

Práctica 1

Vamos a considerar el problema de estimación de un modelo de regresión lineal simple mediante STATGRAPHICS. Con el fin de asegurar el cumplimiento de las hipótesis de dicho modelo, e ilustrar la bondad de las estimaciones mínimo cuadráticas, vamos a simular una nube de puntos según un modelo lineal.

Así, generaremos 100 pares de observaciones de una variable bidimensional (X, Y) , con Y definida a partir de X mediante una ecuación lineal agregándole un término de perturbación ε . Tomaremos:

- (1) 100 observaciones x_i dadas por 100 números aleatorios con distribución $U([1, 5])$.
- (2) 100 términos de perturbación ε_i dados por 100 números aleatorios con distribución $N(0, 0,8)$.
- (3) 100 observaciones y_i definidas por la ecuación

$$y_i = -2 + 3x_i + \varepsilon_i$$

Generación de los datos

En primer lugar nombraremos las tres primeras columnas de la hoja de datos como **X**, **Epsilon**, e **Y**. Para ello debemos hacer click con el ratón sobre el encabezado de cada columna para seleccionarla, y luego hacer click con el botón derecho del ratón para desplegar el menú emergente asociado a la columna. Entonces seleccionaremos la opción **Modify Column** y modificaremos el nombre de la columna en el campo **Name**.

Para generar los datos de cada columna debemos proceder como antes, pero ahora seleccionaremos la opción **Generate Data** en el menú emergente asociado a cada columna. Entonces, debemos completar el campo **Expression** con las expresiones: **RUNIFORM(100,1,5)**, para la columna **X**; **RNORMAL(100,0,0.8)**, para la columna **Epsilon**; y, $-2 + 3 * X + Epsilon$, para la columna **Y**.

Diagrama de dispersión de los datos

Si representamos la nube de puntos definida por las observaciones (x_i, y_i) podemos ver que resulta claramente una nube alargada alrededor de la recta $-2 + 3 * x$.

Para obtener el diagrama de dispersión debemos hacer click en el botón **Scatterplot** de la barra de herramientas de Statgraphics, o seleccionar la opción **X-Y-Plot** del menú **Plot/Scatterplots**. Así accedemos al cuadro de diálogo **X-Y-Plot**; entonces, completaremos los campos **Y** y **X** con las correspondientes variables, y haremos click en el botón **OK** para obtener el diagrama.

Estimación del modelo

A continuación vamos a ajustar la recta de regresión de Y sobre X. Para ello debemos ejecutar la opción **Relate/Simple Regression**, y entonces, establecer la variable dependiente como **Y**, y la independiente como **X**. Al hacer click en el botón **OK** se obtiene como salida dos tablas: la primera se refiere a los parámetros del modelo; y, la segunda, a la descomposición de la variabilidad del modelo (Analysis of Variance).

Las estimaciones de la ordenada en el origen (**Intercept**) y de la pendiente (**Slope**) de la recta de regresión aparecen en la columna **Estimate** de la primera tabla. El alumno puede comprobar como los parámetros estimados se aproximan bastante a los verdaderos valores de los parámetros, -2 y 3.

La varianza residual aparece en la segunda tabla como la intersección entre la columna **Mean Square** y la fila **Residual**. El alumno puede observar como este valor resulta próximo a la verdadera varianza de los errores dada por $0,8^2 = 0,64$.