



## REGRESIÓN LINEAL CON SPSS

### 1.- INTRODUCCIÓN

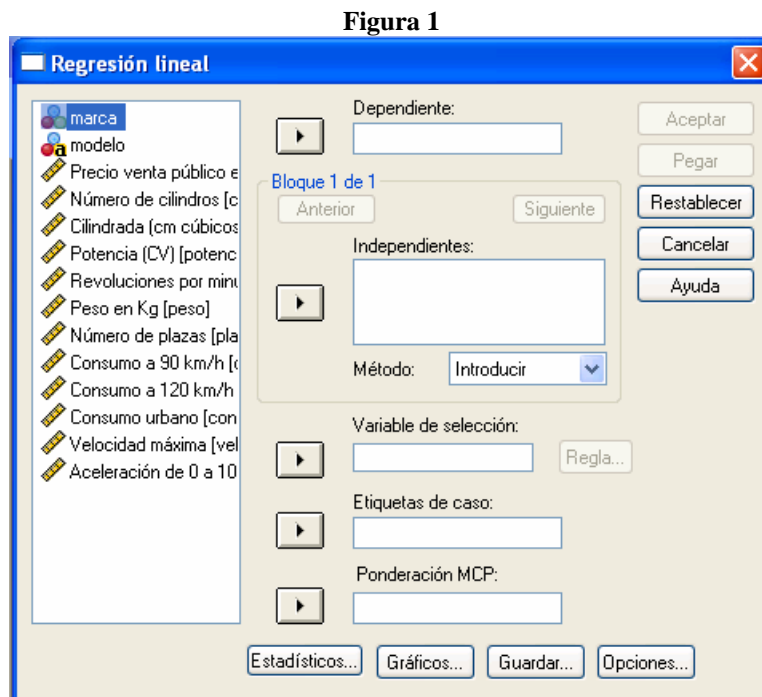
El análisis de regresión lineal es una técnica estadística utilizada para estudiar la relación entre variables cuantitativas. Tanto en el caso de dos variables (regresión *simple*) como en el de más de dos variables (regresión *múltiple*), el análisis de regresión lineal puede utilizarse para explorar y cuantificar la relación entre una variable llamada dependiente o criterio (Y) y una o más variables llamadas independientes o predictoras ( $X_1, X_2, \dots, X_p$ ), así como para desarrollar una ecuación lineal con fines predictivos.

### 2.- REGRESIÓN LINEAL

Para llevar a cabo un análisis de regresión lineal seleccionamos:

**Analizar**  
**Regresión**  
**Lineal...**

Apareciendo el cuadro de diálogo de la figura 1:



En **Dependiente** se traslada la variable cuyos valores se desea predecir o resumir.

**Bloque 1 de 1.** Este recuadro nos sirve para introducir las variables independientes, nos permite hacer varios análisis de regresión a la vez, alternando los botones **anterior** y **siguiente**, y además no deja elegir el método de introducción de las variables independientes.

- **Independientes** se trasladan las variables utilizadas para predecir el valor de la variable dependiente. También se denominan variables predictoras o variables explicativas. Para poder ejecutar este procedimiento, la lista debe contener al menos una variable.
- **Método:** Permite seleccionar el método por el cual se introducen las variables independientes en el análisis. Nos vale para elegir la mejor ecuación de regresión. Permite construir una variedad de modelos de regresión a partir del mismo conjunto de variables:
  - **Introducir (Entry):** Procedimiento para la selección de variables en el que todas las variables de un bloque se introducen en un solo paso. Es el método por defecto.
  - **Pasos sucesivos (Stepwise):** En cada paso se introduce la variable independiente que no se encuentre ya en la ecuación y que tenga la probabilidad para F más pequeña, si esa probabilidad es suficientemente pequeña. Las variables ya introducidas en la ecuación de regresión se eliminan de ella si su probabilidad para F llega a ser suficientemente grande. El método termina cuando ya no hay más variables candidatas a ser incluidas o eliminadas.
  - **Eliminar (Remove):** Procedimiento para la selección de variables en el que las variables de un bloque se eliminan en un solo paso.
  - **Hacia atrás (Backward):** Procedimiento de selección de variables en el que se introducen todas las variables en la ecuación y después se van excluyendo una tras otra. Aquella variable que tenga la menor correlación parcial con la variable dependiente será la primera en ser considerada para su exclusión. Si satisface el criterio de eliminación, será eliminada. Tras haber excluido la primera variable, se pondrá a prueba aquella variable, de las que queden en la ecuación, que presente una correlación parcial más pequeña. El procedimiento termina cuando ya no quedan en la ecuación variables que satisfagan el criterio de exclusión.
  - **Hacia delante (Forward):** Procedimiento de selección de variables en el que éstas son introducidas secuencialmente en el modelo. La primera variable que se considerará para ser introducida en la ecuación será aquella que tenga mayor correlación, positiva o negativa, con la variable dependiente. Dicha variable será introducida en la ecuación sólo si satisface el criterio de entrada. Si ha entrado la primera variable, se considerará como próxima candidata la variable independiente que no esté en la ecuación y cuya correlación parcial sea la mayor. El procedimiento termina cuando ya no quedan variables que satisfagan el criterio de entrada.

En **Variable de selección** se traslada una variable que limite el análisis a un subconjunto de casos que tengan un valor particular para esta variable. Con **Regla...** podemos definir el subconjunto de casos que se emplearán para estimar el modelo de regresión.

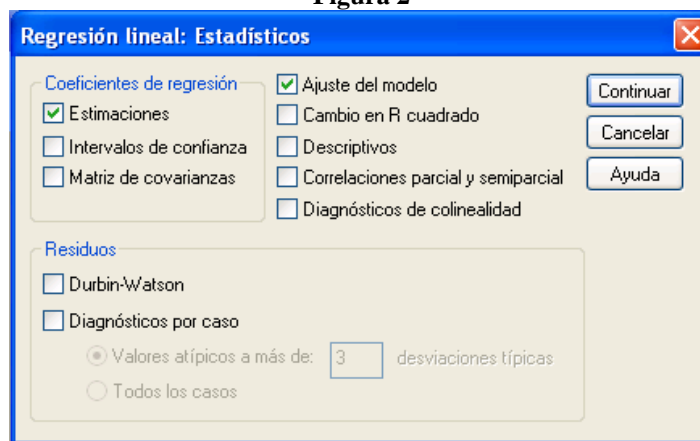
En **Etiquetas de caso** designamos una variable para identificar los puntos de los gráficos. Para cada punto de un diagrama de dispersión podemos utilizar la herramienta de selección de puntos y mostrar el valor de la variable de etiquetas de casos correspondiente al caso seleccionado.

**Ponderación MCP:** Permite obtener un modelo de mínimos cuadrados ponderados. Los puntos de los datos se ponderan por los inversos de sus varianzas. Esto significa que las observaciones con varianzas grandes tienen menor impacto en el análisis que las observaciones asociadas a varianzas pequeñas.

## 2.1.- Estadísticos...

Con el botón **Estadísticos** accedemos al cuadro de diálogo que muestra la figura 2 que nos nos vale para solicitar resultados estadísticos opcionales, incluyendo los coeficientes de regresión, descriptivos, estadísticos de ajuste del modelo, la prueba de Durbin-Watson y diagnósticos de la colinealidad.

Figura 2



**Coefficientes de regresión.** En este recuadro podemos obtener tanto las estimaciones de los coeficientes de regresión, la bondad del ajuste del modelo elegido, los intervalos de confianza de cada coeficiente así como la matriz de covarianzas. Podemos elegir una o más de las opciones:

- **Estimaciones:** nos muestra coeficientes de regresión y medidas relacionadas. Los coeficientes *no estandarizados* (no tipificados) son los coeficientes de regresión parcial que definen la ecuación de regresión en puntuaciones directas. Los coeficientes *estandarizados* ( $\beta$ ) son los coeficientes que definen la ecuación de regresión en puntuaciones típicas. Estos coeficientes estandarizados ayudan a valorar la importancia relativa de cada variable independiente dentro de la ecuación. Muestra las pruebas de significación de cada coeficiente, el estadístico de contraste (t) así como su nivel crítico (Sig.). Una significación pequeña nos permite afirmar que el coeficiente es significativo.
- **Intervalos de confianza:** nos muestra intervalos de confianza al 95% para los coeficientes de regresión.
- **Matriz de covarianzas:** nos muestra la matriz de varianza-covarianza de los coeficientes de regresión, las varianzas en la diagonal y las covarianzas por debajo y por encima de la diagonal.

**Ajuste del modelo.** Muestra el *coeficiente de correlación múltiple* (R múltiple), y su cuadrado ( $R^2$ , *coeficiente de determinación*, que expresa la proporción de varianza de la variable dependiente que está explicada por la variable o variables independientes), la *R cuadrado corregida* y el *error típico de la estimación* (desviación típica de los residuos). También, una tabla de ANOVA muestra las sumas de cuadrados, los grados de libertad, las medias cuadráticas, el valor del estadístico F y el nivel crítico (Sig.) de la F.

**Cambio en R cuadrado.** Nos muestra el cambio en el estadístico R cuadrado que se produce al añadir o eliminar una variable independiente. Si el cambio en R cuadrado asociado a una variable es grande, significa que esa variable es un buen predictor de la variable dependiente.

**Descriptivos.** Muestra las medias de las variables, las desviaciones típicas y la matriz de correlaciones con las probabilidades unilaterales.

**Correlaciones parcial y semiparcial.** Muestra las correlaciones de orden cero, semiparcial y parcial. Los valores del coeficiente de correlación van de -1 a 1. El signo del coeficiente indica la dirección de la relación y su valor absoluto indica la fuerza de la relación. Los valores mayores indican que la relación es más estrecha.

**Diagnósticos de colinealidad.** Muestra las tolerancias para las variables individuales y una variedad de estadísticos para diagnosticar los problemas de colinealidad. La colinealidad (o multicolinealidad) es una situación no deseable en la que una de las variables independientes es una función lineal de otras variables independientes.

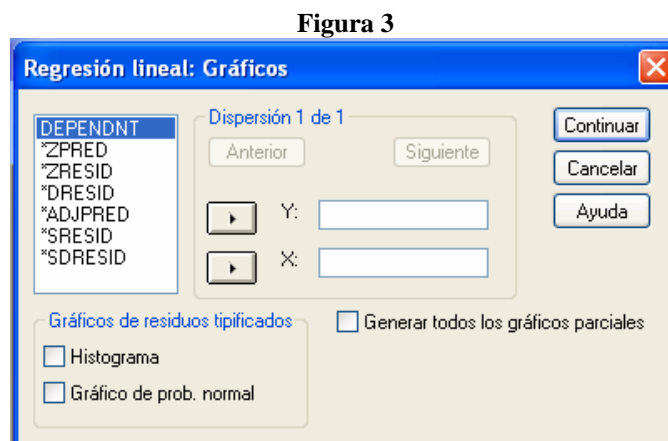
**Residuos.** Este recuadro nos permite seleccionar una de las opciones:

- **Durbin-Watson:** Muestra la prueba de D-W para los residuos correlacionados serialmente. Éste estadístico oscila entre 0 y 4 y toma el valor 2 cuando los residuos son completamente independientes. Los valores mayores de 2 indican autocorrelación positiva y los menores de 2 autocorrelación negativa. Suele asumirse que los residuos son independientes si el estadístico de D-W está entre 1'5 y 2'5. También muestra estadísticos de resumen para los residuos y los valores pronosticados.

- **Diagnósticos por caso:** Genera diagnósticos por casos, para todos los casos que cumplan el criterio de selección (los valores atípicos por encima de n desviaciones típicas).

## 2.2.- Gráficos...

Con el botón **Gráficos...** obtenemos el cuadro de diálogo de la figura 3



En la lista fuente tenemos la variable dependiente (DEPENDT), los valores predichos estandarizados (ZPRED), los residuos estandarizados (ZRESID), los residuos eliminando la puntuación del sujeto (DRESID) y los valores predichos ajustados (SDRESID).

**Dispersión 1 de 1.** Nos muestra los diagramas de dispersión que queramos de la lista de la izquierda, para cada par de variables, alternando **anterior y siguiente**.

**Gráficos de residuos tipificados.** En este recuadro podemos elegir uno de los gráficos:

- **Histograma:** Crea un histograma de los residuos tipificados con una curva normal superpuesta.
- **Gráfico de probabilidad normal:** Muestra un gráfico de probabilidad normal de los residuos tipificados. Se usa para comprobar la normalidad. Si la variable se distribuye normalmente, los puntos representados forman una línea recta diagonal.

**Generar todos los gráficos parciales.** Genera todos los diagramas de dispersión de la variable dependiente con cada una de las variables independientes.

### **2.3.- Guardar...**

El botón **Guardar...** nos permite guardar los valores pronosticados, los residuos y medidas relacionadas como nuevas variables que se añaden al archivo de datos de trabajo. En los resultados, una tabla muestra el nombre de cada nueva variable y su contenido.

### **2.4.- Opciones...**

El botón **Opciones...** nos permite controlar los criterios por los que se eligen las variables para su inclusión o exclusión del modelo de regresión, suprimir el término constante y controlar la manipulación de los valores perdidos.