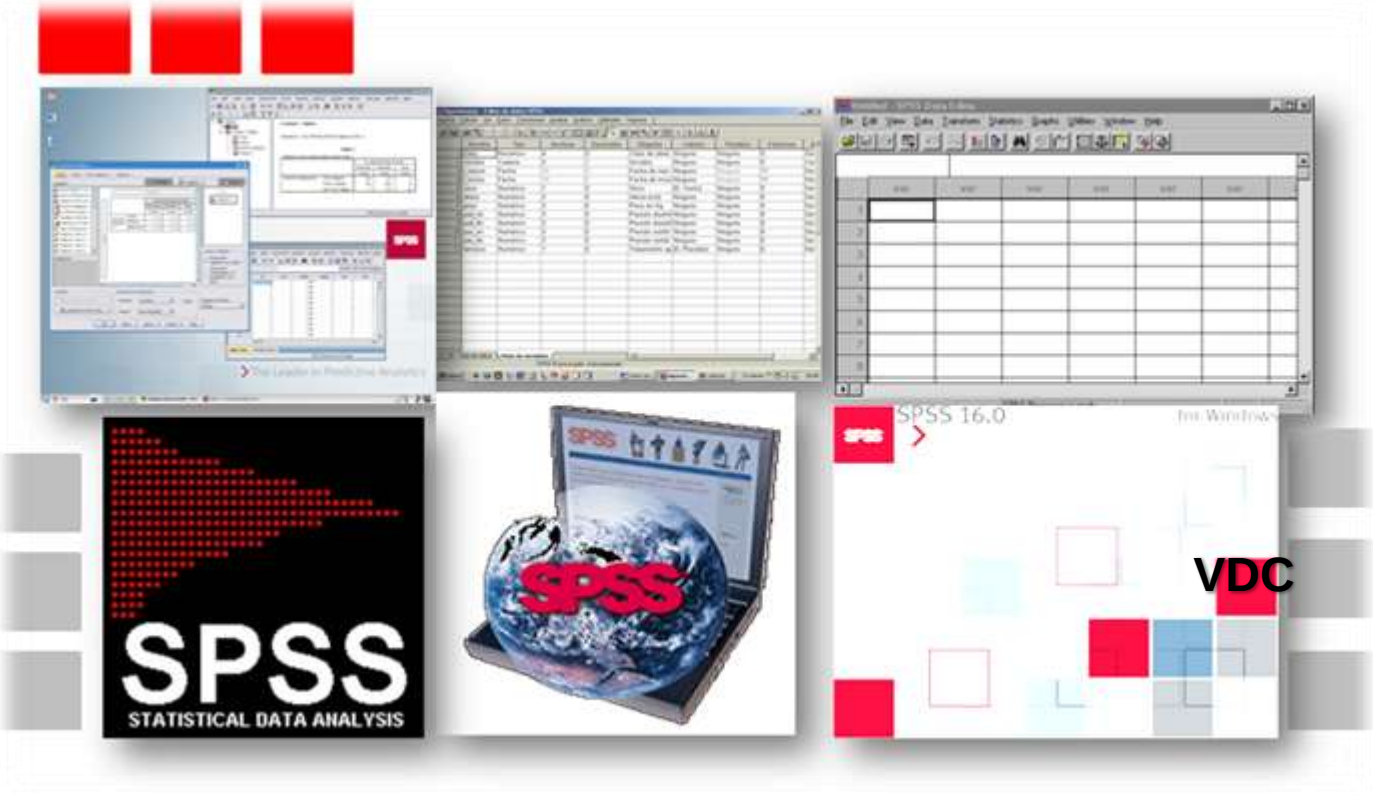
Estrategia de análisis estadístico de los datos. Estadística descriptiva



Prof. Mª JOSÉ PRIETO CASTELLÓ

SPSS

DEFINIR DATOS

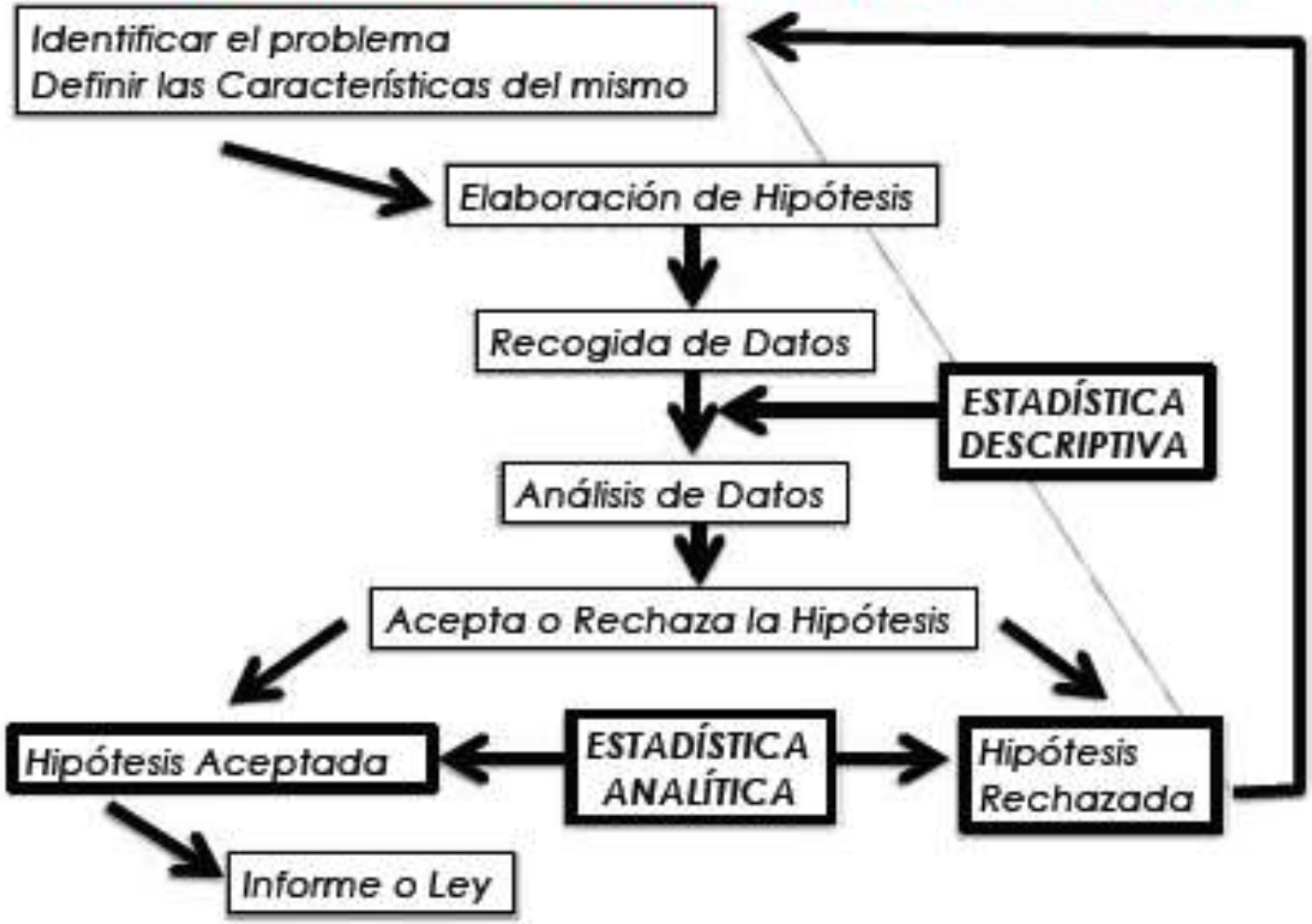
MODIFICAR DATOS

ANÁLISIS ESTADÍSTICO DE DATOS

Estudio descriptivo

Estudio analítico o inferencia estadística

La estadística en el método científico



SPSS

ANÁLISIS ESTADÍSTICO DE DATOS

- **Estadística Descriptiva:**

-Cualitativas: frecuencias, porcentajes

-Cuantitativas: descriptivos, medidas de centralizacíón, dispersión, posición..)

- **Explorar**
- **Tablas de contingencia**

SPSS

ESTADÍSTICA DESCRIPTIVA

- Parte de la estadística cuyo objetivo es identificar, conocer, describir y resumir las variables de la muestra.
  - Variables categóricas.
  - Variables cuantitativas.
  - Outliers.
  - Distribuciones de probabilidad.

SPSS

ESTADÍSTICA DESCRIPTIVA

- Una vez que se han recogido los valores que toman las variables de nuestro estudio (datos), procederemos al análisis descriptivo de los mismos. Para variables categóricas, como el sexo, se quiere conocer el número de casos en cada una de las categorías, reflejando habitualmente el porcentaje que representan del total, y expresándolo en una tabla de frecuencias.

- Para variables numéricas, en las que puede haber un gran número de valores observados distintos, se ha de optar por un método de análisis distinto, respondiendo a las siguientes preguntas:

  1. ¿Alrededor de qué valor se agrupan los datos?

  2. Supuesto que se agrupan alrededor de un número, ¿cómo lo hacen? ¿muy concentrados? ¿muy dispersos?

SPSS

VARIABLES CATEGÓRICAS

- No se realizan cálculos matemáticos (Sexo). Su descripción se basa en tabla de frecuencias y en representaciones gráficas.

  - Tabla de frecuencias.
  - Gráficos de barras.
  - Gráficos de sectores.

SPSS

VARIABLES CATEGÓRICAS

- Fc Absoluta.
- Fc. Relativa.
- Porcentaje.
- Fc. Absoluta acumulada.
- Fc. Relativa acumulada.
- Porcentaje acumulado.

- **Absolutas:** nº de sujetos en cada categoría.
- **Relativas:** % de sujetos

SPSS

VARIABLES CATEGÓRICAS

- La frecuencia absoluta de una variable estadística es el número de veces que aparece en la muestra dicho valor de la variable, la representaremos por n_i

$$f_i = \frac{n_i}{N}$$

- La frecuencia absoluta, es una medida que está influida por el tamaño de la muestra, al aumentar el tamaño de la muestra aumentará también el tamaño de la frecuencia absoluta. Esto hace que no sea una medida útil para poder comparar. Para esto es necesario introducir el concepto de frecuencia relativa, que es el cociente entre la frecuencia absoluta y el tamaño de la muestra. La denotaremos por f_i

SPSS

VARIABLES CATEGÓRICAS

Porcentaje

- La frecuencia relativa es un tanto por uno, sin embargo, hoy día es bastante frecuente hablar siempre en términos de tantos por ciento o porcentajes, por lo que esta medida resulta de multiplicar la frecuencia relativa por 100. La denotaremos por p_i.

$$p_i = f_i \cdot 100\,\%$$

SPSS

VARIABLES CATEGÓRICAS

Fc. Absoluta acumulada

- Para poder calcular este tipo de frecuencias hay que tener en cuenta que la variable estadística ha de ser cuantitativa o cualitativa ordenable. En otro caso no tiene mucho sentido el cálculo de esta frecuencia. La frecuencia absoluta acumulada de un valor de la variable, es el número de veces que ha aparecido en la muestra un valor menor o igual que el de la variable y lo representaremos por N_i.

SPSS

VARIABLES CATEGÓRICAS

Fc. Relativa acumulada

- Al igual que en el caso anterior la frecuencia relativa acumulada es la frecuencia absoluta acumulada dividido por el tamaño de la muestra, y la denotaremos por F_i

Porcentaje acumulado

- Análogamente se define el Porcentaje Acumulado y lo vamos a denotar por P_i como la frecuencia relativa acumulada por 100.

SPSS

VARIABLES CATEGÓRICAS

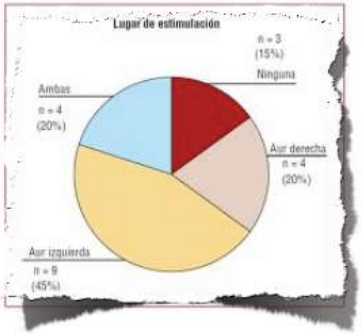
Ejemplo:

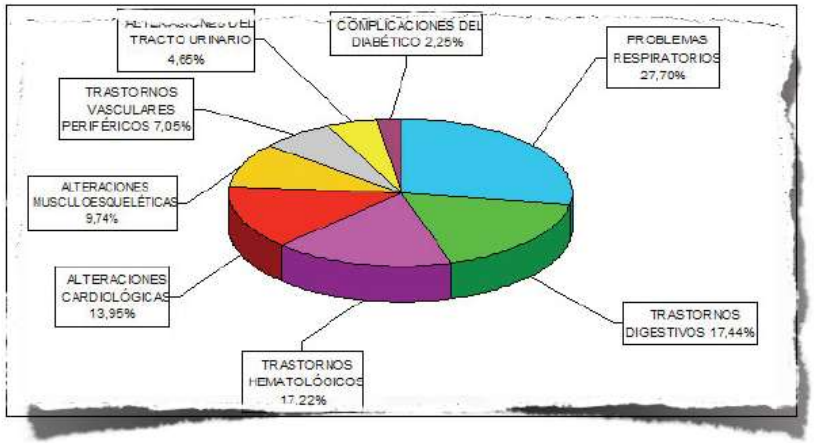
Personas Activas	Nº de Familias					
X_i	n_i	F_i	p_i	N_i	F_i	P_i
1	16	16/50	32%	16	16/50	32%
2	20	20/50	40%	36	36/50	72%
3	9	9/50	18%	45	45/50	90%
4	5	5/50	10%	50	50/50	100%
Total	50					

SPSS

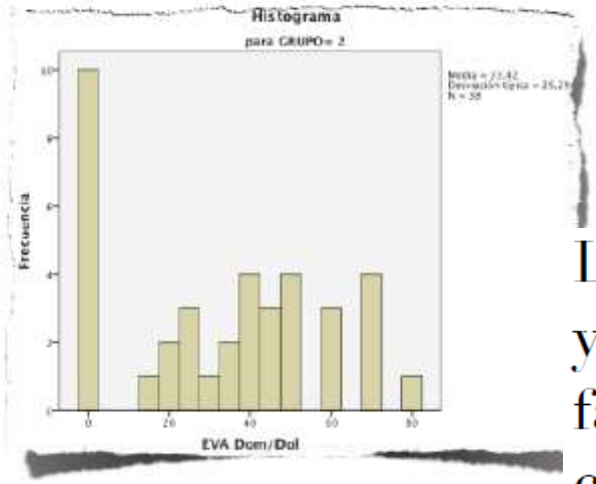
VARIABLES CATEGÓRICAS

Gráficos de sectores:





Gráficos de barras



Los gráficos son útiles porque ponen en relieve y aclaran las tendencias que no se captan fácilmente en la tabla, ayudan a estimar valores con una simple ojeada y brinda una verificación gráfica de la veracidad de las soluciones.

SPSS

VARIABLES CATEGÓRICAS

Tabla de frecuencias: SPSS

1. Menú analizar: Estadísticos descriptivos: opción "Frecuencias".
2. Seleccionar las variables y pasarlas a la ventana "Variables".
3. Si queremos gráficos (de barras, de sectores), seleccionarlos en el botón "gráficos".

SPSS

VARIABLES CUANTITATIVAS

Medidas estadísticas

- **Localización:** cuartiles, deciles, percentiles, puntos de corte.
- **Centralización:** Moda, Mediana y Media.
- **Dispersión:** desviación, desviación media, desviación típica, varianza.
- **Estadísticos de forma:** simetría, curtosis.

SPSS

VARIABLES CUANTITATIVAS

Cuantiles

- Son estadísticos de posición. Nos dan una idea de la distribución general de los valores de esa variable:
  - Percentiles.
  - Cuartiles.
  - Puntos de corte.
- Los cuantiles permiten hacer un análisis minucioso de la distribución, se utilizan generalmente cuando se quiere ubicar un dato dentro del conjunto. Por ejemplo. Pertenece el dato x al 50% superior ?, al 10% inferior? , al 50 % central?, etc.

SPSS

VARIABLES CUANTITATIVAS

Percentiles

- Medida de localización que divide la población o muestra en 100 partes iguales.
- Valor de la variable por debajo del cual se encuentran el porcentaje correspondiente de sujetos.

Puntos de corte

- Calcula los k-1 valores que dividen la muestra en k grupos del mismo tamaño. Por defecto, en SPSS k=10. Son los percentiles 10, 20, 30, 40, etc.

SPSS

VARIABLES CUANTITATIVAS

Cuartiles

- Medida de localización que divide la población o muestra en cuatro partes iguales. Equvale a los percentiles 25, 50 y 75.
  - Q1= Valor de la variable que deja a la izquierda el 25% de la distribución.
  - Q2= Valor de la variable que deja a la izquierda el 50% de la distribución = mediana.
  - Q3= Valor de la variable que deja a la izquierda el 75% de la distribución.

SPSS

VARIABLES CUANTITATIVAS

Estadísticos de tendencia central:

- Índices del "centro" de una distribución. Se suele considerar que el valor central es representativo de toda la variable.
  - Media.
  - Mediana.
  - Moda.

SPSS

VARIABLES CUANTITATIVAS

Media

- Promedio aritmético de los valores observados.
- Es el mejor estadístico de tendencia central si la muestra se distribuye simétricamente.

$X=\Sigma xi/N$

Mediana

- Valor que ocupa una posición central cuando se ordenan los datos de menor a mayor.
- Es el mejor estadístico de tendencia central en una distribución asimétrica.

Moda

- Valor que se presenta con más frecuencia.

SPSS

VARIABLES CUANTITATIVAS

Ejemplo:

Simétrica

- 8, 2, 8, 1, 9, 3, 8, 8, 2
  - Media: 5,44
  - Mediana: 8
  - Moda:8

La MEDIA

- 8, 2, 8, 1, 9, 3, 8, 8, 87
  - Media: 14,89
  - Mediana: 8
  - Moda:8

La MEDIANA

Asimétrica

SPSS

VARIABLES CUANTITATIVAS

Medidas de dispersión

- Nos dan información de cuanto se apartan los datos de la media.
  - Rango.
  - Varianza.
  - Desviación típica.
  - Error típico de la media.
  - Coeficiente de variación.

SPSS

VARIABLES CUANTITATIVAS

Rango

- Diferencia entre el máximo y el mínimo de los valores observados.
- El rango o recorrido da alguna idea del grado de variación que ocurre en la población, pero con frecuencia los resultados pueden ser engañosos, pues este depende de los valores extremos e ignora la variación de las demás observaciones. Está afectado por ocurrencias raras o extraordinarias.

SPSS

VARIABLES CUANTITATIVAS

Varianza

$S^2 = \Sigma(Xi-X)^2/N$

- (s^2) medida de la dispersión en torno a la media.
- Se calcula elevando al cuadrado las desviaciones y sumando los cuadrados, dividiendo la suma por el número de casos.
- Tiene la ventaja de no prescindir de los signos de las desviaciones, pero al igual que la desviación media los valores extremos pueden distorsionarla.
- En inferencia, con una muestra tomada de una población grande se pretende descubrir cuanto varían los datos al rededor de la media poblacional.
- Es considerado un estimador más eficiente para la varianza poblacional.

SPSS

VARIABLES CUANTITATIVAS

Desviación típica

X ± s → 68%de los datos
X ± 1,96s → 95%
X ± 2,58s → 99%

DIST NORMAL

- (s): es la raíz cuadrada de la varianza, es decir, ofrece la misma información que la varianza pero en las mismas unidades que la propia variable.
- Es una buena medida de dispersión para datos medidos simétricamente.
- Es la medida de dispersión más utilizada.

SPSS

VARIABLES CUANTITATIVAS

Coeficiente de variación

CV= S/X * 100 (%)

- C. de V. de Pearson: es una media de dispersión adimensional.
- Es el porcentaje que representa la desviación estándar sobre la media.
- Se usa para comparar dispersiones relativas de variables medidas en las mismas o diferentes unidades.
- Para comparar la variabilidad de dos muestras diferentes NUNCA compararemos sus desviaciones típicas, sino sus coeficientes de variación.

SPSS

VARIABLES CUANTITATIVAS

Error típico de la media

EEM=S/√N

- Es una medida de cuánto puede variar el valor de una media de una muestra a otra, extraídas de la misma población.
- Se calcula dividiendo la desviación estándar de la muestra entre la raíz cuadrada del tamaño muestral.
- Con él se calculan los intervalos de confianza de la media, es decir, los valores entre los cuales puedo afirmar, con un determinado grado de confianza, que estará la media de la población.

Linferior= X – 1,96xEEM
Lsuperior=X+1,96xEEM

IC (95%)

SPSS

VARIABLES CUANTITATIVAS

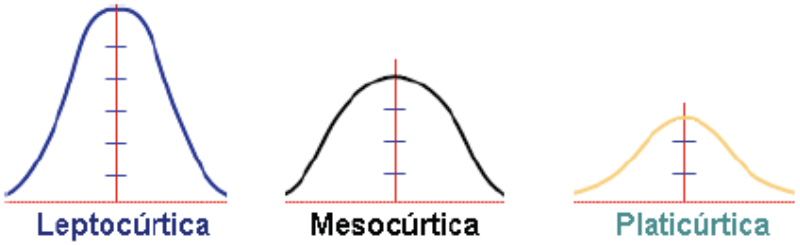
Estadísticos de forma

- Índice de asimetría.
- Curtosis.
- Informan de cuanto se apartan los datos de la media.
- Nos dan una idea de la forma que tendrá la curva que resulta de representar gráficamente los datos frente a su frecuencia.

SPSS

VARIABLES CUANTITATIVAS

Curtosis



- Es una medida del grado en que las observaciones están agrupadas en torno al punto central.
- Puede ser:
  - **Leptocúrtica**: observaciones más agrupadas, colas más largas. >0
  - **Platicúrtica**: observaciones más dispersas, colas más cortas. <0
  - **Mesocúrtica**: concentración normal. =0

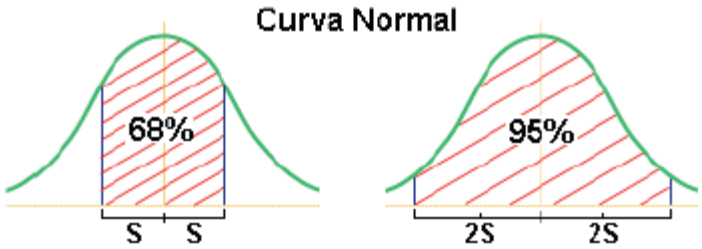
Índice de asimetría

- Indicardor de la simetría de los datos distribuidos a los lados de la media. Simétrica = 0
  - Asimetría negativa: los datos tienden a ser menores que la media. <0
  - Asimetría positiva: los datos tienden a ser mayores que la media. >0

SPSS

VARIABLES CUANTITATIVAS

- Cuando la distribución de los datos cuenta con un coeficiente de asimetría de ±0.5 y un coeficiente de curtosis de ±0.5 de le denomina curva NORMAL.
- Para la mayoría de los procedimientos es necesario que los datos se distribuyan normalmente.
- La ventaja radica en el supuesto que el 95% de los valores se encuentra dentro de una distancia de 2 desviaciones estándar de la media.



SPSS

VARIABLES CUANTITATIVAS

SPSS: Frecuencias

1. Menú analizar: Estadísticos descriptivos: opción "Frecuencias".
2. Seleccionar las variables y pasarlas a la ventana "Variables".
3. Si queremos gráficos, seleccionarlos en el botón "Gráficos" (no nos permite pedir diagramas de caja).
4. Seleccionar los estadísticos que queramos en la ventana "Estadísticos" (podemos pedir puntos de corte y errores típicos de la media).
5. Aceptar

También DESCRIPTIVOS (menos información)

SPSS

VARIABLES CUANTITATIVAS

SPSS Explorar

1. Menú analizar: Estadísticos descriptivos: opción “Explorar”.
2. Seleccionar las variables y pasarlas a la ventana “Dependientes”.
3. Si queremos que nos muestren estadísticos de una variable en función de otra, pasamos la variable de agrupación a la ventana “Factores”.
4. Si queremos gráficos, seleccionarlos en el botón “Gráficos”.
5. Seleccionar los estadísticos que queramos en la ventana “Estadísticos” /están preestablecidos: podemos pedir los percentiles y los valores extremos).
6. Aceptar.

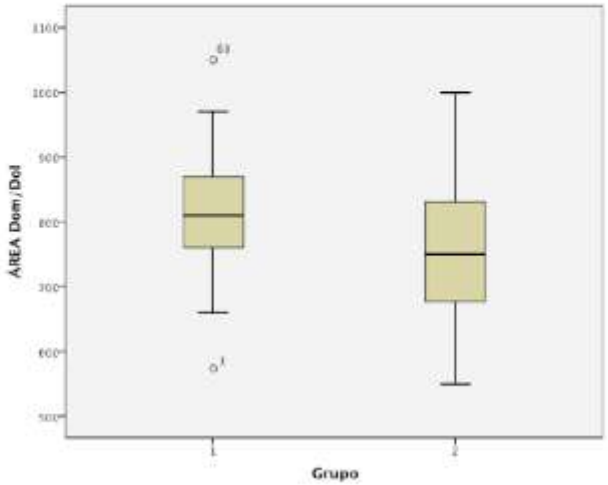
ESTADÍSTICA DESCRIPTIVA

SPSS

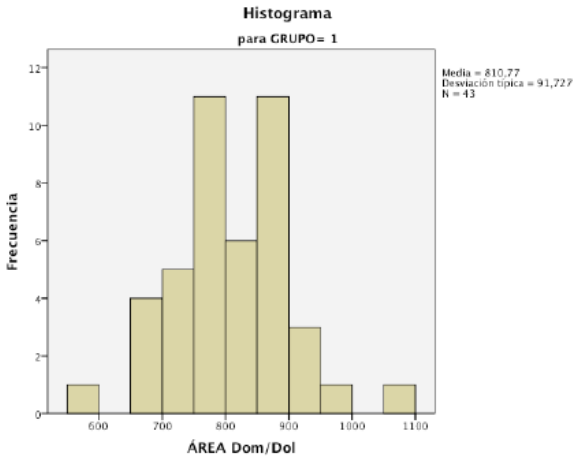
VARIABLES CUANTITATIVAS

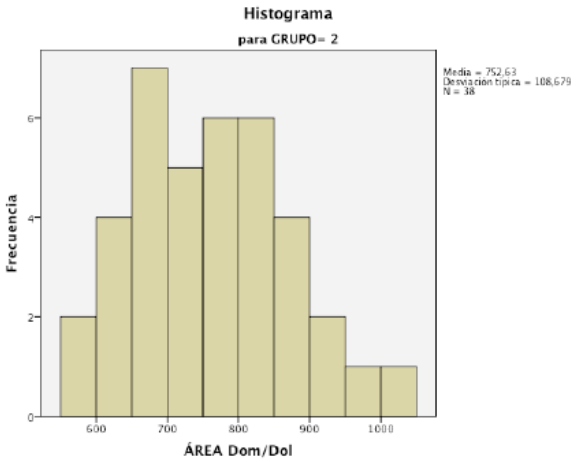
Gráficos

Diagramas de caja:



Histogramas:





Outliers

- Es una observación numéricamente distante del resto de los datos.
- Se trata de una o pocas observaciones. Si hay más, lo más probable es que la técnica de medida no haya sido correcta, lo que muy probablemente invalidará la muestra.
- Un verdadero Outlier puede venir de una avería puntual de un aparato de medida, de un error en la transmisión o transcripción de los datos, o de una contaminación de la muestra con elementos externos a la población examinada.
- También puede deberse a un fallo en la hipótesis de partida, o en las teorías en las que esta hipótesis se basa. Un outlier puede ser la base de un descubrimiento.