Profesor Titular de Escuela Universitaria	Funcionarios de Carrera	Hombres	32994,04	30
Catedrático de Escuela Universitaria	Funcionarios de Carrera	Hombres	37425,60	14
Profesor Titular de Universidad	Funcionarios de Carrera	Hombres	37425,60	469
Catedrático de Universidad	Funcionarios de Carrera	Hombres	47114,30	275

A partir de esa base de datos, puedo usar la herramienta de **Ponderar Casos** disponible al final de Datos para que el programa use la frecuencia de cara a contar cada caso como si los hubiese introducido de forma separada.



Ponderar casos

No ponderar casos
 Ponderar casos mediante

Variable de frecuencia:

Estado actual: Ponderar casos mediante Frecuencia

Una vez aceptada la ponderación, trabajaría con los casos ponderados. Aquí por ejemplo se muestra una salida de resultados para una tabla de frecuencias descriptivas para las categorías.

Frecuencias

Estadísticos

Categoría

N	Válido	1529
	Perdidos	0

Categoría

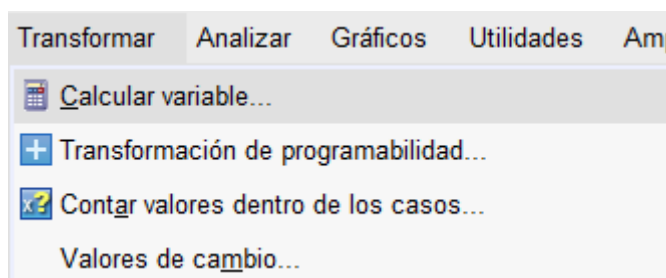
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Profesor Colaborador	25	1,6	1,6	1,6
	Profesor Contratado Doctor	234	15,3	15,3	16,9
	Profesor Titular de Escuela Universitaria	45	2,9	2,9	19,9
	Catedrático de Escuela Universitaria	24	1,6	1,6	21,5
	Profesor Titular de Universidad	852	55,7	55,7	77,2
	Catedrático de Universidad	349	22,8	22,8	100,0
	Total	1529	100,0	100,0	

Transformar

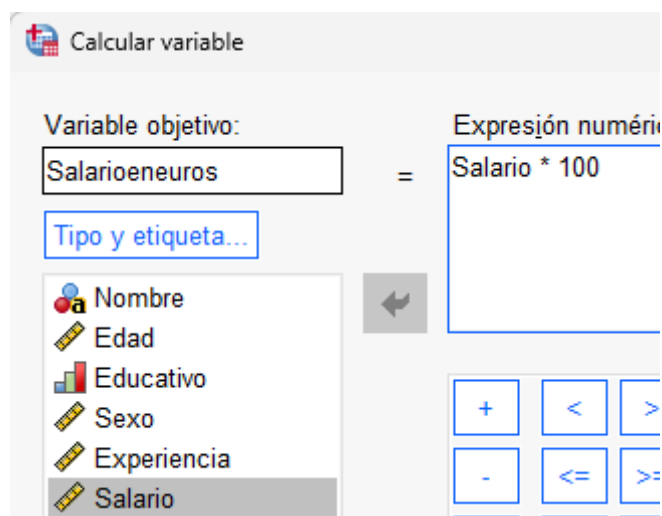
Esta pestaña de herramientas incorpora varias herramientas útiles entre las que cabe destacar: **Calcular variable** y **Recodificar variables**.

- **Calcular variable**

Esta herramienta se utiliza para crear variables nuevas a partir de un cálculo sobre las variables disponibles. Por ejemplo, en la base de datos Ficticias.sav el salario aparece en cientos de euros. Es decir, Gohan no cobra 8€ sino que cobra 800€. Esto se suele hacer por comodidad y simplicidad. Si quisiéramos crear una variable en euros de forma que aparezcan esos 800 podríamos hacerlo con calcular variable.



La ventana de la herramienta es la siguiente:



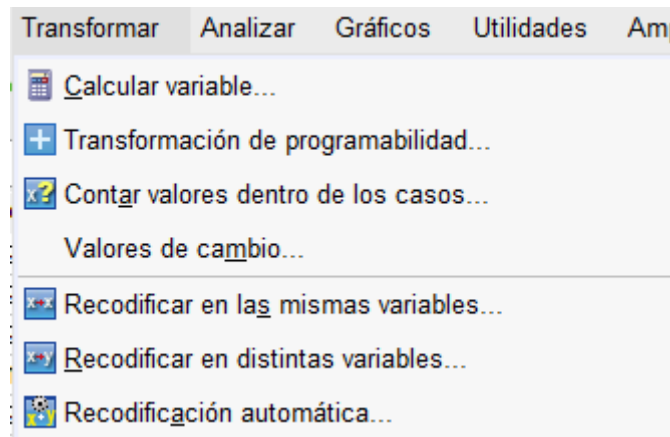
Una vez le dé a aceptar se me agregará la nueva variable calculada a la base de datos.

MÓDULO CUANTITATIVO

	Nombre	Edad	Educativo	Sexo	Experiencia	Salario	Satisfacción	Laboral	SSocial	Tiempo	Vehículo	Salarioeneuros
1	Cheng Xin	28	Grado	Mujer	3	11	Bajo	Empleado	123-45-6789	30	No	1100,00
2	Gohan	24	Grado	Hombre	12	8	Alta	Empleado	987-65-4321	20	Sí	800,00
3	Hermione	25	Posgrado	Mujer	14	13	Bajo	Empleado	456-78-9123	45	No	1300,00
4	Luffy	19	Bachillerato	Hombre	11	6	Muy alta	Empleado	321-98-7654	15	No	600,00
5	Mia Corvere	21	Posgrado	Mujer	16	15	Bajo	Empleado	789-12-3456	40	Sí	1500,00
6	Naruto	20	Bachillerato	Hombre	13	10	Alta	Empleado	234-56-7891	25	Sí	1000,00
7	Ichigo	22	Bachillerato	Hombre	17	14	Bajo	Empleado	567-89-1234	30	No	1400,00
8	Leia	30	Grado	Mujer	5	16	Bajo	Empleado	876-54-3219	50	Sí	1600,00
9	Edward Elric	23	Grado	Hombre	12	9	Alta	Empleado	210-98-7654	35	No	900,00
10	Katniss	26	Bachillerato	Mujer	3	14	Bajo	Empleado	654-32-1987	40	Sí	1400,00
11	Legolas	2931	Grado	Hombre	7	1	Bajo	Empleado	987-65-4320	60	Sí	100,00
12	Malekith	5000	Bachillerato	Hombre	9	2	Bajo	Empleado	123-45-6788	45	Sí	200,00
13	Bruce Wayne	35	Posgrado	Hombre	6	1	Bajo	Empleado	246-80-1357	30	Sí	100,00
14	Peter Parker	23	Posgrado	Hombre	12	11	Alta	Empleado	579-31-4680	25	No	1100,00
15	Mei Ling	27	Grado	Mujer	11	7	Alta	Empleado	801-24-6813	20	Sí	700,00
16	Cloud	24	Bachillerato	Hombre	13	12	Bajo	Empleado	135-79-0246	35	Sí	1200,00
17	Aloy	19	Posgrado	Mujer	16	13	Bajo	Empleado	468-13-5790	45	Sí	1300,00
18	Kratos	400	Bachillerato	Hombre	338	19	Alta	Empleado	246-80-1358	40	Sí	1900,00
19	Ellen Ripley	35	Posgrado	Mujer	8	18	Bajo	Empleado	468-13-5791	50	Sí	1800,00
20	John Connor	30	Posgrado	Hombre	6	20	Bajo	Empleado	791-35-4680	40	Sí	2000,00
21	Geralt of Rivia	40	Posgrado	Hombre	15	17	Bajo	Jubilado	246-80-1359	.	.	1700,00
22	Urza	5000	Posgrado	Hombre	4800	4	Bajo	Empleado	579-31-4681	60	Sí	400,00
23	Eren Jaeger	21	Bachillerato	Hombre	12	9	Media	Empleado	791-35-4681	30	Sí	900,00
24	Shinji Ikari	16	Bachillerato	Hombre	1	5	Media	Desempl...	111-22-3333	.	.	500,00
25	Roberto Larcos	32	Posgrado	Hombre	10	19	Bajo	Empleado	888-99-7777	.	.	1900,00
26	Lara Croft	31	Posgrado	Mujer	6	20	Bajo	Empleado	777-88-9999	30	.	2000,00
27	Once	13	Primaria	Mujer	2	.	Media	Estudiante	000-00-0000	0	.	.
28	Raistlin Majere	44	Posgrado	Hombre	3	50000	Media	Empleado	777-77-7777	30	Sí	5000000,00

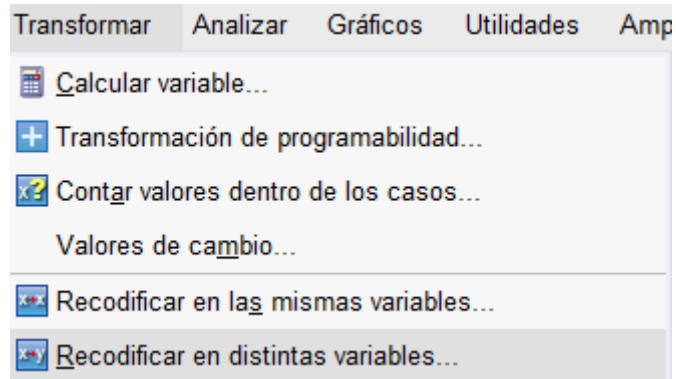
- Recodificar variables

Esta herramienta sirve para reformular los códigos para una variable o establecer códigos distintos. Tiene tres versiones: **Recodificar en las mismas variables** (no la recomiendo salvo que te guste el riesgo), **Recodificar en distintas variables** (es lo mismo que la anterior pero no destruye los datos originales) y **Recodificación automática** útil para convertir variables cadena en códigos numéricos.

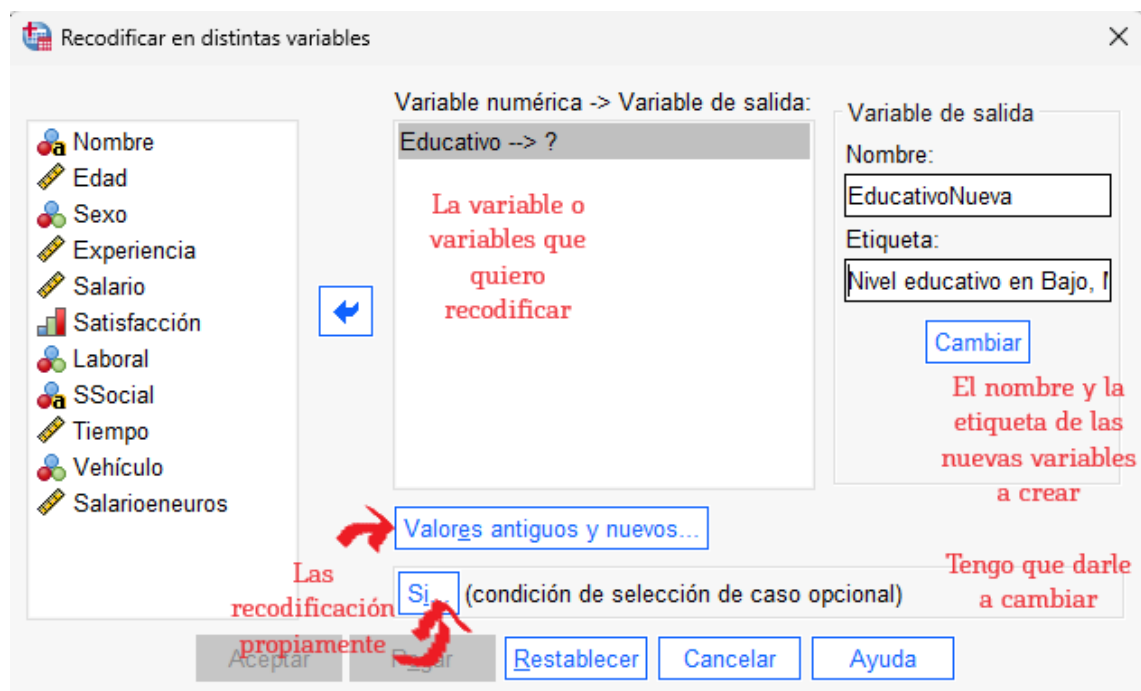


- Recodificar en distintas variables

Supongamos que quiero crear una nueva codificación para la Educación. Bajo, Medio, Alto.



La ventana de esta herramienta quedaría así:



Accedo a **Valores Antiguos y Nuevos** e introduzco siguiente recodificación.

- Sin estudios y Primaria pasa a ser Nivel Bajo
- Secundaria y Bachillerato pasa a ser Nivel Medio
- Grado y Posgrado pasa a ser Nivel Alto

Existe varias formas de introducir la recodificación, por ejemplo:

Recodificar en variables diferentes: valores antiguo y nuevo

Valor antiguo

Valor:

Perdido del sistema

Perdido por el sistema o el usuario

Rango:

hasta

Rango, LOWEST hasta el valor:

Rango, valor hasta HIGHEST:

Valor nuevo

Valor:

Perdido del sistema

Copiar valores antiguos

Antiguo --> Nuevo:

0 thru 1 --> 1

2 thru 3 --> 2

4 thru 5 --> 3

Añadir

Cambiar

Eliminar

Recuerda añadir las etiquetas de valor a la nueva variable en la vista de variables.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Nombre	Cadena	16	0		Ninguna	Ninguna	16	Izquierda	Nominal	Entrada
2	Edad	Númérico	8	0		Ninguna	Ninguna	8	Derecha	Escala	Entrada
3	Educativo	Númérico	8	0		{0, Sin estu...	Ninguna	12	Derecha	Ordinal	Entrada
4	Sexo	Númérico	8	0		{1, Hombre}...	Ninguna	8	Derecha	Nominal	Entrada
5	Experiencia	Númérico	8	0		Ninguna	Ninguna	12	Derecha	Escala	Entrada
6	Salario	Númérico	8	0		Ninguna	Ninguna	10	Derecha	Escala	Entrada
7	Satisfacción	Númérico	8	0		{1, Bajo}...	Ninguna	14	Derecha	Ordinal	Entrada
8	Laboral	Númérico	8	0		{1, Desemp...	Ninguna	8	Derecha	Nominal	Entrada
9	SSocial	Cadena	15	0		Ninguna	Ninguna	15	Izquierda	Nominal	Entrada
10	Tiempo	Númérico	8	0		Ninguna	Ninguna	8	Derecha	Escala	Entrada
11	Vehículo	Númérico	8	0		{1, No}...	Ninguna	11	Derecha	Nominal	Entrada
12	Salarioene...	Númérico	8	2		Ninguna	Ninguna	16	Derecha	Escala	Entrada
13	EducativoN...	Númérico	8	2	Nivel educativo...	Ninguna	Ninguna	16	Derecha	Etiquetas de valor	
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											

Etiquetas de valor

Ortografía...

Etiquetas de valor:

Valor	Etiqueta
1,00	Bajo
2,00	Medio
3,00	Alto

Aceptar Restablecer Cancelar Ayuda

Y el resultado quedaría como:

	Vehículo	Salarioeneuros	EducativoNueva
	No	1100,00	Alto
	Sí	800,00	Alto
	No	1300,00	Alto
	No	600,00	Medio
	Sí	1500,00	Alto
	Sí	1000,00	Medio
	No	1400,00	Medio
	Sí	1600,00	Alto
	No	900,00	Alto
	Sí	1400,00	Medio
	Sí	100,00	Alto
	Sí	200,00	Medio
	Sí	100,00	Alto
	No	1100,00	Alto
	Sí	700,00	Alto
	Sí	1200,00	Medio
	Sí	1300,00	Alto
	Sí	1900,00	Medio
	Sí	1800,00	Alto
	Sí	2000,00	Alto
	.	1700,00	Alto
	Sí	400,00	Alto
	Sí	900,00	Medio
	.	500,00	Medio
	.	1900,00	Alto
	.	2000,00	Alto
	.	.	Bajo
	Sí	5000000,00	Alto

- Recodificación automática

Algunas herramientas de SPSS no reconocen como variables de clasificación las variables cadena. Por eso, es habitual usar códigos numéricos para definir estas variables. Esto se suma a que al establecer códigos numéricos nos ahorramos escribir toda la cadena y reducimos los errores.

Si escribimos rápido muchos códigos de cadena pueden aparecer valores como:

Hombre

Homrbe

Hombe

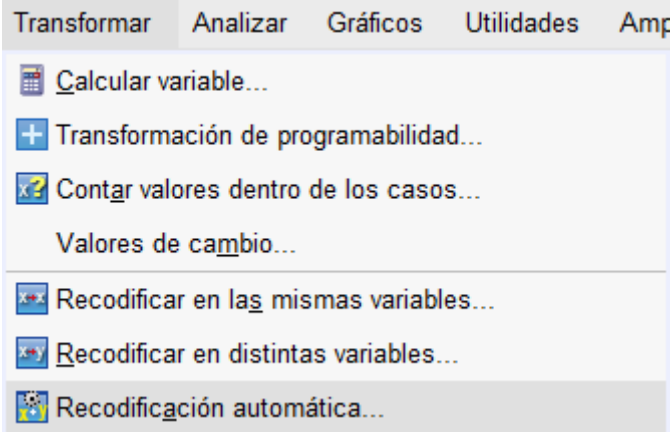
Hobmre

Que el programa no puede entender como iguales. Esos errores se reducen cuando establecemos una codificación.

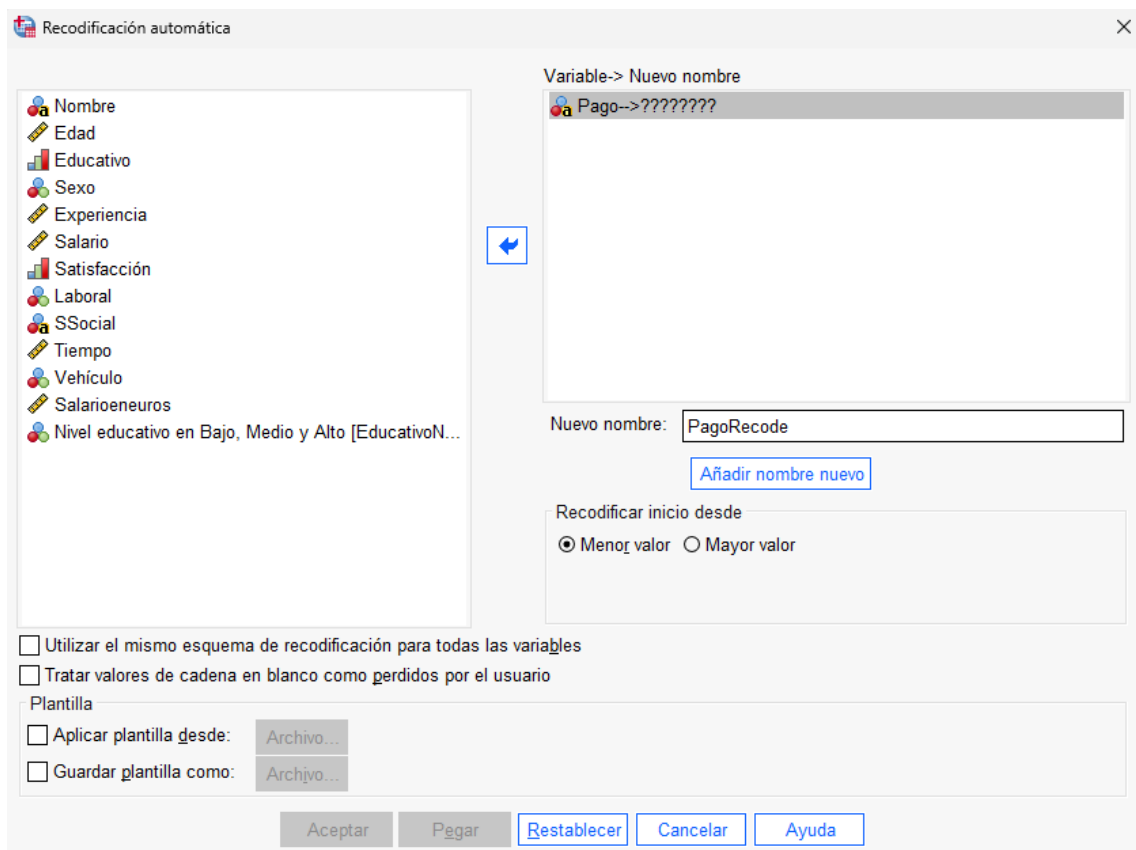
Sin embargo, a menudo creamos variables como cadenas o importamos datos que ya están en formato de cadena, lo cual es bastante común. En estos casos, la recodificación automática puede resultar útil. Usando el ejemplo de la Base de datos Ficticia.sav voy a crear manualmente una nueva variable que representa la forma de pago del trabajador.

Tiempo	Vehículo	Salarioeneuros	EducativoNueva	Pago
30	No	1100,00	Alto	Domiciliación
20	Sí	800,00	Alto	Domiciliación
45	No	1300,00	Alto	Cheque
15	No	600,00	Medio	Domiciliación
40	Sí	1500,00	Alto	Cheque
25	Sí	1000,00	Medio	Domiciliación
30	No	1400,00	Medio	Cheque
50	Sí	1600,00	Alto	Cheque
35	No	900,00	Alto	Cheque
40	Sí	1400,00	Medio	Cheque
60	Sí	100,00	Alto	Cheque
45	Sí	200,00	Medio	Efectivo
30	Sí	100,00	Alto	Efectivo
25	No	1100,00	Alto	Domiciliación
20	Sí	700,00	Alto	Domiciliación
35	Sí	1200,00	Medio	Efectivo
45	Sí	1300,00	Alto	Domiciliación
40	Sí	1900,00	Medio	Cheque
50	Sí	1800,00	Alto	Cheque
40	Sí	2000,00	Alto	Domiciliación
.	.	1700,00	Alto	Cheque
60	Sí	400,00	Alto	Domiciliación
30	Sí	900,00	Medio	Cheque
.	.	500,00	Medio	Domiciliación
.	.	1900,00	Alto	Cheque
30	.	2000,00	Alto	Efectivo
0	.	.	Bajo	Cheque
30	Sí	5000000,00	Alto	Domiciliación

Ahora voy a usar la **Recodificación automática** para que el programa me genere la tabla de códigos.



La ventana de la herramienta se vería así:



Tengo que fijar un nombre para la nueva variable y clicar **Añadir nombre nuevo**. Una vez hecho, los signos de interrogación cambiarán por el nombre de la variable elegido y me habilitará el botón **Aceptar**.

En la vista de variables podéis ver la nueva variable con los códigos automáticos elegidos por orden alfabético.

15	PagoRecode	Numérico	1	0	{1, Cheque}...	Ninguna
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						

Etiquetas de valor ✕

[Ortografía...](#)

Etiquetas de valor:

Valor ▾	Etiqueta
1	Cheque
2	Domiciliación
3	Efectivo

+
✕

Aceptar
Restablecer
Cancelar
Ayuda

En la vista de datos tendrás la variable con la clasificación codificada numéricamente.

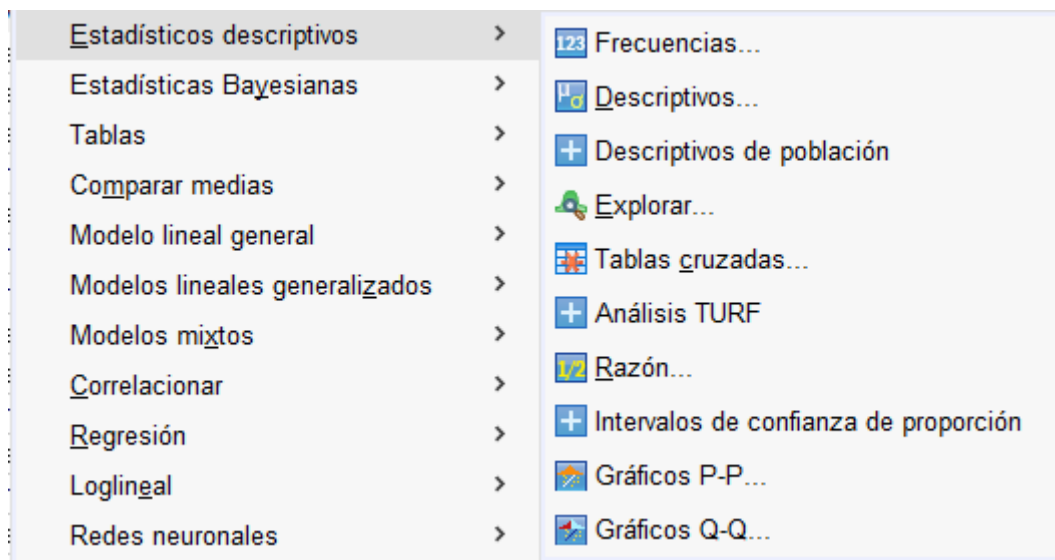
ilo	Salarioeneuros	EducativoNueva	Pago	PagoRecode
No	1100,00	Alto	Domiciliación	Domiciliación
Sí	800,00	Alto	Domiciliación	Domiciliación
No	1300,00	Alto	Cheque	Cheque
No	600,00	Medio	Domiciliación	Domiciliación
Sí	1500,00	Alto	Cheque	Cheque
Sí	1000,00	Medio	Domiciliación	Domiciliación
No	1400,00	Medio	Cheque	Cheque
Sí	1600,00	Alto	Cheque	Cheque
No	900,00	Alto	Cheque	Cheque
Sí	1400,00	Medio	Cheque	Cheque
Sí	100,00	Alto	Cheque	Cheque
Sí	200,00	Medio	Efectivo	Efectivo
Sí	100,00	Alto	Efectivo	Efectivo
No	1100,00	Alto	Domiciliación	Domiciliación
Sí	700,00	Alto	Domiciliación	Domiciliación
Sí	1200,00	Medio	Efectivo	Efectivo
Sí	1300,00	Alto	Domiciliación	Domiciliación
Sí	1900,00	Medio	Cheque	Cheque
Sí	1800,00	Alto	Cheque	Cheque
Sí	2000,00	Alto	Domiciliación	Domiciliación
.	1700,00	Alto	Cheque	Cheque
Sí	400,00	Alto	Domiciliación	Domiciliación
Sí	900,00	Medio	Cheque	Cheque
.	500,00	Medio	Domiciliación	Domiciliación
.	1900,00	Alto	Cheque	Cheque
.	2000,00	Alto	Efectivo	Efectivo
.	.	Bajo	Cheque	Cheque
Sí	5000000,00	Alto	Domiciliación	Domiciliación

Analizar

En esta pestaña es donde se encuentra realmente las herramientas del análisis de datos. Las principales agrupaciones se encuentran en **Estadísticos Descriptivos, Comparar Medias, Correlacionar y Regresión.**

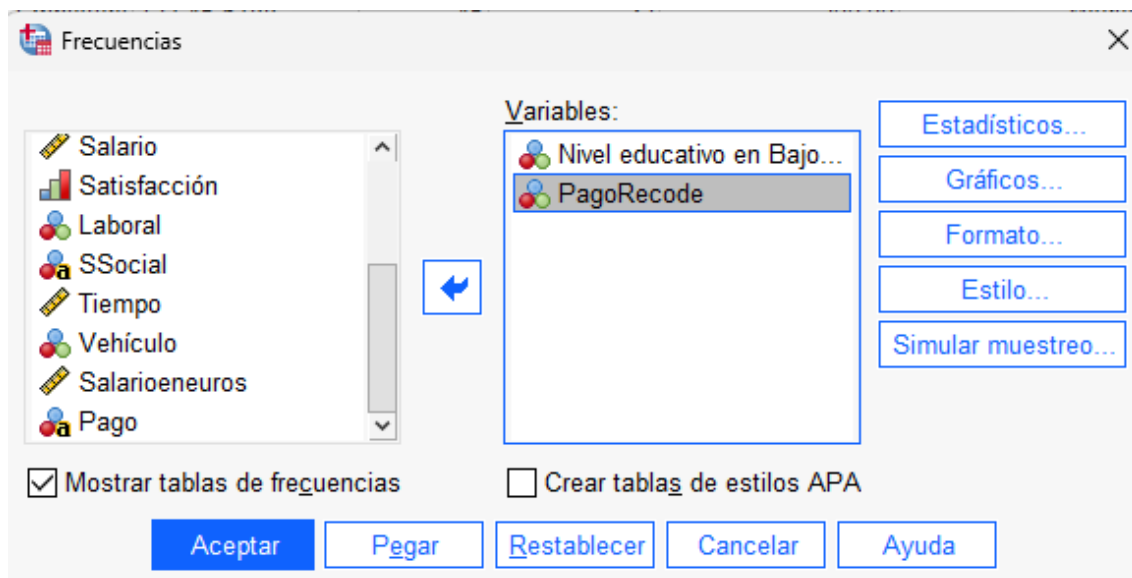
Estadísticos Descriptivos

En esta pestaña usaremos principalmente: **Frecuencias, Descriptivos, Explorar y Tablas Cruzadas.**



- **Frecuencias**

Sirve para generar una tabla de frecuencias con el recuento de casos que cumplen una determinada condición. Es útil para describir variables cualitativas: Nominales u Ordinales. Por ejemplo, usando la base de datos Ficticiasav voy a generar la tabla de frecuencias para Domiciliación y EducativoNuevo. La ventaja de la herramienta Frecuencias debería verse así.



Y el resultado sería:

Frecuencias

		Estadísticos	
		Nivel educativo en Bajo, Medio y Alto	
N			PagoRecod e
	Válido	28	28
	Perdidos	0	0

Tabla de frecuencia

Nivel educativo en Bajo, Medio y Alto

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Bajo	1	3,6	3,6	3,6
	Medio	9	32,1	32,1	35,7
	Alto	18	64,3	64,3	100,0
	Total	28	100,0	100,0	

PagoRecode

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Cheque	13	46,4	46,4	46,4
	Domiciliación	11	39,3	39,3	85,7
	Efectivo	4	14,3	14,3	100,0
	Total	28	100,0	100,0	

Si dentro de la herramienta accedemos a Gráficos podremos generar directamente gráficos muy útiles.

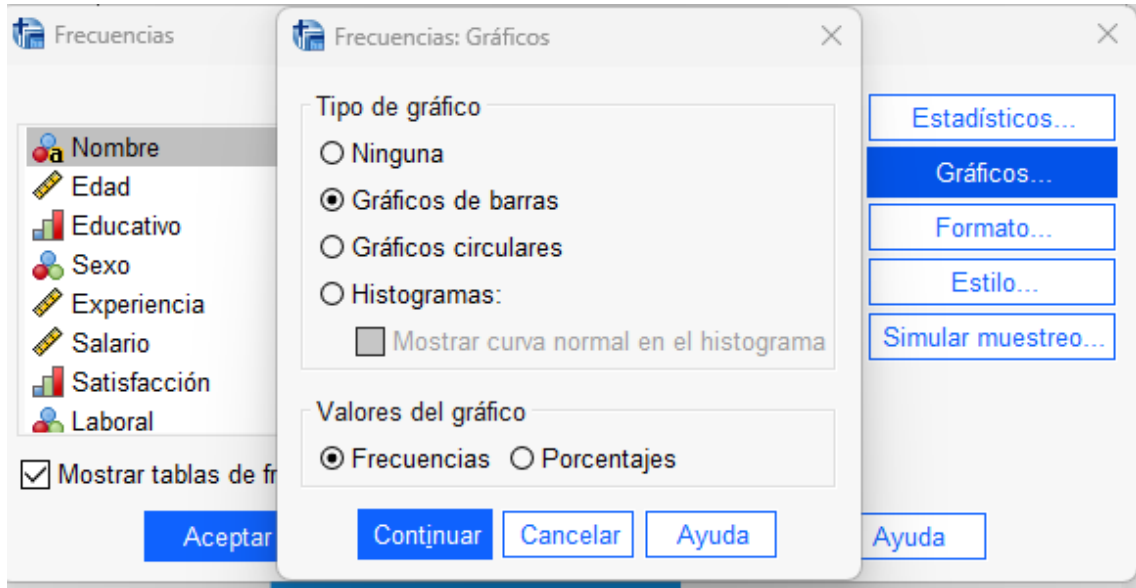
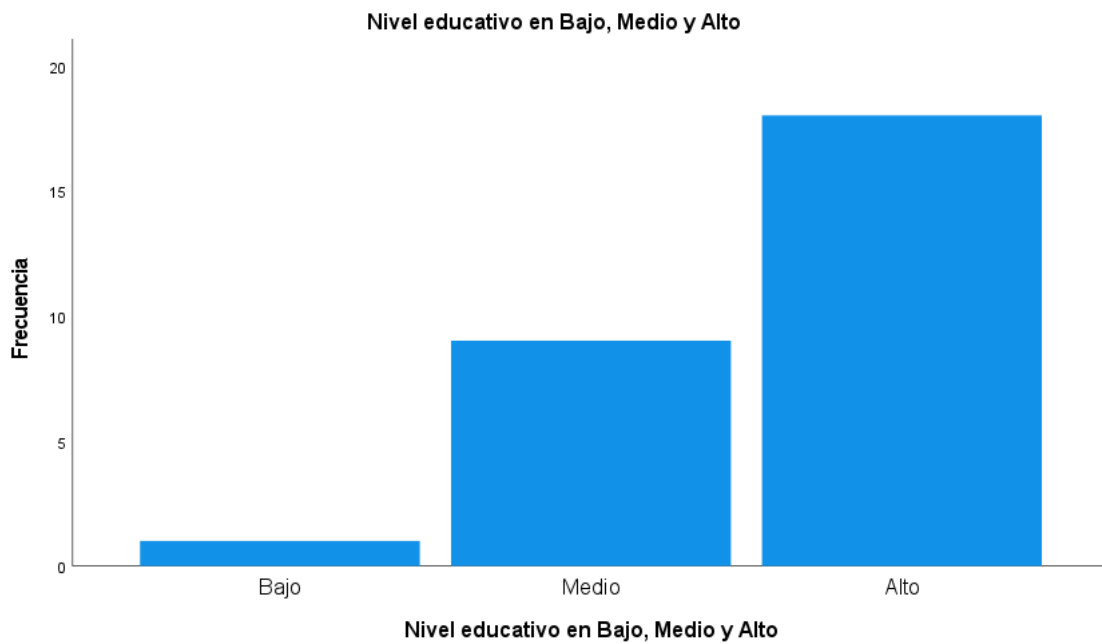
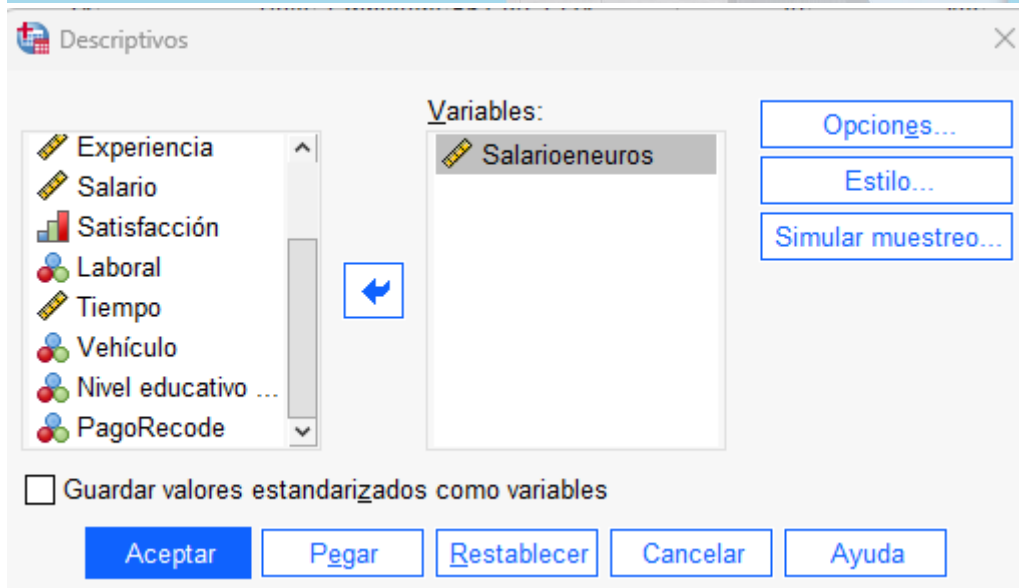


Gráfico de barras

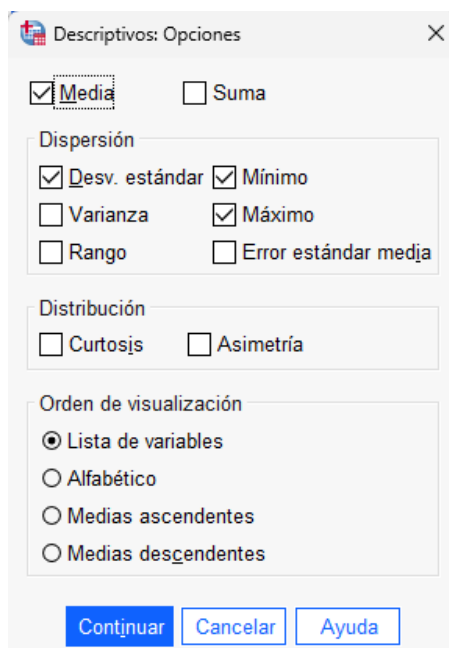


- **Descriptivos**

Sirve fundamentalmente para proporcionar los principales estadísticos descriptivos por lo que se usa fundamentalmente para variables cuantitativas. Por ejemplo, para la variable Salarioseneuros creada anteriormente.



Si clico en Opciones puede editar los estadísticos a incluir entre un conjunto de ellos.



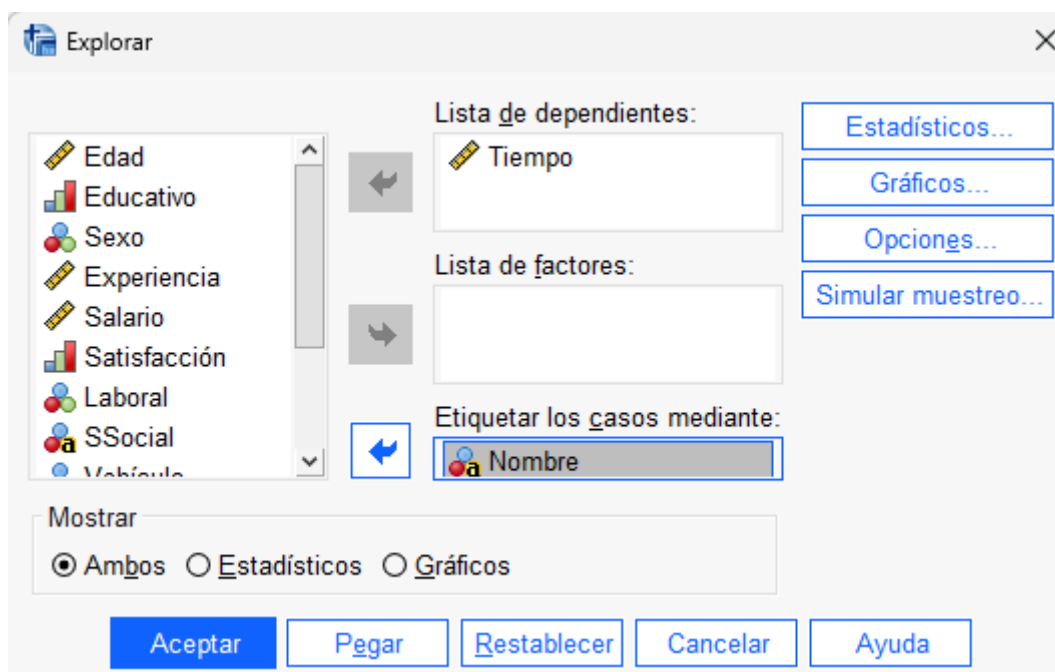
El resultado del análisis por defecto sería:

Estadísticos descriptivos

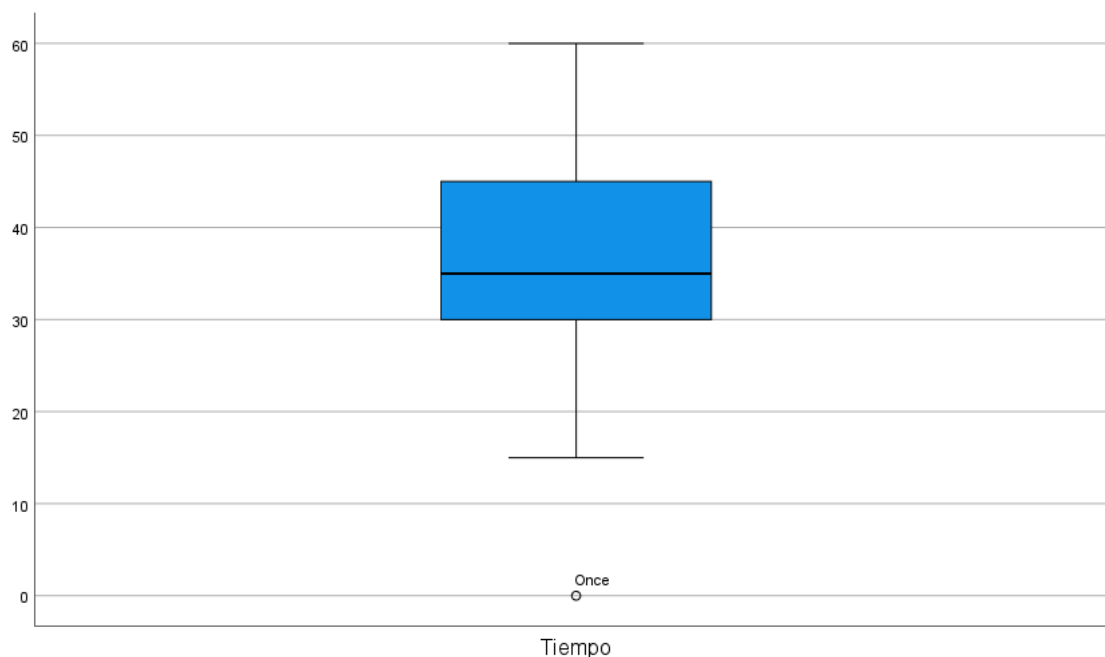
	N	Mínimo	Máximo	Media	Desv. estándar
Salarioeneuros	27	100,00	5000000,00	186274,0741	962033,00503
N válido (por lista)	27				

- Explorar

Explorar es una herramienta muy útil que proporciona información muy diversa sobre las variables. Una de sus utilidades más interesante es generar el diagrama de cajas y bigotes para detectar atípicos de una forma rápida y ágil². La interfaz de la herramienta se vería así:



El gráfico para detectar atípicos más usado sería:



² Hay una opción dentro de la subventana Estadísticos de herramienta Explorar llamada Valores atípicos que no proporciona realmente los atípicos si no que da los 5 valores mayores y los 5 menores.

Según este gráfico el valor de Once en la variable Tiempo podría ser considerado un valor atípico.

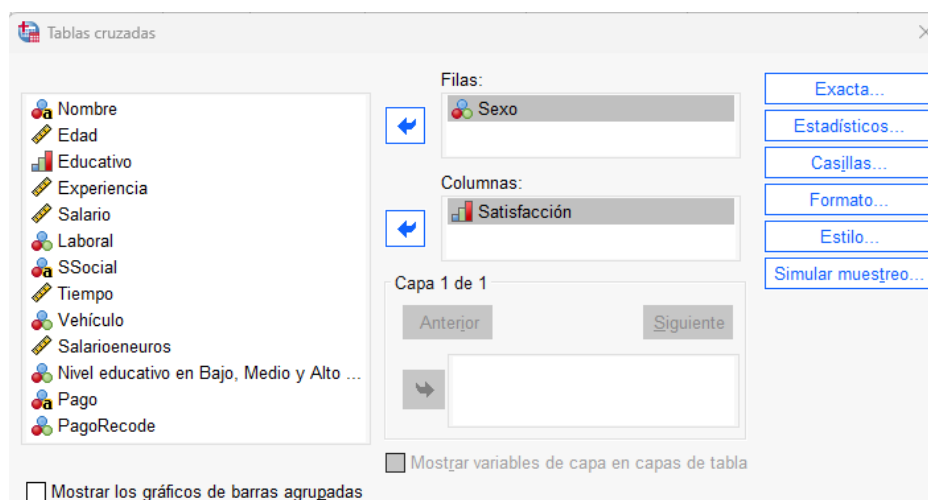
Como posiblemente hayas podido comprobar la herramienta **Explorar** sirve también para proporcionar los principales estadísticos descriptivos de la variable.

Descriptivos

		Estadístico	Error estándar	
Tiempo	Media	34,80	2,746	
	95% de intervalo de confianza para la media	Límite inferior	29,13	
		Límite superior	40,47	
	Media recortada al 5%	35,17		
	Mediana	35,00		
	Varianza	188,500		
	Desv. estándar	13,730		
	Mínimo	0		
	Máximo	60		
	Rango	60		
	Rango intercuartil	18		
	Asimetría	-,283	,464	
	Curtosis	,632	,902	

- Tablas Cruzadas.

Es la herramienta empleada para generar **tablas de doble entrada** y generar los correspondientes estadísticos de asociación **Chi cuadrado**, **Tau-b** y **Tau-c** de **Kendall** entre otros. La ventana de la herramienta sería:



Voy a generar una tabla de doble entrada con las variables Sexo y Laboral. Además, en la sección de Estadísticos voy a marcar la prueba Chi Cuadrado para comprobar si existe independencia entre estas dos variables.

El resultado de este análisis sería:

Tablas cruzadas

Resumen de procesamiento de casos

	Válido		Casos Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Sexo * Satisfacción	28	100,0%	0	0,0%	28	100,0%

Tabla cruzada Sexo*Satisfacción

Recuento

		Satisfacción				Total
		Bajo	Media	Alta	Muy alta	
Sexo	Hombre	9	3	5	1	18
	Mujer	8	1	1	0	10
Total		17	4	6	1	28

Pruebas de chi-cuadrado

	Valor	gl	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	2,657 ^a	3	,448
Razón de verosimilitud	3,085	3	,379
Asociación lineal por lineal	2,497	1	,114
N de casos válidos	28		

a. 6 casillas (75,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es ,36.

Como vemos, no se cumplen los supuestos básicos necesarios de la prueba. Si asumiésemos que sí fuesen ciertos comprobaríamos la Significación asintótica (bilateral) para la prueba Chi-cuadrado. Como es mayor que 0.05 asumiríamos que las variables son independientes.

Comparar Medias

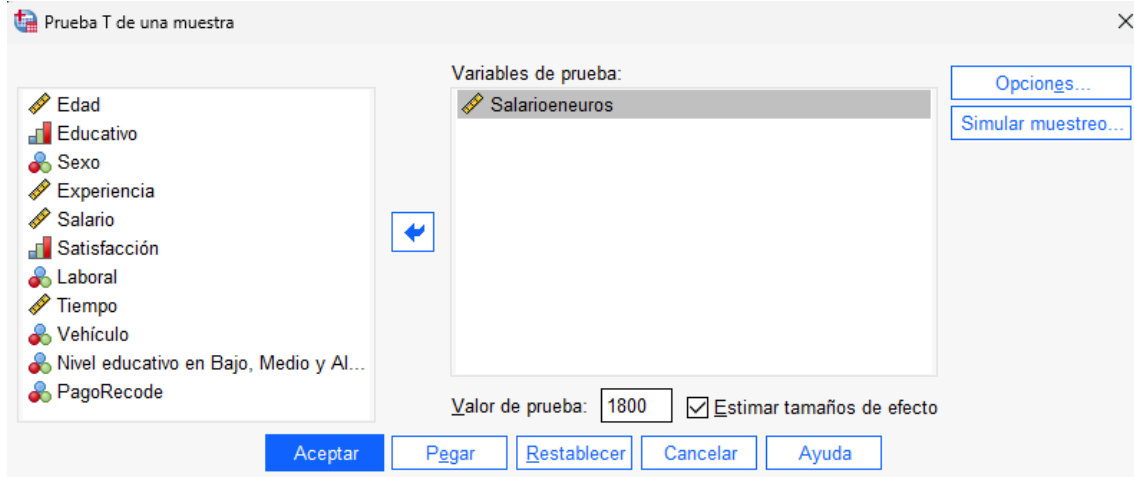
En esta pestaña podemos encontrar las Pruebas T de una muestra, Prueba T de muestras independientes, Pruebas T de muestras emparejadas y ANOVA de un factor, entre otras.

Comparar medias	>	M Medias...
Modelo lineal general	>	t Prueba T de una muestra...
Modelos lineales generalizados	>	i Prueba I de muestras independientes...
Modelos mixtos	>	+ Prueba T de muestras independientes de resumen
Correlacionar	>	t Prueba T de muestras emparejadas...
Regresión	>	F ANOVA de un factor...
Loglineal	>	Z Proporciones de una muestra...
Redes neuronales	>	Z Proporciones de muestras independientes...
Clasificar	>	Z Proporciones de muestras emparejadas...
Reducción de dimensiones	>	

- Pruebas T de una muestra

Sirven para comprobar si la afirmación sobre un determinado valor para la media poblacional tiene validez en base a una muestra.

Por ejemplo, supongamos que los datos disponibles corresponden solo a una muestra de trabajadores y queremos contrastar si la media del salario de todos los trabajadores puede ser 1800€. La ventana sería:



Y los resultados serían:

Prueba T

Estadísticas para una muestra

	N	Media	Desv. estándar	Media de error estándar
Salarioeneuros	27	186274,0741	962033,00503	185143,33814

Prueba para una muestra

Valor de prueba = 1800

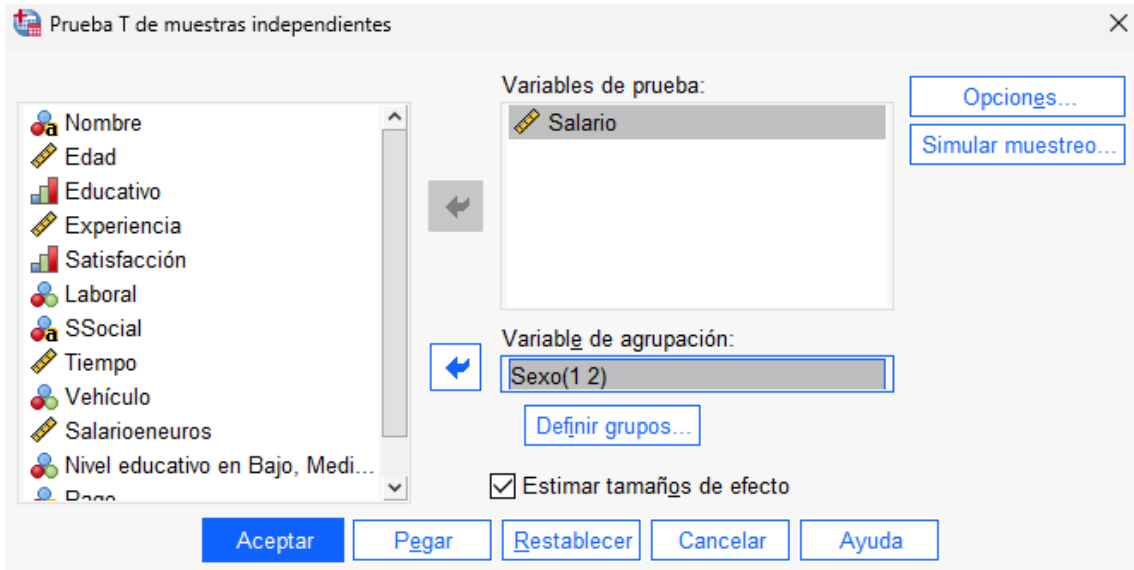
	t	gl	Significación		Diferencia de medias	95% de intervalo de confianza de la diferencia	
			P de un factor	P de dos factores		Inferior	Superior
Salarioeneuros	,996	26	,164	,328	184474,07407	-196093,5078	565041,6560

Como la Significación P de dos factores 0.328 es superior al 0.05 podemos deducir que la afirmación de que el sueldo medio de los empleados es de 1800€ no puede ser refutado y, por lo tanto, tenemos que asumirlo como cierto.

- Prueba T de muestras independientes.

Esta herramienta sirve para comprobar si podemos decidir, a partir de muestras, que dos grupos independientes tienen medias poblacionales iguales.

Por ejemplo, supongamos que los datos disponibles corresponden solo a una muestra de trabajadores y queremos contrastar si la media del salario de hombres y mujeres es igual. La ventana de la herramienta quedaría definida como:



Y los resultados serían:

Prueba T

Estadísticas de grupo

	Sexo	N	Media	Desv. estándar	Media de error estándar
Salario	Hombre	18	2787,06	11782,799	2777,232
	Mujer	9	14,11	3,822	1,274

La siguiente tabla es la que nos interesa para responder y está dividida en dos partes.

		Prueba de Levene de igualdad de varianzas		prueba t para la igualdad de medias							
		F	Sig.	t	gl	Significación		Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
						P de un factor	P de dos factores			Inferior	Superior
Salario	Se asumen varianzas iguales	2,211	,150	,699	25	,245	,491	2772,944	3966,681	-5396,589	10942,478
	No se asumen varianzas iguales			,998	17,000	,166	,332	2772,944	2777,233	-3086,504	8632,393

La primera parte es una hipótesis previa que nos dice que fila debemos leer. Como Sig es mayor a 0,05 asumimos varianzas iguales y seguimos leyendo la primera fila.

		Prueba de Levene de igualdad de varianzas	
		F	Sig.
Salario	Se asumen varianzas iguales	2,211	,150
	No se asumen varianzas iguales		

ba de muestras independientes

		prueba t para la igualdad de medias				95% de intervalo de confianza de la diferencia	
t	gl	Significación		Diferencia de medias	Diferencia de error estándar	Inferior	Superior
		P de un factor	P de dos factores				
,699	25	,245	,491	2772,944	3966,681	-5396,589	10942,478
,998	17,000	,166	,332	2772,944	2777,233	-3086,504	8632,393

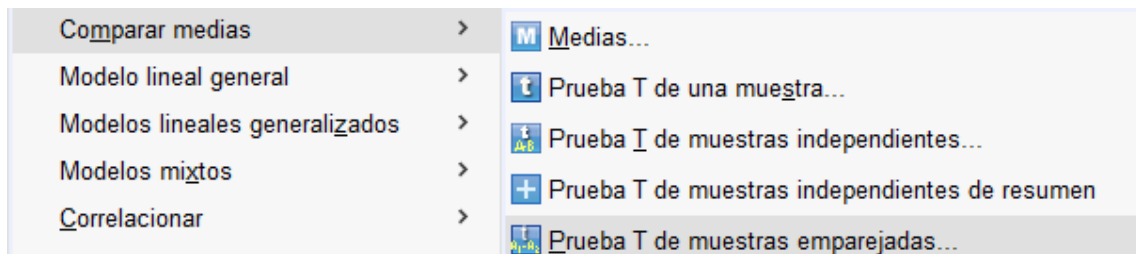
Leemos la Significación P de dos factores de la primera fila. Como 0,491 es mayor que 0,05 asumimos que los salarios de hombres y mujeres son iguales en media.

- **Pruebas T de muestras emparejadas**

Esta herramienta puede ser útil para para comprobar si podemos decidir, a partir de muestras, que dos grupos dependientes tienen medias poblacionales iguales.

Por ejemplo, consideremos una situación en la que los datos disponibles provienen de una muestra de trabajadores y deseamos evaluar si una medida específica implementada por la empresa ha tenido un impacto en la media salarial. Recolectamos datos sobre los salarios de los trabajadores antes y después de la implementación de la medida. Estos datos son pareados, ya que corresponden a los mismos trabajadores, lo que significa que el resultado de cada trabajador antes de la acción está emparejado con su resultado después de la acción.

Teniendo en cuenta que muestras pareadas es sinónimo de muestras emparejadas, la ruta sería:



La ventana quedaría como:

Prueba T de muestras emparejadas

Variables emparejadas:

Par	Variable1	Variable2
1	[Salario]	[SalariosPost]
2		

Estimar tamaños de efecto

Calcular el uso del estandarizador

Desviación estándar de la diferencia

Desviación estándar corregida de la diferencia

Promedio de varianzas

Botones: Opciones..., Simular muestreo..., Aceptar, Pegar, Restablecer, Cancelar, Ayuda

Los resultados del análisis serían:

Prueba T

Estadísticas de muestras emparejadas

		Media	N	Desv. estándar	Media de error estándar
Par 1	Salario	1862,74	27	9620,330	1851,433
	SalariosPost	159,22	27	767,603	147,725

Correlaciones de muestras emparejadas

		N	Correlación	Significación	
			n	P de un factor	P de dos factores
Par 1	Salario & SalariosPost	27	1,000	<,001	<,001

Prueba de muestras emparejadas

		Diferencias emparejadas					Significación			
		Media	Desv. estándar	Media de error estándar	95% de intervalo de confianza de la diferencia		t	gl	P de un factor	P de dos factores
					Inferior	Superior				
Par 1	Salario - SalariosPost	1703,519	8852,743	1703,711	-1798,510	5205,547	1,000	26	,163	,327

Tamaños de efecto de muestras emparejadas

Par 1	Salario - SalariosPost	d de Cohen	Standardizer ^a	Estimación de puntos	Intervalo de confianza al 95%	
					Inferior	Superior
			8852,743	,192	-,190	,571
		corrección de Hedges	9118,771	,187	-,185	,555

a. El denominador utilizado en la estimación de tamaños del efecto.

La d de Cohen utiliza la desviación estándar de muestra de la diferencia de medias.

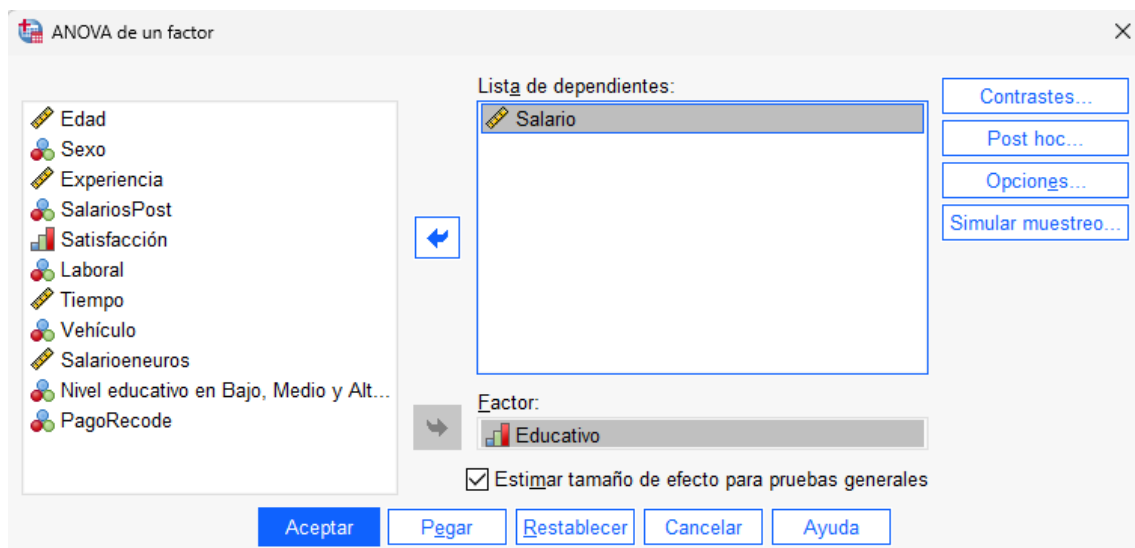
La corrección de Hedges utiliza la desviación estándar de muestra de la diferencia de medias, más un factor de corrección.

La tabla más importante es la Prueba de muestras emparejadas. Nos fijaríamos en Significación P de dos factores. Como el valor es superior al alfa de 0,05 debemos asumir que los salarios siguen en media igual y que la medida no ha tenido un efecto significativo.

En la ventana de la herramienta podríamos haber cambiado el orden de las variables. Eso no va a suponer ningún cambio en ese P-valor.

- ANOVA de un factor.

Es la herramienta que nos permite comparar las medias cuando hay más de dos casos. Por ejemplo, si comparamos las medias para el nivel educativo no podríamos hacerlo por ninguna de las herramientas anteriores. La ventana para comparar los salarios medios entre los distintos niveles educativos quedaría como:



Y los resultados proporcionados se visualizarían de la siguiente manera:

Unidireccional

ANOVA

Salario

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	115910254,7 13	2	57955127,35 6	,607	,553
Dentro de grupos	2290409252, 472	24	95433718,85 3		
Total	2406319507, 185	26			

Tamaños de efecto ANOVA^{a,b}

	Estimación de puntos	Intervalo de confianza al 95%	
		Inferior	Superior
Salario Eta cuadrado	,048	,000	,223
Epsilon cuadrado	-,031	-,083	,158
Omega cuadrado efecto fijo	-,030	-,080	,153
Omega cuadrado efecto aleatorio	-,015	-,038	,083

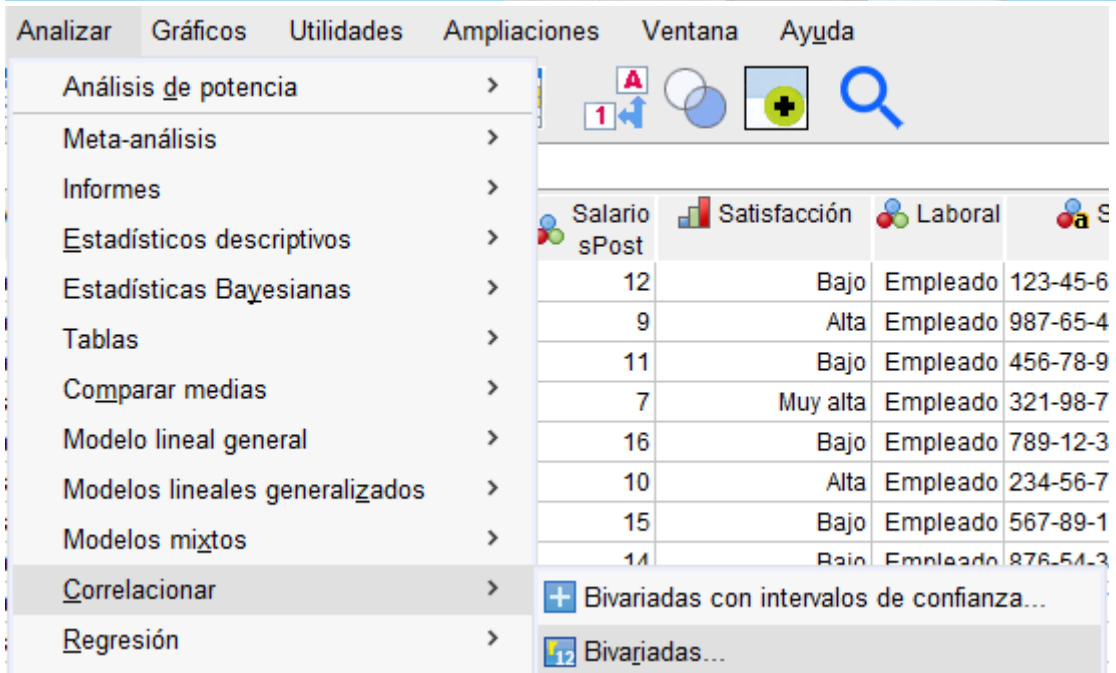
a. Eta cuadrado y Epsilon cuadrado se estiman basándose en el modelo de efecto fijo.

b. Las estimaciones negativas pero menos sesgadas se conservan, no se redondean a cero.

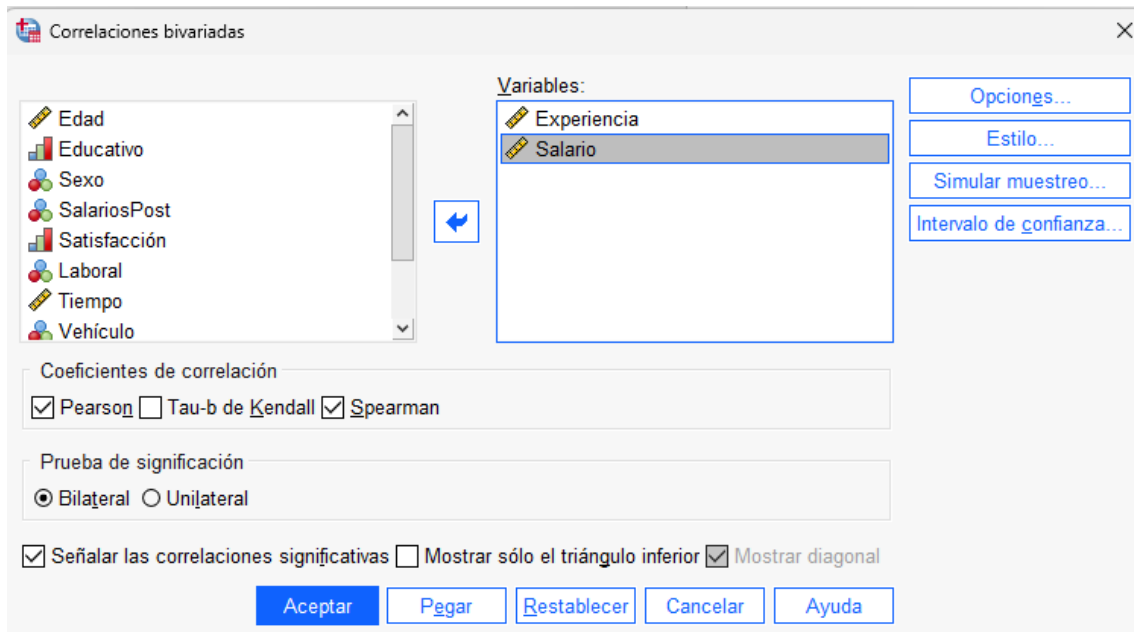
Al comprobar el Sig. de la tabla de ANOVA y ser este valor superior al alfa de 0,05 podemos concluir que los salarios medios en este ejemplo son iguales para todos los niveles educativos. Es decir, en este ejemplo, el nivel educativo no ejerce influencia en el salario.

Correlacionar

Las principales medidas de correlación las podemos encontrar rápidamente en:



La ventana de herramientas nos da acceso haciendo clic a las principales medidas. Por ejemplo, en esta captura he marcado el coeficiente de correlación de Pearson y el coeficiente de asociación de Spearman. No he marcado el coeficiente Tau-b de Kendall.



Los resultados de ese análisis serían:

Correlaciones

Correlaciones

		Experiencia	Salario
Experiencia	Correlación de Pearson	1	-,043
	Sig. (bilateral)		,833
	N	28	27
Salario	Correlación de Pearson	-,043	1
	Sig. (bilateral)	,833	
	N	27	27

Correlaciones no paramétricas

Correlaciones

			Experiencia	Salario
Rho de Spearman	Experiencia	Coefficiente de correlación	1,000	-,019
		Sig. (bilateral)	.	,926
		N	28	27
	Salario	Coefficiente de correlación	-,019	1,000
		Sig. (bilateral)	,926	.
		N	27	27

Si miramos la Sig. (bilateral) de ambas pruebas podemos concluir que partiendo de la muestra no podemos afirmar que exista asociación entre la Experiencia y el Salario a nivel población, ya que ambos valores son superiores al alfa del 0,05 fijado.

Regresión

Para realizar la regresión debemos de asumir qué variables van a ser independientes y qué variable será la dependiente. La ruta se encuentra en:

The screenshot shows the 'Análisis de potencia' menu with various statistical options. Below the menu, a table displays data for 'Salario sPost', 'Satisfacción', and 'Laboral'.

Salario sPost	Satisfacción	Laboral	
12	Bajo	Empleado	123
9	Alta	Empleado	987
11	Bajo	Empleado	456
7	Muy alta	Empleado	321
16	Bajo	Empleado	789
10	Alta	Empleado	234
15	Bajo	Empleado	567
14	Bajo	Empleado	876
13	Alta	Empleado	210

En la siguiente captura de la ventana de la herramienta se simula un ejemplo donde intentamos estimar el salario en función de la Edad, la Experiencia y el Tiempo.

The 'Regresión lineal' dialog box shows the following configuration:

- Dependientes:** Salario
- Bloque 1 de 1:**
 - Independientes:** Edad, Experiencia, Tiempo
 - Método:** Intro
- Variable de selección:** (Empty field)
- Etiquetas de caso:** (Empty field)
- Ponderación MCP:** (Empty field)

Buttons on the right: Estadísticos..., Gráficos..., Guardar..., Opciones..., Estilo..., Simular muestreo...
Buttons at the bottom: Aceptar, Pegar, Restablecer, Cancelar, Ayuda

Los resultados de esta regresión serían:

Regresión

Variables entradas/eliminadas^a

Modelo	Variables entradas	Variables eliminadas	Método
1	Tiempo, Experiencia, Edad ^b		Introducir

a. Variable dependiente: Salario

b. Todas las variables solicitadas introducidas.

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,113 ^a	,013	-,135	10872,103

a. Predictores: (Constante), Tiempo, Experiencia, Edad

ANOVA^a

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	30727626,515	3	10242542,172	,087	,967 ^b
	Residuo	2364052388,110	20	118202619,406		
	Total	2394780014,625	23			

a. Variable dependiente: Salario

b. Predictores: (Constante), Tiempo, Experiencia, Edad

Coefficientes^a

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Desv. Error	Beta		
1	(Constante)	5337,789	8196,807		,651	,522
	Edad	-,181	2,179	-,026	-,083	,935
	Experiencia	,129	3,012	,012	,043	,966
	Tiempo	-87,396	233,148	-,102	-,375	,712

a. Variable dependiente: Salario

Este ejemplo ilustra un modelo deficiente, ya que su coeficiente de determinación R cuadrado (R^2) es de 0.013 (indicando un ajuste muy pobre), y su R cuadrado ajustado es negativo (indicando un ajuste aún peor). A nivel global, el valor de significación (Sig.) en la tabla de ANOVA, superior a 0.05, sugiere que el modelo no es útil para la predicción. Al analizar los valores Sig. en la tabla de coeficientes, observamos que todos ellos superan el umbral de 0.05, lo que implica que ninguna variable es significativamente diferente de cero. En consecuencia, ninguna de las variables puede explicar de manera individual los cambios en el modelo.