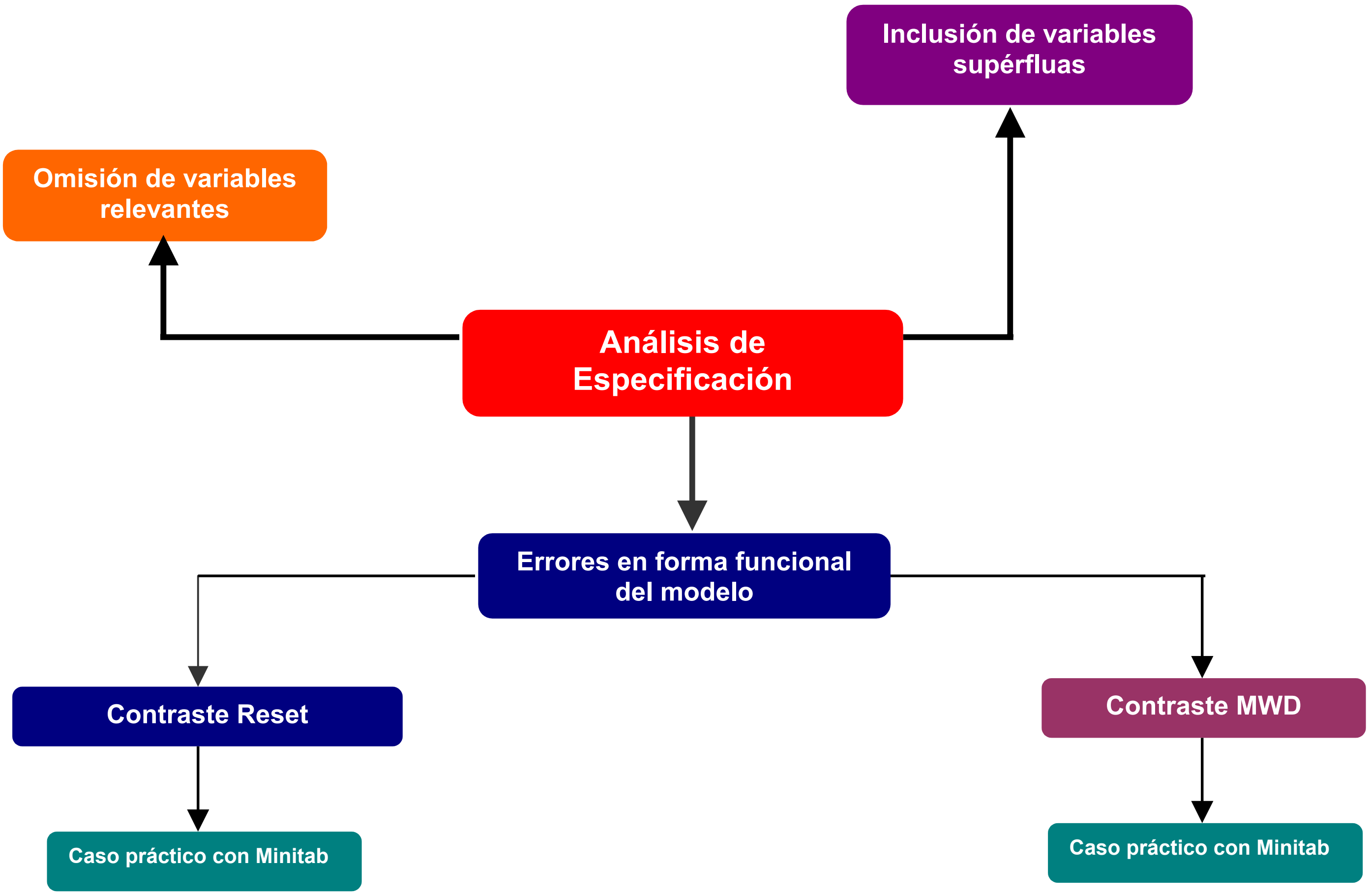
ANÁLISIS DE ESPECIFICACIÓN

Autores: Renatas Kizys (rkizys@uoc.edu), Anna López (alopezrat@uoc.edu)

ESQUEMA DE CONTENIDOS



INTRODUCCIÓN

En el *math-block* MRLM trata la especificación del MRLM y postula las hipótesis que subyacen esta especificación. En concreto, si se cumplen todas las hipótesis básicas, los estimadores MCO de los componentes del vector son insesgados, eficientes y consistentes, y el estimador de la varianza del término de error es insesgado. Pero ¿qué ocurre con estas propiedades ante los errores de especificación? Con finalidad de contestar a esta pregunta, en el presente math-block analizamos distintas situaciones que pueden surgir como consecuencia de uno u otro tipo de errores de especificación. En el presente math-block planteamos la mencionada pregunta contestar a la pregunta, situándonos ante las situaciones que surgen al cometer un error de especificaciónm considerando las distintas situaciones, cuando se incluyen las variables superfluas, o bien, se excluyen las variables relevantes.

En este *math-block* analizaremos las propiedades teóricas del MRLM cuando la parte determinista del modelo (i.e. el conjunto de las variables explicativas) no cumple las hipótesis básicas. En concreto consideraremos cuáles son las consecuencias de la inclusión de las variables superfluas, de la exclusión de las variables relevantes, así como de la incorrecta especificación de la forma funcional del modelo. Por último, analizaremos un método de selección entre dos modelos no añadidos.

OBJETIVOS

- Conocer las propiedades de las estimaciones de MCO ante una especificación errónea del modelo.
- Conocer las diferencias y las analogías entre las consecuencias de un problema de omisión de variables relevantes y uno de inclusión de variables irrelevantes.
- Saber detectar un problema de una mala especificación, tanto por lo que respecta al conjunto de regresores como la forma funcional del modelo.
- Familiarizarse con el contraste de selección entre los modelo no añadidos.

CONOCIMIENTOS PREVIOS

Aparte de estar iniciado en el uso del paquete estadístico Minitab, resulta muy conveniente haber leído con profundidad los siguientes *math-blocks* relacionados con Estadística e Introducción a la Econometría:

- Intervalos de confianza y contraste de hipótesis para 1 y 2 poblaciones
- Análisis de regresión y correlación lineal
- Modelo de Regresión Lineal Múltiple

CONCEPTOS FUNDAMENTALES

La lista de variables que el directivo de una empresa o el investigador de un centro de investigaciones decide incluir como explicativas puede ser errónea por distintas razones:

- Omisión de variables relevantes: en algún caso puede omitirse una variable explicativa relevante por carecer de datos fiables acerca de la misma o por ignorar su relevancia.
- Inclusión de variables irrelevantes: ocurre cuando se incluyen, equivocadamente, como explicativas variables que no lo son.
- Forma funcional incorrecta: se da cuando la relación especificada entre la variables endógena y las variables explicativas no es la adecuada.

Omisión de variables relevantes

Supongamos que el verdadero modelo que describe el comportamiento de la variable **Y** es, en forma matricial:

$$Y = X \cdot B + U; \text{ con } E(U) = 0_{nx1},\ Var(U) = \sigma^2 I_{nxn}.$$

El investigador, sin embargo, se propone a estimar el modelo:

$$Y = X_0 \cdot B_0 + U_0,$$

con X_0 una matriz con dimensión n x r, siendo r < k. Estamos, pues, ante la omisión de las variables relevantes.

La estimación MCO de los coeficientes de este último modelo será:

$$B_0 = (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot Y = (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot (X \cdot B + U) = (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot X \cdot B + (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot U,$$

por lo que se tiene:

$$E(B_0) = (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot X \cdot B,$$

que es, en general, distinto del vector B, y que el error cometido al escribir la lista de variables explicativas puede conducir a sesgos en el estimador MCO [8].

Por otra parte, si denotamos por e los residuos del modelo estimado, se tiene:

$$e' \cdot e = Y' \cdot M_0 \cdot Y = (X \cdot B + U)' \cdot M_0 \cdot (X \cdot B + U) = U' \cdot M_0 \cdot U + B' \cdot X' \cdot M_0 \cdot X \cdot B,$$

donde $M_0 = I_N - X_0 \cdot (X_0' \cdot X_0)^{-1} \cdot X_0'$.

Tomando la esperanza matemática de ambos miembros de esta expresión e explorando las propiedades de la traza, queda:

$$E[e' \cdot e] = E[U' \cdot M_0 \cdot U] + B' \cdot X' \cdot M_0 \cdot X \cdot B;\ E[U' \cdot M_0 \cdot U] = E[tr(U' \cdot M_0 \cdot U)] = E[tr(M_0 U \cdot U')] = tr[E(M_0 U \cdot U')];$$
$$E[e' \cdot e] = tr[M_0 E(U \cdot U')] = tr[M_0 \sigma^2 \cdot I_N] = \sigma^2 tr M_0.$$

Por tanto,

$$E[e' \cdot e] = \sigma^2 tr M_0 + B' \cdot X' \cdot M_0 \cdot X \cdot B.$$

Es fácil comprobar que si no hubiese cometido ningún error de especificación, la matriz M_0 sería simplemente M, $M = I_n - X \cdot (X' \cdot X)^{-1} \cdot X'$, y entonces $E[e' \cdot e] = \sigma^2 \cdot trM = (n - k) \cdot \sigma^2$. En todos los demás casos, la forma cuadrática $B' \cdot X' \cdot M_0 \cdot X \cdot B$ es positiva, por lo que $E[e' \cdot e]$ excederá a $\sigma^2 \cdot trM_0$, siendo $e' \cdot e/(n - r)$ un estimador de σ^2 con sesgo positivo [8].

Ahora bien, supongamos que la matriz X está formada por $X = (X_0; Z)$ donde X_0 es una matriz n x r y Z es una matriz n x (k – r), es decir, que el modelo que se estima excluye alguna de las variables explicativas del modelo verdadero. En tal caso:

$$(X_0' \cdot X_0)^{-1} \cdot X_0' \cdot X = (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot [X_0 ; Z] = (X_0' \cdot X_0)^{-1} [X_0' \cdot X_0; X_0' \cdot Z] = [I_r; (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot Z].$$

En consecuencia, sustituyendo esta última expresión en la esperanza del estimador, se tiene

$$E(B_0) = [I_r; (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot Z] \cdot B = B_r + (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot Z \cdot B_{k-r},$$

siendo vector de dimensión r x 1, donde B_r y B_{k-r} denotan, respectivamente, los verdaderos vectores de coeficientes asociados a las variables en X_0 y Z. De este modo las estimaciones MCO de los coeficientes de las variables X_0 resultan sesgadas.

En conclusión, debido a la omisión de las variables explicativas las estimaciones MCO de los coeficientes de las variables en X_0 resultan sesgadas e inconsistentes [1], [8].

Inclusión de variables superfluas

Supongamos ahora que se comete error de incluir como explicativas variables que no pertenecen al modelo.

Sea $X_0 = (X; Z)$ donde X es la matriz habitual n x k de observaciones de variables explicativas, y Z es una matriz n x s. En este caso, en lugar de especificar el modelo correcto

$$Y = X \cdot B_1 + U,$$

se contempla la siguiente especificación:

$$Y = X \cdot B_1 + Z \cdot B_2 + U.$$

Parecería que podríamos tener la misma clase de problemas que los considerados bajo la omisión de las variables relevantes. Sin embargo, en realidad no es así. Podemos interpretar la omisión de un conjunto de variables relevantes como un resultado de imponer una restricción incorrecta. Alternativamente, la omisión de variables relevantes equivale a la incorporación de información incorrecta a la estimación del modelo. En concreto, la omisión de Z equivale a la estimación *incorrecta* de la última ecuación de regresión sujeta a la restricción

$$B_2 = 0.$$

El previo análisis nos muestra que la imposición de una restricción incorrecta produce sesgos. No obstante, supongamos que nuestro error es, simplemente la omisión de información que es correcta. La inclusión de variable irrelevantes Z en la regresión equivale a no imponer la restricción anterior en el último modelo con las k+s variables explicativas [3]. Pero la especificación anterior no es incorrecta; simplemente no incorpora a la restricción. Razonando así, podemos afirmar que la estimación de los componentes del vector B_1 es insesgada incluso bajo la restricción de exclusión de Z. Para verlo más formalmente, estimamos el modelo erróneo, obteniéndose [8]:

$$B_0 = (B_1; B_2)' = (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot Y = (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot (X \cdot B_1 + U) = (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot X \cdot B_1 + (X_0' \cdot X_0)^{-1} \cdot X_0' \cdot U.$$

Por otra parte, se tiene que $(X_0'\cdot X_0)^{-1}\cdot X_0'\cdot X_0 = I_{k+s} = (X_0'\cdot X_0)^{-1}\cdot[X_0'\cdot X; X_0'\cdot Z]$.

Entonces, $(X_0'\cdot X_0)^{-1}\cdot X_0'\cdot X = (I_{kxk}\ 0_{sxk})'$. En consecuencia, tomando la esperanza del estimador, queda:

$$E(B_0) = (I_{kxk}\ 0_{sxk})'\cdot B_1 = (B_{kx1}\ 0_{sx1})' \Rightarrow E(B_2) = 0_{sx1},$$

lo cual demuestra la afirmación anterior. Es más, si la variable explicativa no contribuye en explicar la variabilidad de la variable endógena, su efecto correspondiente será cero. Por consiguiente, debería esperarse que, al estimar el modelo, los coeficientes de variables respectivas no resultasen significativas en términos absolutos, ni tampoco en términos estadísticos.

Por otra parte, una vez estimado el modelo $Y = X\cdot B_1 + Z\cdot B_2 + U$, el investigador procedería a estimar σ^2, dividiendo la suma cuadrática de los errores por el número de grados de libertad del modelo estimado, es decir:

$$s^2 = e'\cdot e/(n - k - s).$$

Ahora bien, las propiedades de la matriz M_0 implican $M_0\cdot X = M_0\cdot Z = 0_N$, por lo que, tomando la esperanza, queda:

$$E(e'\cdot e) = \sigma^2\cdot trM_0 = (n - k - s)\cdot\sigma^2.$$

Así pues, la varianza del término de error también es insesgado [8]:

$$E[e'\cdot e/(n - k - s)] = \sigma^2.$$

En conclusión, el estimador de MCO de los coeficientes asociados las variables X es insesgado, mientras que el estimador de los parámetros que acompañan a las variables Z (incorrectamente incluidas) tiene esperanza cero. Entonces, ¿no sería deseable, generalmente, "sobreajustar" el modelo? La inclusión de variables irrelevantes no sesga el estimador MCO, pero sí que hace que aumente la varianza con que se estiman los coeficientes de las variables explicativas verdaderamente relevantes, sobre los que perdemos, por consiguiente, la precisión. Por último, aumentar la varianza de los coeficientes podríamos creer equivocadamente que dichas variables no son importantes en explicar la evolución de la variable endógena cuando lo que sucede es que la incorporación de variables irrelevantes nos lleva a perder la precisión en la estimación de los coeficientes de todas ellas las variables.

Errores en forma funcional

Diremos que se comete un error en la forma funcional cuando se especifica una relación (que puede ser lineal, cuadrática, cúbica, exponencial, logarítmica, etc.) y la verdadera relación es diferente de la especificada.

Una especificación incorrecta en la forma funcional del modelo puede considerarse, en algunos casos, como la omisión de variables relevantes. Entonces, las consecuencias son las mismas que las que provoca la omisión de variables relevantes, es decir, los estimadores serán sesgados e inconsistentes (ver ejemplo 1 de la parte práctica con software).

En general, un error en la forma funcional nos puede llevar a obtener término de perturbación no esférico (i.e., con heteroscedasticidad y/o autocorrelación), así como al hecho de que la distribución se aleje de la distribución del término de perturbación del modelo correctamente especificado. En consecuencia, es importante disponer de algún método para detectar un posible error en la especificación de la forma funcional. Uno de los contrastes más utilizados es el contraste **Reset**, propuesto por *Anscombe y Ramsey* en los años sesenta [1].

Contraste Reset:

Para realizar el contraste, partimos del modelo de regresión especificado en forma lineal:

$$Y_i = \beta_1 + \beta_2 \cdot X_{2i} + ... + \beta_k \cdot X_{ki} + u_i, \; \forall \; i = 1, \ldots, n.$$

A fin de comprobar si la forma funcional lineal es la adecuada seguiremos las etapas siguientes:

1. Estimamos el modelo en la forma funcional lineal y obtenemos la variable endógena ajustada la cual elevamos al cuadrado.

2. A continuación se especifica la regresión auxiliar siguiente:

$$Y_i = \delta_1 + \delta_2 \cdot X_{2i} + ... + \delta_k \cdot X_{ki} + \gamma \cdot \hat{Y}_i^2 + u_i, \; \forall \; i = 1, \ldots, n;$$

   donde se añade un regresor adicional al modelo original, el valor al cuadrado correspondiente a la variable endógena estimada en la etapa anterior.

3. Finalmente, estimamos el modelo de regresión auxiliar y contrastamos si el coeficiente asociado a la variable adicional, $\hat{Y}_i^2$, es significativamente diferente de cero, en cuyo caso rechazamos la linealidad de la relación.

Contraste de MWD

Por otro lado, el análisis de especificación cuenta con la notable contribución por parte de por *MacKinnon, White y Davidson* [7]. En concreto, se trata de un contraste de selección entre los modelos no añadidos (non-nested models), según el cual bajo la hipótesis nula se especifica el modelo lineal, mientras que bajo alternativa, el no lineal. El dilema de dicha selección ha sido la eterna pregunta en el análisis empírico. La elaboración de la prueba de *MacKinnon, White y Davidson* (lo cual denominaremos, por brevedad, la prueba de **MWD**), permitió comparar dos modelos cualquiera y proponer aquel que mejor se ajustaba al criterio. Consideremos la siguiente situación. Deseamos contrastar la hipótesis nula de que la relación entre la variable endógena Y, y las variables explicativas puede definirse por la siguiente especificación (posiblemente lineal)

$$H_0: Y_t = f_t(X_t, A) + u_{1t},$$

Y la hipótesis alternativa que postula una relación mediante una especificación distinta

$$H_A: Y_t = g_t(Z_t, B) + u_{2t},$$

Siendo X_t y Z_t las t-ésimas observaciones de los vectores de variables explicativas. Para realizar el contraste, se construye, de forma artificial el siguiente modelo:

$$H_C: Y_t = (1 - \alpha) \cdot f_t(X_t, A) + \alpha \cdot g_t(Z_t, B) + u_t, \text{ o bien } Y_t = f_t(X_t, A) + \alpha \cdot [g_t(Z_t, B) - f_t(X_t, A)] + u_t.$$

Pero el modelo en sí no es válido, ya que, en general, los parámetros A, B y α no son identificables. Por tanto, los autores propusieron estimar los parámetros del modelo compuesto en dos etapas.

En la primera etapa, se trata de estimar los dos modelos, según la hipótesis nula y la alternativa, por separado.

Posteriormente, en la segunda etapa, se sustituyen los parámetros de A y B, obtenidos en la primera etapa, en el modelo compuesto a fin de realizar la estimación del parámetro α. A tales efectos, generamos los residuos $\hat{u}_{1t}$, así como la diferencia $w_t = g_t(Z_t,B) - f_t(X_t,A)$. A continuación, estimamos el modelo:

$$\hat{u}_{1t} = \alpha \cdot w_t + \varepsilon_t.$$

Si la estimación de α resulta estadísticamente significativa, rechazamos la hipótesis nula de la validez del modelo lineal. En caso contrario, la aceptamos. El contraste se basa en la premisa de que si el modelo lineal es válido, entonces no puede predecirse a partir de los términos procedentes del modelo no lineal.

CASOS PRÁCTICOS CON SOFTWARE

Errores en forma funcional

Ejemplo 1. Especificación incorrecta de la forma funcional del modelo. Se sabe que en un determinado sector de la economía la función de costes totales depende, de forma cúbica, del volumen de output que se produce en ese sector:

$$Y_i = \beta_1 + \beta_2 \cdot X_i + \beta_3 \cdot X_i^2 + \beta_4 \cdot X_i^3 + u_i, \ \forall\, i = 1, \ldots, n,$$

donde Y_i denote coste total y X_i es output de la empresa i.

Sin embargo, un grupo de investigación decide la relación existente entre el coste y el output erróneamente especificar en forma lineal:

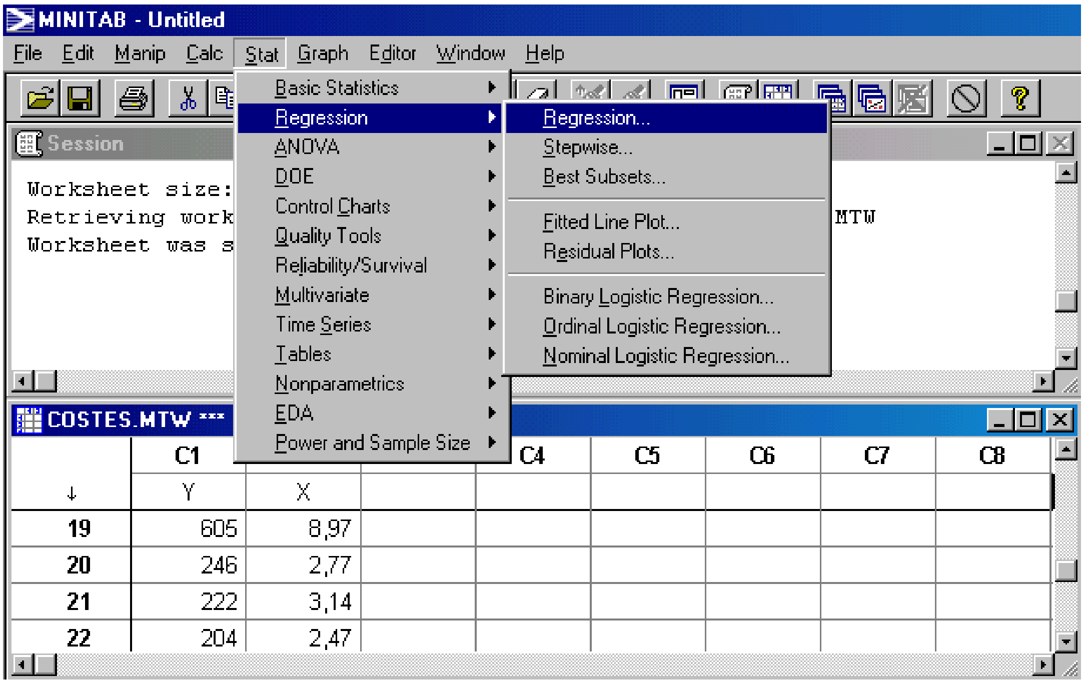
$$Y_i = \beta_1 + \beta_2 \cdot X_i + u_i \ \forall\, i = 1, ..., N.$$

A fin de detectar el error cometido por el grupo de investigación a la hora de especificar la forma funcional del modelo, recurrimos al contraste Reset. Se dispone de la información sobre 28 empresas. Los datos se muestran en la siguiente tabla:

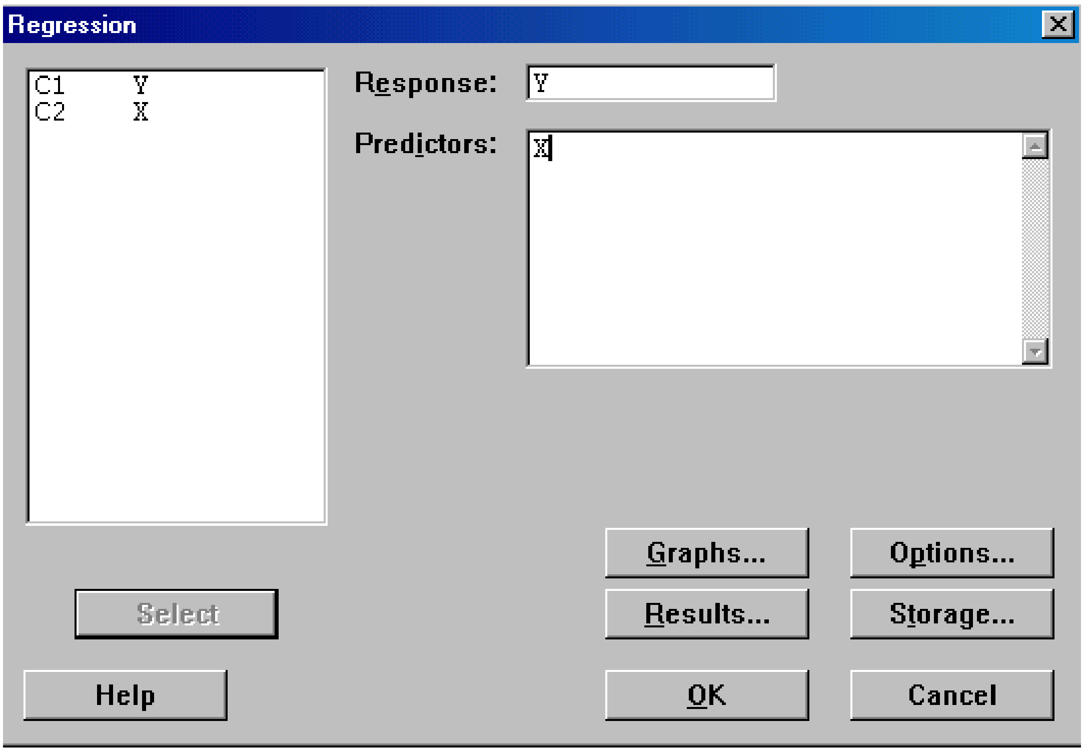
Observación	Coste	Output
1	493	8,20
2	410	7,39
3	451	7,68
4	723	9,88
5	329	5,65
6	432	7,10
7	294	5,17
8	270	3,34
9	311	5,63
10	194	1,39
11	640	9,30
12	217	2,21
13	272	2,88
14	401	6,94
15	196	3,17
16	238	2,36
17	269	2,33
18	256	2,76
19	605	8,97
20	246	2,77
21	222	3,14
22	204	2,47
23	356	6,77
24	378	7,00

25	177	1,69
26	263	4,41
27	549	8,60
28	267	4,71

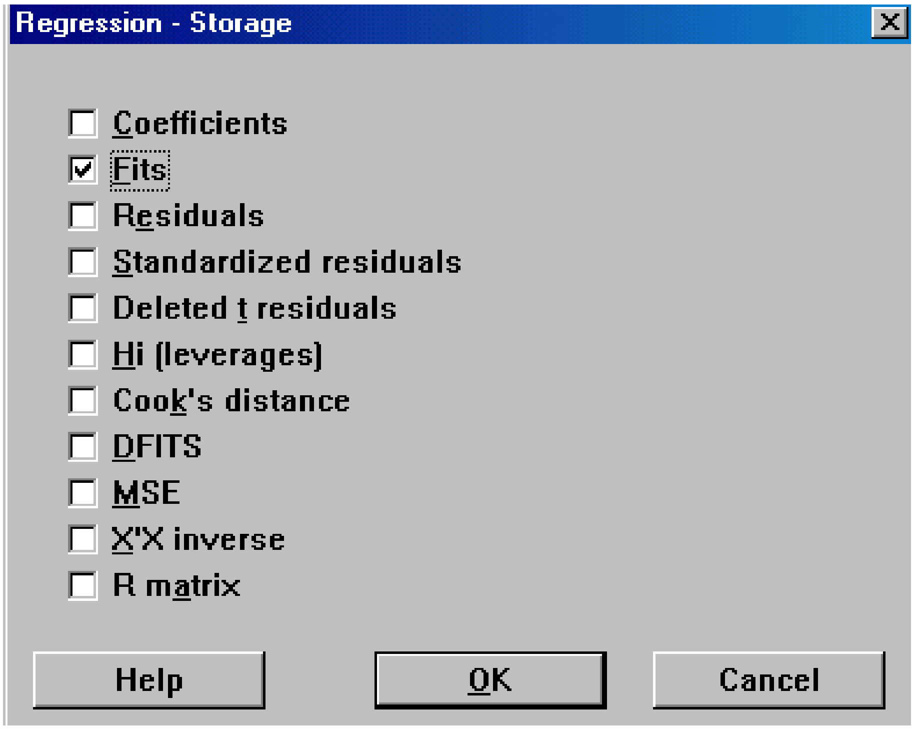
A continuación pasaremos a realizar las siguientes operaciones en el entorno de Minitab. Para ello, seleccionamos ***Stat > Regression > Regression:***



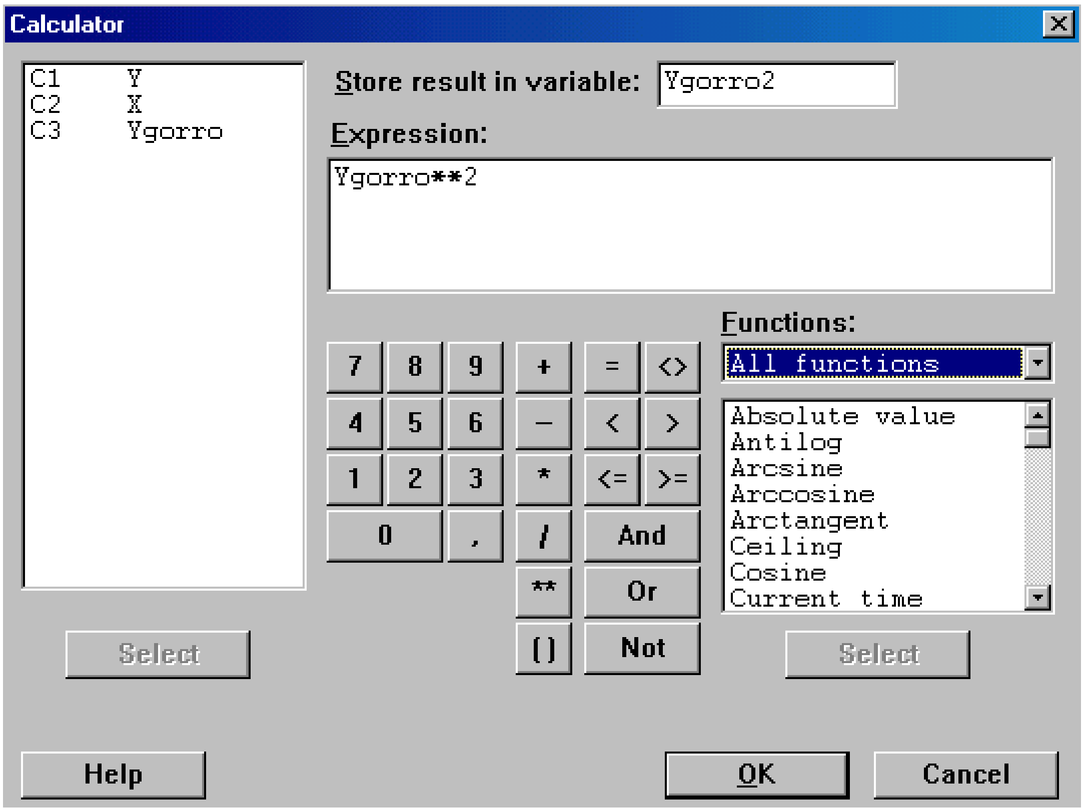
A continuación completamos los campos según se indica en la siguiente imagen:



Para guardar la variable endógena ajustada en la ventana anterior seleccionamos la opción **> *Storage:*** y marcamos **> *Fits*** tal y como se indica:



Esta opción nos permite guardar el vector de valores ajustados de la variable endógena en la columna C3 de la hoja de trabajo ("Worksheet") con el nombre "FITS1", lo cual por el motivo de comodidad, cambiaremos al "*Ygorrro*". A continuación, generaremos dos nuevas variables elevando *Ygorro al cuadrado y al cubo*. Para ello seleccionamos ***Calc* > *Calculator*** y completamos los campos de manera siguiente:



Así, hemos generado la variable endógena al cuadrado y la hemos denotado por "*Ygorro2*". Para generar el coste total ajustado al cubo, repetimos el mismo procedimiento. Una vez generadas nuevas variables, pasaremos a estimar la siguiente regresión auxiliaria por MCO:

$$Y_i = \delta_1 + \delta_2 X_i + \delta_3 \hat{Y}_i^2 + \delta_4 \hat{Y}_i^3 + v_i, \forall\, i = 1, \ldots, n,$$

Los resultados de la estimación se presentan en el cuadro siguiente:

Regression Analysis

```
* Ygorro is highly correlated with other X variables
* Ygorro has been removed from the equation


The regression equation is
Y = 167 + 98,6 X - 0,00597 Ygorro2 +0,000008 Ygorro3

Predictor         Coef         StDev          T         P
Constant        166,98         31,19       5,35     0,000
X                98,64         49,26       2,00     0,057
Ygorro2      -0,005974      0,002684      -2,23     0,036
Ygorro3     0,00000813    0,00000239       3,40     0,002

S = 21,93       R-Sq = 98,0%     R-Sq(adj) = 97,7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      564209      188070    391,22    0,000
Residual Error    24       11537         481
Total             27      575747
```

El contraste Reset consiste en estudiar si en la regresión auxiliaria los parámetros δ_3 y δ_4 son estadísticamente distintos de cero. Atendiendo los resultados obtenidos destacamos la significación estadística de las variables "Ygorro2" y de "Ygorro3"; consecuentemente, podemos rechazar la hipótesis nula de la linealidad del modelo. Por tanto, el investigador ha cometido un error en especificar una forma lineal cuando la verdadera relación es cúbica.

Ejemplo 1. Selección entre el modelo lineal y el modelo log-lineal. Se desea a estudiar la demanda de rosas en una determinada región de España. Para ello, se utilizan las siguientes variables: $\mathbf{Y}_t$ la cantidad de rosas vendidas (en docenas), $\mathbf{X}_{2t}$ el precio promedio al por mayor de las rosas (euros, por docena), $\mathbf{X}_{3t}$ el precio promedio al por mayor de los claveles, (en euros, por docena), $\mathbf{X}_{4t}$ la renta media disponible familiar semanal (en euros) y $\mathbf{X}_{5t}$ la variable tendencia que toma valores 1, 2, y así sucesivamente, durante el período 1991:3T – 1995:2T . Se dispone de 16 observaciones trimestrales correspondientes al mismo período 1991:3T – 1995:2T. Los datos aparecen en el la siguiente tabla:

Obs. (X5)	Año	Trimestre	Y	X2	X3	X4
1	1991	3	11484	2,26	3,49	158,11
2	1991	4	9348	2,54	2,85	173,36
3	1992	1	8429	3,07	4,06	165,26
4	1992	2	10079	2,91	3,64	172,92
5	1992	3	9240	2,73	3,21	178,46
6	1992	4	8862	2,77	3,66	198,62
7	1993	1	6216	3,59	3,76	186,28
8	1993	2	8253	3,23	3,49	188,98

9	1993	3	8038	2,60	3,13	180,49
10	1993	4	7476	2,89	3,20	183,33
11	1994	1	5911	3,77	3,65	181,87
12	1994	2	7950	3,64	3,60	185,00
13	1994	3	6134	2,82	2,94	184,00
14	1994	4	5868	2,96	3,12	188,20
15	1995	1	3160	4,24	3,58	175,67
16	1995	2	5872	3,69	3,53	188,00

Se pide considerar las siguientes funciones de demanda:

$$Y_t = \alpha_1 + \alpha_2 \cdot X_{2t} + \alpha_3 \cdot X_{3t} + \alpha_4 \cdot X_{4t} + \alpha_5 \cdot X_{5t} + u_t,\ t = 1, \ldots, T;$$

$$Ln(Y_t) = \beta_1 + \beta_2 \cdot ln(X_{2t}) + \beta_3 \cdot ln(X_{3t}) + \beta_4 \cdot ln(X_{4t}) + \beta_5 \cdot ln(X_{5t}) + w_t,\ t = 1, \ldots, T;$$

El objetivo de este ejercicio es familiarizarse con el contraste de forma funcional de la regresión. En concreto, se trata de la selección entre los modelos de regresión **lineal** y **log-lineal**. Utilizaremos la prueba de **MWD** para escoger entre los dos modelos.

Para ilustrar esta prueba, especificamos la hipótesis nula y la alternativa:

H_0: Modelo lineal: Y es una función lineal de las variables explicativas, X
H_A: Modelo Log-lineal: log(Y) es una función lineal de los logaritmos de los regresores, los logaritmos de las X.

A efectos del presente contraste, prefijamos el nivel de significación α = 0,1. Ahora bien, la prueba **MWD** comprende los siguientes pasos:

Paso I: Estimamos el modelo lineal y obtenemos los valores Y estimados. Llamarlos $\hat{Y}$ (es decir, *Ygorro*).

Paso II: Generamos los residuos $\hat{u}_{1t} = Y_t - \hat{Y}_t$.
Paso III: Estimamos el modelo log-lineal y obtenemos los valores *Log(Y)* estimados; los denominamos *LYgorro*.

Paso IV: Calculamos *ALYgorro* = *Antilog(Lygorro)*.

Paso V: Generamos *w* = *ALYgorro* - *Ygorro*.

Paso VI: Efectuamos la regresión de $\hat{u}_{1t}$ sobre la variable w_t obtenida en la etapa anterior. Rechazamos H_0 si el la estimación de α es estadísticamente significativa mediante la prueba t usual.

A continuación procedemos a realizar las operaciones relativas mediante el software.

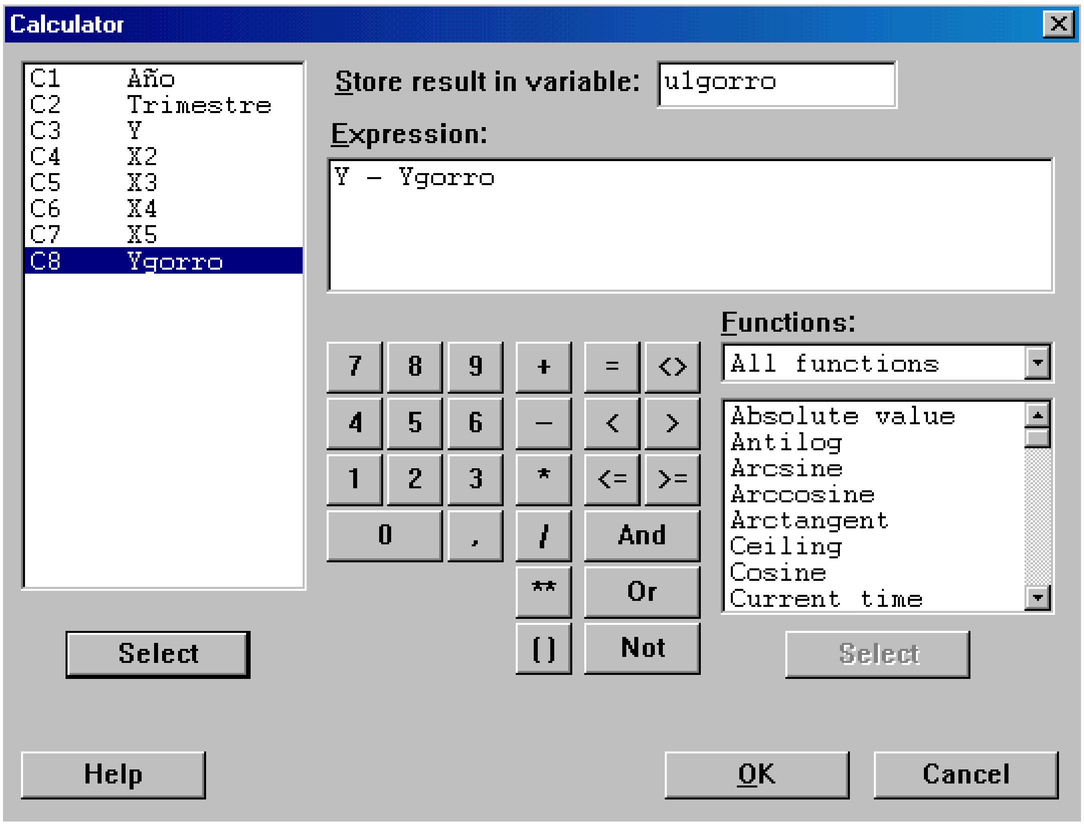
Paso I. La primera etapa consiste, pues, en estimar el modelo lineal; la estimación MCO vía el Minitab nos proporciona los siguientes resultados:

Regression Analysis

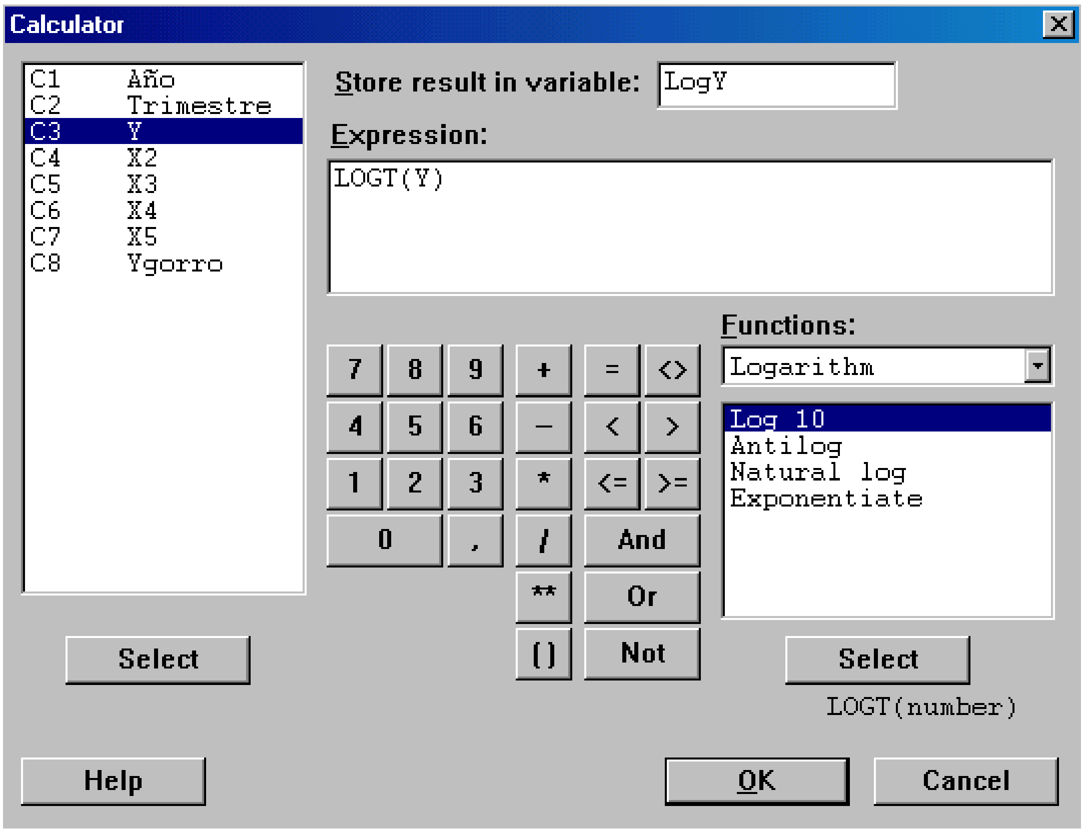
```
The regression equation is
Y = 10816 - 2228 X2 + 1251 X3 + 6,3 X4 - 197 X5

Predictor        Coef       StDev          T        P
Constant        10816        5988       1,81    0,098
X2            -2227,7       920,5      -2,42    0,034
X3               1251        1157       1,08    0,303
X4               6,28       30,62       0,21    0,841
X5             -197,4       101,6      -1,94    0,078

S = 969,9       R-Sq = 83,5%     R-Sq(adj) = 77,5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         4    52249136    13062284     13,89    0,000
Residual Error    11    10347220      940656
Total             15    62596356
```

De forma simultánea, generamos , en la hoja de cálculo de Minitab, los valores ajustados de la variable endógena, denotando por "Ygorro".

Paso II: Generamos, mediante la opción ***> Calc > Calculator*** los residuos $\hat{u}_{1t} = Y_t - \hat{Y}_t$ (u1gorro = Y – Ygorro):



Paso III. Para estimar el modelo log-lineal, realizamos la transformación logarítmica del modelo lineal. A tales efectos, utilizamos la opción Minitab **> *Calc* > *Calculator***:



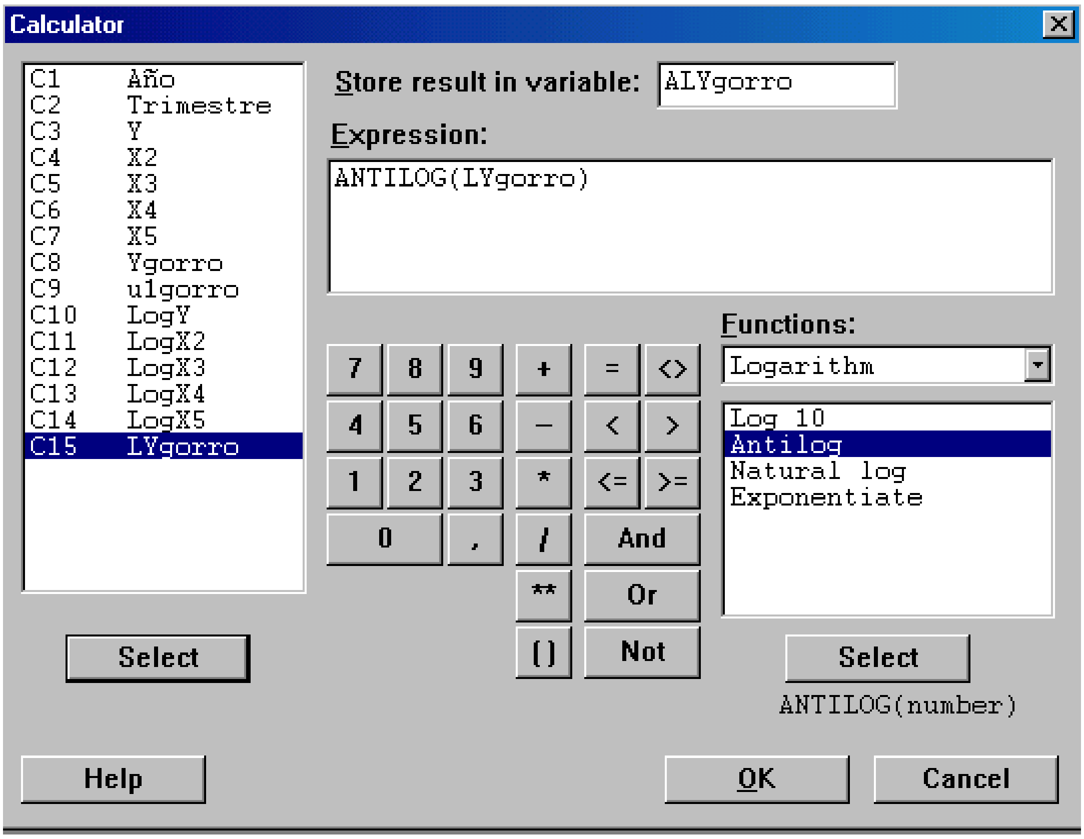
Así, obtenemos la variable endógena en términos logarítmicos. De forma análoga, se transforman el resto de las variables. A continuación, se estima el modelo log-lineal, los resultados mostrándose en el cuadro siguiente:

Regression Analysis

```
The regression equation is
LogY = 0,27 - 1,27 LogX2 + 0,937 LogX3 + 1,71 LogX4 - 0,182 LogX5

Predictor        Coef       StDev          T        P
Constant        0,272       2,670       0,10    0,921
LogX2         -1,2736      0,5266      -2,42    0,034
LogX3          0,9373      0,6592       1,42    0,183
LogX4           1,713       1,201       1,43    0,181
LogX5         -0,1816      0,1279      -1,42    0,183

S = 0,07334     R-Sq = 77,8%     R-Sq(adj) = 69,7%

Analysis of Variance

Source           DF          SS          MS          F        P
Regression        4    0,207272    0,051818       9,63    0,001
Residual Error   11    0,059161    0,005378
Total            15    0,266433
```

Simultáneamente generamos los valores ajustados de la variable endógena, denotando por **LYgorro**.

Paso IV. Generamos la variable "ALYgorro" como antilogaritmo de la variable obtenida en la etapa anterior, "LYgorro", por medio de > ***Calc > Calculator***:



Paso V: Generamos w_t = ALYgorro$_t$ - Ygorro$_t$.

Paso VI. Regresamos los residuos del modelo lineal sobre la variable w. La estimación nos proporciona los siguientes resultados:

Regression Analysis

```
The regression equation is
u1gorro = 0,118 w

Predictor         Coef        StDev          T        P
Noconstant
w               0,1183       0,3090       0,38    0,707

S = 826,5
Analysis of Variance

Source            DF           SS           MS          F        P
Regression         1       100089       100089       0,15    0,707
Residual Error    15     10247131       683142
Total             16     10347220
```

A fin de contrastar la validez del modelo bajo la hipótesis nula, pasamos a realizar el contraste de significación individual del término w. El p-valor asociado al estadístico de contraste es de 0,707 lo cual es muy superior al nivel $\alpha = 0,1$; por lo que el término no es significativo individualmente. Consecuentemente, no podemos rechazar la validez del modelo lineal. En definitiva, la evidencia empírica parece indicar que el **modelo lineal** es el preferible a la hora de explicar el comportamiento de la demanda de rosas.

BIBLIOGRAFÍA

[1] Artís, M.; Suriñach, J.; et al (2001): "Introducción a la Econometría". Ed. Fundació per a la Universitat Oberta de Catalunya. Barcelona.

[2] Doran, H. (1989): "Applied Regression Analysis in Econometrics". Ed. Marcel Dekker, Inc. ISBN: 0-8247-8049-3

[3] Green, W. H. (1999): "Análisis Econométrico". Prentice Hall Iberia. Madrid.
ISBN: 84-8322-007-5

[4] Gujarati, D. (1997): "Econometría básica". McGraw-Hill. ISBN 958-600-585-2

[5] Johnston, J. (2001): "Métodos de econometría". Ed. Vicens Vives. Barcelona.
ISBN 84-316-6116-X

[6] Kennedy, P. (1998): "A Guide to Econometrics". Ed. MIT Press. ISBN: 0262611406

[7] MacKinnon, J.G; White, H., Davidson, R. (1983): "Tests for Model Specification in the Presence of Alternative Hypotheses". Journal of Econometrics, 21, 53-70.

[8] Novales, A. (1993): "Econometría". McGraw-Hill. ISBN 84-481-0128-6

[9] Pulido, A. (2001): "Modelos econométricos". Ed. Pirámide. Madrid. ISBN 84-368-1534-3

[10] Uriel, E. (1990): "Econometría: el modelo lineal". Ed. AC. Madrid. ISBN 84-7288-150-4

[11] Wooldridge, J. (2001): "Introducción a la conometría: un enfoque moderno". Ed. Thomson Learning. ISBN: 970-686-054-1

ENLACES

- http://www.feweb.vu.nl/econometriclinks/index.html
  The Econometrics Journal On-Line
- http://www.elsevier.com/hes/books/02/menu02.htm
  Libro on-line: Handbook of Econometrics Vols. 1-5
- http://elsa.berkeley.edu/users/mcfadden/discrete.html
  Libro on-line: Structural Analysis of Discrete Data and Econometric Applications
- http://www.oswego.edu/~kane/econometrics/stud_resources.htm
  Online Resources for Econometric Students
- http://www.econ.uiuc.edu/~morillo/links.html
  Econometric Sources: a collection of links in econometrics and computing. University of Illinois
- http://www.econometrics.net/
  Econometrics, Statistics, Mathematics, and Forecasting
- http://ideas.uqam.ca/EDIRC/ectrix.html
  Economics Departments, Institutes and Research Centers in the World: Econometrics, Mathematical Economics

Financiado por la Secretaría de Estado de Educación y Universidades (MECD)