

***ACTIVIDADES ESTADÍSTICA EN
PSICOLOGÍA***

PEDRO M. VALERO MORA

1. Comentarios al Informe Pisa

- El informe PISA 2003 es un diagnóstico del nivel educativo de los estudiantes. A partir del documento de resumen responde de manera razonada a las siguientes preguntas:

EJERCICIO 1.1 **Comenta los resultados en Matemáticas indicando qué países obtienen mejores resultados y qué países obtienen peores resultados ofreciendo si puedes alguna explicación de las diferencias**

EJERCICIO 1.2 **En la página 6 se contemplan diferentes áreas de Matemáticas. ¿Es consistente la posición de España en todas esas áreas? ¿Hay países que tienen posiciones muy diferentes según el area?**

EJERCICIO 1.3 Comenta los resultados para el resto de las áreas estudiadas en el informe PISA (Lectura, Ciencias y Solución de Problemas). ¿Es consistente la posición de España? ¿Hay países que no son consistentes y ocupan posiciones muy diferentes?

EJERCICIO 1.4 Comenta las diferencias entre hombres y mujeres. Comenta si hay algún país que resulte especial y por qué.

EJERCICIO 1.5 Comenta qué tipo de colegio consigue mejores resultados en el informe PISA, público o privado. Indica las razones que explican estas diferencias

Análisis de variables categóricas con el SPSS

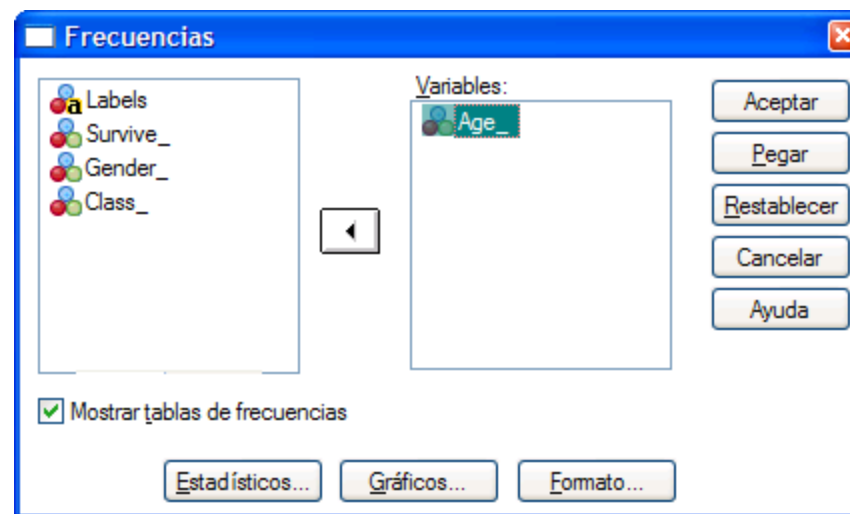
- Esta actividad introduce al uso del paquete estadístico SPSS para realizar gráficos para 1 variable categórica. En primer lugar hay una introducción a como realizarlo utilizando un ejemplo y luego hay una serie de preguntas que tienes que contestar utilizando otros archivos de datos.

- Haciendo click en el archivo Titanic que está en mi página web verás más o menos esto..

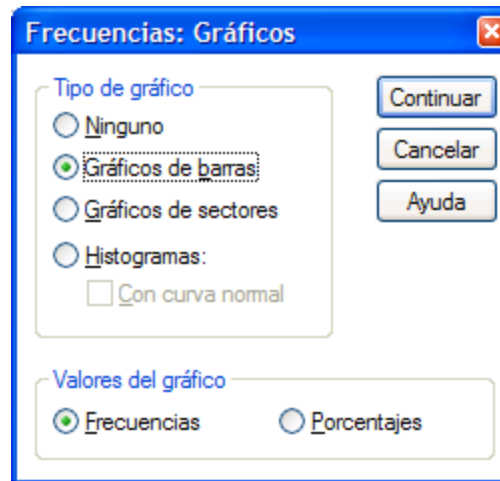
	Labels	Survive	Age	Gender	Class
1	Lived*Adult*Male*1st	Lived	Adult	Male	1st
2	Lived*Adult*Male*1st	Lived	Adult	Male	1st
3	Lived*Adult*Male*1st	Lived	Adult	Male	1st
4	Lived*Adult*Male*1st	Lived	Adult	Male	1st
5	Lived*Adult*Male*1st	Lived	Adult	Male	1st
6	Lived*Adult*Male*1st	Lived	Adult	Male	1st
7	Lived*Adult*Male*1st	Lived	Adult	Male	1st
8	Lived*Adult*Male*1st	Lived	Adult	Male	1st
9	Lived*Adult*Male*1st	Lived	Adult	Male	1st
10	Lived*Adult*Male*1st	Lived	Adult	Male	1st
11	Lived*Adult*Male*1st	Lived	Adult	Male	1st
12	Lived*Adult*Male*1st	Lived	Adult	Male	1st
13	Lived*Adult*Male*1st	Lived	Adult	Male	1st

Si ves **esto** diferente haz click **aquí**

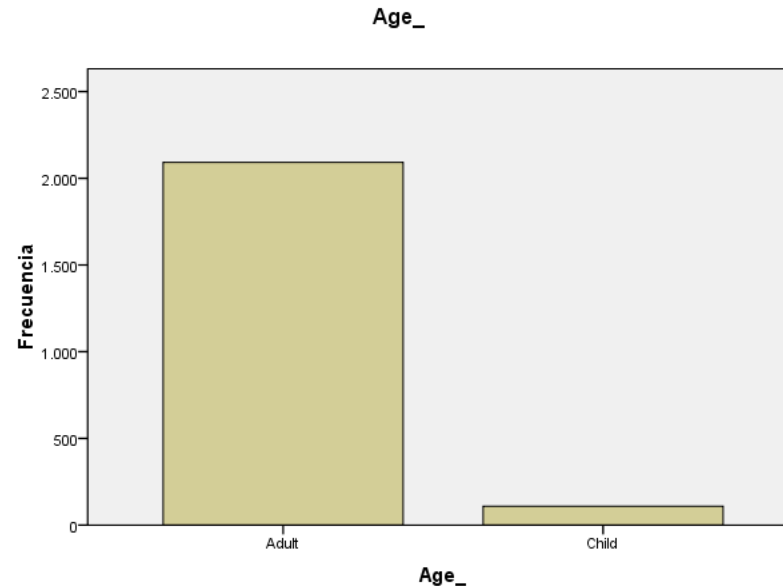
- Para hacer recuentos de las variables individualmente utiliza el subcomando Frecuencias que está en el subcomando Estadísticos Descriptivos que está en el menú Analizar. Eso da lugar al siguiente cuadro de diálogo en donde tienes que elegir la variable a recontar



- Si quieres gráficos utiliza el botón Gráficos y tienes las siguientes opciones.

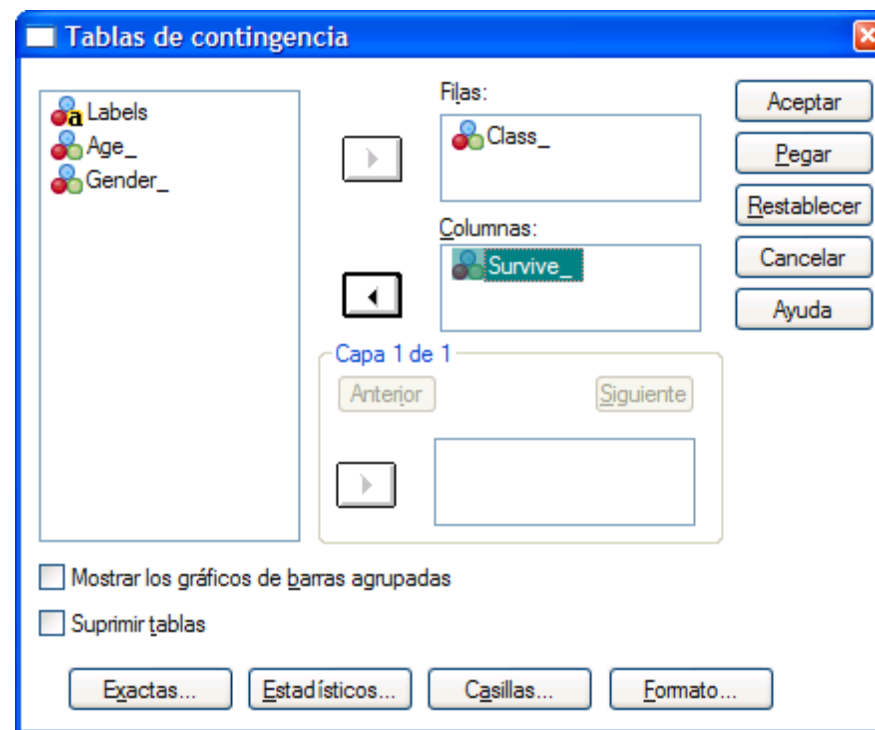


- Para la variable Age tienes que obtener el siguiente gráfico:

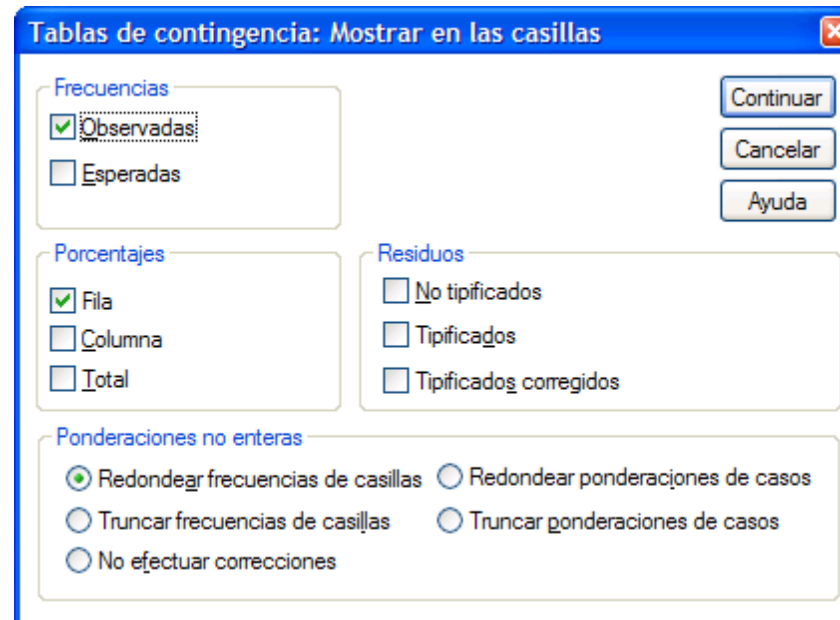


- Para pasar este documento a un documento de Word, haceis click en el gráfico y seleccionais Copiar en el menú edición. Luego os vais al documento de Word y haceis Pegar.

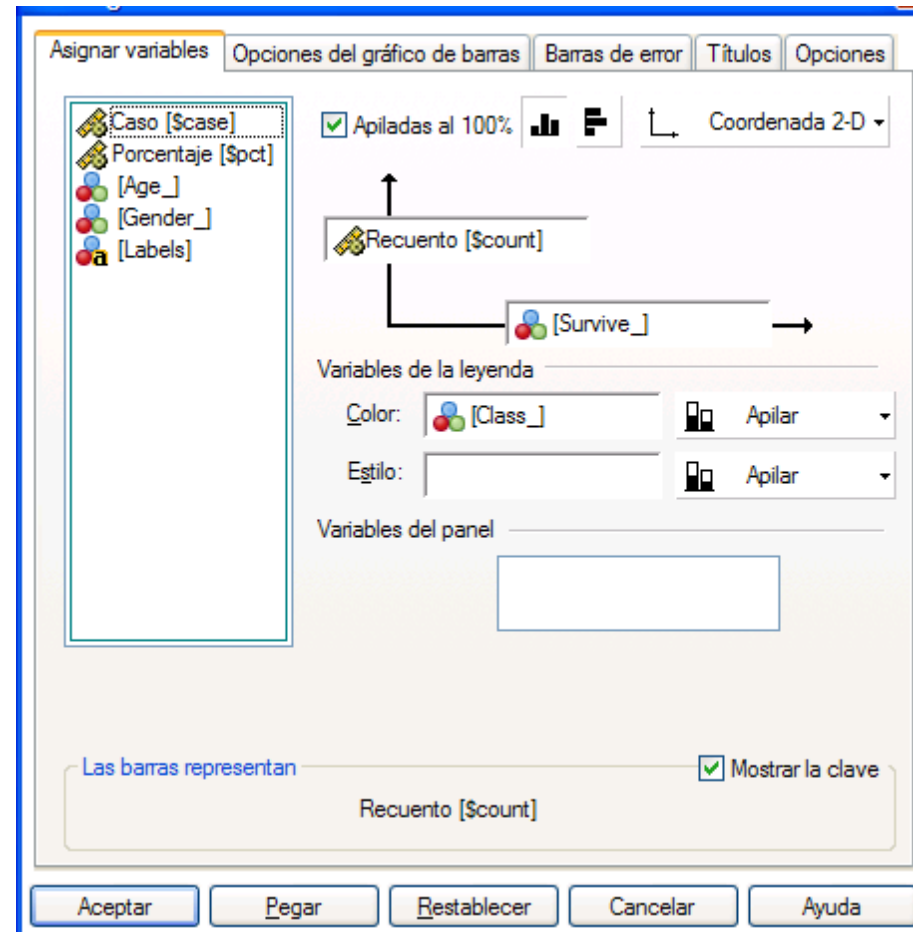
- Para cruces de dos variables hay que ir al comando Tablas de Contingencia en el comando Estadísticos Descriptivos dentro del menú Analizar. Eso da lugar al siguiente cuadro de diálogo



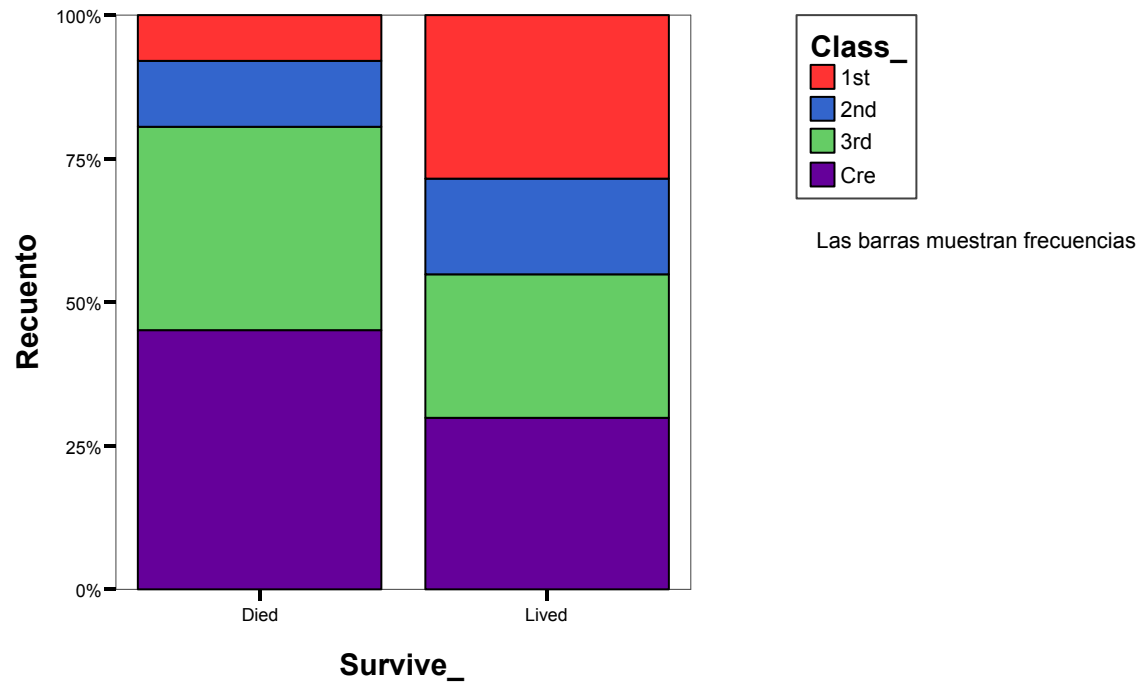
- Para calcular porcentajes por filas o por columnas puedes usar el botón Casillas del cuadro anterior



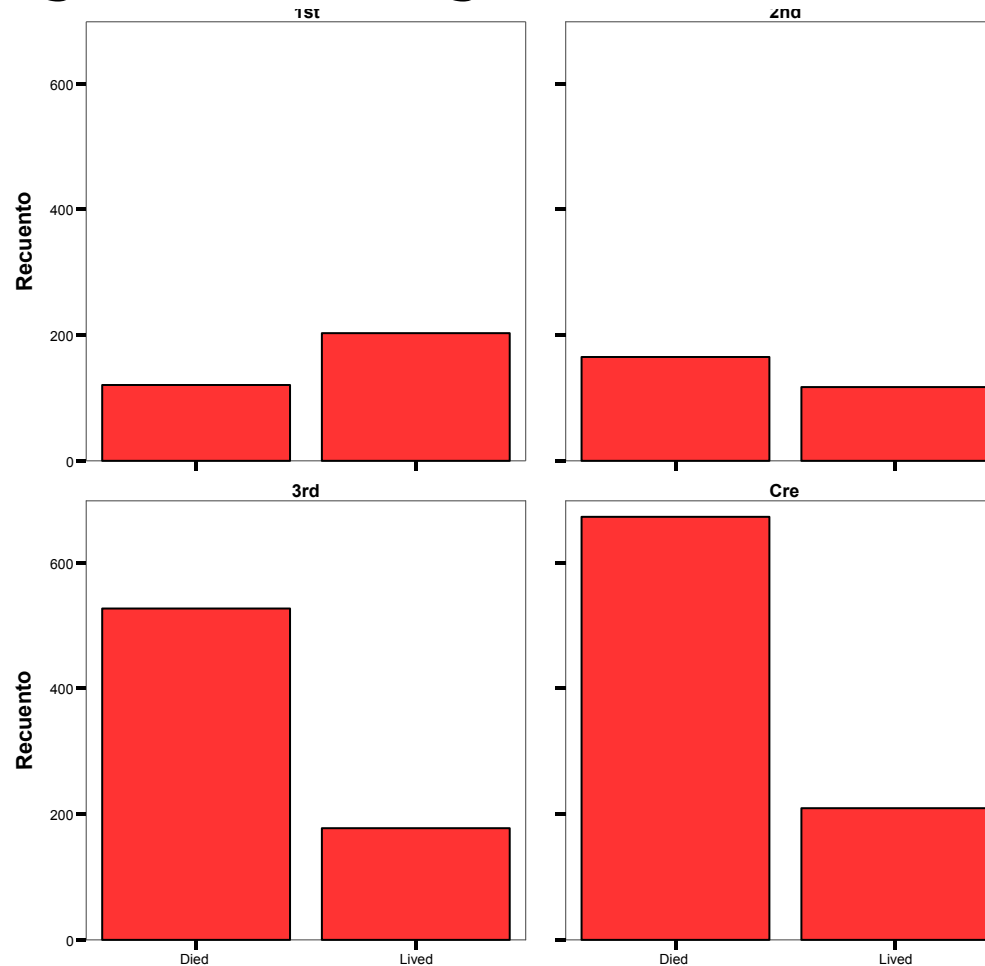
- Para hacer gráficos hay que ir al menú gráficos, submenú interactivos y elegir barras o sectores. Aquí explicaremos barras.



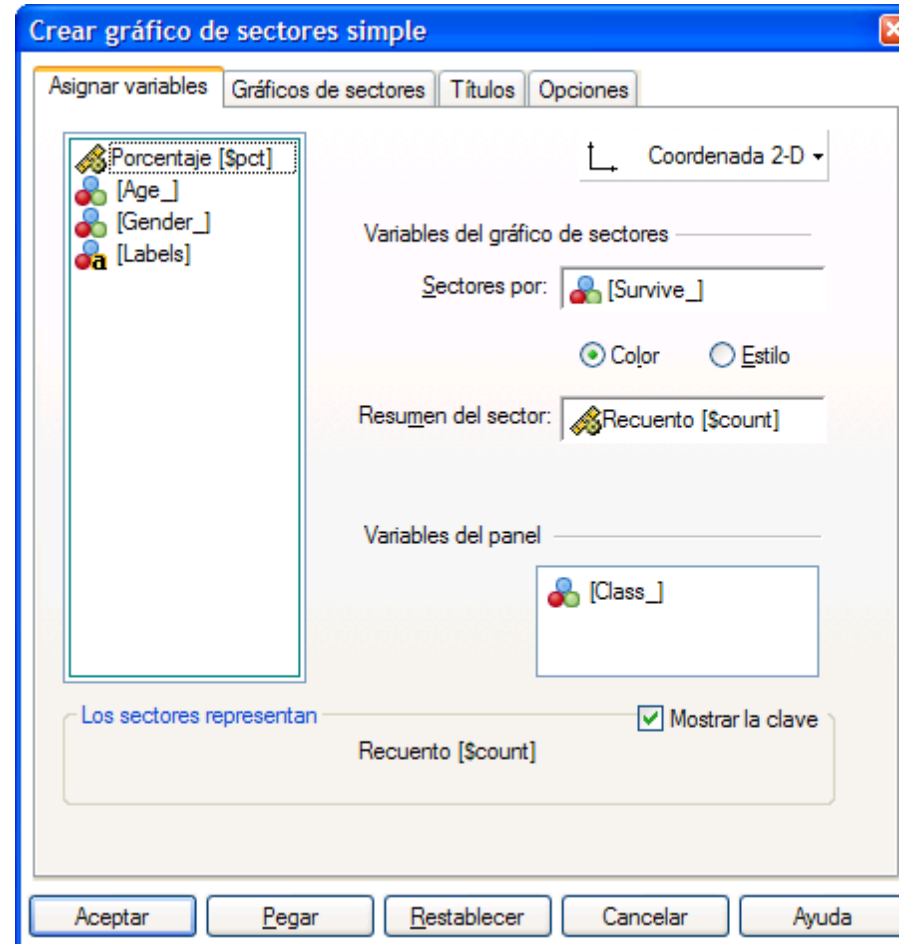
- Las opciones que ves en el cuadro de diálogo anterior dan lugar a esto



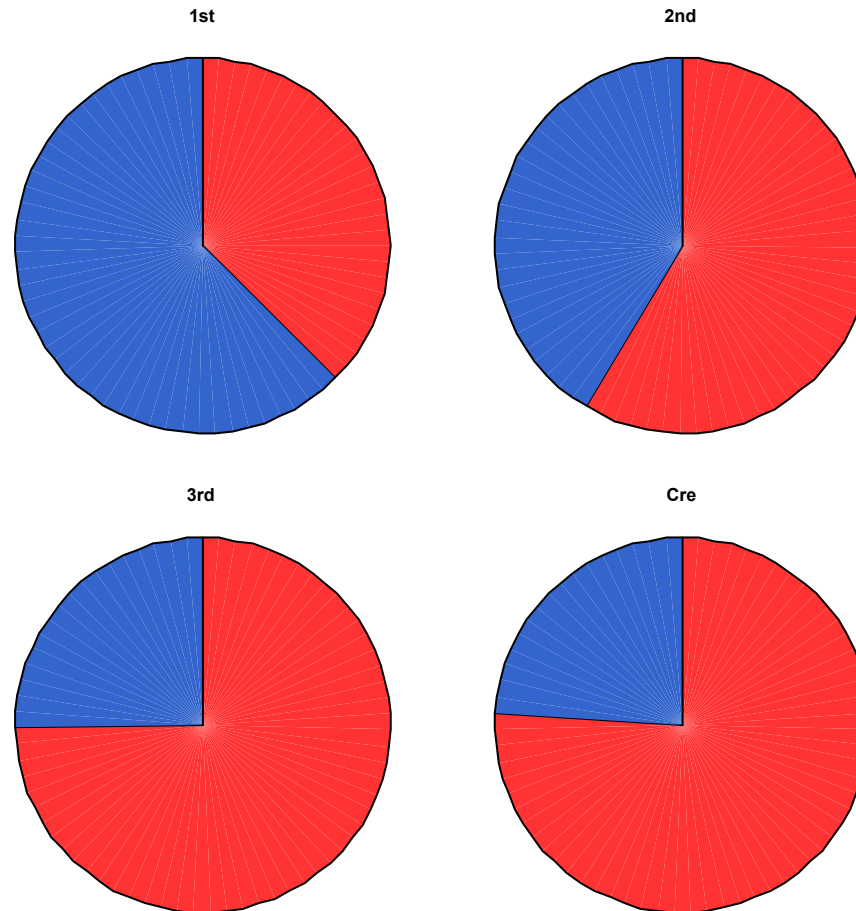
- Hay varias posibilidades que se pueden tocar aquí. Por ejemplo, si la variable clase la pones en variables del panel el gráfico es el siguiente



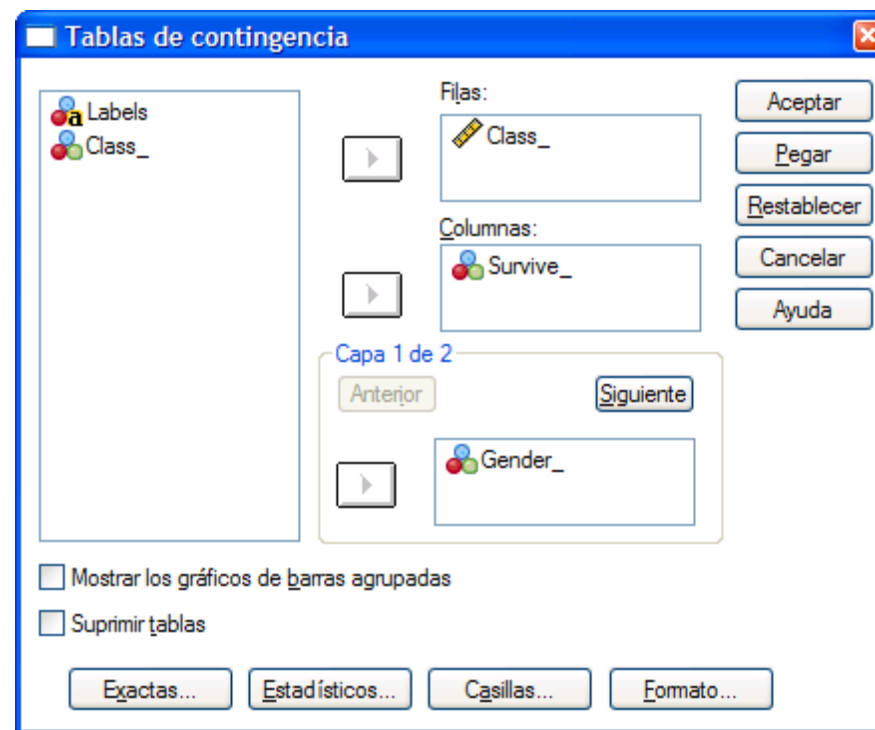
- Seleccionando los gráficos de sectores simples tienes lo siguiente



- Ese cuadro de diálogo da lugar a lo siguiente



- Si queremos hacer una tabla con tres variables simultáneamente podemos hacerla en el mismo sitio que las tablas de dos variables pero poniendo una variable en la capa



- Eso da lugar a lo siguiente

Tabla de contingencia Class_ * Survive_ * Gender_

% de Class_

Gender_			Survive_		Total
			Died	Lived	
Fema	Class_	1st	2.8%	97.2%	100.0%
		2nd	12.3%	87.7%	100.0%
		3rd	54.1%	45.9%	100.0%
		Cre	13.0%	87.0%	100.0%
	Total	26.8%	73.2%	100.0%	
Male	Class_	1st	65.6%	34.4%	100.0%
		2nd	86.0%	14.0%	100.0%
		3rd	82.7%	17.3%	100.0%
		Cre	77.7%	22.3%	100.0%
	Total	78.8%	21.2%	100.0%	

- Si pones más variables en otras capas puedes conseguir lo siguiente (cuatro variables)

Tabla de contingencia Class_ * Survive_ * Gender_ * Age_

% de Class_

Age_	Gender_	Class_		Survive_		Total
				Died	Lived	
Adult	Fema	Class_	1st	2.8%	97.2%	100.0%
			2nd	14.0%	86.0%	100.0%
			3rd	53.9%	46.1%	100.0%
			Cre	13.0%	87.0%	100.0%
		Total	25.6%	74.4%	100.0%	
	Male	Class_	1st	67.4%	32.6%	100.0%
			2nd	91.7%	8.3%	100.0%
			3rd	83.8%	16.2%	100.0%
			Cre	77.7%	22.3%	100.0%
		Total	79.7%	20.3%	100.0%	
Child	Fema	Class_	1st		100.0%	100.0%
			2nd		100.0%	100.0%
			3rd	54.8%	45.2%	100.0%
			Total	37.8%	62.2%	100.0%
		Male	Class_	1st		100.0%
	2nd				100.0%	100.0%
	3rd			72.9%	27.1%	100.0%
	Total			54.7%	45.3%	100.0%

2. Gráficos para variables categóricas con el SPSS

***Fecha de Entrega: Grupo B 8 de Octubre, Grupos G
y H 16 de Octubre.***

EJERCICIO 2.1 Haz un gráfico de barras para cada una de las variables del archivo Sexo y Matrimonio. Pega los gráficos y comentalos.

EJERCICIO 2.2 Haz cuatro gráficos de sectores para las variables de Sexo y Matrimonio. Ponlos en el documento de Word y comentalos.

EJERCICIO 2.3 **Haz una tabla para las variables Extramarital y Divorce. Explica si tener relaciones extramatrimoniales lleva a un porcentaje mayor de divorcios e indica los valores relevantes que apoyan ese resultado. Pega la tabla en el documento de Word.**

EJERCICIO 2.4 **Haz una tabla para las variables Premarital y Divorce. Explica si tener relaciones Prematrimoniales lleva a un porcentaje mayor de divorcios e indica los valores relevantes que apoyan ese resultado. Pega la tabla en el documento de Word.**

EJERCICIO 2.5 **Haz una tabla para las variables Gender y Divorce. Explica si ser varón o mujer lleva a un porcentaje mayor de divorcios e indica los valores relevantes que apoyan ese resultado. Pega la tabla en el documento de Word.**

EJERCICIO 2.6 Haz una tabla para las variables Extramarital y Premarital y Divorce. ¿Qué combinación de categorías (por ejemplo gente sin relaciones premaritales y sin relaciones extramatrimoniales) tiene mayor porcentaje de divorcio. Pon la tabla en la respuesta

EJERCICIO 2.7 Haz una tabla para Extramarital, Gender y Divorce. Indica qué combinación de categorías (por ejemplo mujer con relaciones extramatrimoniales) tiene más porcentaje de divorcios) y qué valores aparece. Pon la tabla en la respuesta.

EJERCICIO 2.8 **Haz una tabla para las cuatro categorías. Qué combinación de categorías produce más divorcios? Pega la tabla y comenta.**

Interpretación de tablas con datos categóricos (actividades 3, 4, 5 y 6) Fecha de entrega Grupo B 22/10; Grupo G, H 23/10)

3. Sexo y matrimonio

- Usaremos los datos sobre Sexo y matrimonio que están en la página web (aunque no es necesario bajarse los datos para realizar la práctica)

EJERCICIO 3.1 **Cuál es el porcentaje de sujetos que no tuvieron relaciones prematrimoniales? ¿Cuál es el total de sujetos encuestados?**

Premarital

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	N	795	76.7	76.7	76.7
	Y	241	23.3	23.3	100.0
	Total	1036	100.0	100.0	

EJERCICIO 3.2 ¿Cuál es el total de sujetos que tuvieron relaciones extramatrimoniales? ¿Cuál es el total que no tuvieron relaciones? ¿Cuál es el porcentaje que no tuvieron relaciones extramatrimoniales? ¿Cuál es la proporción de los que sí tuvieron relaciones extramatrimoniales?

Extramarital

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	N	915	88.3	88.3	88.3
	Y	121	11.7	11.7	100.0
	Total	1036	100.0	100.0	

EJERCICIO 3.3 Usando las tablas de los ejercicios ejercicio 3.1 y ejercicio 3.2 indica cual el porcentaje de sujetos que tuvieron tanto relaciones prematrimoniales como extramatrimoniales. Indica como has hecho ese cálculo.

EJERCICIO 3.4 Indica qué género tiene una tendencia a tener relaciones extramatrimoniales en proporción. Indica como has llegado a esa conclusión

Tabla de contingencia Gender * Extramarital

Recuento

		Extramarital		Total
		N	Y	
Gender	Fema	615	61	676
	Male	300	60	360
Total		915	121	1036

EJERCICIO 3.5 Responde a la pregunta del ejercicio 3.4 utilizando esta tabla. Indica si esta tabla es mejor o peor que la anterior.

Tabla de contingencia Gender * Extramarital

% de Gender

		Extramarital		Total
		N	Y	
Gender	Fema	91.0%	9.0%	100.0%
	Male	83.3%	16.7%	100.0%
Total		88.3%	11.7%	100.0%

EJERCICIO 3.6 Indica qué género de los sujetos está relacionado más con divorcio en la muestra considerada ¿Qué diferencia hay con el otro género? ¿Crees que la diferencia es muy grande?

Tabla de contingencia Gender * Married

% de Gender

		Married		Total
		Divorced	Married	
Gender	Fema	47.5%	52.5%	100.0%
	Male	48.1%	51.9%	100.0%
Total		47.7%	52.3%	100.0%

EJERCICIO 3.7 ¿Se podría decir a partir de esta tabla si haber tenido relaciones prematrimoniales conduce a tener relaciones extramatrimoniales?

Tabla de contingencia Premarital * Extramarital

% de Extramarital

		Extramarital		Total
		N	Y	
Premarital	N	80.2%	50.4%	76.7%
	Y	19.8%	49.6%	23.3%
Total		100.0%	100.0%	100.0%

**EJERCICIO 3.8 ¿Se respondería mejor a la pregunta del ejercicio 3.7 mediante esta tabla?
¿Hay información que sobra en esta tabla? ¿Por qué?**

Tabla de contingencia Premarital * Extramarital

			Extramarital		Total
			N	Y	
Premarital	N	% de Premarital	92.3%	7.7%	100.0%
		% de Extramarital	80.2%	50.4%	76.7%
	Y	% de Premarital	75.1%	24.9%	100.0%
		% de Extramarital	19.8%	49.6%	23.3%
Total		% de Premarital	88.3%	11.7%	100.0%
		% de Extramarital	100.0%	100.0%	100.0%

EJERCICIO 3.9 A partir de estas dos tablas indica qué género (mujeres o hombres) tiene más tendencia a divorciarse cuando ha tenido relaciones extramatrimoniales. Justifica tu respuesta ¿Qué tabla te proporciona la información más adecuada para este propósito? ¿Ambas son igualmente válidas?

Tabla de contingencia Gender * Married * Extramarital

% de Married

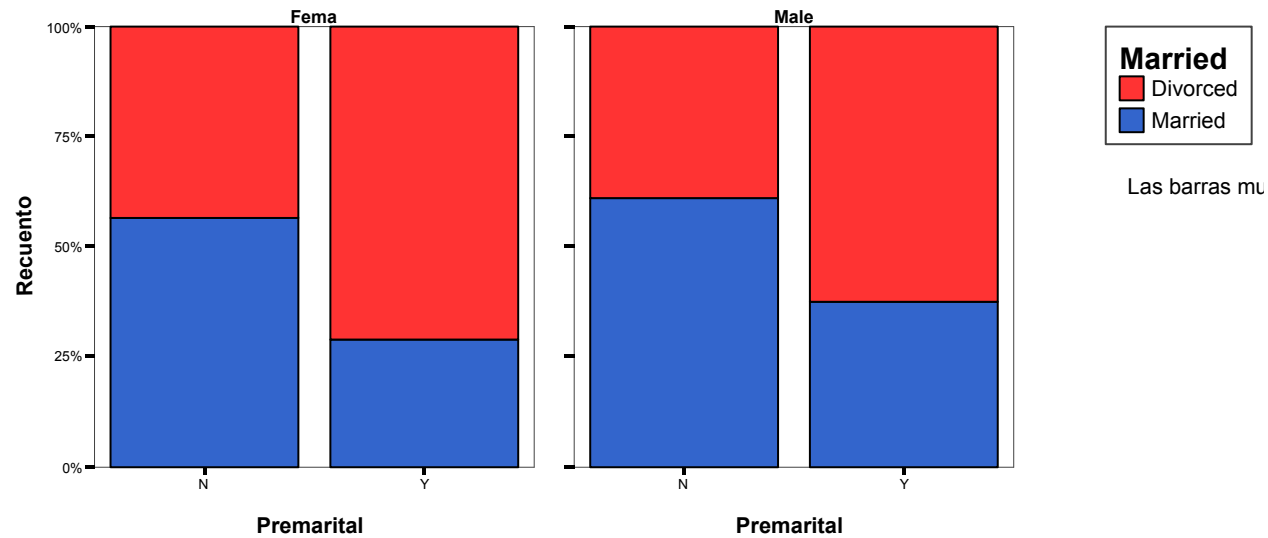
Extramarital			Married		Total
			Divorced	Married	
N	Gender	Fema	67.7%	66.9%	67.2%
		Male	32.3%	33.1%	32.8%
	Total		100.0%	100.0%	100.0%
Y	Gender	Fema	54.1%	34.8%	50.4%
		Male	45.9%	65.2%	49.6%
	Total		100.0%	100.0%	100.0%

Tabla de contingencia Gender * Married * Extramarital

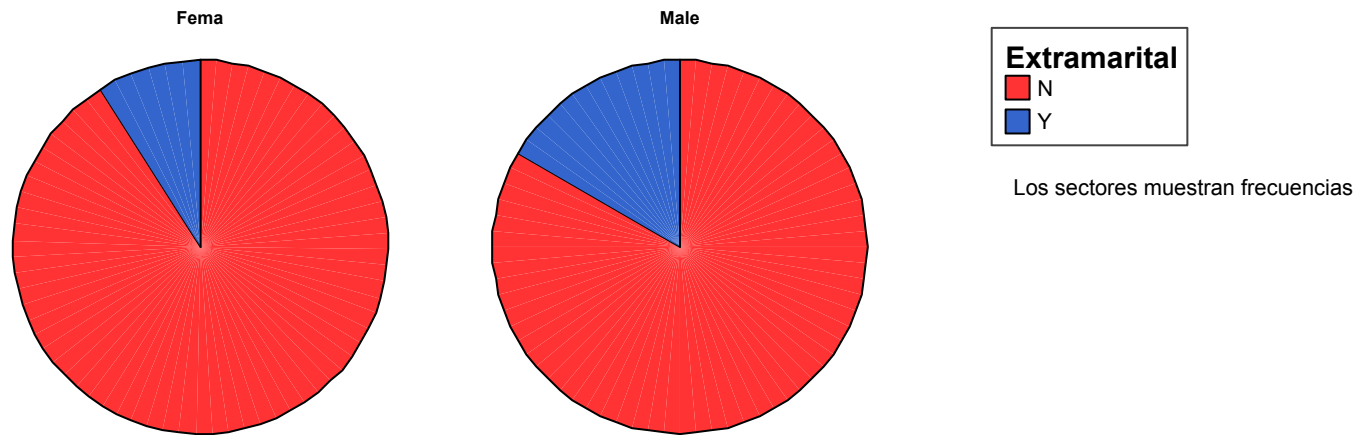
% de Gender

Extramarital			Married		Total
			Divorced	Married	
N	Gender	Fema	43.6%	56.4%	100.0%
		Male	42.7%	57.3%	100.0%
	Total		43.3%	56.7%	100.0%
Y	Gender	Fema	86.9%	13.1%	100.0%
		Male	75.0%	25.0%	100.0%
	Total		81.0%	19.0%	100.0%

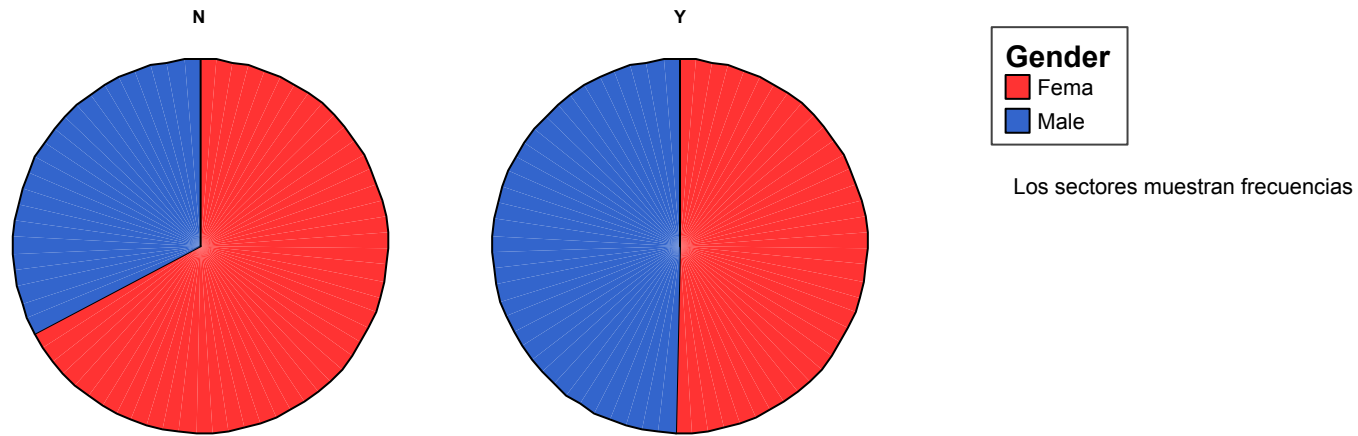
EJERCICIO 3.10 ¿Crees que el haber mantenido relaciones prematrimoniales conduce más al divorcio entre los hombres que entre las mujeres en estos datos según este gráfico? ¿Por qué?



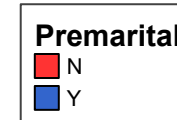
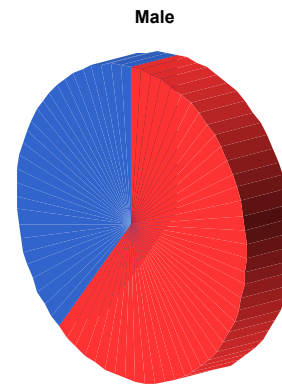
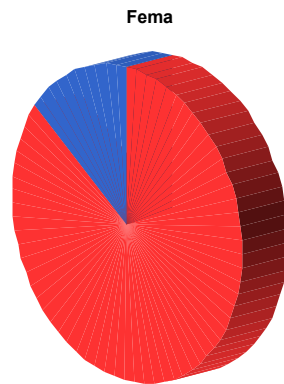
EJERCICIO 3.11 ¿Qué interpretación podría hacerse de este gráfico?



EJERCICIO 3.12 ¿Qué interpretación podría hacerse de este gráfico? (las variables son tener relaciones extramatrimoniales y género)

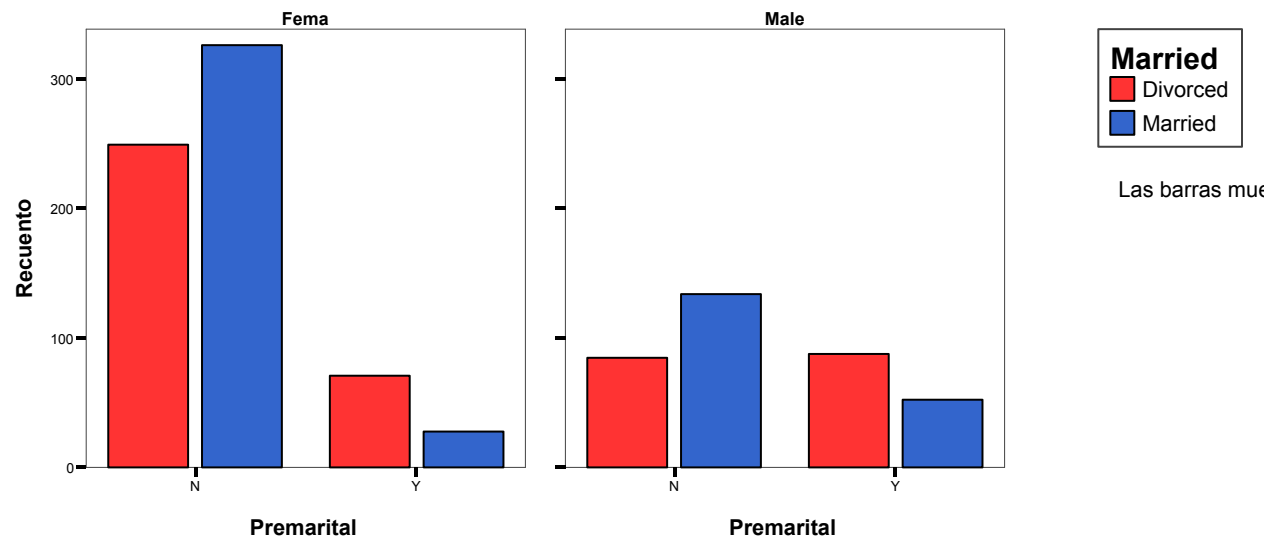


EJERCICIO 3.13 Compara este gráfico con el del Ejercicio 3.12. ¿Cuál es más recomendable?
¿Por qué?



Los sectores muestran frecuencias

EJERCICIO 3.14 ¿Podrías decir que las mujeres se divorcian más que los hombres cuando han tenido relaciones prematrimoniales según este gráfico? ¿Por qué?



EJERCICIO 3.15 Si eres un varón y crees en la frase “hasta que la muerte nos separe” ¿qué deberías evitar según esta tabla? ¿Y si fueras mujer? ¿Por qué?

Tabla de contingencia Gender * Married * Extramarital * Premarital

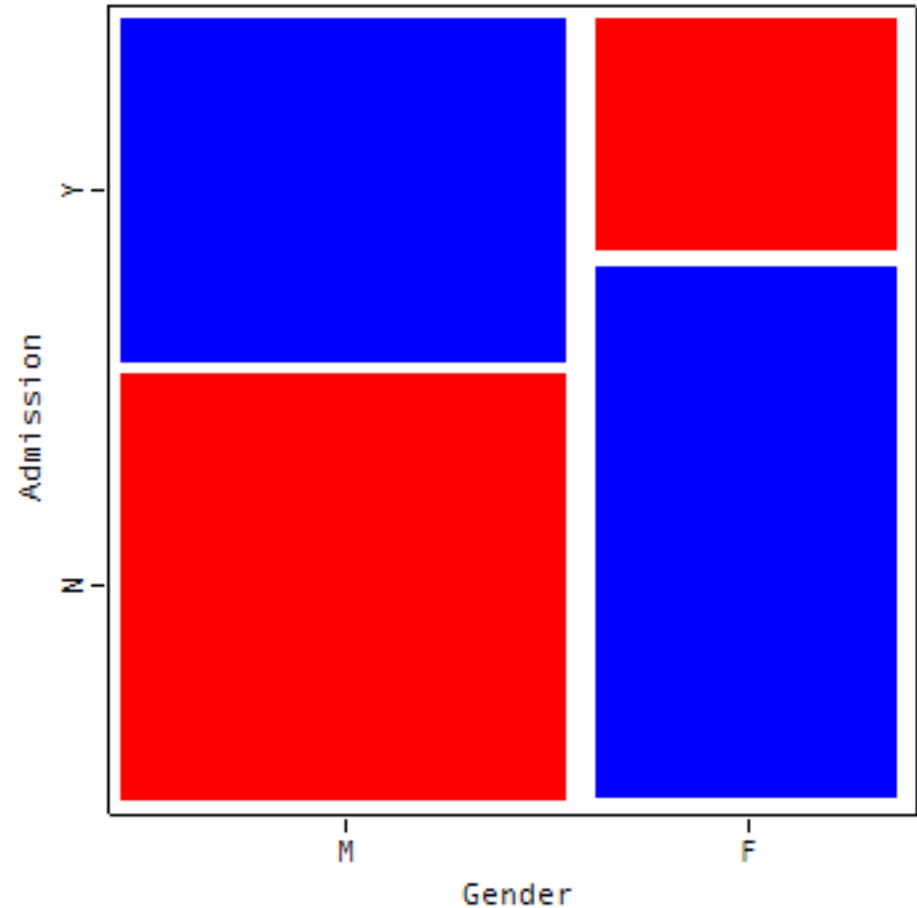
% de Gender

Premarital	Extramarital	Gender	Fema	Married		Total
				Divorced	Married	
N	N	Gender	Fema	39.9%	60.1%	100.0%
			Male	34.3%	65.7%	100.0%
		Total		38.4%	61.6%	100.0%
	Y	Gender	Fema	90.0%	10.0%	100.0%
			Male	81.0%	19.0%	100.0%
		Total		86.9%	13.1%	100.0%
Y	N	Gender	Fema	68.4%	31.6%	100.0%
			Male	58.8%	41.2%	100.0%
		Total		63.0%	37.0%	100.0%
	Y	Gender	Fema	81.0%	19.0%	100.0%
			Male	71.8%	28.2%	100.0%
		Total		75.0%	25.0%	100.0%

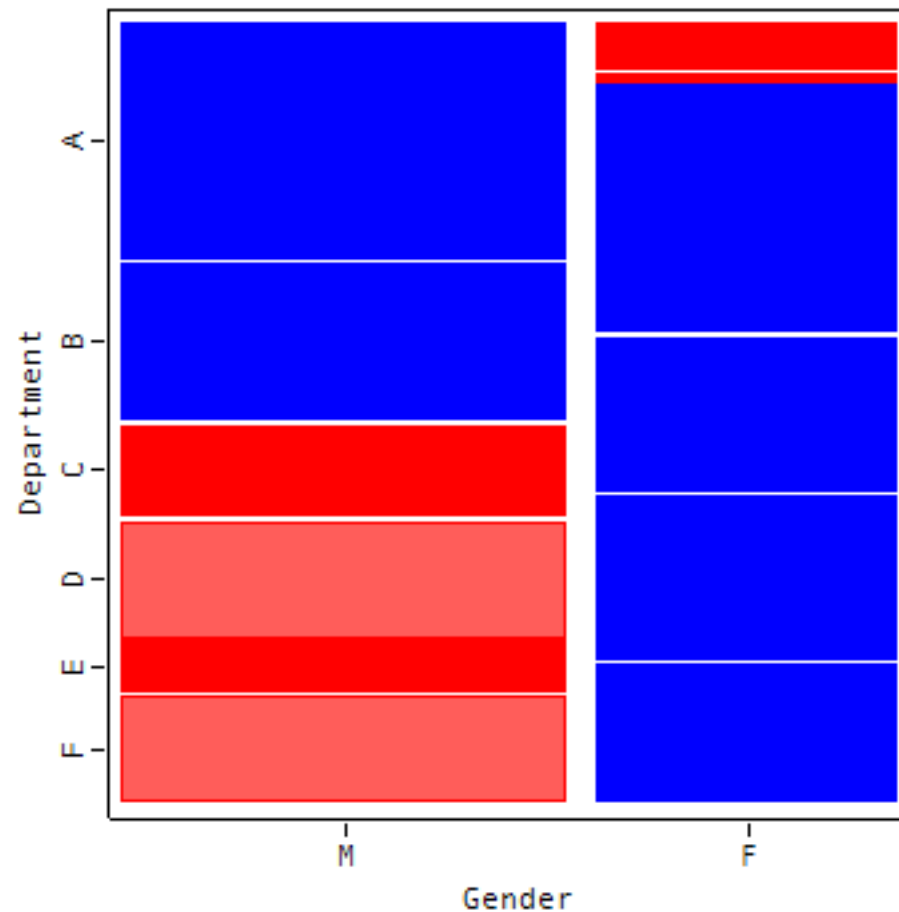
4. Admisión en Berkeley

- En la clase de teoría ya hablamos acerca de los datos de admisión en una universidad y como aparentemente existía un caso de discriminación contra las mujeres.
- En este ejercicio utilizaremos gráficos de mosaico, los cuales el SPSS no calcula así que no es posible para vosotros repetir los ejercicios aquí puestos.

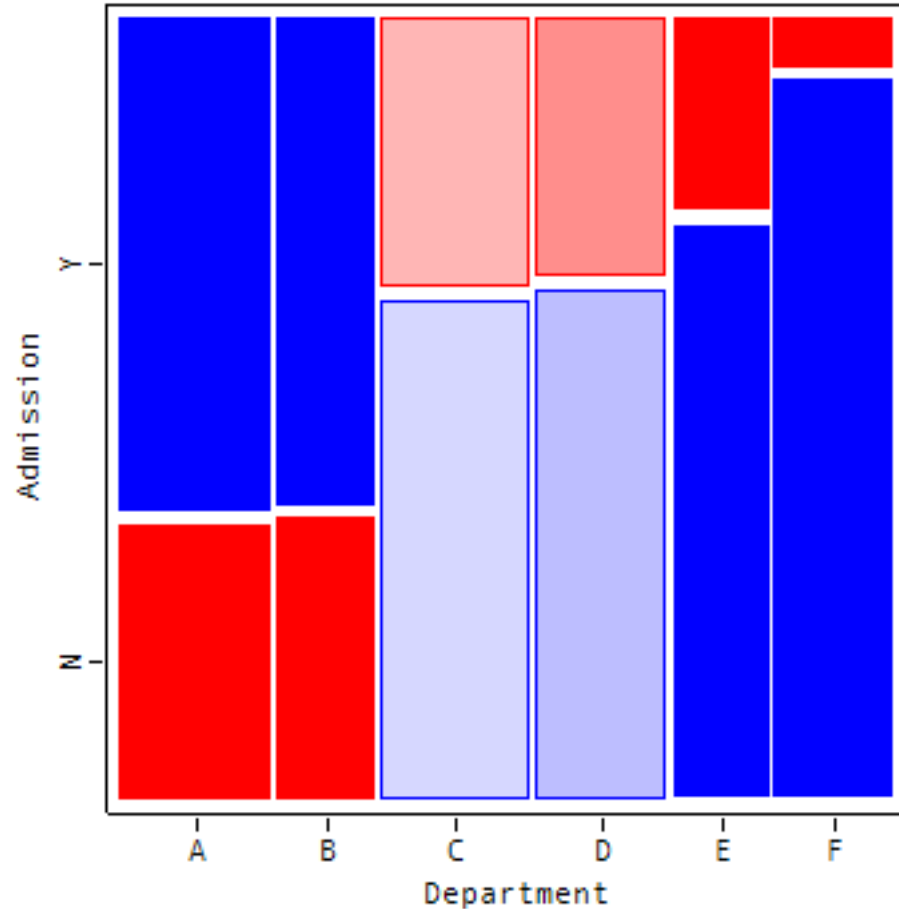
EJERCICIO 4.1 ¿Podría decirse a partir de este gráfico que se puede sospechar que había discriminación contra las mujeres en la admisión y que se las rechazaba más que a los hombres? (Recuerda Gender= Género: M=Male-Hombres; F=Female-Mujeres) También, ¿en esa universidad había más solicitudes de hombres o de mujeres? ¿Cómo lo sabes?

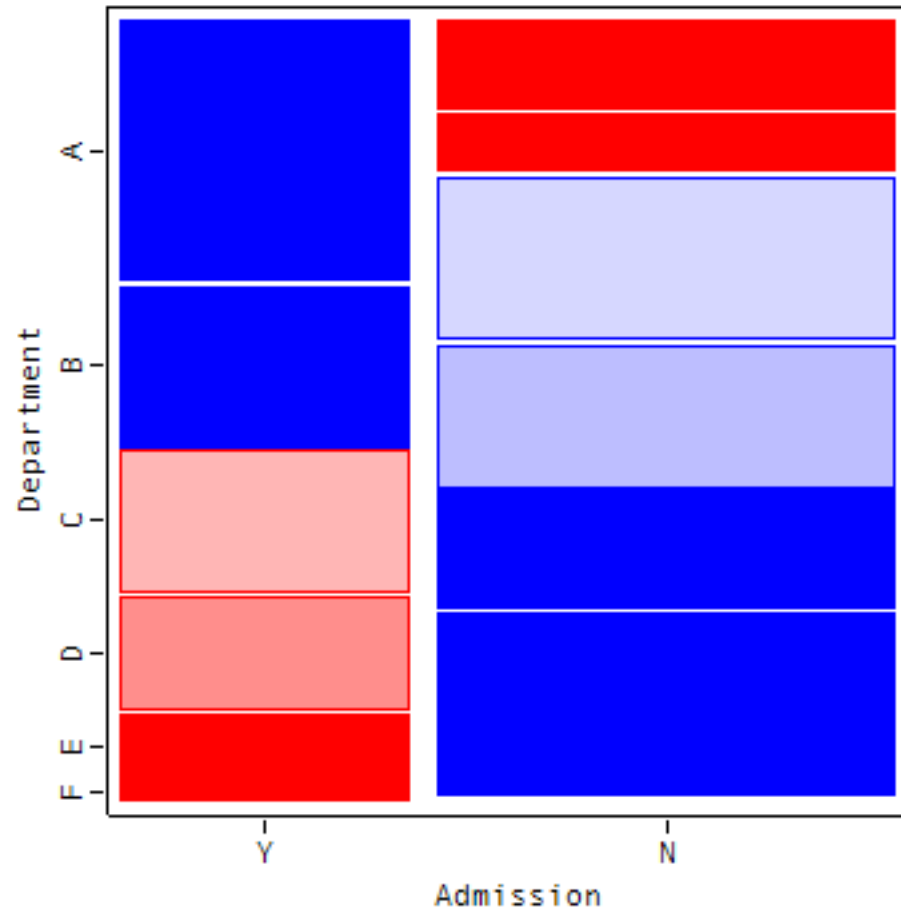


EJERCICIO 4.2 Mirando el gráfico de abajo. ¿Cuál era el departamento favorito de las mujeres? ¿Cuál es el de los hombres? ¿Las preferencias en cuanto a departamentos son similares para hombres y para mujeres? ¿Cuáles son las diferencias más importantes?



EJERCICIO 4.3 De los dos gráficos siguientes ¿cuál dirías que es más interesante para responder a la pregunta sobre si qué departamentos admiten una proporción de gente más grande con respecto a los que lo solicitan? ¿Por qué? ¿En qué departamentos es más fácil entrar según estos gráficos? ¿Por qué?

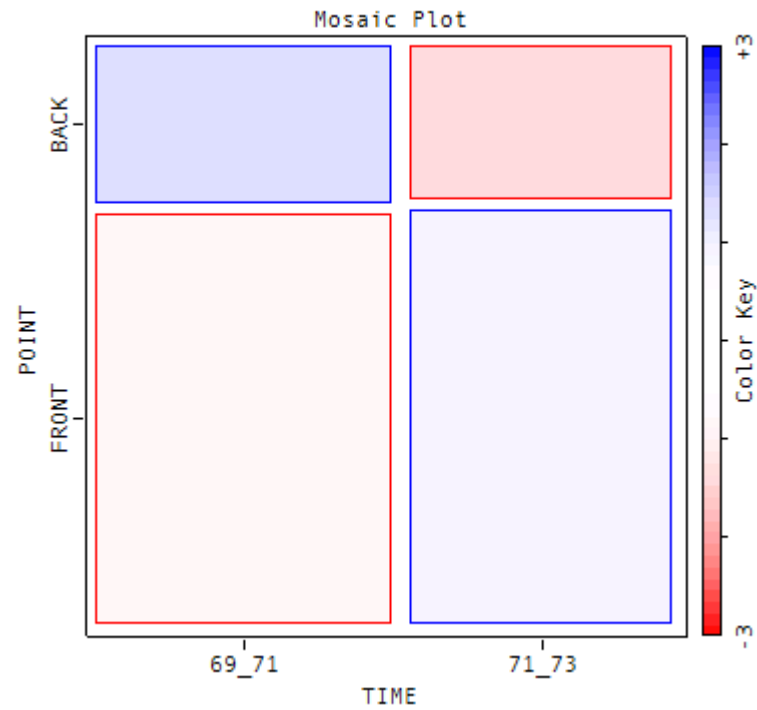




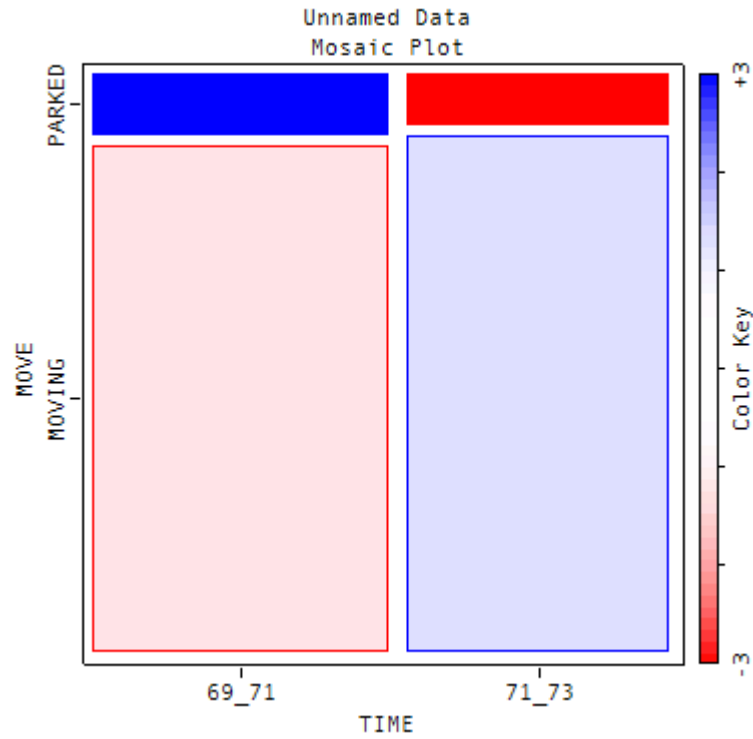
5. Accidentes con camiones

- En unos datos de Anderson se analizan accidentes con camiones en dos periodos de tiempo. El primer periodo es el inmediatamente anterior a la introducción de una medida para aumentar la visibilidad en camiones que se introdujo en 1971.
- Los datos incluyen variables acerca de si el accidente fue por la noche (con carretera iluminada o no) o por el día, si el camión estaba en marcha o parado, y donde fue el golpe (Frente-lados o por detrás)

EJERCICIO 5.1 ¿Dirías que la proporción de accidentes por delante o por detrás en el periodo 71-73 fue mayor o menor que en el periodo 69-71? ¿Podríamos decir que la medida introducida en el año 1971 redujo los accidentes con choque por detrás?



EJERCICIO 5.2 ¿Dirías que ha habido una reducción de accidentes en los que los camiones están aparcados?



EJERCICIO 5.3 Con respecto a las conducciones de luz (DAYLIGHT=Luz del día; Night_Illum=Noche iluminada; Night_Dark =Noche sin iluminar), ¿cuándo dirías que hubo proporcionalmente más reducción en los accidentes? ¿Por qué?

DataType:Frcncy	69 71	71 73	RowSum
Size:4 X 3	Numeric	Numeric	Numeric
DAYLIGHT	14210.	12933.	27143.
NIGHT_ILLUM	2310.	1736.	4046.
NIGHT_DARK	1988.	1597.	3585.
ColSum	18508.	16266.	34774.

EJERCICIO 5.4 ¿Comparativamente hablando, cuándo se dan más accidentes de frente (FRONT) que por detrás (BACK) según esta tabla? ¿Por qué?

Data Type: Frqncy	DAYLIGH	NIGHT I	NIGHT D	RowSum
Size: 3 X 4	Numeric	Numeric	Numeric	Numeric
BACK	6255.	1653.	1597.	9505.
FRONT	20888.	2393.	1988.	25269.
ColSum	27143.	4046.	3585.	34774.

EJERCICIO 5.5 En la tabla de abajo tienes el recuento del cruce entre si estaba aparcado el camión o no y los dos periodos de tiempo considerados. ¿Dirías que hay una reducción en los accidentes en el segundo periodo? ¿En qué caso? ¿Por qué?

DataType:Frqncy	69 71	71 73	RowSum
Size:3 X 3	Numeric	Numeric	Numeric
PARKED	2024.	1476.	3500.
MOVING	16484.	14790.	31274.
ColSum	18508.	16266.	34774.

6. Composición del claustro de la universitat de València

- En la página sobre información de los resultados de elecciones se puede encontrar la información sobre la composición del claustro de la universidad por género en el año 2003 (la tabla está a continuación)
- Las abreviaturas significan: PFD (Personal Funcionario Docente), RPDI (Resto de Personal Docente Investigador), Estudiantes y PAS (Personal de Servicios)

COMPOSICIÓ DEL CLAUSTRÉ 2002 (28 DE MAIG) ESPECIFICADA PER SEXE

	PFD		RFDI		ESTUDIANTS		PAS	
	M	F	M	F	M	F	M	F
Facultat de Geografia i Història	8	2	1	1	2	4		
Facultat de Filologia	7	4	2	0	3	3		
Facultat de Filosofia i Ciències de l'Educació	5	3	0	1	2	2		
Facultat de Psicologia	5	6	0	2	3	2		
Facultat de Medicina i Odontologia	16	1	7	0	1	3		
Escola Universitària d'Infermeria	1	0	0	1	1	0		
Escola Universitària de Fisioteràpia	1	0	0	1	1	0		
Escola Universitària de Magisteri "Ausias March"	3	1	0	1	1	2		
Facultat de Ciències de l'Activitat Física i l'Esport	2	0	1	0	0	0		
Facultat de Química	7	4	0	1	2	1		
Facultat de Física	10	0	4	0	4	0		
Facultat de Ciències Matemàtiques	3	4	1	0	1	0		
Facultat de Ciències Biològiques	8	3	1	1	3	1		
Facultat de Farmàcia	9	3	1	1	3	2		
Facultat de Dret	9	4	2	2	3	7		
Facultat d'Economia	11	7	2	2	8	5		
Facultat de Ciències Socials	5	0	0	2	1	5		
Campus de Blasco Ibañez							6	2
Campus de Burjassot-Paterna							4	2
Campus dels Tarongers							2	1
Servei d'Informació Bibliogràfica							2	2
Altres Centres i la resta de Serv. Gnrls de la Univ.							3	1
Serveis Centrals Administratius i Econòmics							1	4
TOTAL	110	42	22	16	39	37	18	12

TOTAL CLAUSTRALS	M	F	
	189 (63,85%)	107 (36,15%)	296

EJERCICIO 6.1 Indica como organizarías la tabla anterior para comprobar si el estamento influye en la composición por sexos del claustro de la universidad. En este caso no tenemos interés en los resultados por facultad. Usar solamente los totales de la universidad.

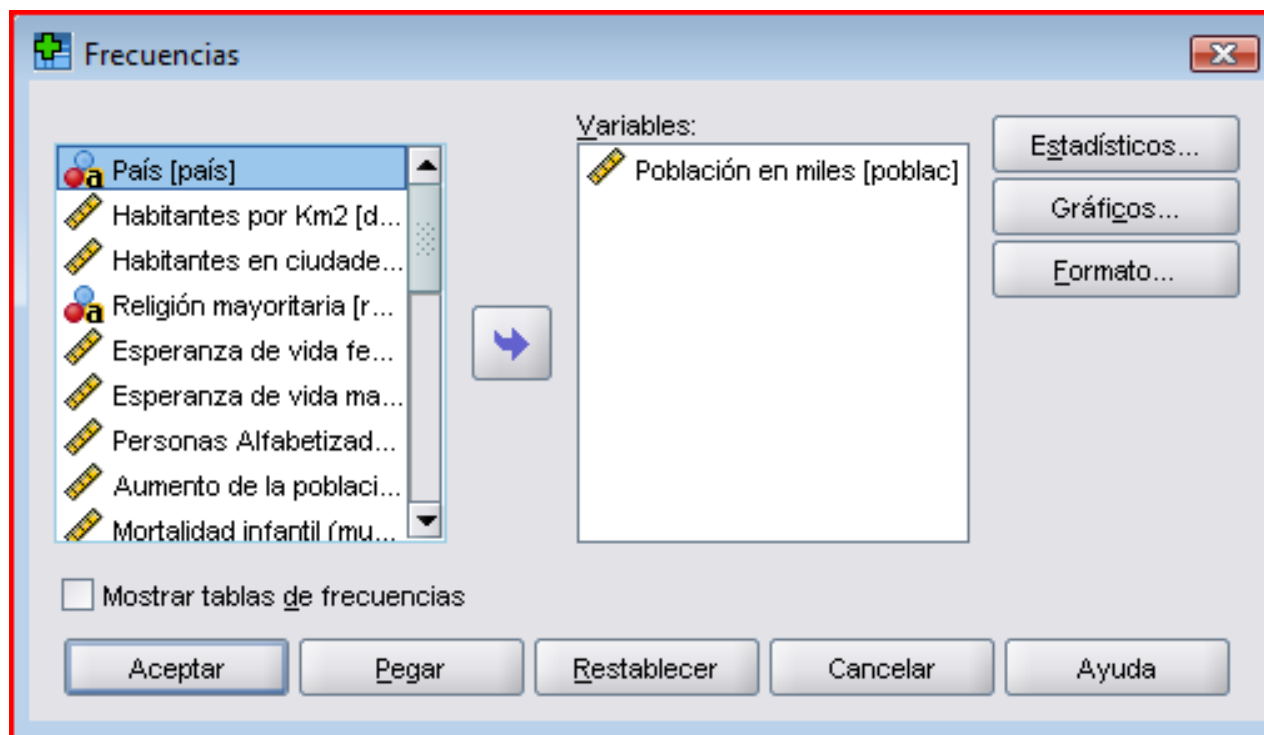
EJERCICIO 6.2 Justifica basándote en los datos si existe aparentemente algún tipo de discriminación en los resultados de composición del claustro sin tener en cuenta los resultados por facultad. Muestra qué cálculos has hecho para llegar a esa conclusión.

EJERCICIO 6.3 Busca una representación gráfica apropiada para describir la posible discriminación. ¿Qué estamento es el que hay más diferencias? Haz que la representación gráfica muestre claramente qué estamento es el que hay más diferencias.

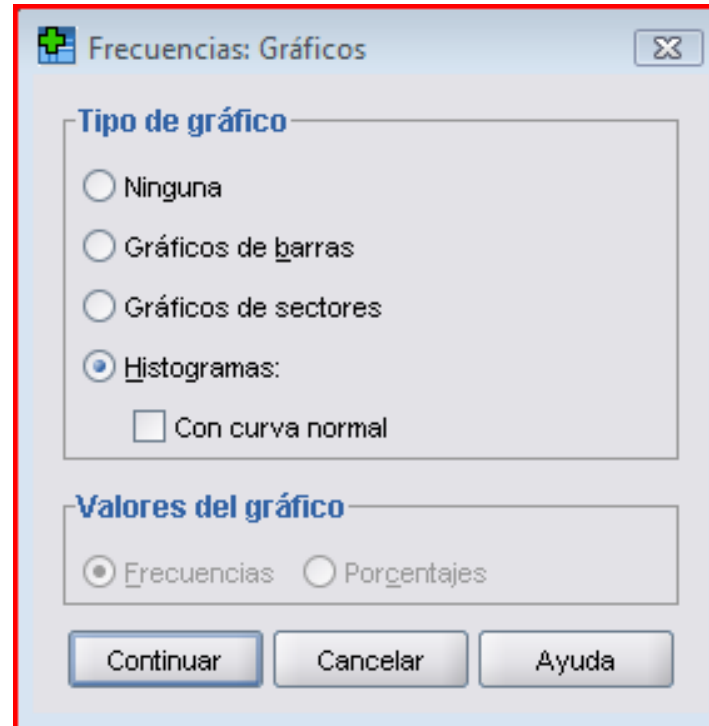
Histogramas y medias con el SPSS (PASW)

- Esta actividad introduce como calcular histogramas y medias con el PASW

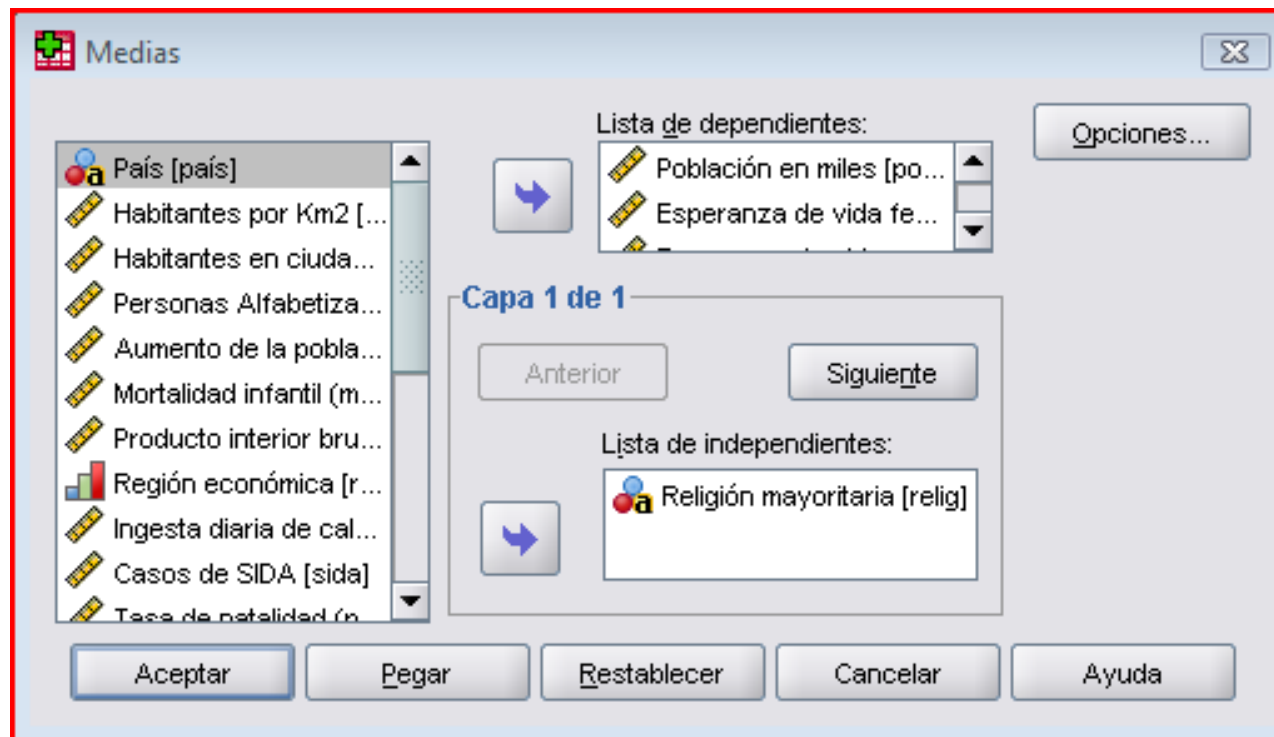
- Para calcular histogramas hay que ir después de abrir el archivo correspondiente al menú Analizar, subcomando Estadísticos Descriptivos y de nuevo subcomando Frecuencias. En el cuadro de diálogo que aparece ir al



botón Gráficos y allí elegir Histogramas.



- Para calcular medias podeis hacerlo en Descriptivos que está dentro de Estadísticos Descriptivos o si se os pide que calculeis medias en función de una variable categórica en el menú Medias que está dentro del submenú de Comparar Medias.



-
-
- En ese cuadro de diálogo poneis la variable categórica en las independientes y la numérica o numéricas en la dependiente.

Cálculo con el PASW e interpretación de histogramas y medias

***Fecha de Entrega Grupo B 29/10, Grupos G, H 30/
10***

7. Datos del mundo 95

EJERCICIO 7.1 **Calcula con el PASW para *seis variables de estos datos (puedes elegir las 6 que quieras)* los histogramas para todas las variables que sea posible, pega los gráficos y comentalos teniendo en cuenta la explicación de las clases teóricas al respecto. Indica qué variables no es apropiado calcular histogramas y la razón a que se debe esto. Intenta también identificar individualmente los países que aparecen como más destacados en los gráficos e indica las razones por las que destacan.**

EJERCICIO 7.2 Calcula las medias y las desviaciones típicas de las 6 variables que hayas elegido anteriormente

EJERCICIO 7.3 **Calcula las medias de de los países en las 6 variables que has elegido del archivo de datos en función del clima que tienen en esos países, pega las tablas correspondientes en el documento y comenta donde se dan más diferencias.**

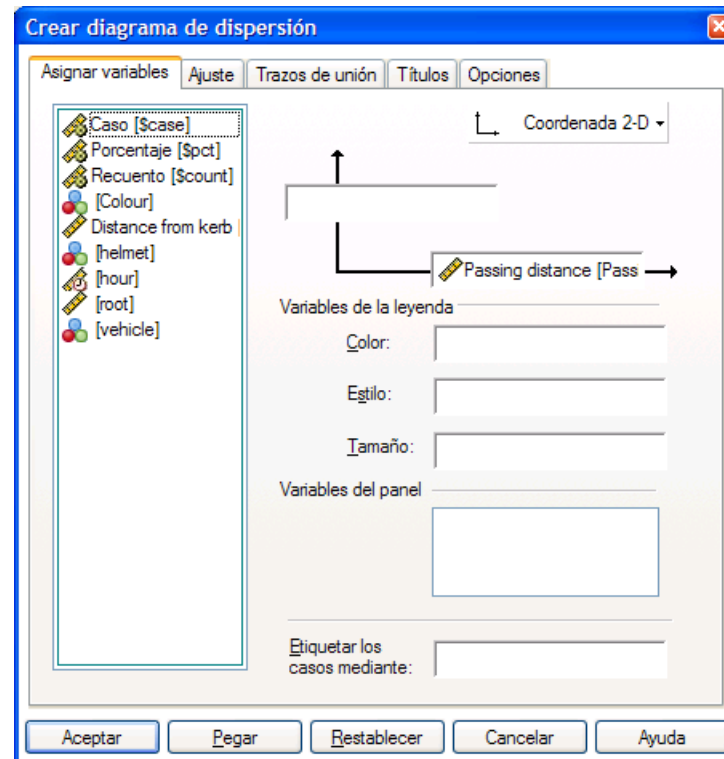
EJERCICIO 7.4 Elige otra variable categórica que puedas usar como variable independiente y analiza todas las variables (no sólo 6) en función de esa variable categórica. Sólo pega en el archivo de respuesta *las variables que muestren resultados relevantes y comentalas.*

Más sobre gráficos y análisis en el SPSS

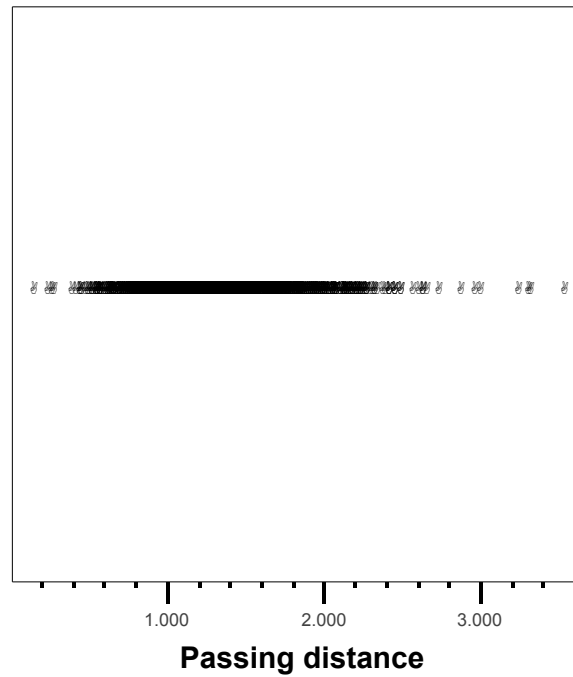
- Utilizaremos el archivo de paseando en bicicleta. Abriendo ese archivo vemos una imagen de este estilo.

	vehicle	Colour	hour	helmet	Distance_fr	Passing_di	root	var
1	car	blue	16:00:00	0	.5	1.864	1.37	
2	HGV	red	16:00:00	0	.5	.748	.86	
3	LGV	blue	16:00:00	0	.5	1.567	1.25	
4	car	unknown	16:00:00	0	.5	1.390	1.18	
5	bus	other	16:00:00	0	.5	1.294	1.14	
6	car	silver/grey	16:00:00	0	.5	1.259	1.12	
7	LGV	white	7:00:00	1	1.0	1.040	1.02	
8	car	red	7:00:00	1	1.0	1.262	1.12	
9	bus	red	7:00:00	1	1.0	.799	.89	
10	PTW	unknown	7:00:00	1	1.0	1.682	1.30	
11	LGV	white	7:00:00	1	1.0	.895	.95	
12	SUV	white	7:00:00	1	1.0	1.160	1.08	

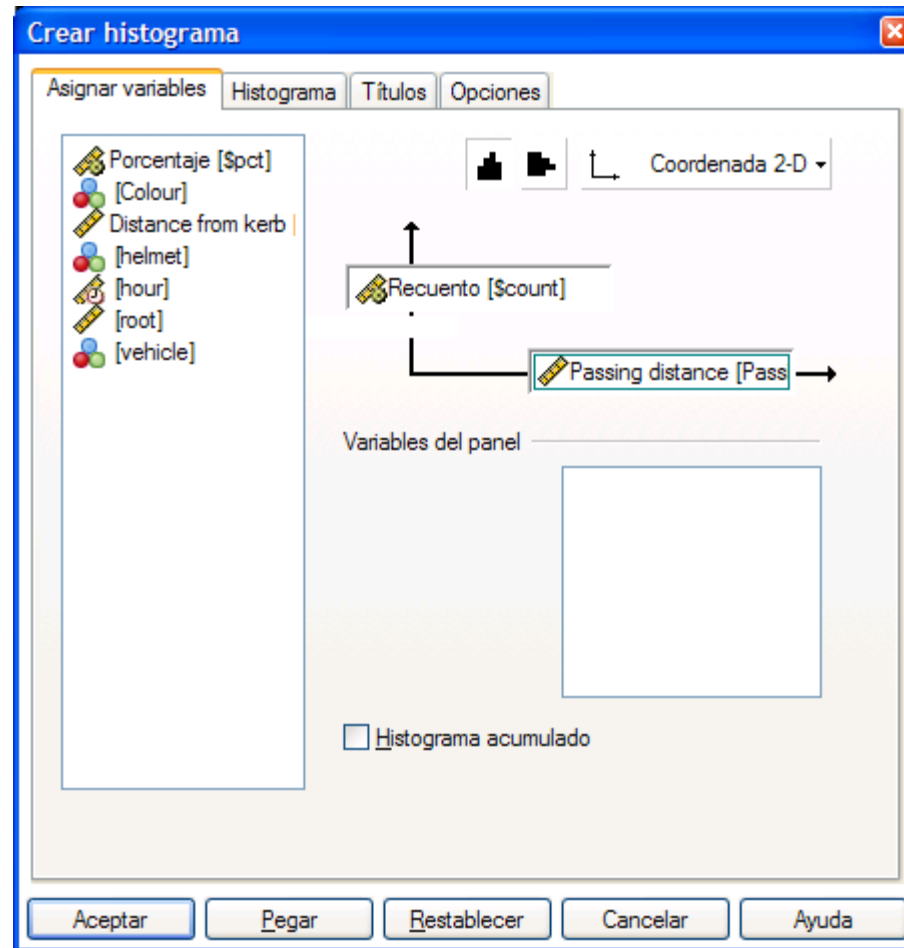
- Para hacer un diagrama de puntos podeis ir a Graficos>Cuadros de diálogo antiguos>interactivos>diagramas de dispersión y poner la variable passing distance en el eje X



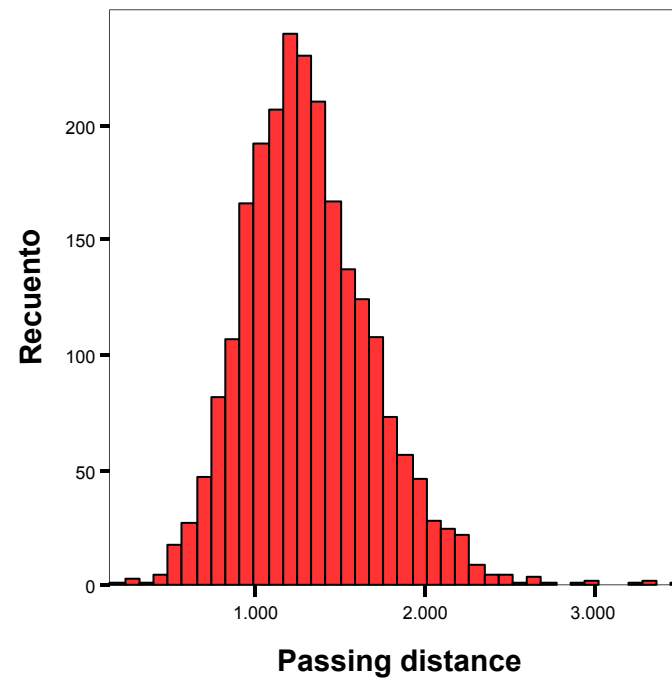
- El resultado es el siguiente



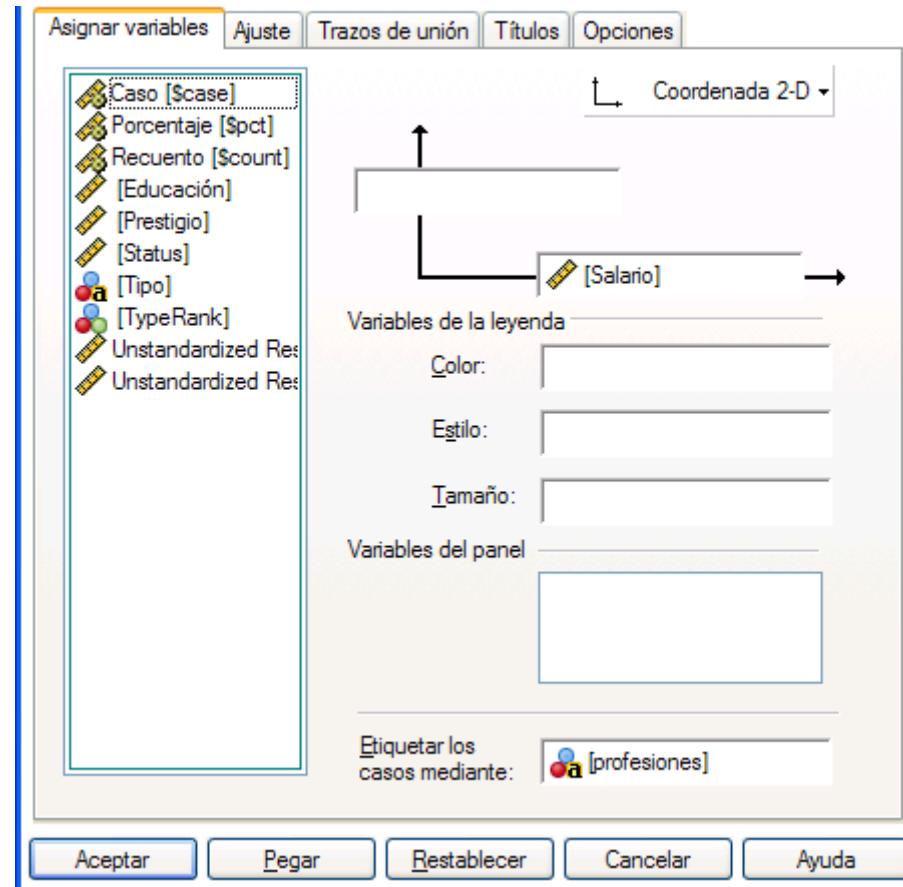
- Para hacer un histograma de "passing distance" podemos ir al menu graficos>interactivos>histograma



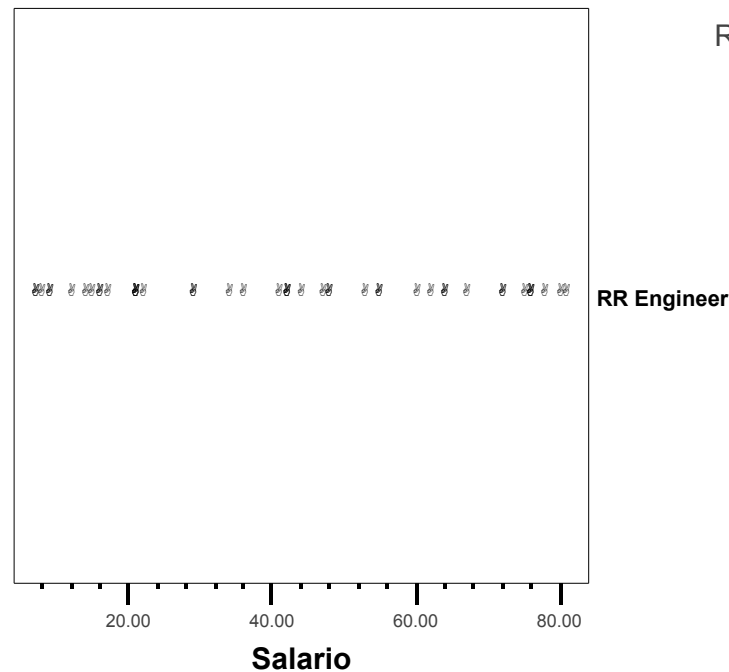
- El resultado es el siguiente



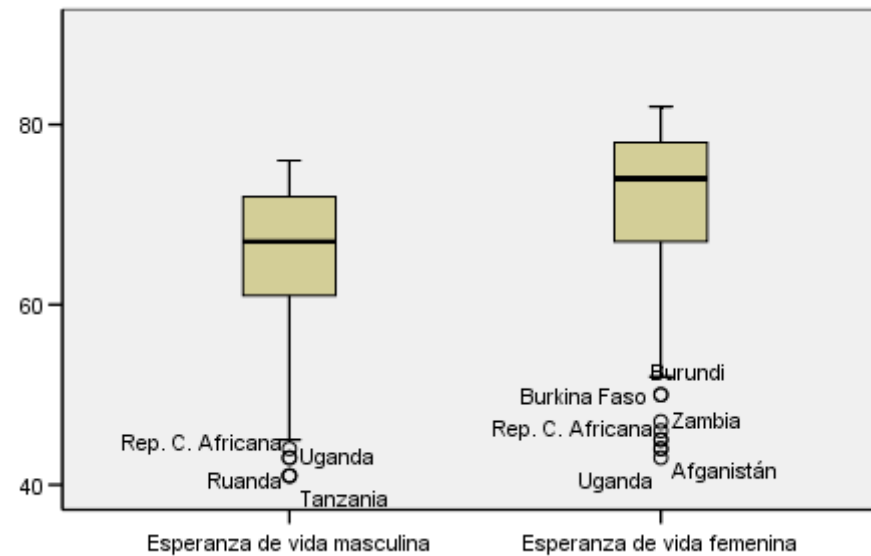
- Para ponerle etiquetas a los puntos hay que añadir una variable que tenga las etiquetas. Un buen ejemplo es el archivo de profesiones que está en la página del curso



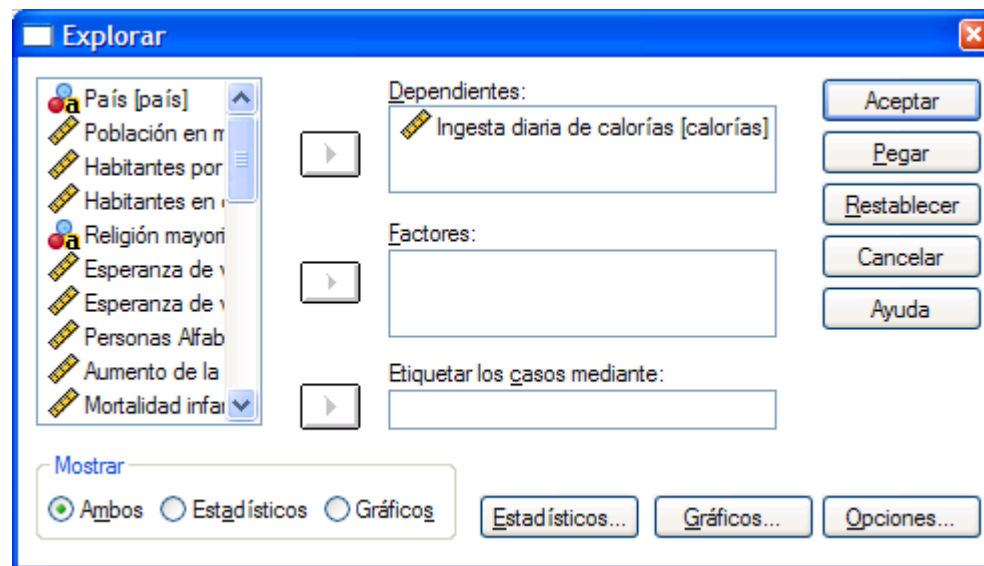
- Con bastante esfuerzo y usando una herramienta que está a la izquierda se puede conseguir esto. Vemos que la profesión mejor pagada es RR Engineer



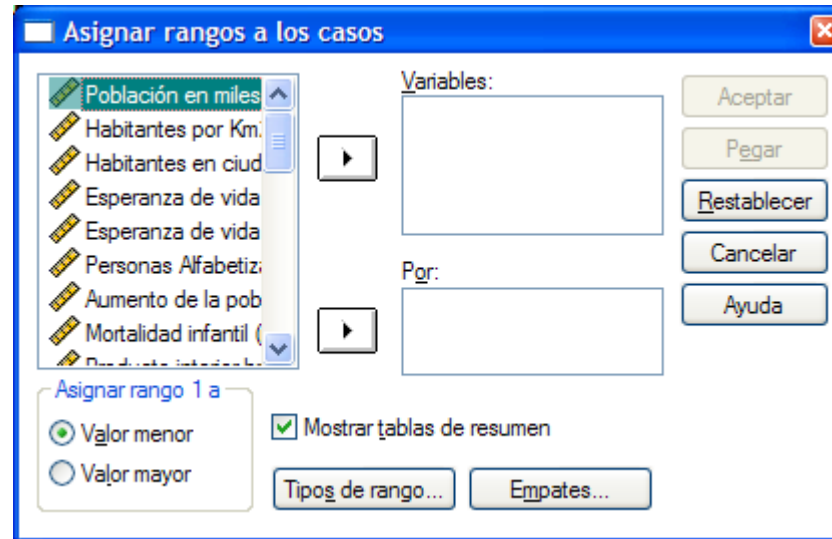
- Para comparar dos variables podemos usar el archivo de datos de Mundo95. Los gráficos de cajas y bigotes están en Gráficos>Cuadros de diálogo antiguos>Diagramas de caja



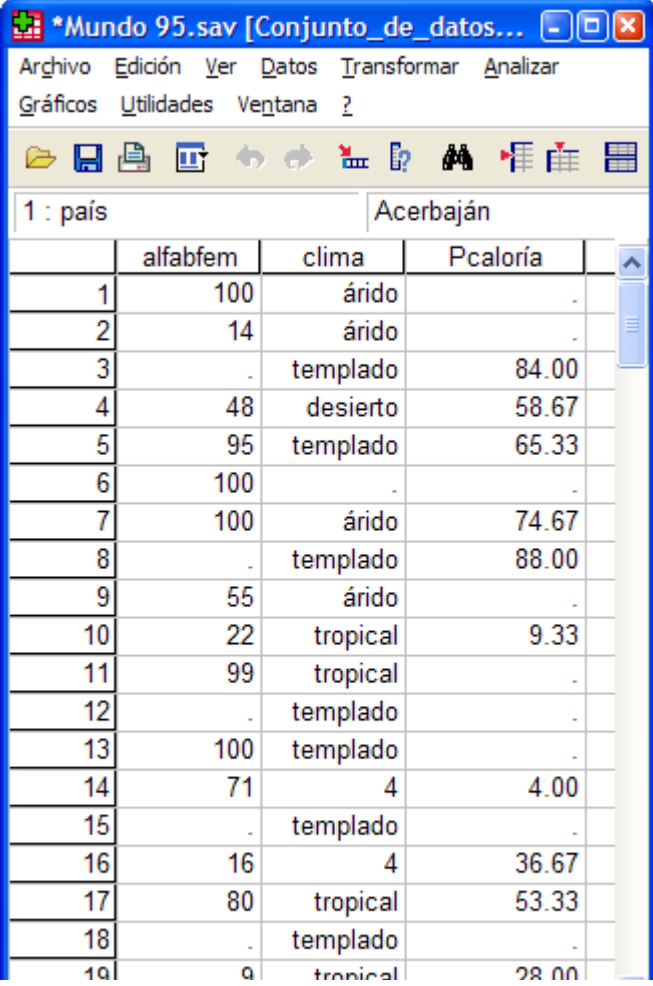
- Para ver los resultados numéricos ir a Analizar>estadísticos descriptivos>Explorar (tener en cuenta que el resultado incluye más cosas que las que usamos en clase)



- Para calcular rangos percentiles vamos a transformar y asignar rangos a casos. En ese cuadro vamos a tipo de rango y elegimos rango fraccional como porcentaje



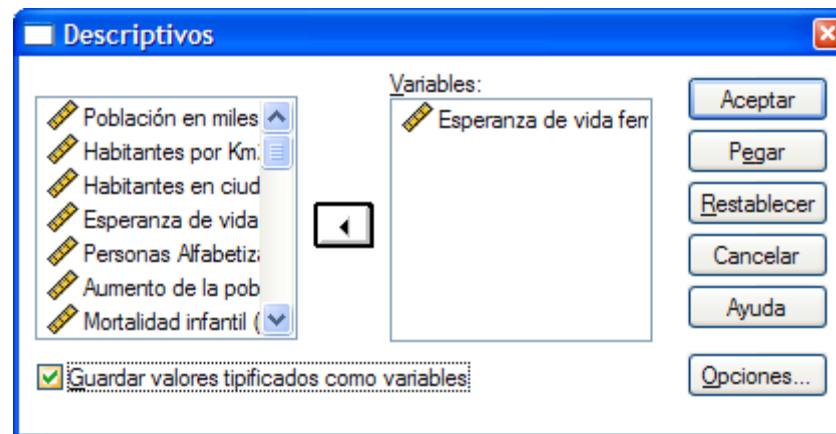
- Tener en cuenta que el resultado aparece como una nueva variable en la hoja de datos (se llama Pcaloría)



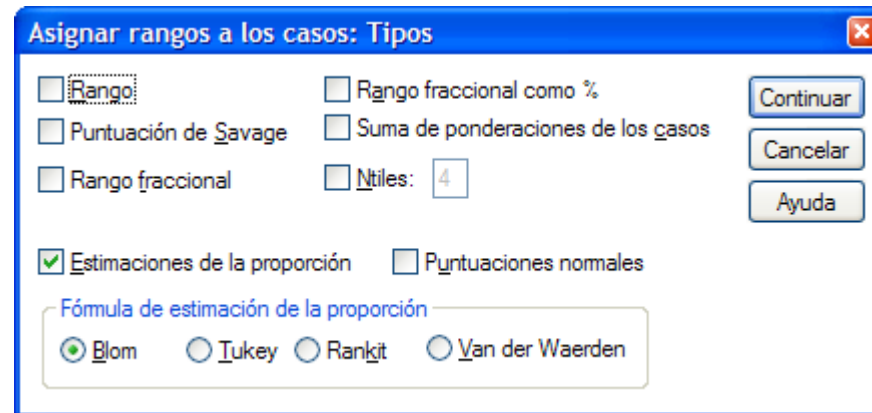
The screenshot shows the SPSS data editor window for a file named '*Mundo 95.sav [Conjunto_de_datos...'. The window title bar includes standard OS window controls. The menu bar contains 'Archivo', 'Edición', 'Ver', 'Datos', 'Transformar', and 'Analizar'. Below the menu bar is a toolbar with various icons for file operations and data manipulation. The main area displays a data grid with the following columns: '1 : país', 'alfabfem', 'clima', and 'Pcaloría'. The data is organized into a table with 19 rows and 4 columns. The '1 : país' column contains values from 1 to 19. The 'alfabfem' column contains numerical values. The 'clima' column contains categorical values like 'árido', 'templado', 'desierto', and 'tropical'. The 'Pcaloría' column contains numerical values, some of which are zero or a small number, and some are decimal values.

1 : país	alfabfem	clima	Pcaloría
1	100	árido	.
2	14	árido	.
3	.	templado	84.00
4	48	desierto	58.67
5	95	templado	65.33
6	100	.	.
7	100	árido	74.67
8	.	templado	88.00
9	55	árido	.
10	22	tropical	9.33
11	99	tropical	.
12	.	templado	.
13	100	templado	.
14	71	4	4.00
15	.	templado	.
16	16	4	36.67
17	80	tropical	53.33
18	.	templado	.
19	9	tropical	28.00

- Para calcular puntuaciones típicas puedes ir a Analizar>Estadísticos descriptivos>Descriptivos y en el cuadro de diálogo marcar "Guardar valores tipificados como variables"



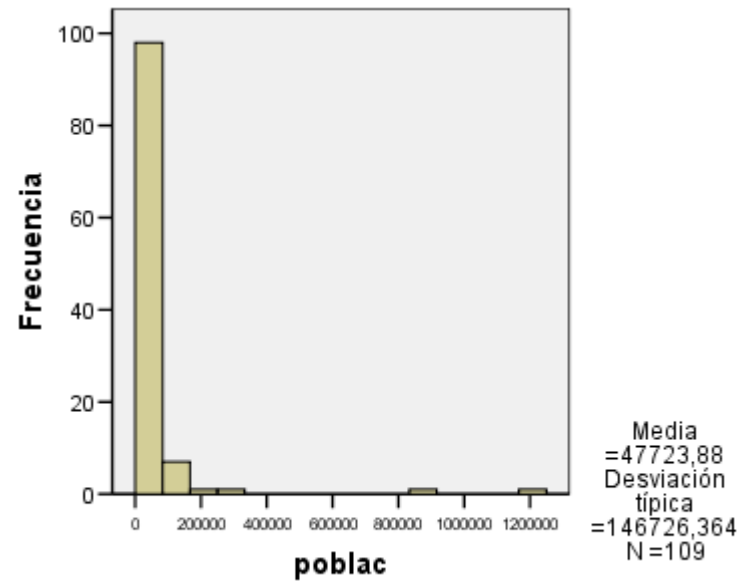
- Para calcular puntuaciones normalizadas ir a Transformar>Asignar rangos a casos>tipos de rango>estimaciones de la proporción



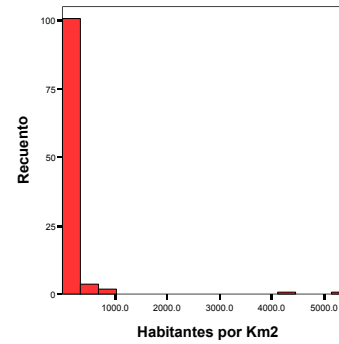
8. Histogramas para las variables del Mundo

- Los datos de Mundo95 tienen diferentes variables acerca de países del mundo. En este caso vamos a evaluar los histogramas de esas variables

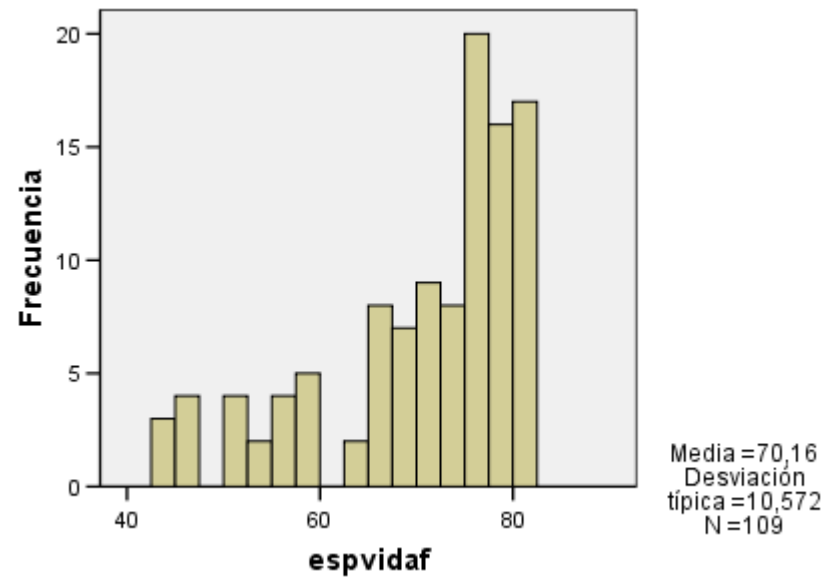
EJERCICIO 8.1 Describe las características principales del histograma de abajo e intenta explicar (si puedes) cual puede ser la explicación de la forma que tiene. En este caso tenemos el gráfico de la población en los países.



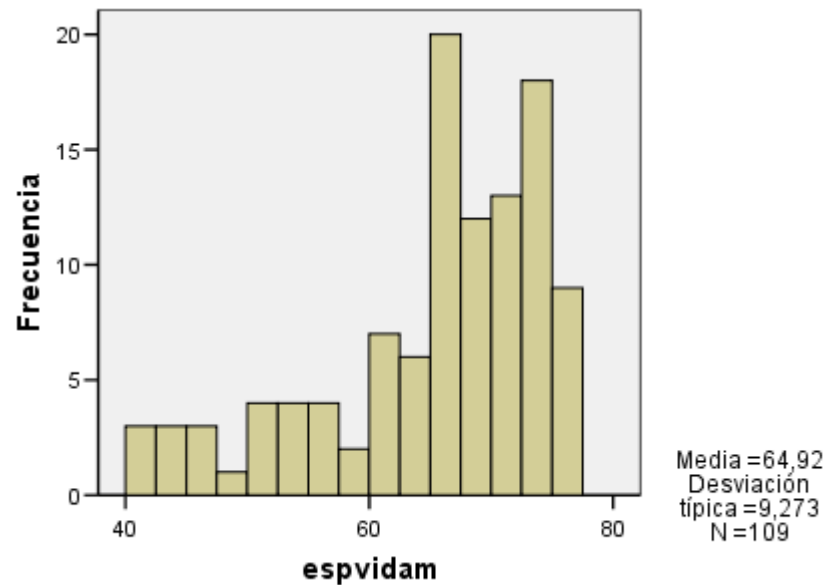
EJERCICIO 8.2 Describe las características e intenta explicarlas. En este caso tenemos la densidad de habitantes (habitantes por km cuadrado)



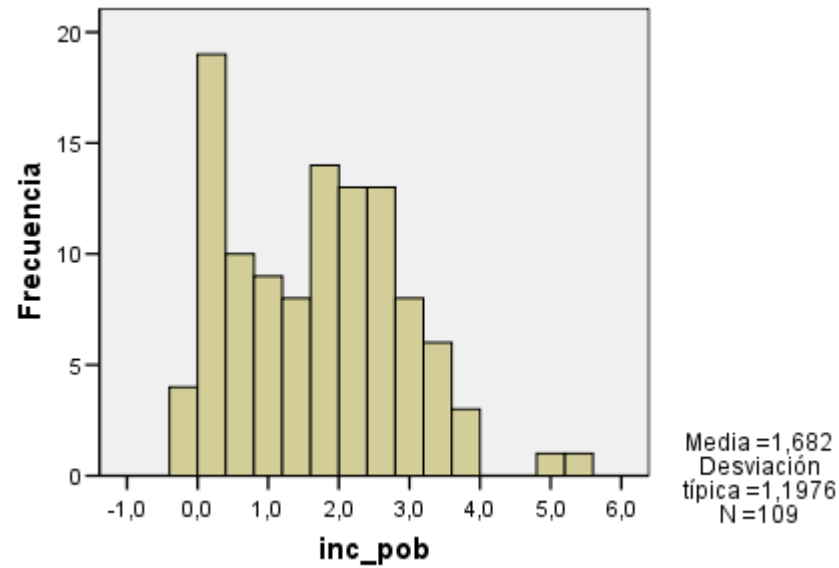
EJERCICIO 8.3 Describe la esperanza de vida femenina



EJERCICIO 8.4 Describe la esperanza de vida masculina. Compara este gráfico con el de la esperanza de vida femenina. ¿Qué método de los que hemos aprendido en clase sería mejor para hacer esta comparación?

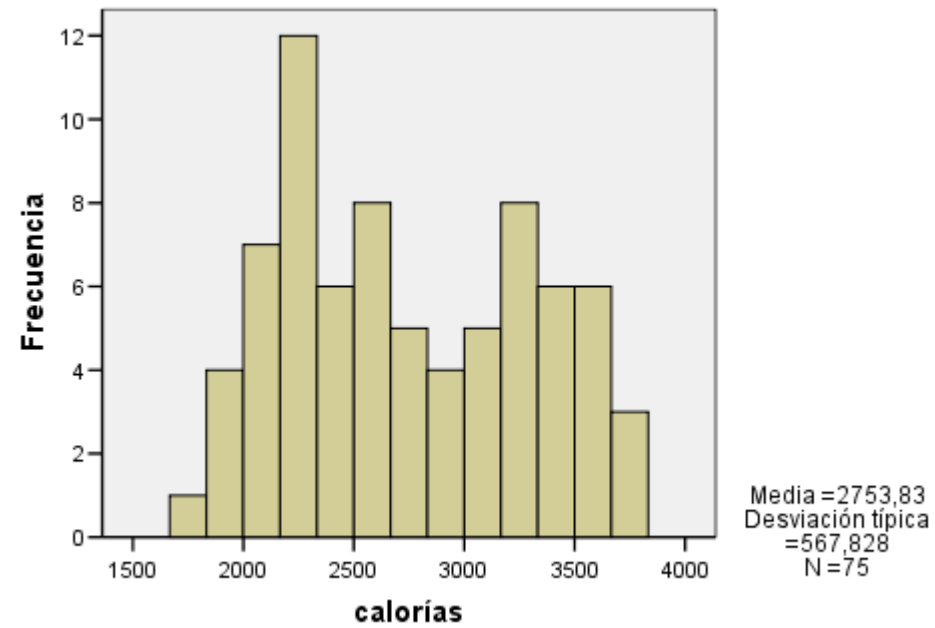


EJERCICIO 8.5 Describe el gráfico para el incremento de la población en tantos por ciento anual. Usando el SPSS y el archivo de datos que esta en la página del curso averigua qué países son los que tienen un incremento cercano al 5%.

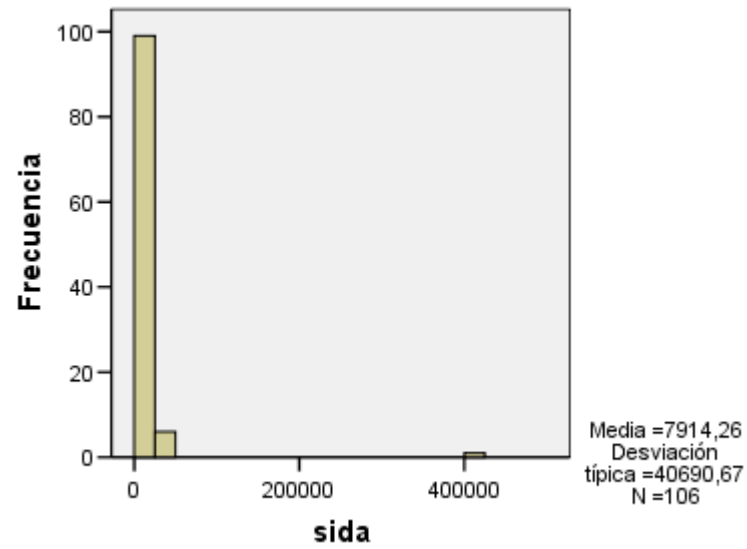


EJERCICIO 8.6 Comenta el gráfico de la mortalidad infantil (muertos por 1000 nacimientos vivos)

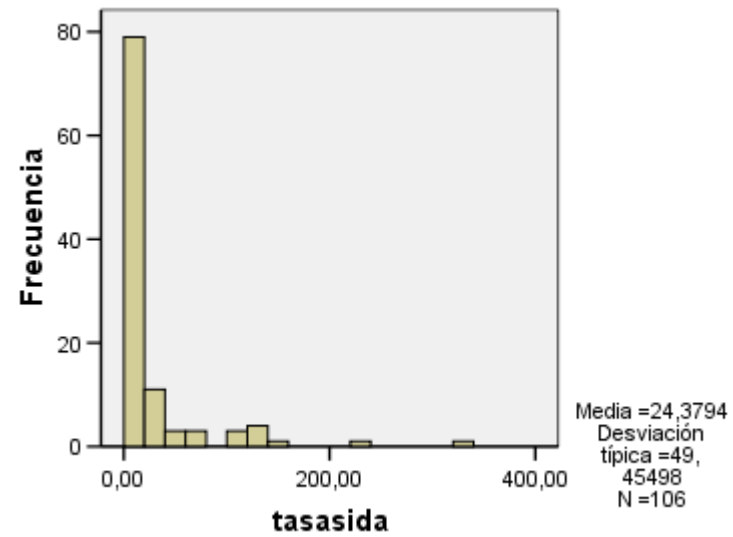
EJERCICIO 8.7 Comenta la ingesta media de calorías



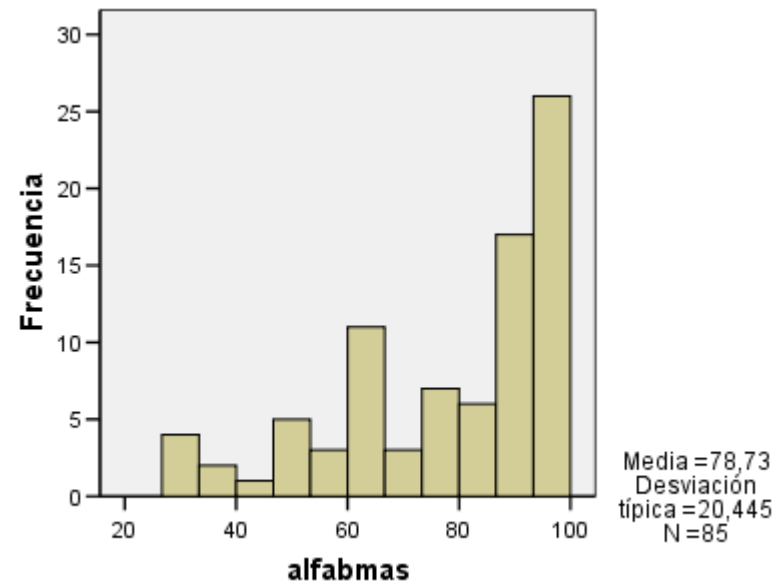
EJERCICIO 8.8 Comenta los casos de sida totales.



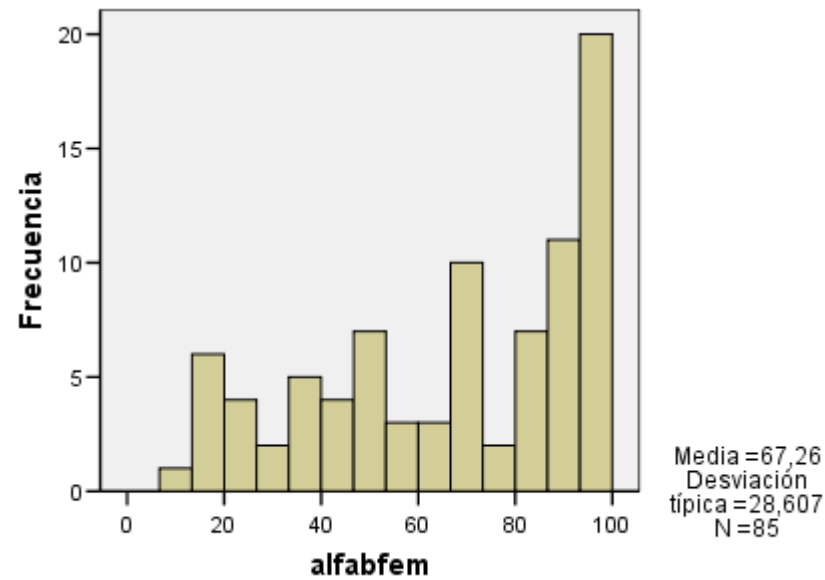
EJERCICIO 8.9 Comenta los casos de sida por 100000 habitantes. ¿Dirías que el valor extremo en este gráfico es el mismo valor que en el ejercicio 8.8.



EJERCICIO 8.10 Comenta la alfabetización masculina



EJERCICIO 8.11 Comenta la alfabetización femenina. Comparala con la alfabetización masculina. ¿Qué gráfico sería mas interesante para hacer esta comparación de los que hemos visto en clase?



EJERCICIO 8.12 Comenta los siguientes estadísticos descriptivos para la variable Población de cada país (en miles). ¿Puedes explicar la diferencia entre la media y la mediana encontrada en este resultado?

Descriptivos

		Estadístico	Error ttp.
Población en miles	Media	47723.88	14053.837
	Intervalo de confianza para la media al 95%	Límite inferior 19866.74 Límite superior 75581.02	
	Media recortada al 5%	24432.96	
	Mediana	10400.00	
	Varianza	2.2E+010	
	Desv. ttp.	146726.364	
	Mínimo	256	
	Máximo	1205200	
	Rango	1204944	
	Amplitud intercuartil	32100	
	Asimetría	6.592	.231
	Curtosis	46.651	.459

EJERCICIO 8.13 Comenta los siguientes estadísticos descriptivos para la variable densidad de habitantes por kilómetro cuadrado de cada país. ¿Puedes explicar la diferencia entre la media y la mediana encontrada en este resultado?

Descriptivos

		Estadístico	Error típ.	
Habitantes por Km2	Media	203.415	64.7208	
	Intervalo de confianza para la media al 95%	Límite inferior	75.127	
		Límite superior	331.703	
	Media recortada al 5%	99.007		
	Mediana	64.000		
	Varianza	456577.508		
	Desv. típ.	675.7052		
	Mínimo	2.3		
	Máximo	5494.0		
	Rango	5491.7		
	Amplitud intercuartil	100.0		
	Asimetría	6.887	.231	
	Curtosis	49.076	.459	

EJERCICIO 8.14 Comenta los siguientes estadísticos descriptivos para la variable esperanza de vida femenina de cada país. ¿Puedes explicar la diferencia entre la media y la mediana encontrada en este resultado?

Descriptivos

			Estadístico	Error típ.
Esperanza de vida femenina	Media		70.16	1.013
	Intervalo de confianza para la media al 95%	Límite inferior	68.15	
		Límite superior	72.16	
	Media recortada al 5%		70.96	
	Mediana		74.00	
	Varianza		111.762	
	Desv. típ.		10.572	
	Mínimo		43	
	Máximo		82	
	Rango		39	
	Amplitud intercuartil		12	
	Asimetría		-1.109	.231
	Curtosis		.213	.459

EJERCICIO 8.15 Comenta los siguientes estadísticos descriptivos para la variable esperanza de vida masculina de cada país. ¿Puedes explicar la diferencia entre la media y la mediana encontrada en este resultado?

Descriptivos

			Estadístico	Error típ.
Esperanza de vida masculina	Media		64.92	.888
	Intervalo de confianza para la media al 95%	Límite inferior	63.16	
		Límite superior	66.68	
	Media recortada al 5%		65.59	
	Mediana		67.00	
	Varianza		85.984	
	Desv. típ.		9.273	
	Mínimo		41	
	Máximo		76	
	Rango		35	
	Amplitud intercuartil		12	
	Asimetría		-1.080	.231
	Curtosis		.336	.459

EJERCICIO 8.16 Comenta los resultados numéricos para la mortalidad infantil. ¿Puedes explicar la diferencia entre la media y la mediana?

Descriptivos

			Estadístico	Error típ.
Mortalidad infantil (muertes por 1000 nacimientos vivos)	Media		42.313	3.6473
	Intervalo de confianza para la media al 95%	Límite inferior	35.083	
		Límite superior	49.543	
	Media recortada al 5%		39.341	
	Mediana		27.700	
	Varianza		1450.027	
	Desv. típ.		38.0792	
	Mínimo		4.0	
	Máximo		168.0	
	Rango		164.0	
	Amplitud intercuartil		55.3	
	Asimetría		1.090	.231
	Curtosis		.365	.459

EJERCICIO 8.17 Comenta los resultados para la ingesta media de calorías. ¿Que valores dirías que representan mejor la tendencia central en este caso? Para contestar ten en cuenta el histograma de esta variable mostrado en el ejercicio 8.7.

Descriptivos

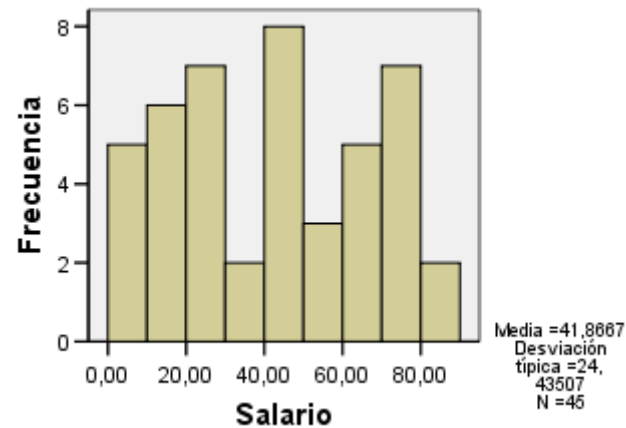
		Estadístico	Error típ.
Ingesta diaria de calorías	Media	2753.83	65.567
	Intervalo de confianza para la media al 95%		
	Límite inferior	2623.18	
	Límite superior	2884.47	
	Media recortada al 5%	2749.51	
	Mediana	2653.00	
	Varianza	322428.334	
	Desv. típ.	567.828	
	Mínimo	1667	
	Máximo	3825	
	Rango	2158	
	Amplitud intercuartil	989	
	Asimetría	.170	.277
	Curtosis	-1.207	.548

9. Profesiones

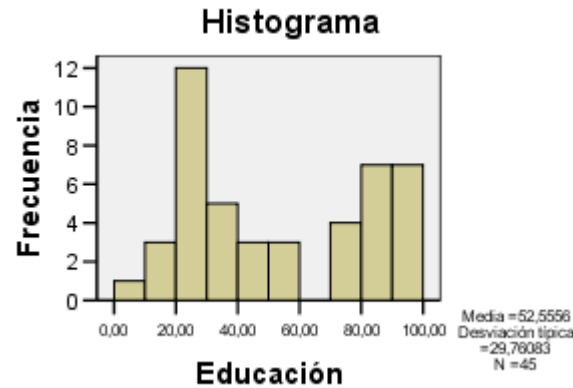
- Utilizaremos los datos acerca del prestigio, salario y nivel de educación de profesiones.

EJERCICIO 9.1 Utilizando la información mostrada abajo comenta qué valor te parece una buena indicación de la tendencia central de la variable salario y por qué.

		Estadístico	Error típ.
Salario	Media	41.8667	3.64257
	Intervalo de confianza para la media al 95%	Límite inferior 34.5256	
		Límite superior 49.2078	
	Media recortada al 5%	41.6667	
	Mediana	42.0000	
	Varianza	597.073	
	Desv. típ.	24.43507	
	Mínimo	7.00	
	Máximo	81.0	
	Rango	74.00	
	Amplitud intercuartil	45.00	

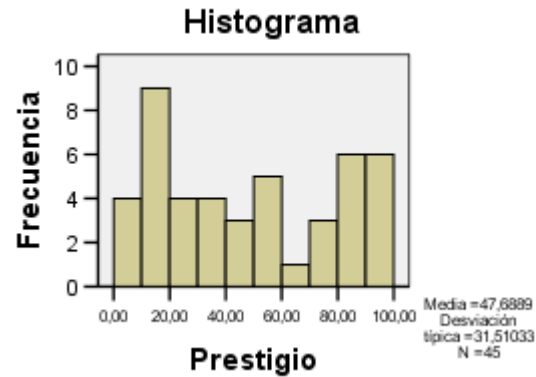


EJERCICIO 9.2 Utilizando la información mostrada abajo comenta cual sería una buena estimación de la tendencia central de la variable educación y por qué.



Educación			
Media		52.5556	4.43648
Intervalo de confianza para la media al 95%	Límite inferior	43.6144	
	Límite superior	61.4967	
Media recortada al 5%		52.2593	
Mediana		45.0000	
Varianza		885.707	
Desv. típ.		29.76083	
Mínimo		7.00	
Máximo		100.0	
Rango		93.00	
Amplitud intercuartil		59.50	
Asimetría		.234	.354
Curtosis		-1.589	.605

EJERCICIO 9.3 Utilizando la información mostrada abajo comenta cual sería una buena estimación de la tendencia central de la variable prestigio profesional y por qué.

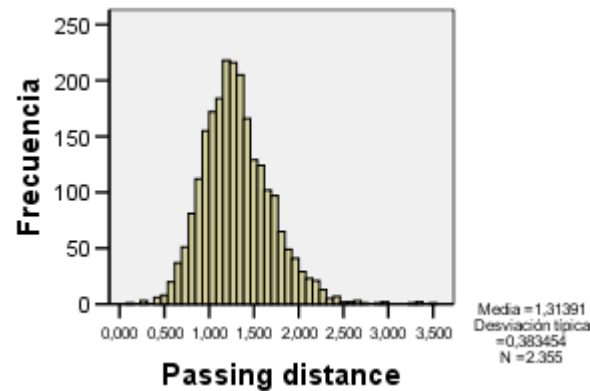


			Estadístico	Error típ.
Prestigio	Media		47.6889	4.69728
	Intervalo de confianza para la media al 95%	Límite inferior	38.2221	
		Límite superior	57.1556	
	Media recortada al 5%		47.4630	
	Mediana		41.0000	
	Varianza		992.901	
	Desv. típ.		31.51033	
	Mínimo		3.00	
	Máximo		97.0	
	Rango		94.00	
	Amplitud intercuartil		65.50	
	Asimetría		.147	.354

10. Paseando en bicicleta

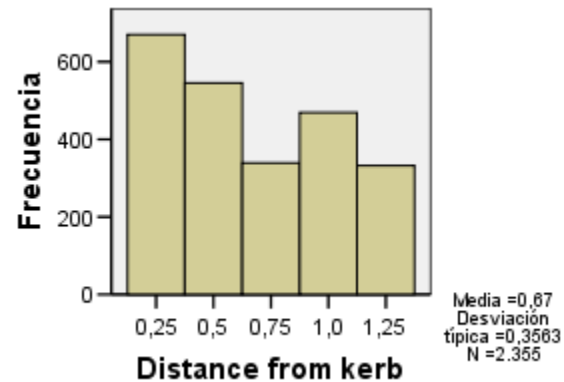
- A continuación analizaremos variables que están en el archivo sobre paseando en bicicleta.

EJERCICIO 10.1 La variable "Passing distance" tiene la distancia en metros a la que pasaron los vehículos al lado del ciclista. A partir de la información de abajo, ¿cuál dirías que es una buena estimación de la tendencia central?



		Estadístico	Error típ.
Passing distance	Media	1.31391	.007902
	Intervalo de confianza para la media al 95%	Límite inferior 1.29842 Límite superior 1.32941	
	Media recortada al 5%	1.30008	
	Mediana	1.27900	
	Varianza	.147	
	Desv. típ.	.383454	
	Mínimo	.144	
	Máximo	3.537	
	Rango	3.393	
	Amplitud intercuartil	.489	
	Asimetría	.722	.050
	Curtosis	1.659	.101

EJERCICIO 10.2 La variable "Distance from the kerb" tiene la distancia en metros a la que estaba el ciclista del bordillo. A partir de la información de abajo, ¿cuál dirías que es una buena estimación de la media de esa distancia?

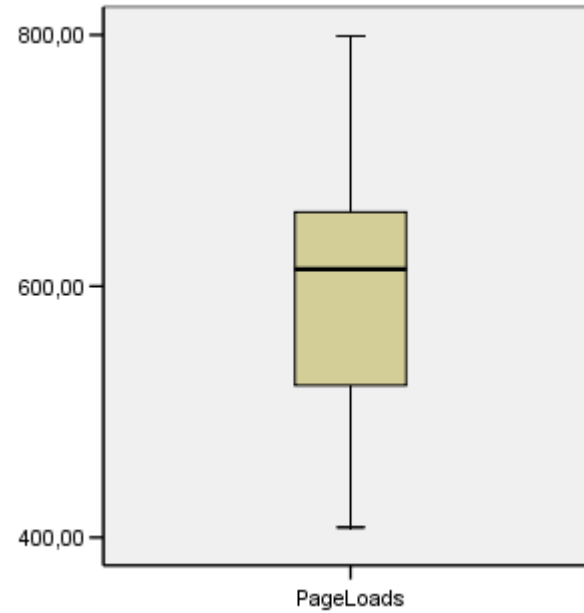


		Estadístico	Error típ.
Distance from kerb	Media	.670	.0073
	Intervalo de confianza para la media al 95%		
	Límite inferior	.656	
	Límite superior	.685	
	Media recortada al 5%	.661	
	Mediana	.500	
	Varianza	.127	
	Desv. típ.	.3563	
	Mínimo	.3	
	Máximo	1.3	
	Rango	1.0	
	Amplitud intercuartil	.8	
	Asimetría	.276	.050

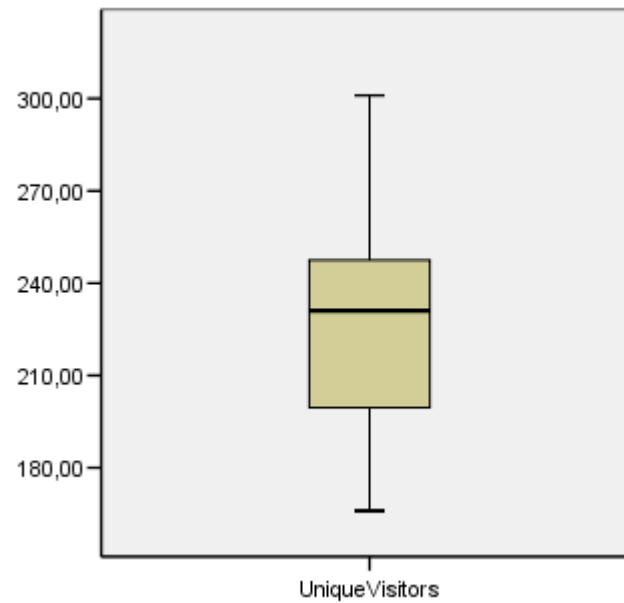
11. Visitas a mi página web

- En este caso analizaremos las visitas realizadas a mi página web sobre el libro Visual Statistics desde Octubre 2006 hasta Septiembre del 2007

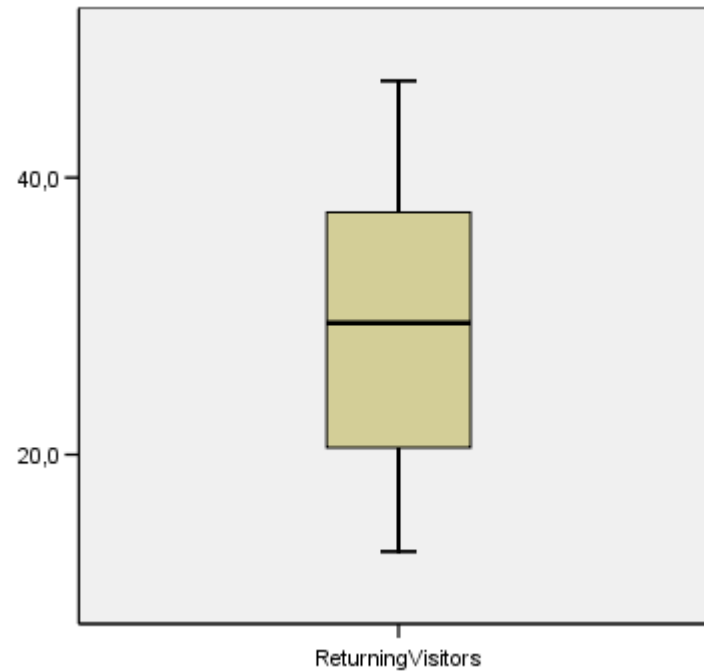
EJERCICIO 11.1 A continuación tienes la información acerca de las páginas cargadas o visitadas cada mes durante los meses indicados anteriormente ¿Qué puedes decir al respecto a partir de este gráfico? Menciona todos los valores que puedes extraer de este gráfico



EJERCICIO 11.2 ¿Qué puedes decir sobre el número de visitantes por mes durante esos meses? Menciona todos los valores que puedes extraer de este gráfico

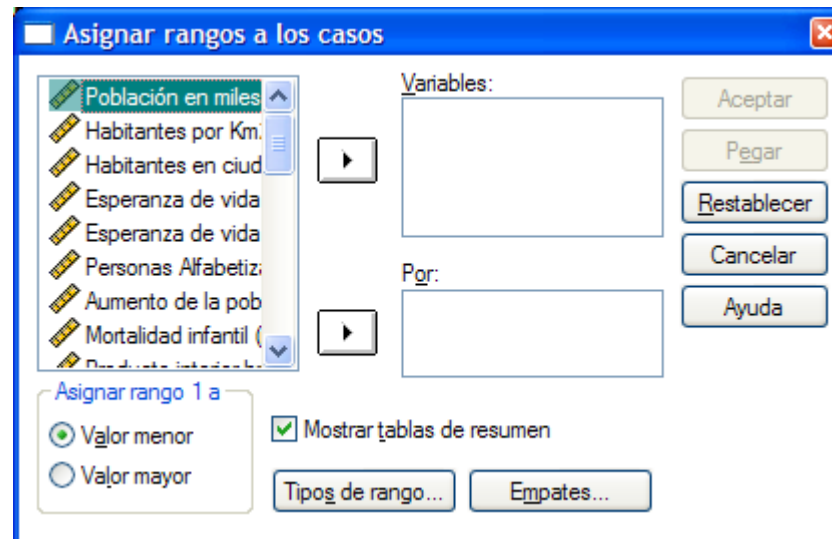


EJERCICIO 11.3 ¿Y sobre el número de visitantes que repitieron visita y volvieron a visitar la página? Menciona todos los valores que puedes extraer de este gráfico

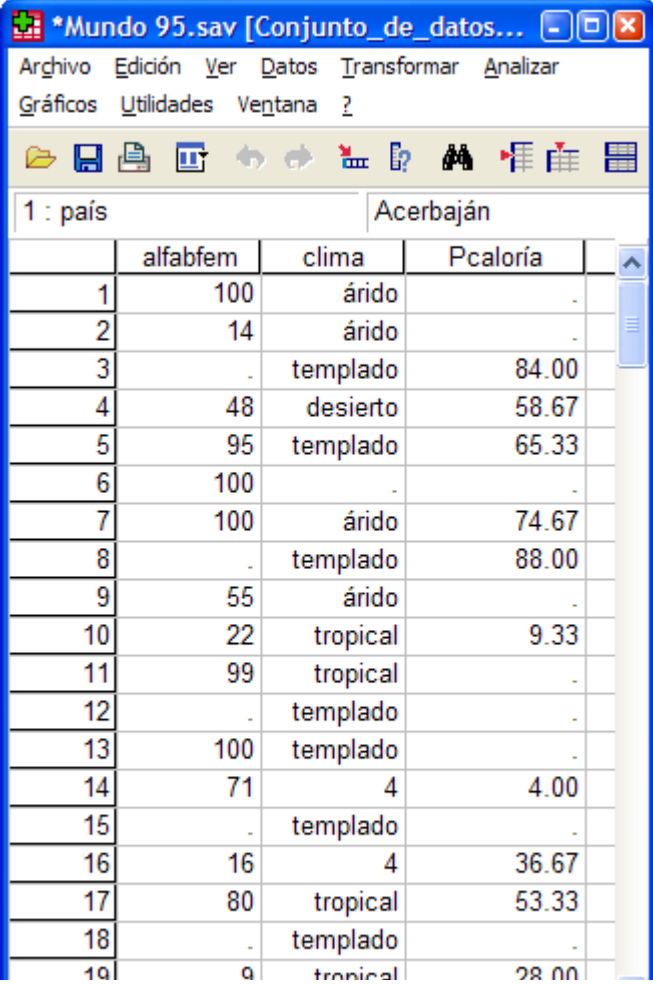


Calculo de puntuaciones especiales con SPSS (rangos percentiles, puntuaciones típicas y puntuaciones normalizadas)

- Para calcular rangos percentiles vamos a transformar y asignar rangos a casos. En ese cuadro vamos a tipo de rango y elegimos rango fraccional como porcentaje



- Tener en cuenta que el resultado aparece como una nueva variable en la hoja de datos (se llama Pcaloría)



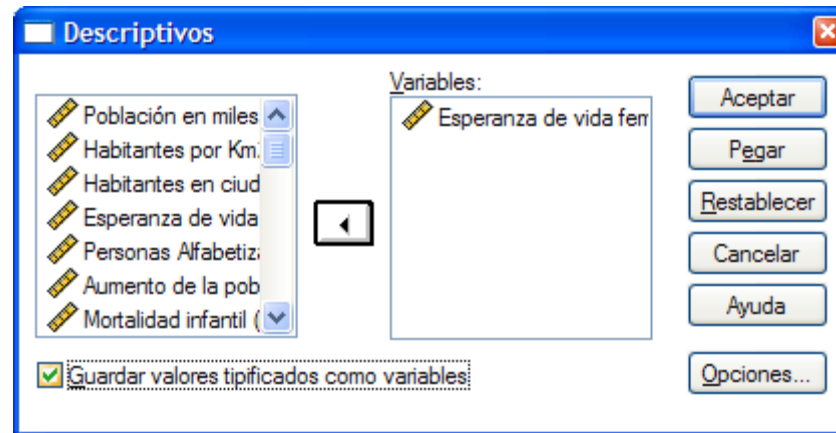
*Mundo 95.sav [Conjunto_de_datos...]

Archivo Edición Ver Datos Transformar Analizar
Gráficos Utilidades Ventana ?

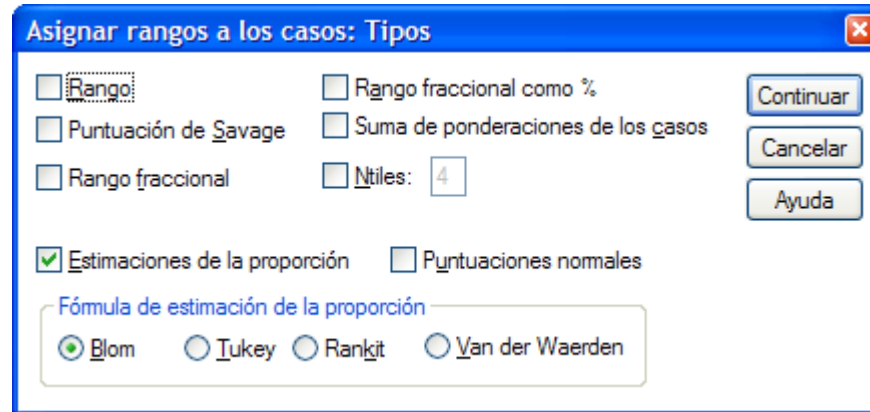
1 : país Acerbaján

	alfabfem	clima	Pcaloría
1	100	árido	.
2	14	árido	.
3	.	templado	84.00
4	48	desierto	58.67
5	95	templado	65.33
6	100	.	.
7	100	árido	74.67
8	.	templado	88.00
9	55	árido	.
10	22	tropical	9.33
11	99	tropical	.
12	.	templado	.
13	100	templado	.
14	71	4	4.00
15	.	templado	.
16	16	4	36.67
17	80	tropical	53.33
18	.	templado	.
19	9	tropical	28.00

- Para calcular puntuaciones típicas puedes ir a Analizar>Estadísticos descriptivos>Descriptivos y en el cuadro de diálogo marcar "Guardar valores tipificados como variables"



- Para calcular puntuaciones normalizadas ir a Transformar>Asignar rangos a casos>tipos de rango>estimaciones de la proporción



Lo mejor y lo peor

- Los siguientes ejercicios están dirigidos a determinar, teniendo en cuenta varias variables simultáneamente, quién o qué es mejor o peor.
- Estos ejercicios se realizarán en clase y se defenderán individualmente. Los alumnos que expongan respuestas razonables tendrán recompensa.
- Antes de la clase habrá una exposición sobre como realizar esos análisis mediante el programa ViSta, desarrollado parcialmente por el profesor de la asignatura.
- Los ejercicios están a continuación

12. Qué ciudad es la mejor y cual es la peor para jubilarse

- Esto se refiere al archivo Retire.vdf que está en File>Open Data>princomp>retire.vdf
- Las variables son: Población, porcentaje de la población mayor de 65 años, Distancia a una gran ciudad, Coste de la vida, alquiler medio, Precio de las casas, Impuestos sobre la propiedad, Impuestos sobre los ingresos, Nivel de ingresos medios, y días soleados.

13. Ciudades para trabajar en el mundo

- Este archivo está en File>Open
Data>regress>bigmac.vdf
- De nuevo se trata de elegir la ciudad más apropiada para vivir y para trabajar.
- Los factores a tener en cuenta son (datos de 1991):
Bigmac (minutos de trabajo para comprar una Bigmac con patatas), Pan (minutos para comprar 1 k. de pan), Horas de trabajo por año, Dias de vacaciones por año, precio de 10 viajes en autobús, precio de 19 servicios al año, salario de un profesor de primaria, impuestos

pagados por un profesor de primaria, sueldo de un ingeniero electrico, impuestos de un ingeniero eléctrico, Nombre de la ciudad.

- Hay que elegir la mejor y la peor de las ciudades.

14. Coches

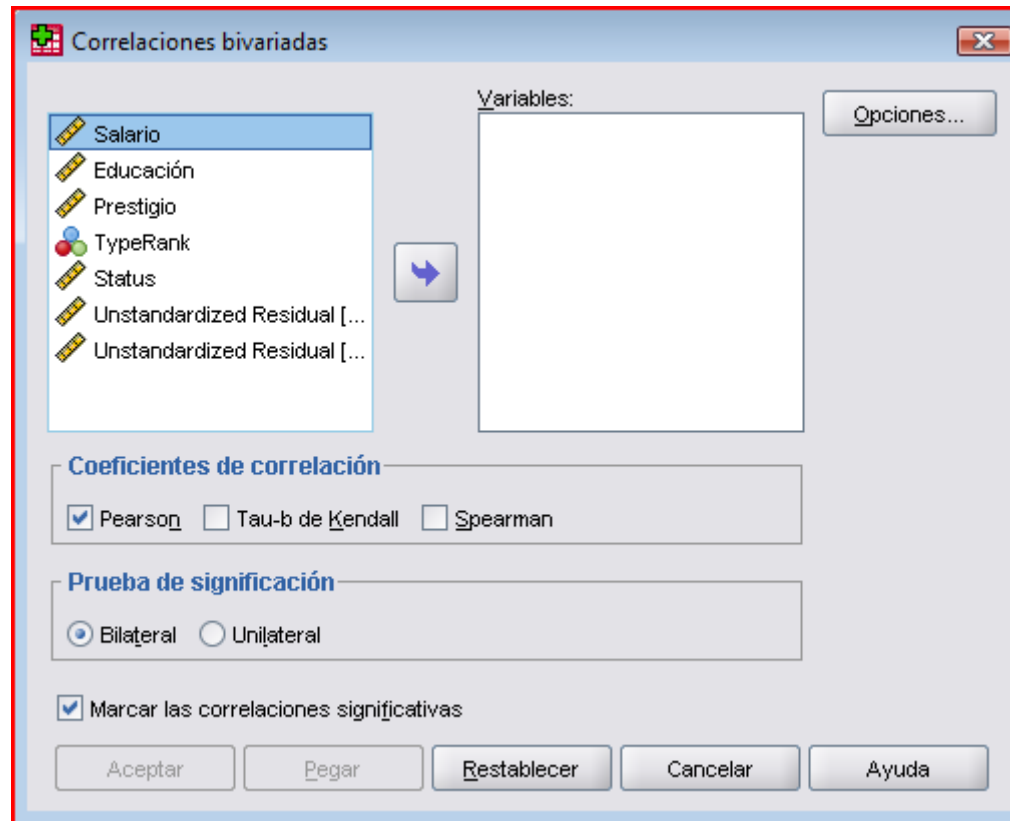
- Este archivo está en File>Open Data>regress>cars.vdf
- Las variables son: Millas por Galón (cuanto más alto, menos consume), Peso, Drive ratio medida de las revoluciones del motor (cuanto más alto en general más potencia de arrancada pero más gasto de combustible), caballos de potencia, capacidad de movimiento
- Intenta elegir un coche que tenga buenas características de potencia pero que aun así no gaste demasiado. ¿Es eso posible?

15. Crímenes

- Este archivo está en File>Open
Data>princomp>crime.vdf
- Las variables son (por 100.000 hbts en 1980)
Asesinatos, Violaciones, Robos, Asaltos, Allanamiento
de morada, Hurto y Robo de coches
- Intenta elegir estados en USA que sean los más seguros
en tu opinión, los más peligrosos, y también los más
seguros para las personas independientemente de otros
crímenes
-

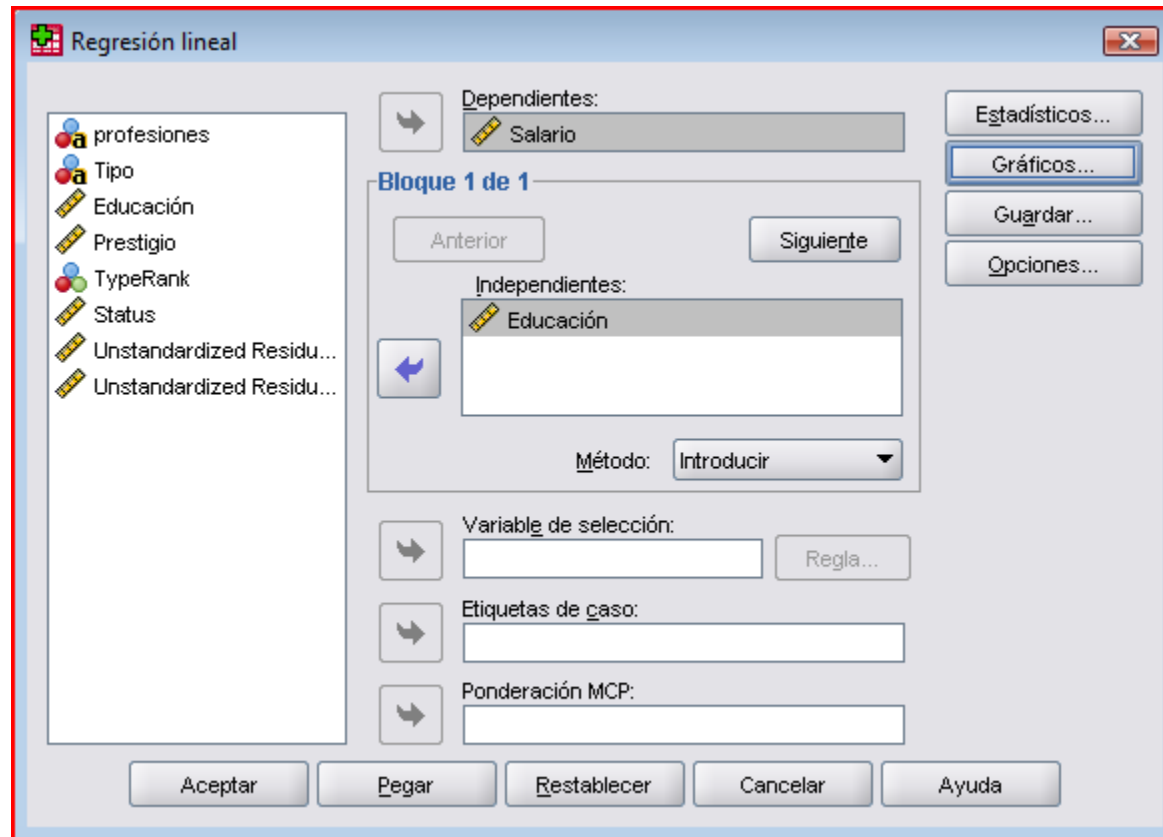
Diagramas de dispersión y análisis de regresion en el SPSS

- Para calcular correlaciones hay que ir al menú Analizar>Correlaciones>Bivariadas. El cuadro de diálogo para las variables de educación es el siguiente:



– Se seleccionan las variables y se aprieta aceptar

- Para un análisis de regresión se va a Analizar<Regresión<Lineales. Si queremos ver el efecto de la educación sobre el salario lo hacemos así.



16. Correlaciones para datos de encuesta

- Entre los archivos de ejemplo del SPSS hay uno que se denomina GSS93.sav. En este archivo hay una serie de contestaciones de encuestados a preguntas sobre estilo de vida, comportamientos, etc. El ejercicio en este caso consistirá en encontrar variables relacionadas con los gustos musicales. ¿Podeis encontrar qué características diferencia los que declaran que les gustan un estilo musical frente a otros?

Ejercicios de Regresión

***Fechas de entrega: Grupo B 10 de Diciembre,
Grupos H, G 11 de Diciembre***

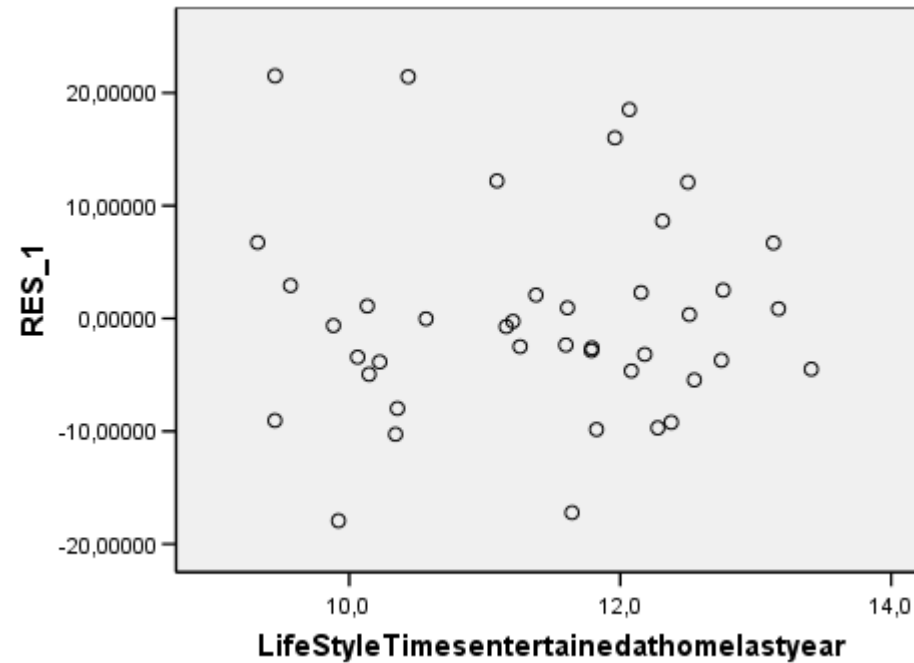
- Los siguientes ejercicios consisten en una serie de preguntas acerca de análisis de regresión, comprobación de supuestos, etc.
- Se trata de contestar a las preguntas haciendo referencia a los contenidos teóricos trabajados en clase

17. Bowling Alone

- En el libro titulado "Bowling Alone" de Robert Putnam (2000) se plantea que el comportamiento social en la sociedad en Estados Unidos ha ido descendiendo a lo largo del siglo pasado hasta llegar a un punto en el que la mayoría de la gente actúa de un modo más individualista en una gran cantidad de esferas (el título del libro hace referencia a que cada vez hay más gente que juega a los bolos pero las ligas de equipos de bolos casi han desaparecido por lo que la gente ahora juega sola).

-
-
- El libro está basado en una serie de datos recogidos por el autor y otros colaboradores que están disponibles en Internet. Usaremos uno de los conjuntos de datos para esta práctica.
 - En uno de los archivos se dan las medidas a nivel de estado de lo que se denomina capital social, una medida de la forma en que la gente participa, se relaciona, y está en general cohesionada socialmente por estado. Además se dan valores medios de las contestaciones de sujetos encuestados a una serie de preguntas que en principio están relacionadas con los aspectos de cohesión social. Analizaremos estos datos a continuación.

EJERCICIO 17.3 ¿Dirías que los residuales de la regresión para las variables “número de veces que se tuvo a gente de visita en casa el último año” (eje X) que es la media de las respuestas a esa pregunta de los sujetos de cada estado y la variable “La mayoría de la gente es de fiar” (eje Y) sugieren curvilinealidad en la relación?



EJERCICIO 17.8 Supongamos que queremos estudiar si el número medio de veces que los individuos de cada estado atendieron a reuniones de Clubes es un buen predictor del número de organizaciones sociales y cívicas que hay en cada estado y empezamos haciendo un diagrama de dispersión de ambas variables. ¿Cuál pondríamos en el eje X y cual en Y? ¿Por qué?

EJERCICIO 17.9 Si queremos estudiar si el número medio de veces que los individuos de cada estado atendieron a reuniones de Clubes es un buen predictor del número de organizaciones sociales y cívicas que hay en cada estado y empezamos haciendo un diagrama de dispersión de ambas variables. ¿Cuál de ellas diríamos que es la variable independiente? ¿Cuál es la variable dependiente? ¿Y la predictora y la predicha?

18. Consumo de cigarrillos por estado

- En el siguiente ejemplo analizaremos el consumo de cigarrillos por estado en USA y su relación con una serie de cánceres bastante comunes (los datos son de 1960).
- Las variables disponibles son el número de cigarrillos medio per capita, las muertes por 100000 habitantes por cancer de próstata, pulmón, riñón y leucemia
- Obviamente, el cancer más relacionado con el consumo de cigarrillos es el de pulmón pero es posible que otros cánceres también estén relacionados con este consumo.

EJERCICIO 18.1 De las variables descritas anteriores...¿cuál o cuáles tendrían el papel de variable dependiente y cuál o cuáles tendrían el papel de variable independiente?

EJERCICIO 18.2 Examinando los resultados de abajo, ¿dirías que el consumo de cigarrillos (CIG) está relacionado con algún tipo de cancer?

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
Pulmón	.697 ^a	.486	.474	3.06607

a. Variables predictoras: (Constante), CIG

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
Riñón	.487 ^a	.238	.219	.45862

a. Variables predictoras: (Constante), CIG

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
Leucemia	.068 ^a	.005	-.019	.64430

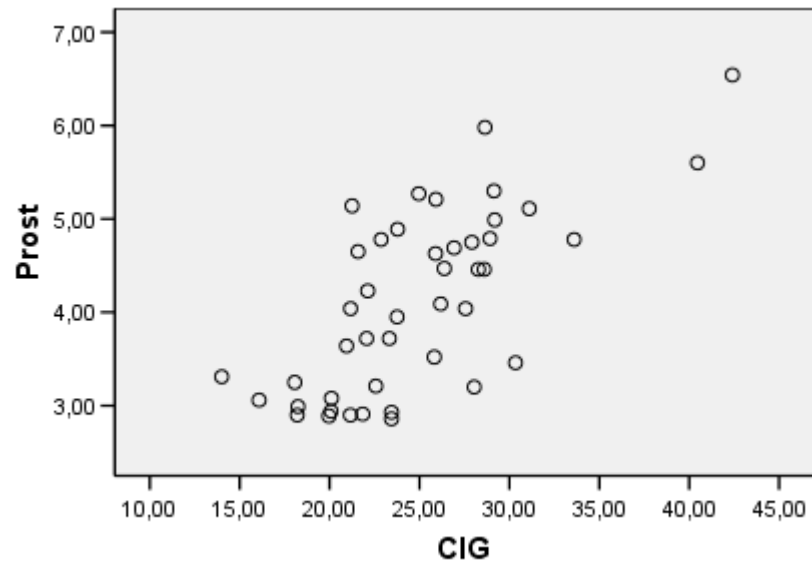
a. Variables predictoras: (Constante), CIG

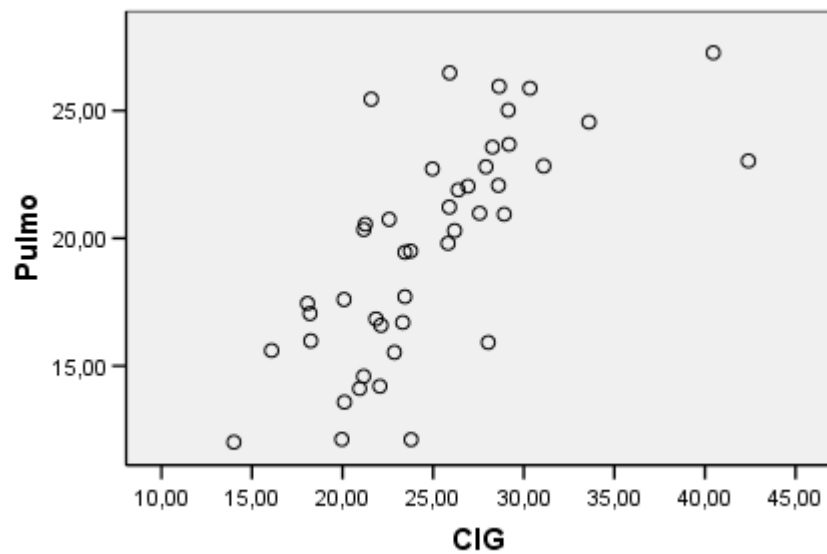
Resumen del modelo

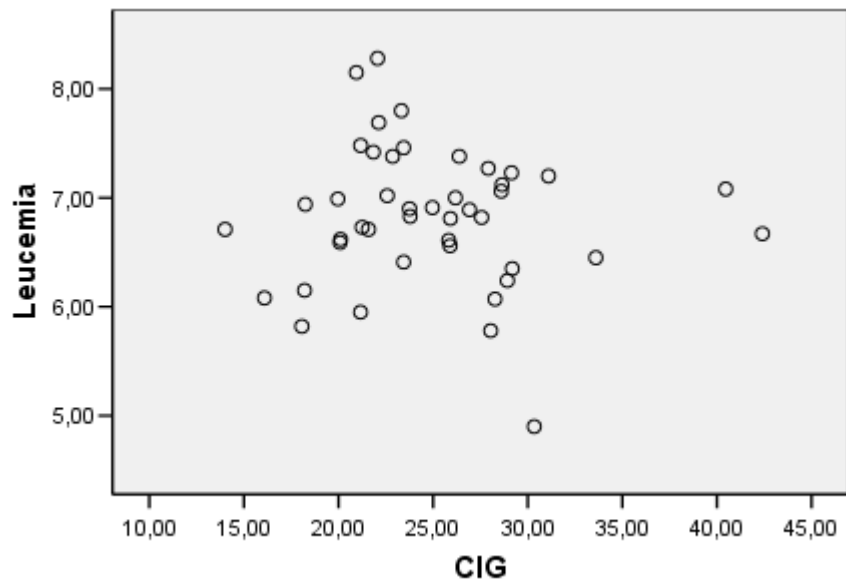
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
Próstata	.704 ^a	.495	.483	.69377

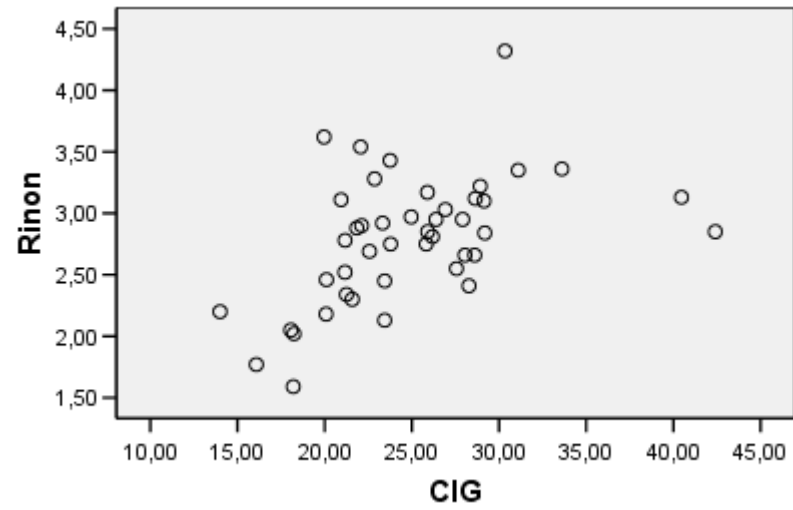
a. Variables predictoras: (Constante), CIG

EJERCICIO 18.3 Examinando los gráficos de abajo ¿qué podrías decir sobre ellos que ayudarán a entender la relación entre consumo de cigarrillos y cancer?





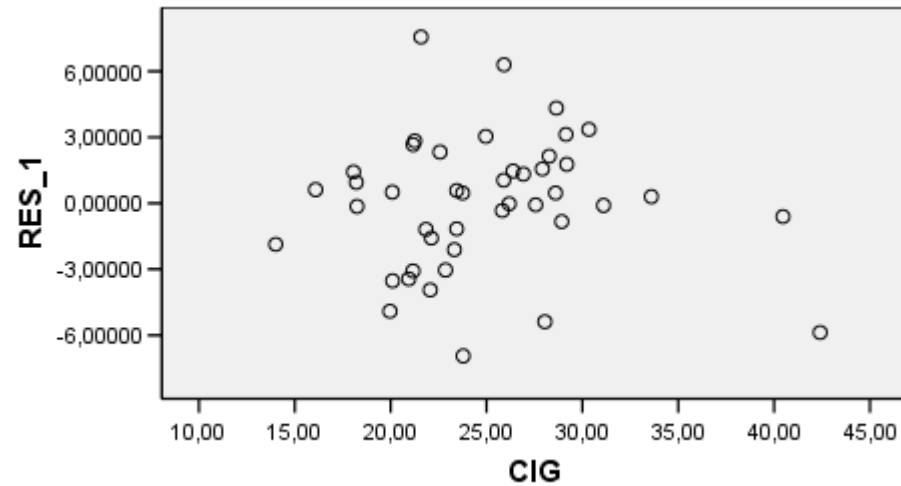




EJERCICIO 18.4 En los gráficos del Ejercicio 18.3 aparecen dos valores destacados, ¿podrías indicar qué caracteriza a esos valores?

EJERCICIO 18.5 ¿Crees que hay evidencia de curvilinealidad en alguno de los gráficos del Ejercicio 18.3.

EJERCICIO 18.6 En el gráfico siguiente se muestra los residuales de la regresión del consumo de cigarrillos y los cánceres de pulmón por estado. ¿Qué información obtenemos mirando este gráfico?



EJERCICIO 18.7 En la siguiente tabla se muestran las correlaciones entre las variables analizadas en este ejercicio. ¿Dirías que hay relación entre los diferentes tipos de cancer y el consumo de cigarrillos? ¿Hay algún tipo de cancer que parezca diferente a los otros? ¿Por qué?

Correlaciones

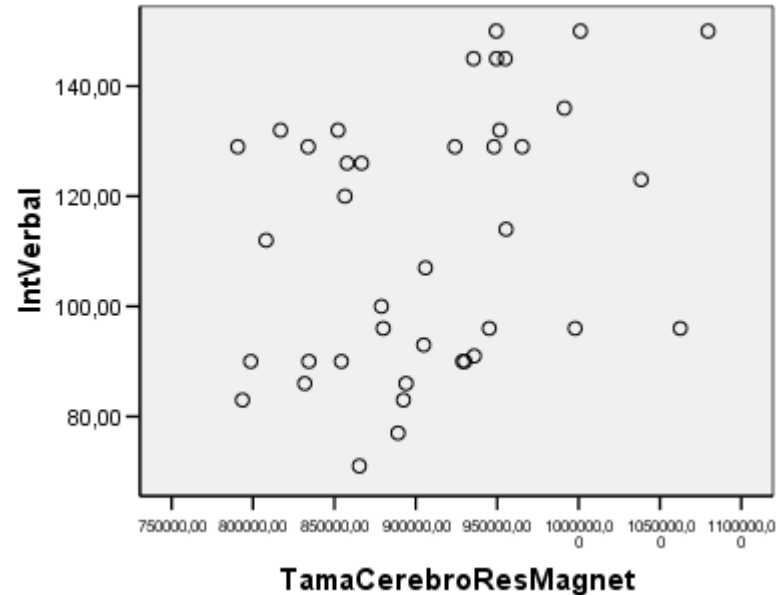
		CIG	Prost	Pulmo	Rinon	Leucemia
CIG	Correlación de Pearson	1	.704	.697	.487	-.068
	Sig. (bilateral)		.000	.000	.001	.659
	N	44	44	44	44	44
Prost	Correlación de Pearson	.704	1	.659	.359	.162
	Sig. (bilateral)	.000		.000	.017	.293
	N	44	44	44	44	44
Pulmo	Correlación de Pearson	.697	.659	1	.283	-.152
	Sig. (bilateral)	.000	.000		.063	.326
	N	44	44	44	44	44
Rinon	Correlación de Pearson	.487	.359	.283	1	.189
	Sig. (bilateral)	.001	.017	.063		.220
	N	44	44	44	44	44
Leucemia	Correlación de Pearson	-.068	.162	-.152	.189	1
	Sig. (bilateral)	.659	.293	.326	.220	
	N	44	44	44	44	44

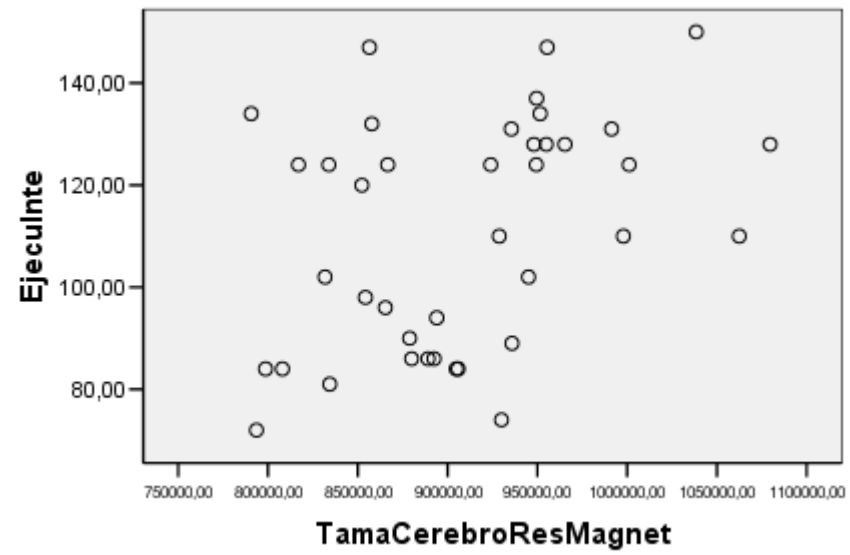
19. Tamaño Cerebro

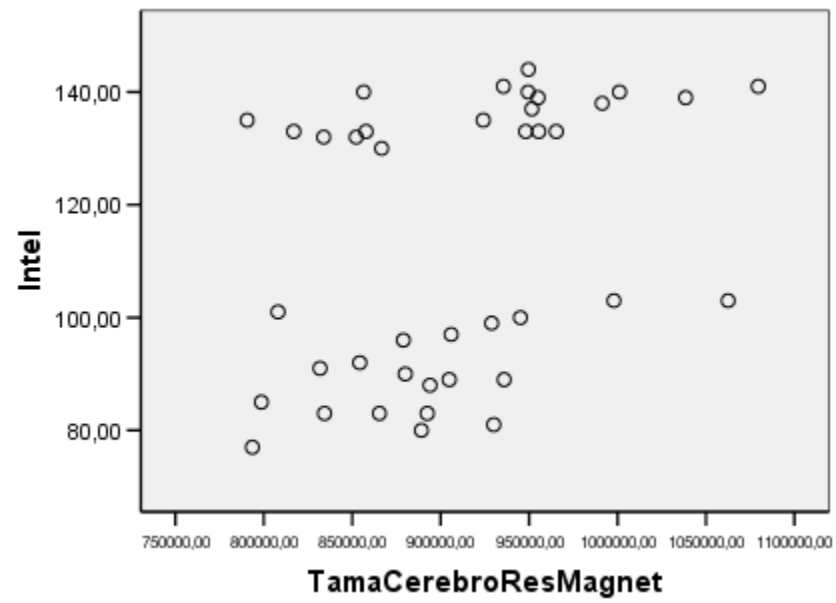
- Willerman et al. (1991) recogieron una muestra de 40 estudiantes de psicología de una universidad y les hicieron unos tests de inteligencia (escalas del Wechsler). Luego les hicieron una resonancia magnética para determinar el tamaño del cerebro de esos sujetos. También tenemos información acerca del género de los sujetos y su peso y su altura.
- La idea del estudio es examinar si tener el cerebro más grande está relacionado con tener más inteligencia según los tests.

-
-
- Las variables son: TamaCerebroResMagnet (tamaño cerebro según la resonancia), IntVerbal (Inteligencia verbal), EjecuInte (ejecución en una tarea de similitudes), Intel (Inteligencia General), PesoCorporal (peso), AlturaPulgadas (altura) y Genero

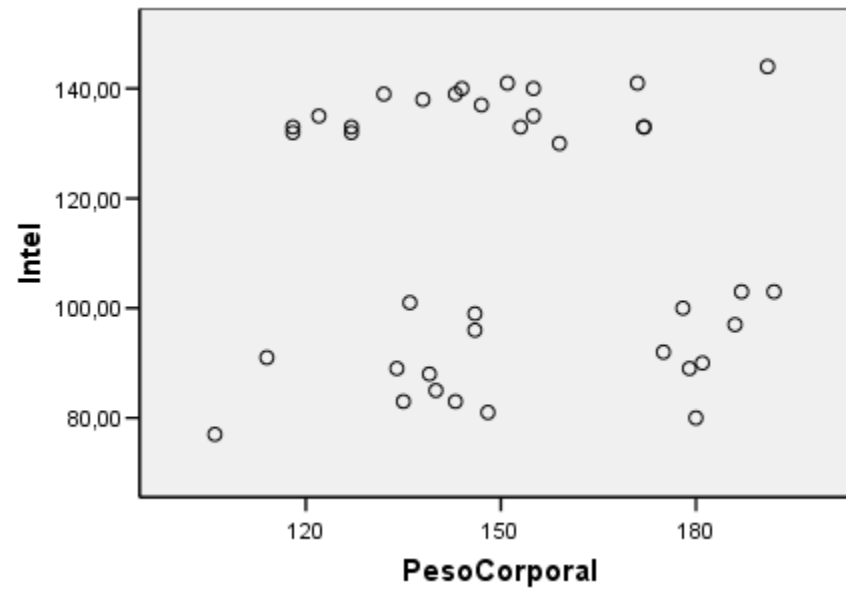
EJERCICIO 19.1 Empezaremos examinando si hay relación entre el tamaño del cerebro y los tipos de inteligencia utilizando diagramas de dispersión. ¿Qué dirías al respecto? Indica también si hay alguna característica en los gráficos que te llame la atención en especial.

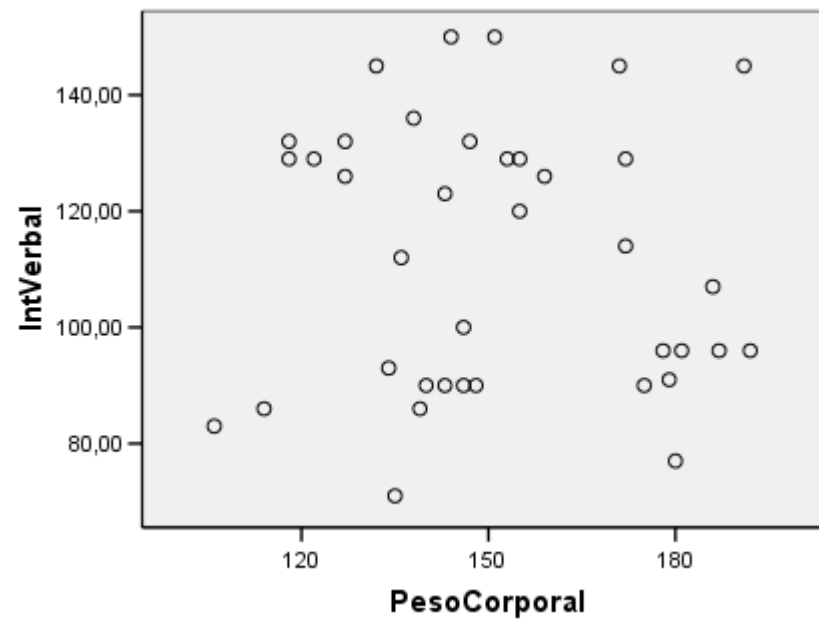


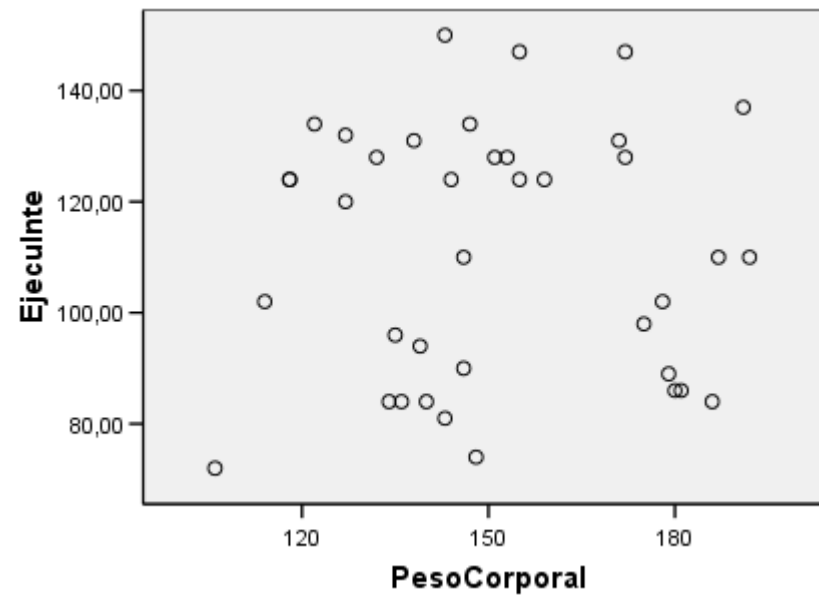




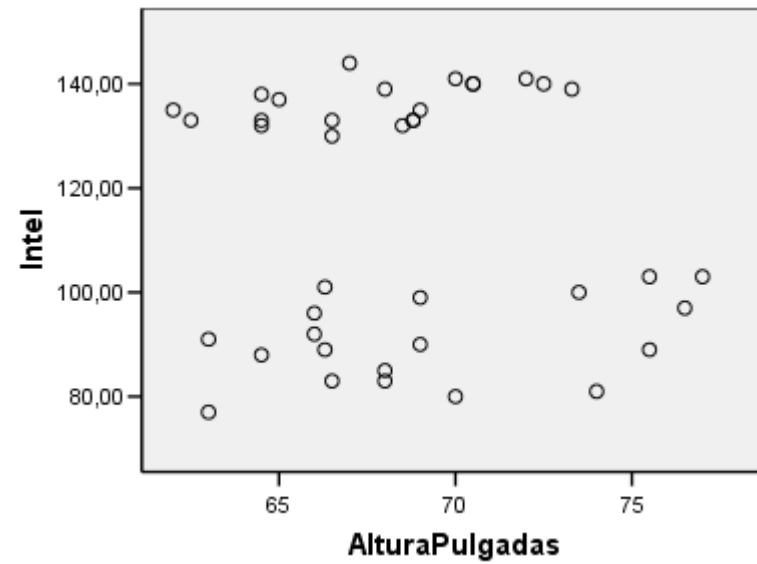
EJERCICIO 19.2 ¿Dirías que el peso está relacionado con los distintos tipos de inteligencia a partir de estos gráficos?

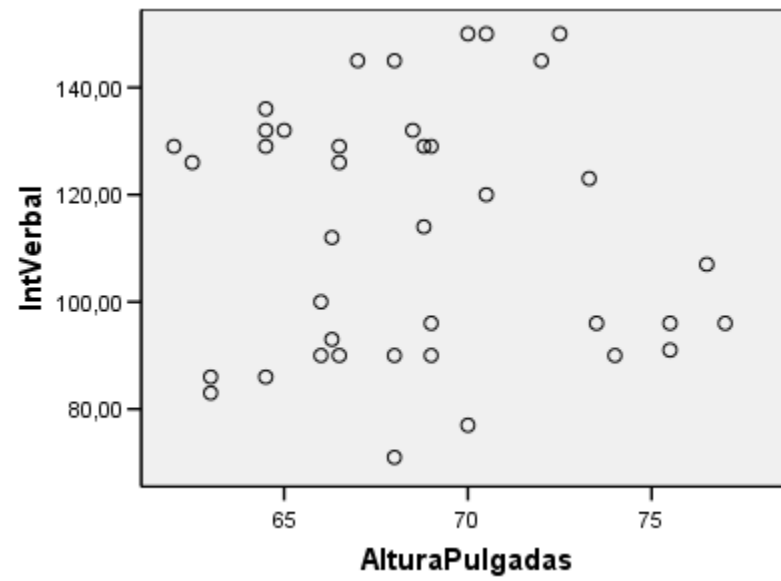


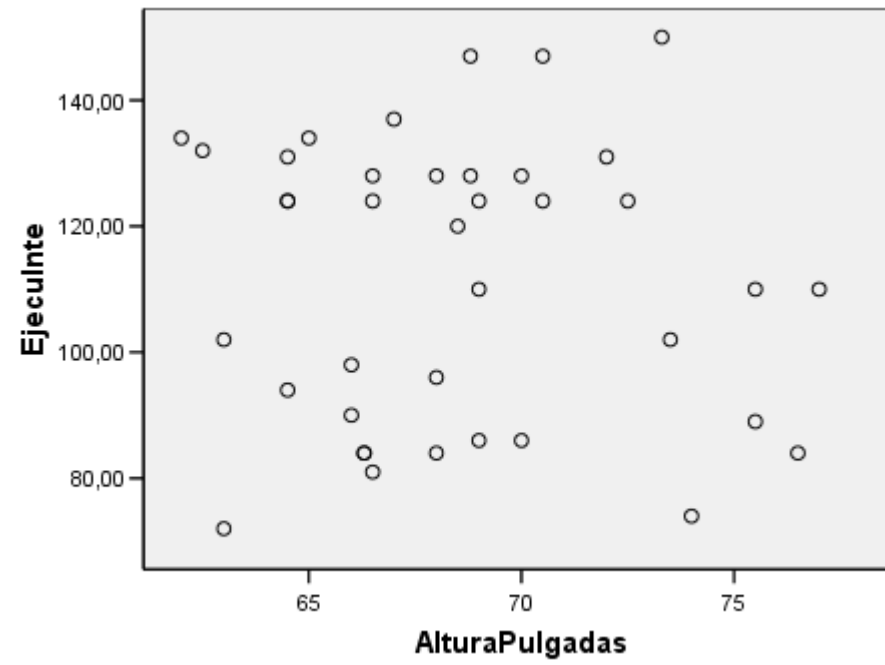




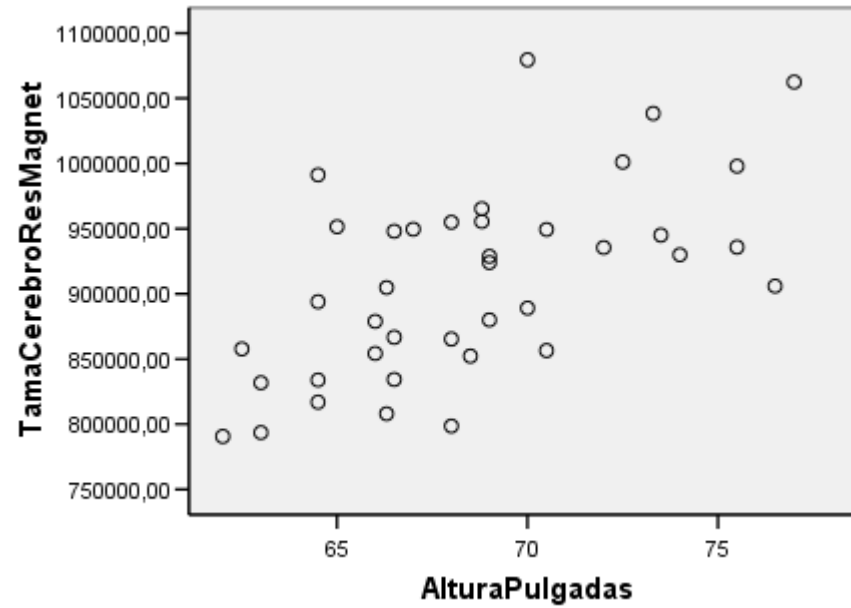
EJERCICIO 19.3 ¿Qué dirías de la altura? ¿Dirías que hay relación entre la altura de los individuos y su inteligencia?

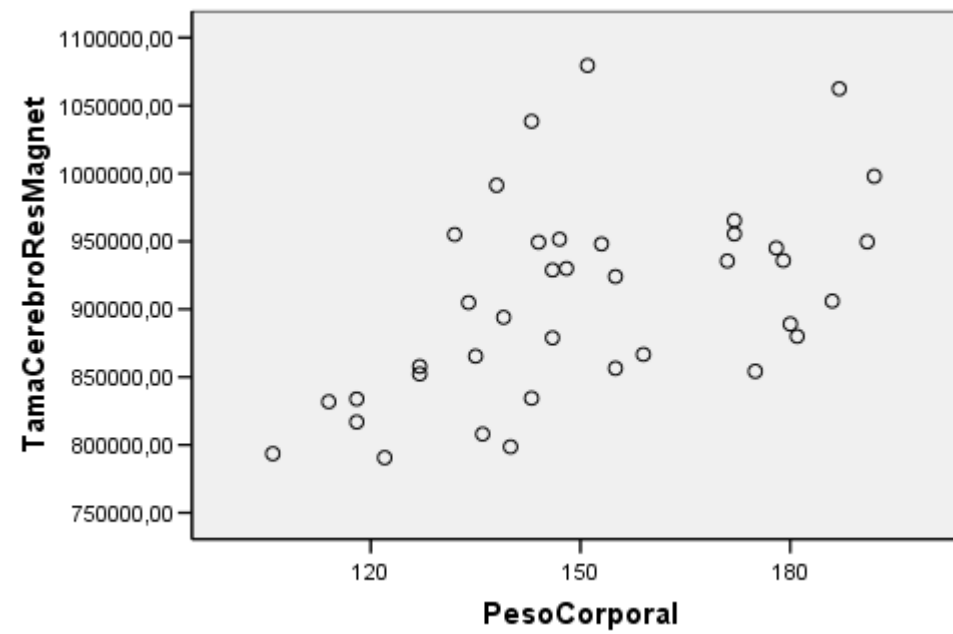






EJERCICIO 19.4 ¿Existe relación entre el tamaño del cerebro y el tamaño del cuerpo medido en peso y altura?





EJERCICIO 19.5 Si la inteligencia de varios tipos está relacionada con el tamaño del cerebro, y el tamaño del cerebro está relacionado con la altura y el peso, ¿no sería lógico pensar que la altura y el peso están relacionados con la inteligencia? Comenta este tema utilizando la tabla de abajo.

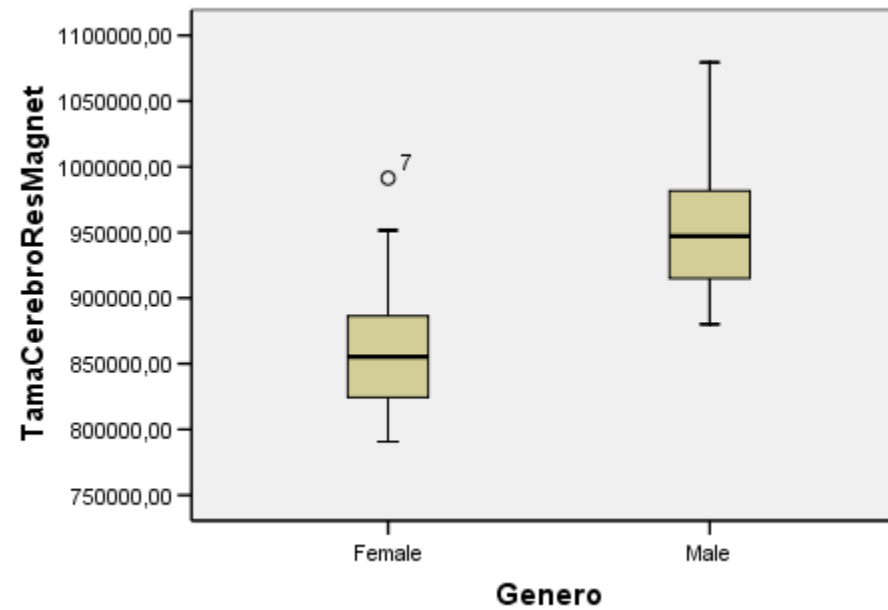
Correlaciones

		Intel	IntVerbal	Ejeculnte	PesoCorporal	Altura Pulgadas	Tama CerebroRes Magnet
Intel	Correlación de Pearson	1	.947**	.934**	-.051	-.086	.358*
	Sig. (bilateral)		.000	.000	.759	.603	.023
	N	40	40	40	38	39	40
IntVerbal	Correlación de Pearson	.947**	1	.778**	-.076	-.071	.337*
	Sig. (bilateral)	.000		.000	.650	.667	.033
	N	40	40	40	38	39	40
Ejeculnte	Correlación de Pearson	.934**	.778**	1	.003	-.077	.387*
	Sig. (bilateral)	.000	.000		.988	.642	.014
	N	40	40	40	38	39	40
PesoCorporal	Correlación de Pearson	-.051	-.076	.003	1	.700**	.513**
	Sig. (bilateral)	.759	.650	.988		.000	.001
	N	38	38	38	38	38	38
AlturaPulgadas	Correlación de Pearson	-.086	-.071	-.077	.700**	1	.602**
	Sig. (bilateral)	.603	.667	.642	.000		.000
	N	39	39	39	38	39	39
TamaCerebroResMagnet	Correlación de Pearson	.358*	.337*	.387*	.513**	.602**	1
	Sig. (bilateral)	.023	.033	.014	.001	.000	
	N	40	40	40	38	39	40

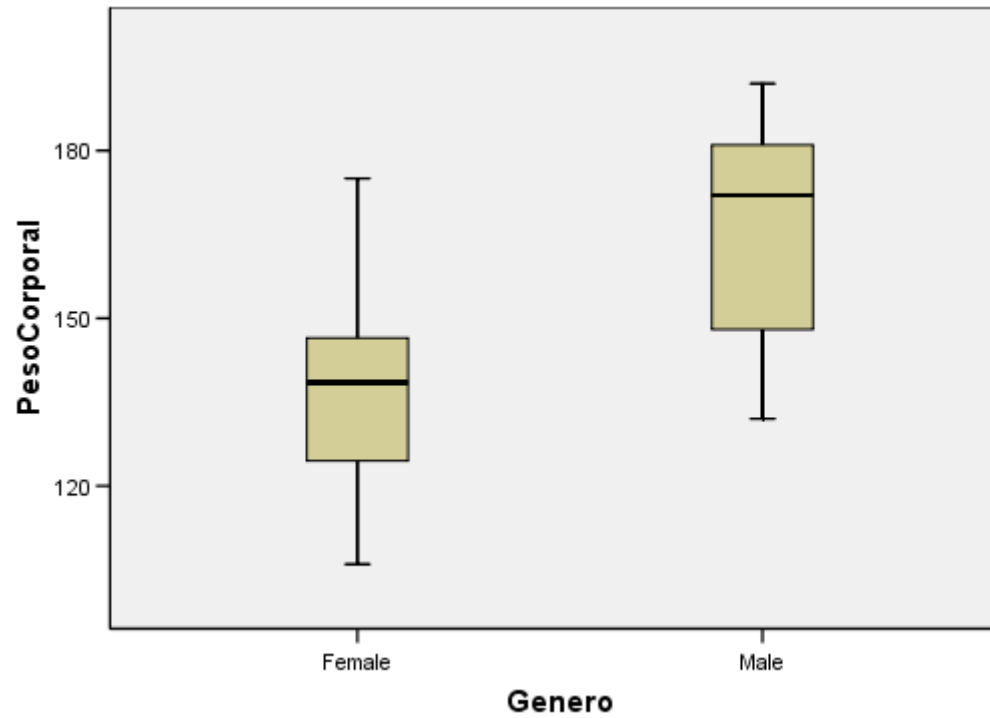
** . La correlación es significativa al nivel 0,01 (bilateral).

* . La correlación es significante al nivel 0,05 (bilateral).

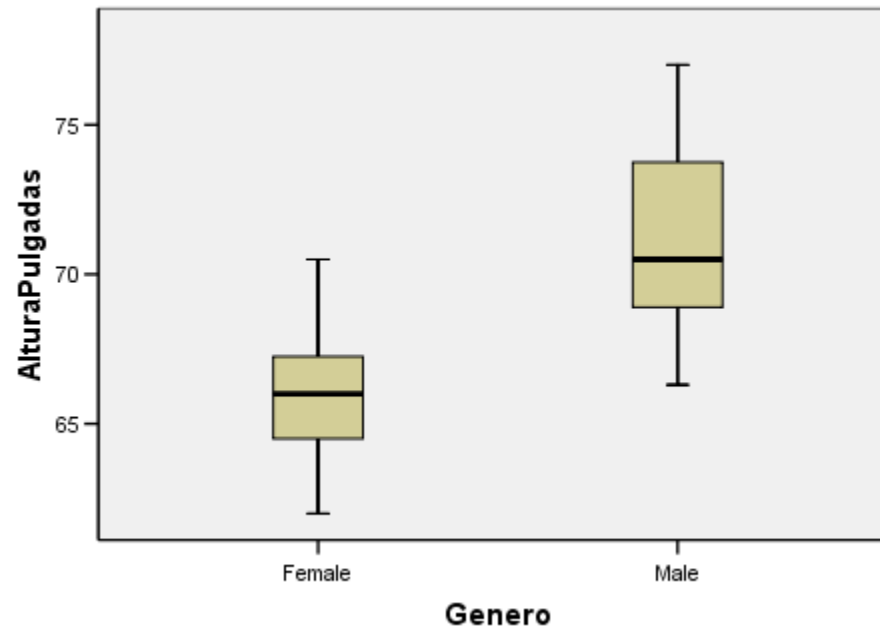
EJERCICIO 19.6 ¿Existen diferencias en el tamaño del cerebro entre hombres y mujeres en este estudio? ¿En qué consisten?



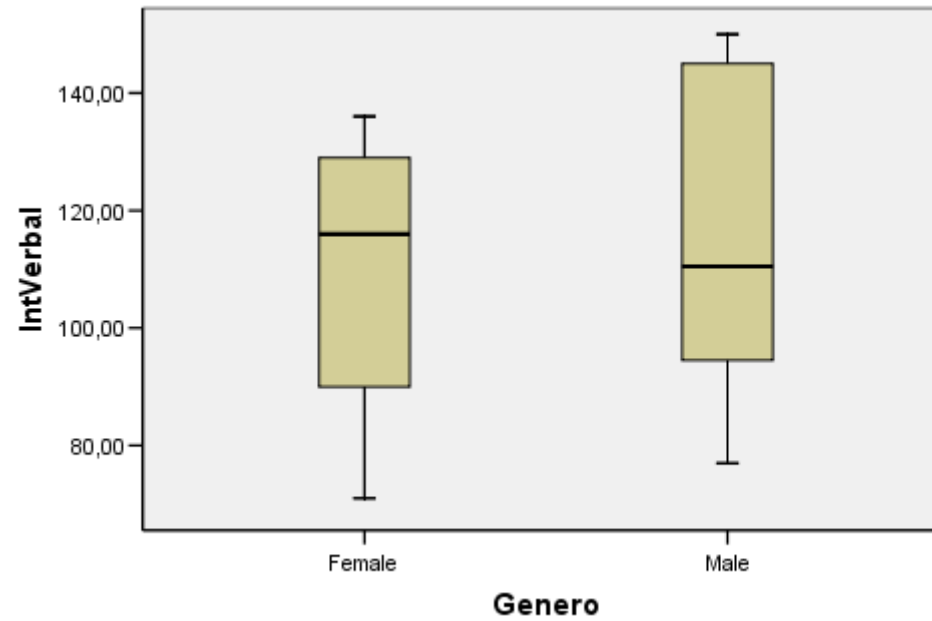
EJERCICIO 19.7 ¿Existen diferencias entre hombres y mujeres en peso en este estudio? ¿En qué consisten?



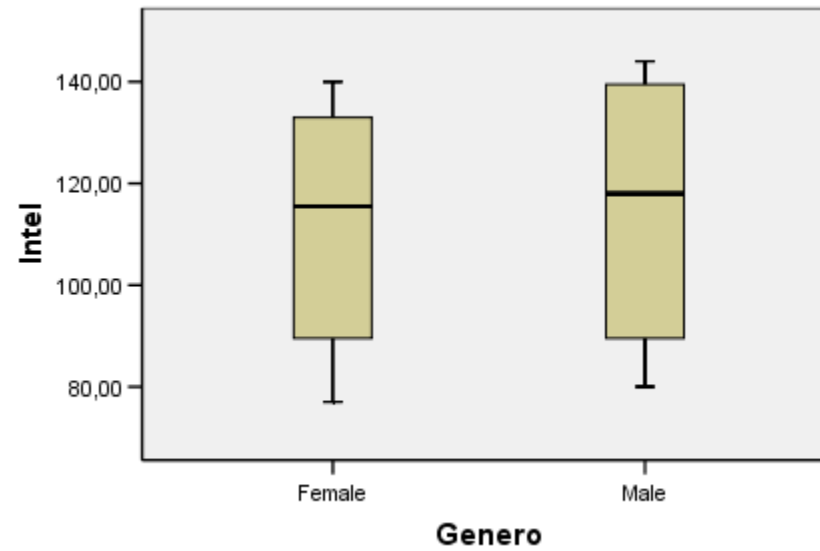
**EJERCICIO 19.8 ¿Existen diferencias entre hombres y mujeres en altura en este estudio?
¿En qué consisten?**



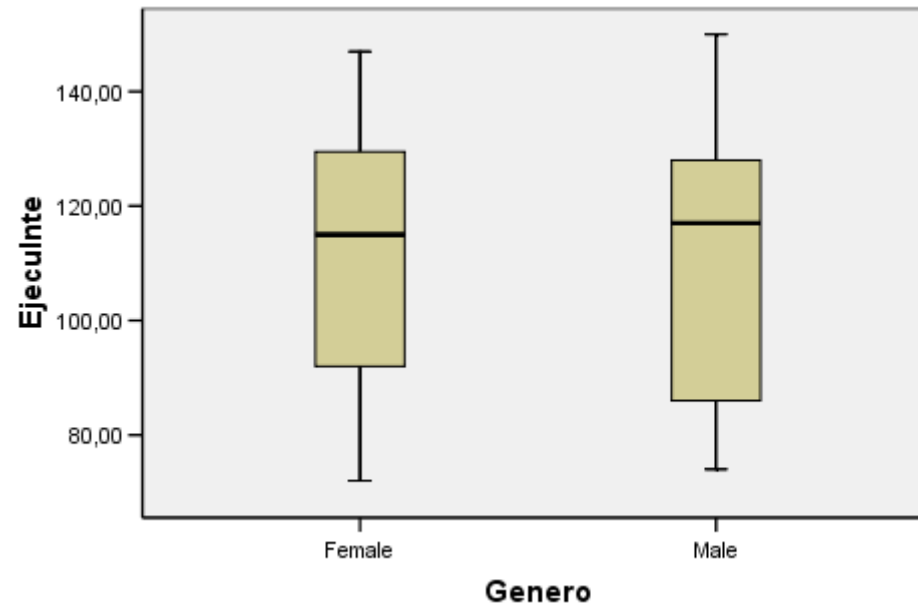
EJERCICIO 19.9 ¿Existen diferencias en inteligencia verbal entre hombres y mujeres en los sujetos de este estudio?



EJERCICIO 19.10 Hay diferencias en Inteligencia general entre hombres y mujeres en los sujetos de este estudio?



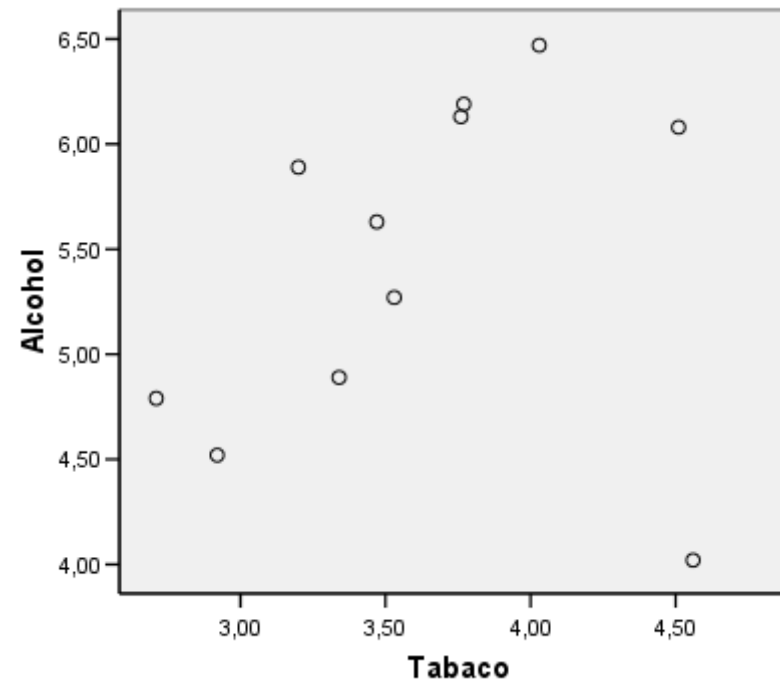
EJERCICIO 19.11 ¿Existen diferencias entre los géneros en este tipo de inteligencia?



20. Tabaco y alcohol

- En los datos siguientes se examina la relación entre el gasto de alcohol y tabaco por hogar promedio en 11 regiones de Gran Bretaña
- Este consumo podría pensarse que está relacionado así que examinaremos esta relación utilizando diversas técnicas

EJERCICIO 20.1 ¿Qué dirías de la relación entre las variables de consumo de alcohol y de tabaco según este diagrama de dispersión? ¿Hay relación o no la hay? ¿Hay algún valor que sea especial?



EJERCICIO 20.2 Si quisieras dar una medida de la relación entre ambas variables, ¿qué coeficiente de correlación utilizarías? ¿Por qué?

EJERCICIO 20.3 A continuación tienes los resultados del coeficiente de correlación de Pearson, Kendall y Spearman, ¿cuál te parece que refleja mejor la relación entre el gasto en alcohol y en tabaco?

Correlaciones

		Alcohol	Tabaco
Alcohol	Correlación de Pearson	1	.224
	Sig. (bilateral)		.509
	N	11	11
Tabaco	Correlación de Pearson	.224	1
	Sig. (bilateral)	.509	
	N	11	11

Correlaciones

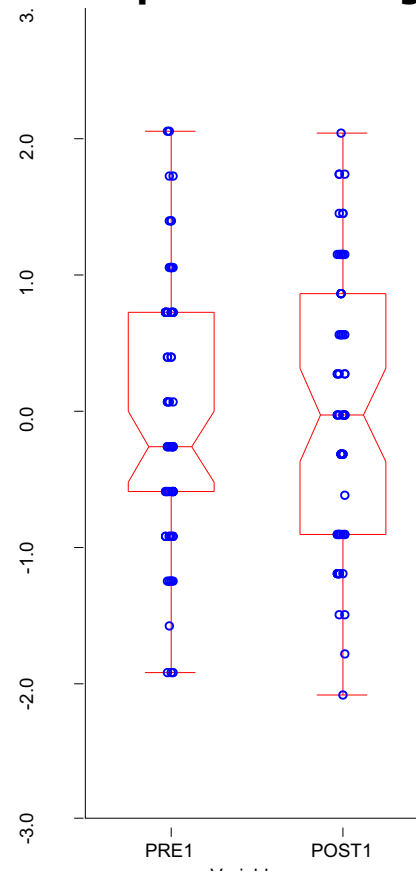
			Alcohol	Tabaco
Tau_b de Kendall	Alcohol	Coeficiente de correlación	1.000	.345
		Sig. (bilateral)	.	.139
		N	11	11
	Tabaco	Coeficiente de correlación	.345	1.000
		Sig. (bilateral)	.139	.
		N	11	11
Rho de Spearman	Alcohol	Coeficiente de correlación	1.000	.373
		Sig. (bilateral)	.	.259
		N	11	11
	Tabaco	Coeficiente de correlación	.373	1.000
		Sig. (bilateral)	.259	.
		N	11	11

EJERCICIO 20.4 Si eliminamos el caso con un residual más grande de los datos anteriores, la correlación de Pearson pasa de ser .224 a .784 ¿dirías que ese valor refleja adecuadamente la relación entre las variables? ¿Por qué?

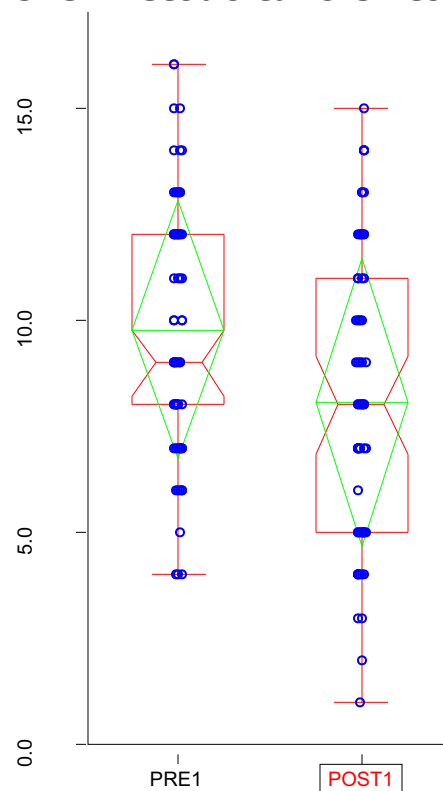
21. Método de lectura

- A 66 niños se les enseñó tres métodos diferentes para leer. Antes y después de enseñarles el método se les dio un test de comprensión lectora.
- En este estudio queremos examinar si los niños tienen diferencias en cuanto a las medidas de comprensión lectora antes y después de enseñarles el método

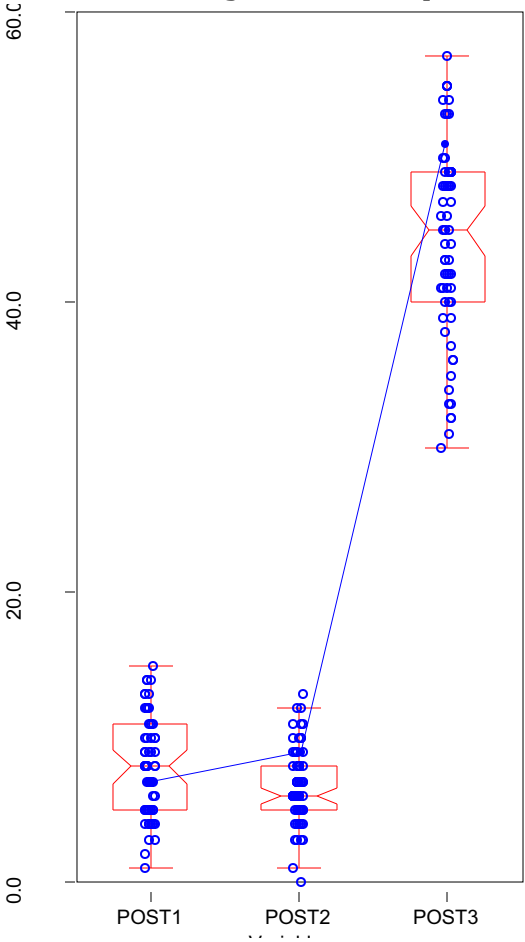
EJERCICIO 21.1 En este gráfico se comparan las puntuaciones típicas de las variables Pre1 y Post1 que son las medidas de comprensión lectora 1 antes y después del método. ¿Podrías decir a partir de este gráfico si el método tuvo éxito?



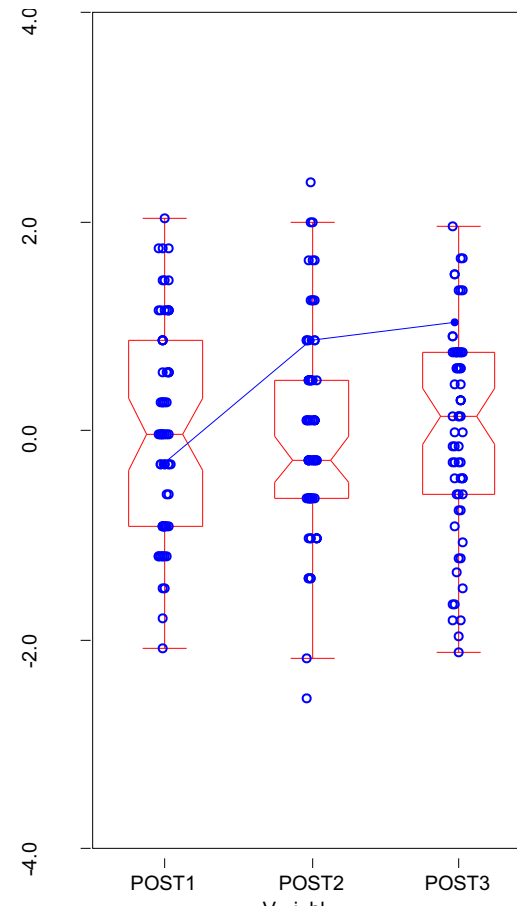
EJERCICIO 21.3 En este gráfico se comparan las puntuaciones de las variables Pre1 y Post1 que son las medidas de comprensión lectora 1 antes y después del método (en este caso las puntuaciones no están normalizadas) ¿Podrías decir a partir de este gráfico si el método tuvo éxito?



EJERCICIO 21.6 Después de enseñar el nuevo método de lectura se hicieron tres tests para medir la comprensión lectora. En este gráfico, el sujeto señalado, ¿puedes decir en qué test lo hizo mejor? ¿Por qué?



EJERCICIO 21.7 Después de enseñar el nuevo método de lectura se hicieron tres tests para medir la comprensión lectora. En este gráfico se muestran las puntuaciones típicas en los tests, ¿puedes decir en qué test lo hizo mejor el sujeto señalado? ¿Por qué?

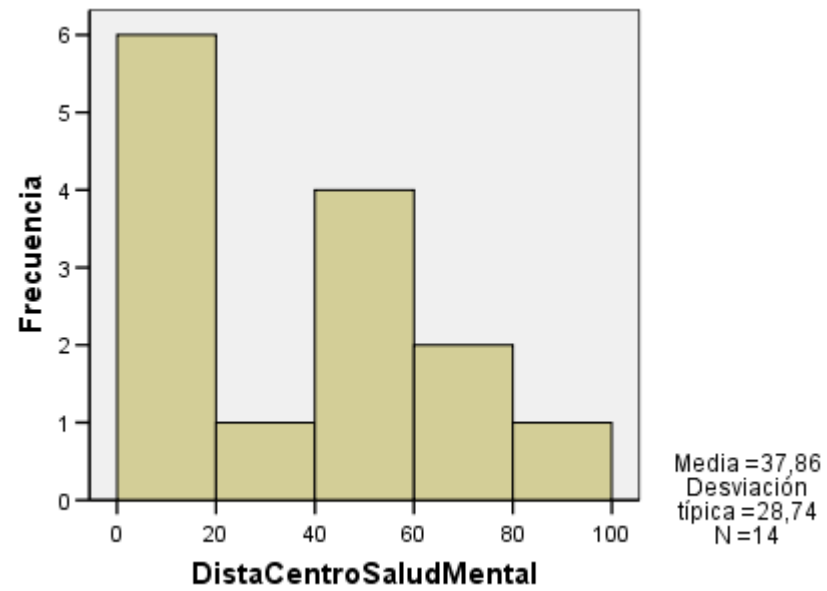


22. Lunáticos

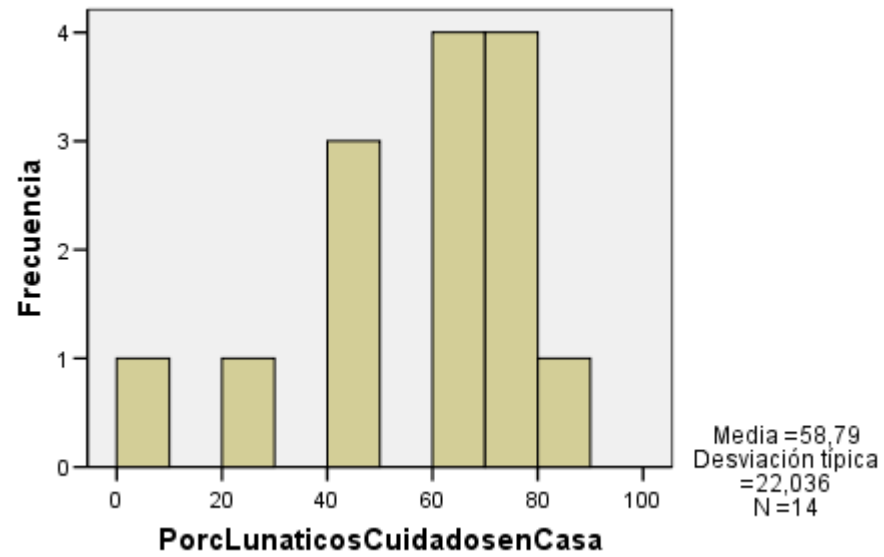
- El Dr. Jarvis, que fue presidente de la asociación de estadística americana desde 1852 hasta 1882, recibió el encargo en 1854 de la comisión de Massachusetts sobre ***Lunaticismo*** (un nombre para enfermedad mental que hoy en día claramente está en desuso).
- En ese estudio se recogieron las siguientes variables, para cada uno de los condados de Masschusetts:
 - Número de enfermos mentales por condado
 - La distancia al centro mental más cercano
 - La población del condado (en miles)
 - La densidad de población (en millas al cuadrado)

-
- El porcentaje de enfermos mentales cuidados en casa
 - Uno de los objetivos principales del estudio era examinar las razones por las que las familias optaban por cuidar a los enfermos mentales en casa o no (el porcentaje de enfermos cuidados en casa era mayor o menor)
 - Es interesante comentar que los métodos de regresión que podíamos usar para este caso todavía no estaban formalizados (esto ocurrió unos 30 o 40 años después)

EJERCICIO 22.1 ¿Qué dirías del histograma de la distancia al centro de salud mental?

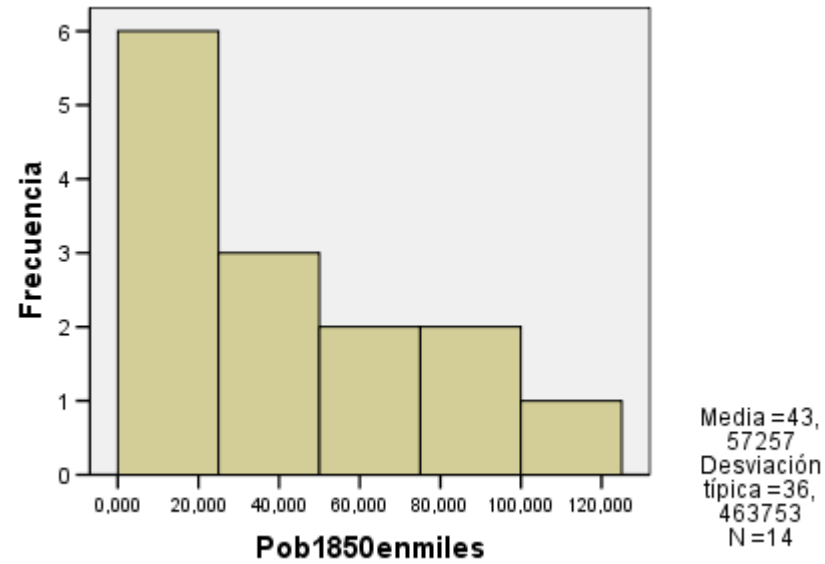


EJERCICIO 22.2 ¿Qué dirías del porcentaje de lunáticos cuidados en casa?

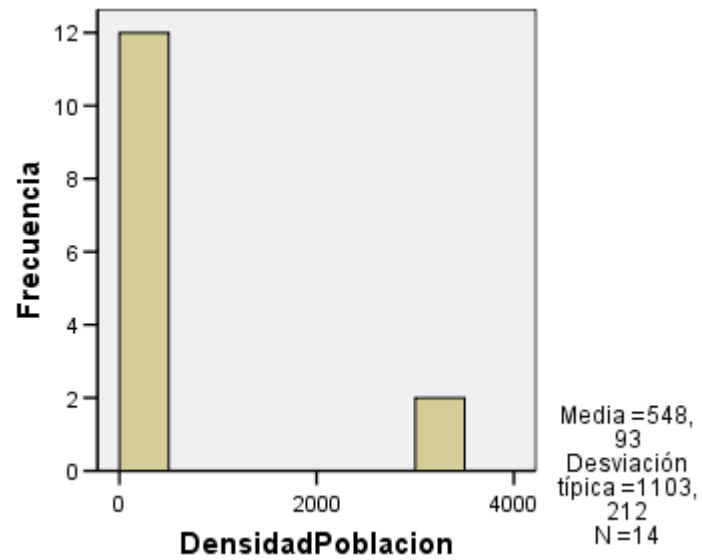


**EJERCICIO 22.3 Partiendo de los objetivos descritos en la presentación de esta pregunta?
¿Qué variable utilizarías como variable dependiente en un análisis de regresión?
¿Por qué? ¿Qué variables utilizarías como independientes? ¿Por qué?**

EJERCICIO 22.4 ¿Qué dirías del histograma de la población en miles?



EJERCICIO 22.5 ¿Qué dirías de la densidad de la población?

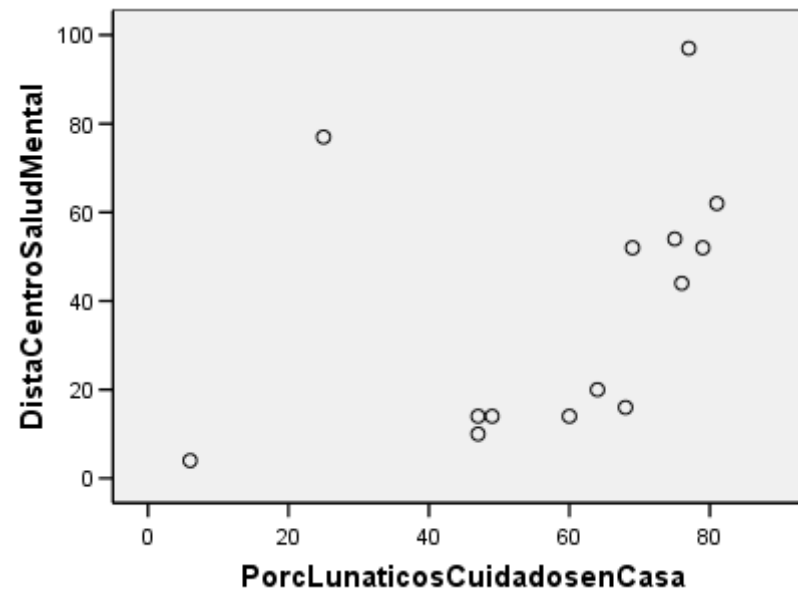


EJERCICIO 22.6 La media de densidad de la población es de 548.93. ¿Qué dirías de ese valor? ¿Es problemático de alguna manera? ¿Qué valor sería mas interesante en este caso? ¿Por qué?

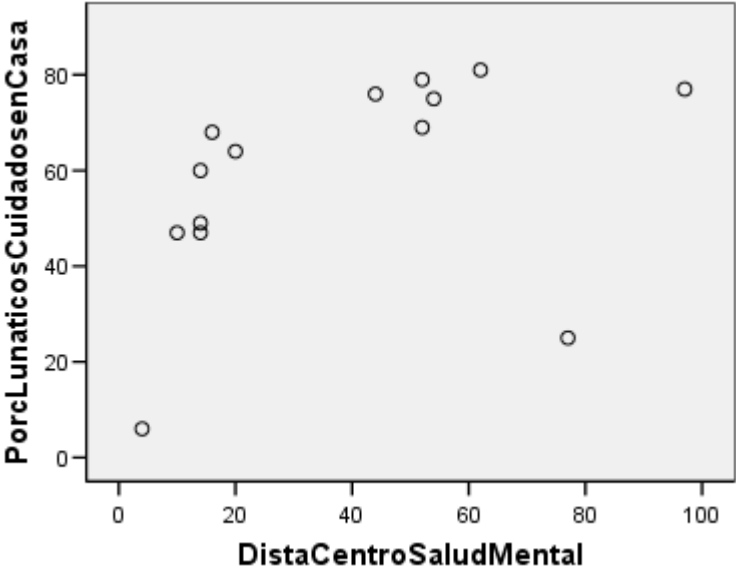
EJERCICIO 22.7 La desviación típica de la densidad de la población es **1103.212**. ¿Qué inconveniente tiene este valor? ¿Qué otro valor sería interesante utilizar en este caso? ¿Por qué?

EJERCICIO 22.8 Si quisieramos valorar la gravedad del problema de los enfermos mentales por condado ¿cómo lo podríamos expresar? ¿qué cantidad nos indicaría como de importante es ese problema en un condado dado con respecto a otros?

EJERCICIO 22.9 A continuación se muestra el diagrama de dispersión del porcentaje de enfermos mentales cuidados en casa y la distancia a un centro mental. ¿Qué problemas de índole estadística ves en este gráfico?



EJERCICIO 22.10 A continuación se muestra el diagrama de dispersión del porcentaje de enfermos mentales cuidados en casa y la distancia a un centro mental. ¿Qué problemas de índole estadística ves en este gráfico ¿Cómo podrias solucionarlos?



EJERCICIO 22.11 Utilizando la información mostrada en el ejercicio 22.10 una comisión decidió recomendar la construcción de hospitales mentales pequeños dispersos por la región en lugar de grandes centros de internamiento como se hacía hasta el momento. ¿Crees que esa recomendación está justificada teniendo en cuenta la información disponible? ¿Por qué?

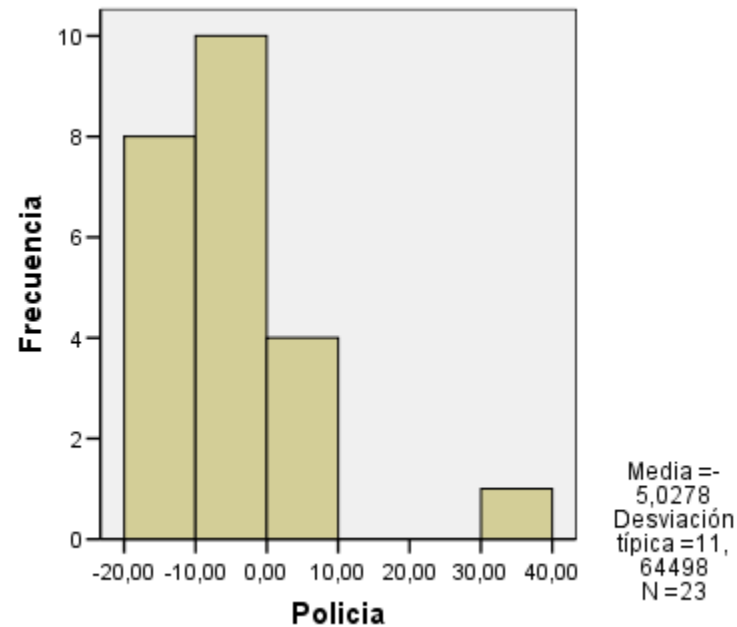
23. Mas policia para que haya menos crímenes

- En este ejemplo tenemos los datos de unos barrios de Nueva York en cuanto a robos de coches y hurtos en el año 1966-1967. En uno de los barrios se hizo una prueba en la que se aumentó el 40% el personal de policia para comprobar el efecto que esto tenía sobre el crimen. Los otros barrios también tuvieron variaciones debidas a las fluctuaciones normales de personal.
- Las variables son Policia: variación en el porcentaje de policias en ese barrio con respecto al periodo anterior. RoboCoches: Robos medios por semana. Hurtos: Hurtos medios por semana. Distrito: Distrito en el que se encuentra el barrio (hay 3 distritos)

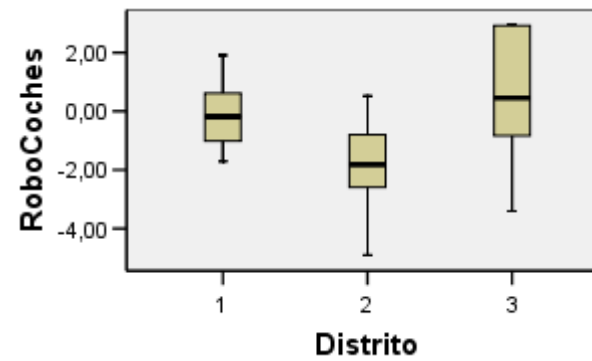
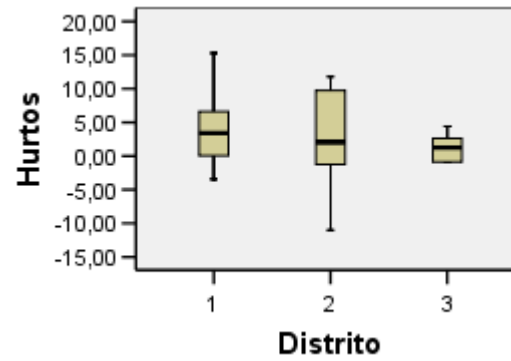
EJERCICIO 23.1 Describe los cambios ocurridos en los barrios en cuanto a personal, indicando la tendencia y los valores excepcionales.

Estadísticos descriptivos

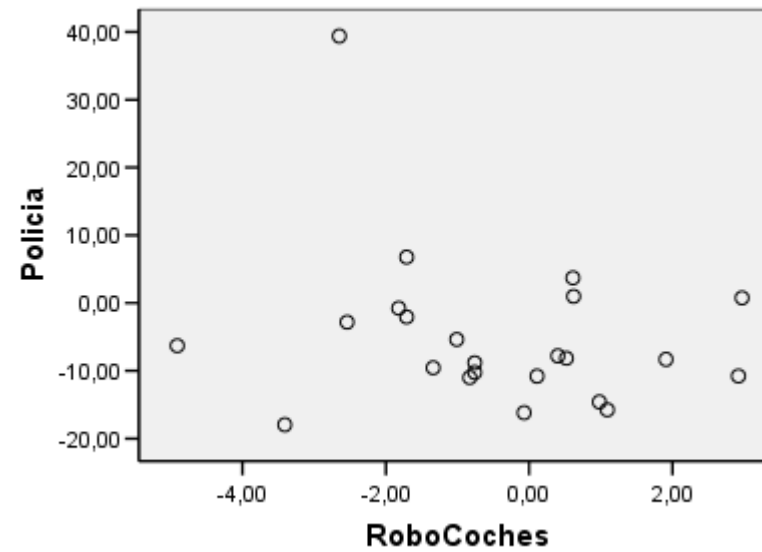
	N	Mínimo	Máximo	Media	Desv. típ.
Policia	23	-17.96	39.40	-5.0278	11.64498
N válido (según lista)	23				



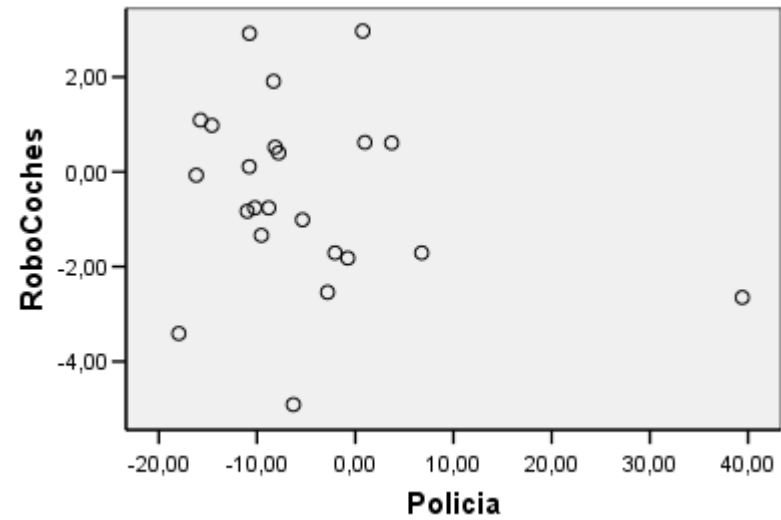
EJERCICIO 23.2 ¿Qué distrito te gustaría más para vivir a partir de estos gráficos?



EJERCICIO 23.3 ¿Qué dirías de este gráfico de la relación entre Policía y Robos de coches?



EJERCICIO 23.4 ¿Qué dirías de este gráfico de la relación entre Policia y Robos de coches?



EJERCICIO 23.5 Indica si hay valores especiales en el gráfico del ejercicio 23.4 desde el punto de vista estadístico.

EJERCICIO 23.6 ¿Que dirías de la relación entre policia y Hurtos según este gráfico?

