

Análisis Exploratorio

Analizar cada una de las variables continuas, observar su distribución junto con los resúmenes numéricos descriptivos. Para conseguir una representación conjunta de más de una variable, recordar que deben tener la misma unidad de medida. De lo contrario recurrir a la tipificación de las mismas para poder comparlas y representarlas conjuntamente. Interpretación de gráficos y resultados

• Cuadro de Diálogo del Análisis Exploratorio.

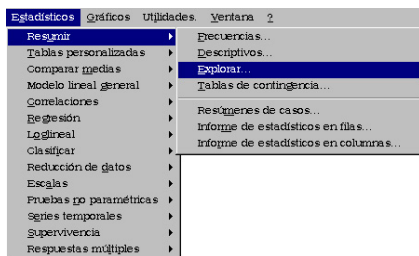


Figura 1

Estadística: Resumir: Explorar (*figura 1*): en lista de dependiente, incluir las variables a explorar (generalmente no más de 9 variables). Seleccionamos las variables que queramos explorar a la vez. Hay que tener en cuenta, que sólo podremos comparar a la vez, variables que tengan el mismo rango de escala, en caso contrario, previamente a este paso, debemos tipificar las variables, procedimiento que explicamos en el siguiente punto.

Una vez que tenemos las variables que queremos analizar y una vez tipificadas si fuera necesario, seleccionamos de la lista de variables de la izquierda (*figura 2*) la variables, marcandolas y pasandolas a la ventana de Dependientes.

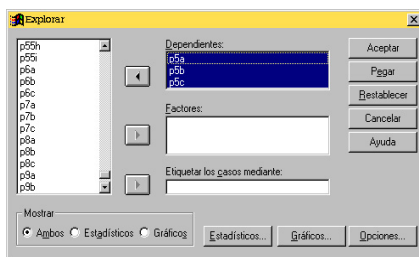


Figura 2

Posteriormente comprobamos que en la sección de mostrar, esté marcada la opción de Ambos, que supone que queremos que aparezcan tanto los estadísticos como los gráficos del análisis. Dentro de los estadísticos tendremos los estadísticos robustos, esenciales para la detección de una distribución normal. Por otro lado, los gráficos, se refieren básicamente al diagrama de tallos y hojas (stem and leaf), posteriormente al gráfico de “caja con bigotes” (box and whisker) y en ocasiones podemos pedirle que nos represente el histograma de dicha distribución.

A continuación entramos en los Estadísticos, si queremos especificar alguno de ellos en concreto. Si esta opción la dejamos como está, el propio ordenador calculará los más característicos y los necesarios, sólo en el caso que queramos especificar un estadístico concreto, entraremos en esta opción.

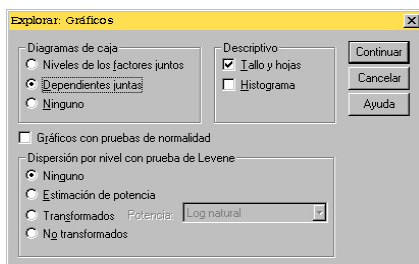


Figura 3

Lo siguiente será, pedirle al ordenador la representación gráfica que nos interesa (*figura 3*). Cuando tengamos varias variables que estén tipificadas o que posean la misma escala, y que queramos comparar entre ellas, marcaremos la opción dentro de “Diagrama de cajas”: “Dependientes juntas”. Esto nos permitirá representar en un mismo gráfico cada una de las “cajas con bigotes” juntas y de este modo poderlas comparar. En el caso de que solo tengamos una variable o no estén tipificadas, dejaremos la que viene por defecto: “Niveles de los factores juntos”. También marcamos los *gráficos con pruebas de normalidad*, que incorpora el test de Kolmogorov-Smirnov con la significación de Lilliefors que establece la probabilidad de cometer un error si rechazamos la hipótesis nula del contraste y es un buen indicativo de si la distribución es normal. Si la sig. es superior a 0,2000

la distribución de la población es normal.

Más a la derecha, en la opción de los “*Descriptivos*” le pedimos que

• Cuadro de Diálogo para la tipificación de variables.

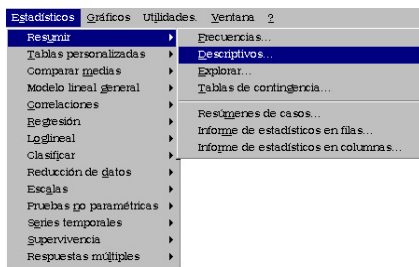


Figura 4

represente el diagrama de tallos y hojas y con esto tenemos todo lo necesario.

Estadística: Resumir: Descriptivos (*figura 4*): a partir de este momento debemos seleccionar la variable que queremos tipificar (*figura 5*) y trasladarla a la ventana de variables, seguidamente seleccionar la opción guardar valores tipificados como variables. En este momento aparecerán en el fichero de datos las nuevas variables tipificadas pero no se perderán las originales (sin normalizar). De este modo en cualquier momento y según las necesidades, podremos optar por utilizar unas u otras.

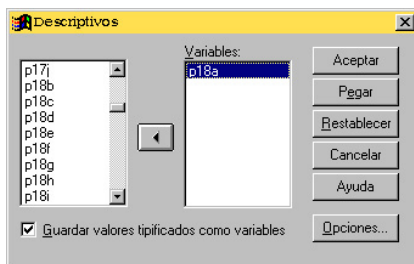


Figura 5

Como decíamos anteriormente, se genera una nueva variable la cual automáticamente es nombrada por el ordenador con la letra “z” delante del nombre de la variable, para de este modo identificar rápidamente la variable que ha sido estandarizada. La variable será colocada al final de la matriz de datos con dicho nombre, en nuestro ejemplo será: “zp18a”.

Sólo a partir de este momento, podemos representar las cajas con bigotes de las variables, por ejemplo, con diferentes unidades de medidas, en un mismo gráfico que nos permita compararlas y sacar conclusiones de su interpretación. Nunca, podremos estar trabajando con variables que tengan unidades de medida

Transformaciones.

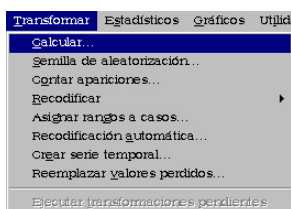


Figura 6

diferentes, porque en ningún caso, podremos comparar los resultados obtenidos con dichas variables.

Sabemos que la mayor parte de los análisis que vamos a utilizar a partir de ahora, son análisis paramétricos que exigen unos requisitos de distribución de los datos como los de normalidad, homocedasticidad... Por ello un paso previo a cualquier análisis de este tipo, consiste en realizar las transformaciones oportunas para que las variables adopten una forma normal o al menos simétrica. Las transformaciones más comúnmente utilizadas son lo “ln (ln)” “raíz cuadrada (SQRT)” y “potencias (*2)”. El favorecer la normalidad en la distribución de una variable no es sólo una característica deseable en si misma, sino que además propicia la linealidad en su relación con el resto de las variables.

Transformar: Calcular (*figura 6*): indicar en variable de destino el nombre de la nueva variable (*figura 7*), que va a crearse a partir de la transformación que indiquemos. A continuación, seleccionar del cuadro de funciones la operación matemática adecuada a la transformación y llevarla al cuadro expresión numérica.

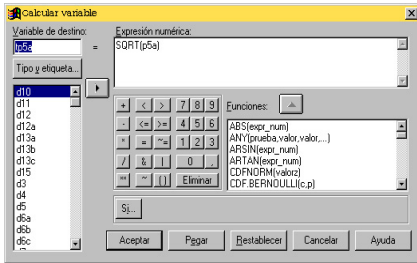


Figura 7

Luego, y para completar la expresión numérica anterior, indicar la variable original que quiere transformarse mediante esa función llevándola a este mismo cuadro.

La transformación nos permitirá obtener de una variable que en principio no sigue una distribución normal, una variables que se aproxime a la distribución normal. Las posibilidades de transformación son varias, según que su distribución sea asimétrica positiva (mediana mas cerca de la parte inferior de la caja; concentración de casos en los valores inferiores de la distribución y cola extendida hacia los valores grandes) o negativa (mediana más cerca de la parte superior de la caja; casos concentrados en los valores altos y cola alargada hacia los valores inferiores de la distribución). *Tukey* ofrece lo que él llama la “Escalera de las transformaciones”, donde muestra el tipo de transformación recomendada según sea la intensidad de la asimetría o la dirección en la que van los casos extremos.

Si la distribución es asimétrica negativa (x^2 , x^3 , antilog x) (la mediana más cerca de la parte inferior de la caja, concentración de los casos en los valores inferiores de la distribución y cola extendida hacia los valores grandes): elevar al cuadrado o al cubo, así como calcular antilogaritmos, tiene el efecto de corregir la asimetría (más fuerte la corrección cuanto mayor sea la exponenciación).

Para corregir distribuciones asimétricas positivas ($\log x$, \sqrt{x} , $-1/x$, $-1/x^2$) (Mediana más cerca de la parte superior de la caja, concentración de los casos en los valores altos de la distribución y la cola alargada hacia los valores inferiores de la distribución): es conveniente utilizar raíces cuadradas, logaritmos, etc.

Análisis Exploratorio

• Resumen de los Datos.

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
P5A	1200	100,0%	0	,0%	1200	100,0%
P5B	1200	100,0%	0	,0%	1200	100,0%
P5C	1200	100,0%	0	,0%	1200	100,0%

Descriptivos

		Estadístico	Error típ.
P5A	Media	6,12	5,50E-02
	Intervalo de confianza para la media al 95%	Límite inferior	6,01
		Límite superior	6,22
	Media recortada al 5%	6,16	
	Mediana	6,00	
	Varianza	3,627	
	Desv. típ.	1,90	
	Mínimo	0	
	Máximo	11	
	Rango	11	
	Amplitud intercuartil	2,00	
	Asimetría	-,428	,071
	Curtosis	1,091	,141

Descriptivos

			Estadístico	Error típ.
P5B	Media		6,07	5,79E-02
	Intervalo de confianza para la media al 95%	Límite inferior	5,96	
		Límite superior	6,19	
	Media recortada al 5%		6,13	
	Mediana		6,00	
	Varianza		4,018	
	Desv. típ.		2,00	
	Mínimo		0	
	Máximo		11	
	Rango		11	
	Amplitud intercuartil		2,00	
	Asimetría		-,477	,071
	Curtosis		,678	,141
P5C	Media		7,24	7,79E-02
	Intervalo de confianza para la media al 95%	Límite inferior	7,09	
		Límite superior	7,40	
	Media recortada al 5%		7,38	
	Mediana		7,00	
	Varianza		7,289	
	Desv. típ.		2,70	
	Mínimo		0	
	Máximo		11	
	Rango		11	
	Amplitud intercuartil		5,00	
	Asimetría		-,374	,071
	Curtosis		-,285	,141

Stem width: 1
Each leaf: 7 case(s)

Stem width: 1
Each leaf: 6 case(s)

[illegible]

- *Representación Gráfica. Gráfico de Caja con Bigotes (Box and Wisper).*

