Modelo de regresión de dos etapas aplicado a un cargador de roca en una operación minera superficial

A two-stage least square regression model applied to a surface mining rock loader

Oswaldo Ortiz Sánchez*

Recibido: 15/10/2012 - Aprobado: 20/11/2012

RESUMEN

La productividad de un cargador de roca que opera en un frente de minado superficial es influenciada por muchas variables. Se identificó las que se consideran más importantes y para un tipo y modelo de cargador se tomaron varias mediciones en el campo. Utilizando regresión lineal se dedujeron expresiones que permiten pronosticar la variable dependiente en función de las variables independientes. Las variables dependientes que se identificaron fueron: tonelaje de producción, factor de llenado, demoras y características de la excavación. Las variables independientes medidas fueron: altura de banco, tamaño de cargador y fragmentación. Se emplea el sistema de mínimos cuadrados para predecir la performance y el costo dados ciertos valores de las variables restantes en la ecuación de regresión pero existe el problema de seleccionar la variable dependiente. Dos o más variables en el sistema podrían parecer igualmente dependientes, ya que hasta el proceso de prueba de significancia de la variable puede ser dudoso.

Palabras clave: Variable dependiente e independiente, fragmentación, predecir, rendimiento, regresión

ABSTRACT

Productivity of a loader in surface mining is dependent on many variables. In this study, we try to identify those variables considered the most important and for a specific loader model and type, field data was taken and analyzed. Linear models were obtained by applying a two-stage least square regression analysis. Models forecast a dependent variable as a function of independent variables and some other dependent variables. Important dependent variables identified are: tonnage production per unit time, fill factor, delays and digging patterns. Independent variables measured were: bench height, loader (bucket) size and fragmentation. The least square regression analysis is used for predicting performance and cost figures given certain values of the remaining variables in the regression equation.

It seems however, that there exists a problem in selecting the dependent variable. Two or even more variables in the system may seem to be dependent. Under certain conditions the process of testing the significance of a given variable in the equation may even be questionable.

Keywords: Dependente and independente variables, fragmentation, forecast, performance, regression.

* Docente de la EAP de Ingeniería de Minas de la UNMSM. E-mail: osoos1990@gmail.com

I. INTRODUCCIÓN

Al analizar un problema complejo generalmente lo primero que puede hacerse es trazar un diagrama de flujo para establecer posibles correlaciones entre variables del sistema. Inicialmente se podría enumerar las variables que pueden afectar el sistema y luego hacer correspondencias para encontrar algún orden de importancia entre ellas. Un ejemplo es la construcción de un diagrama para encontrar los factores que influencian el consumo de energía de un cargador de mineral o desmonte en un frente de avance de un nivel en una operación superficial.

Luego de enumerar las variables y acomodarlos en un diagrama se puede aplicar análisis de regresión que nos da una medida cuantitativa del grado de interrelación entre variables. Se efectúan pruebas para ver si algunas variables en la ecuación son significativas o si la propia ecuación de regresión es significativa. El análisis de regresión se usa para predecir ciertas variables, como la performance y el costo o valor dadas otras variables en la ecuación de regresión. Es típica la predicción de las cotizaciones futuras de los metales basados en cotizaciones históricas reales. Dada la complejidad de variables que entran, estos pronósticos son solo de tipo referencial. Otro caso típico es la predicción de esfuerzos en pilares de sostenimiento subterráneo cuyas resistencias o esfuerzos dependen de muchas variables, como el nivel de extracción (R), ancho, alto, espaciamiento, profundidad, módulos de Young y de Poisson, relación de compresibilidad entre el pilar y la caja techo, el pilar y la caja piso, que puede expresarse como la relación los módulos de elasticidad del pilar y de las cajas techo y piso, etc. El análisis de correlación puede cuantificar la importancia de cada variable habiendo encontrado que la influencia del R en el esfuerzo de un pilar es de 80% y las demás variables contribuyen con el 20% cumpliendo con la regla de Pareto del 80/20.

El objetivo de este trabajo de investigación es encontrar expresiones que relacionen las variables dependientes con las independientes en la operación de un cargador de roca en un frente de avance de un nivel de una operación minera superficial, con el objeto de conocer la contribución cuantitativa en orden de importancia de cada variable dependiente e independiente, que nos conduzca al cálculo de la productividad real del equipo.

II. REGRESIÓN LINEAL SIMPLE

Todos conocemos el modelo simple Y = A + B.X, donde A y B son constantes que representan la intersección y la pendiente de la función. Los valores exactos de A y B son desconocidos pero pueden estimarse observando las variables X, Y. Para esto, seleccionamos ciertos valores de “X” y se observa el valor correspondiente de “Y”. Los valores pronosticados de Y se designan por Y’ y su desviación del valor verdadero “Y” lo llamamos “U”. Estos efectos pueden deberse a otros factores que no fueron considerados al formular la interrelación entre variables. Debido a que puede existir la influencia de muchos factores y sus efectos pueden balacearse o desaparecer, se espera que U tenga valores pequeños más frecuentemente que valores grandes. De esta forma U es una variable con distribución probabilística con media cero y varianza finita.

Los valores de A’ y B’ que son aproximaciones a los valores verdaderos A y B, se estiman por estadística (Scheaffer y otros, 2007):

$$B' = \sum_{l=1}^{n} xi.\, yi / \Sigma xi \qquad \text{y} \qquad A' = y' - B'.\, x'$$

Si el modelo de mínimos cuadrados es válido, las discrepancias “U” son correctas si son independientes, caso contrario debería usarse algún otro modelo.

III. RELACIÓN FUNCIONAL

Para el caso de predecir la productividad de un cargador de roca en un frente de avance, el análisis preliminar conduce a trazar primero un diagrama de flujo como el mostrado en la Figura N.° 1.

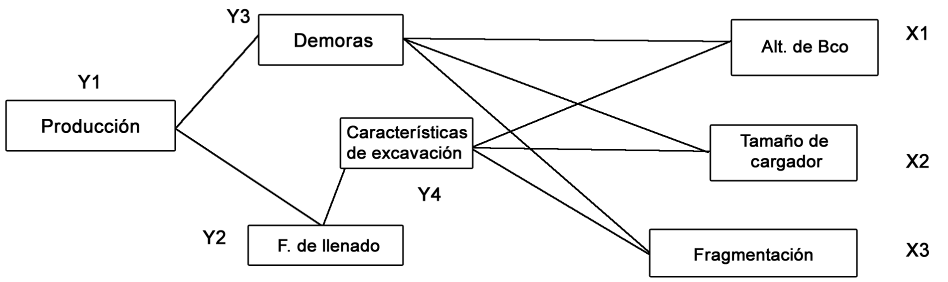

Figura N.° 1. Diagrama preliminar en la investigación de la productividad de un cargador en la excavación de un banco cielo abierto.

No se ha definido la variable dependiente. En la Fig. N.° 1 solo se ha establecido una relación entre variables mediante líneas. En el lado derecho se han colocado aquellas variables que intuitivamente podemos considerarlos como importantes. En el centro están otras variables que se diferencian de las anteriores y tienen otras características.

Se puede seleccionar las relaciones funcionales entre variables interpretando la Fig. N.° 1, lo que puede corroborarse mediante las pruebas de correlación entre las variables elegidas para conocer los grados de dependencia. Estas relaciones se dan en la Tabla N.° 1.

Tabla N.° 1. Relaciones funcionales interpretados de la Figura N.° 1.

$$Y1 = f_1(Y2,\ Y3,\ Y4,\ X1,\ X2,\ X3)$$
$$Y2 = f_2(Y4,\ X1,\ X3)$$
$$Y3 = f_3(Y4,\ X3)$$
$$Y4 = f_4(X1,\ X2,\ X3)$$

Donde:
Y1 = Productividad en la unidad de tiempo t o n/h o m^3/h
Y2 = Factor de llenado del balde del cargador en por ciento
Y3 = Demoras expresadas en minutos o segundos
Y4 = Características de la excavación expresadas numéricamente
X1 = Altura de banco metros o pies
X2 = Tamaño de cargador (balde) m^3 o yd^3.
X3 = Fragmentación expresada en por ciento.

Podemos deducir que las variables "Y" son dependientes y las "X" son independientes. Las variables "X" no son funciones de ninguna otra variable, es decir son externas o exógenas. Las variables "Y" son todas dependientes en alguna fase, es decir son endógenas.

Cada una de las funciones dadas pueden expresarse como una ecuación lineal y pueden deducirse de las relaciones anteriores. Estas ecuaciones se muestran en la Tabla N.° 2.

Tabla N.° 2. Ecuaciones lineales de interrelación funcional entre variables.

	Conste.		Prodción		F. de llenado		Demoras		Caract. de exc.		Alt. Bnco		Tamaño carg.		Fragmtcn.		Discrep.
1.	A1	+	B11.Y1	+	B12.Y2	+	B13.Y3	+	B14.Y4	+	C11.x1	+	C12.x2	+	C13.x3	=	U1
2.	A2	+			B22.Y2	+			B24.Y4	+	C11.x1	+			C23.x3	=	U2
3.	A3	+					B33.Y3	+	B43.Y4	+					C33.x3	=	U3
4.	A4	+							B44.Y4	+	C41.x1	+	C42.x2	+	C43.x3	=	U4
5.	A5	+									C51.x1					=	0
6.	A6	+											C52.x2			=	0
7.	A7	+													C53.x3	=	0

Las tres últimas ecuaciones se han incluido con el objeto de enfatizar que las variables X son predeterminadas y no dependen de ninguna otra variable del sistema. Adicionalmente, estas tres ecuaciones completan el sistema donde se tienen siete ecuaciones con siete variables.

El ordenamiento de las variables se obtiene del siguiente razonamiento: Si X1, X2 y X3 se conocen y existen físicamente, entonces podemos sustituir sus valores en la cuarta ecuación para obtener una solución para Y4, ya que la cuarta ecuación tendría una variable desconocida. Podemos decir que Y4 depende directamente de X1, X2 y X3.

Si Y4 es conocida (se calculó en el paso anterior), puede obtenerse una solución para Y3 en la tercera ecuación. Este proceso puede continuar hasta que se evalúen todas las ecuaciones.

Con el objeto de evaluar las variables X se puede ver que se requiere sustituciones cero, de los cuales uno es para Y4 y otras dos son para las otras variables (X1, X2 y X3 para Y4, y Y4 y X3 para Y3). De esta forma se tiene un orden definido de sustituciones para la solución del sistema y entonces todo el sistema de variables puede reemplazarse por un sistema de ecuaciones simultáneas como se muestra en la Figura N.° 2.

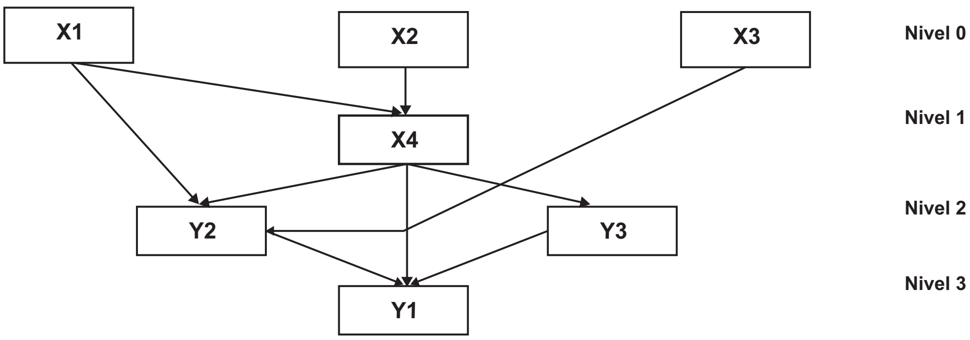

Figura N.° 2. Niveles de dependencia entre variables en la productividad de un cargador de roca.

El caso real en una operación superficial usando este método es más complicado y se muestra en la Figura N.° 3. En este diagrama, el factor de llenado debe interpretarse como una medida de la fragmentación de tal forma que el factor de llenado es una variable influyente en el tamaño del cargador.

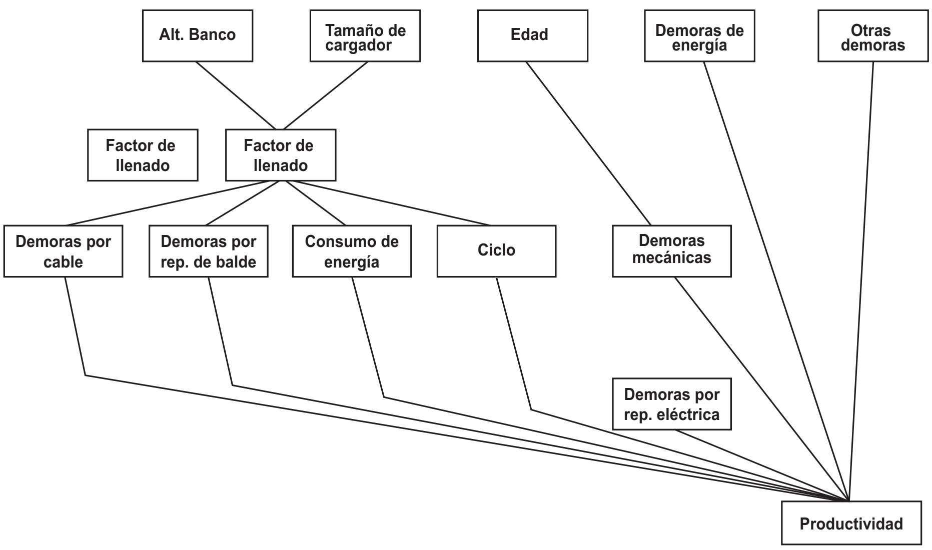

Figura N.° 3. Diagrama de variables de producción de un cargador en una operación minera superficial.

IV. SOLUCIÓN ESTRUCTURAL

Es conveniente estimar los coeficientes de las variables del problema estructural sin interferir los supuestos para el cálculo de mínimos cuadrados. Con la excepción de la variable Y1 (productividad), todas las demás variables dependientes pasan en algún momento por ella independiente de las funciones descritas anteriormente, lo que significa que el modelo considera el sistema de minado en su conjunto y no hace selección de dos o tres variables para analizarlos por separado.

En la solución, se aplica los mínimos cuadrados de dos etapas que proporciona la máxima aproximación al resultado real del problema. Para esto se ejecuta mínimos cuadrados simples en dos etapas. Primero estima los coeficientes de la ecuación de regresión para cada variable dependiente, o sea las Y en términos de las variables pre-determinadas (independientes) o sea las X solamente. De esta manera, para el caso de la cuarta ecuación dada en la Tabla N.° 2, se tiene:

$$Y_4' = P' + 0.1'.X1 + 0.2'.X2 + 0.3'.X3 \quad \text{equivalente a:}$$

$$Y4 = C41.x1 + C42.x2 + C43.x + U4 \quad \text{haciendo } B44 = 1$$

Para una muestra de 10 pruebas, se calculó 10 valores nuevos de Y_4 en términos de las variables pre-determinadas (independientes). Estos valores nuevos pueden luego usarse en lugar de los valores Y_4 en la segunda etapa o sea en la tercera ecuación de la Tabla N.° 2. Haciendo B33 = 1 y cambiando signos apropiados se tiene:

$$Y_3 = A3' + B43'.Y4' + C33.X3 + U3$$

O sea que se calcula Y_3 en términos de la variable Y_4, que a su vez es una combinación lineal de las variables pre-determinadas (X1, X2 y X3) y una de las variables pre-determinadas X3, todas las cuales son independientes de la discrepancia U_3.

En el cálculo real la computación puede efectuarse en un solo paso usando matrices y un programa computacional (Chapra y otros, 2006).

Debe indicarse que un requerimiento para llegar a una solución única de una ecuación en el sistema es que se cumpla la condición (K – K') ≥ G – 1 (Baseman, 1967). Donde: K y K* son números de variables pre-determinadas en la estructura del sistema y en la ecuación, respectivamente. G es número de variables dependientes en la ecuación.

Para el caso de la tercera ecuación de la Tabla N.° 2 se tiene: K = 3, K* = 1, G = 2 con lo que la condición se satisface. De no satisfacerse la condición dada, la estructura desarrollada es incompleta y cualquier solución corresponderá a alguna combinación lineal de la solución real desconocida, por lo que será necesario analizar más a fondo el problema para descubrir otras variables pre-determinadas (independientes).

4.1. Solución mediante mínimos cuadrados de dos etapas

La solución de un sistema lineal de ecuaciones por mínimos cuadrados se simplifica usando matrices. Partiendo de la ecuación general: $Y = \beta_o + \beta_1.X + e$ se tiene la ecuación matricial:

$(X^t.X).\beta' = X^t.Y$ (Carrasco, 2007)

Las incógnitas β_o, β_1 se obtienen despejando β':

$$\beta' = \begin{vmatrix} \beta_o' \\ \beta_1' \end{vmatrix} = (X^t.X)^{-1} \cdot Y$$

Donde X^t es la traspuesta de X.

La solución para este tipo de mínimos cuadrados se obtiene resolviendo el siguiente determinante:

$$\begin{vmatrix} B \\ A \end{vmatrix} = - \begin{vmatrix} (Y^t.X.(X^t.X)^{-1} . X^t.y & Y^t.X^{*}) \\ X^{*t}.Y & X^{*t}.X^{*} \end{vmatrix}^{-1} . \begin{vmatrix} y^t.X.(X^t.X)^{-1} X^t.y \\ X^{*t}.y \end{vmatrix}$$

Dónde: B = Coeficientes de las variables dependientes,
A = Coeficientes de las variables independientes
Y = Variables dependientes observadas de todo el sistema.
X = Matriz de observaciones de todas las variables independientes del sistema
X^* = Matriz de observaciones de las variables independientes de la ecuación
y = Matriz de observaciones de las variables dependientes de la ecuación
Y^t = Matriz traspuesta de Y
X^t = Matriz traspuesta de X
X^{*t} = Matriz traspuesta de X^*

V. APLICACIÓN DEL SISTEMA

La Tabla N.° 3 corresponde a 10 observaciones de campo de las variables enumeradas. La notación es la misma empleada en la descripción, con la excepción de la constante A, que ha sido reemplazada por $x_{o.}$

Tabla N.° 3. Observaciones de variables dependientes e independientes de un cargador a cables

Y2	Y4	X_o	X1	X3	X4
80.16	4	1	16.3	0.7	28
83.56	6	1	17.1	0.9	25
94.69	7	1	16.4	0.2	25
78.60	10	1	17.0	0.6	25
63.62	4	1	30.9	0.1	36
78.29	8	1	13.9	0.5	13
62.00	9	1	17.2	0.8	28
76.38	6	1	14.0	0.7	25
72.29	6	1	29.4	0.5	36
59.29	9	1	12.8	0.5	13

En la Tabla N.° 3 se tiene las siguientes variables equivalentes a las usadas en la matriz de mínimos cuadrados.

$Y = y_2$, $Y = Y_4$, $X^* = (X_o, X1, X3)$, $X = (X_{o,}$ $X1, X2, X3)$

Para el caso de la ecuación 2 de la Tabla N.° 2 se encontrará el modelo para estimar el factor de llenado (Y2).

Los resultados intermedios de las computaciones matriciales son las siguientes matrices:

$$(X^{*t}.Y) = \begin{vmatrix} 69.00 \\ 3991.73 \\ 38.90 \\ 3013.00 \end{vmatrix} \qquad (X^{*t}.Y) = \begin{vmatrix} 69.00 \\ 3991.73 \\ 38.90 \end{vmatrix}$$

$$X^t.X = \begin{vmatrix} 10.00 & 606.25 & 5.50 & 458.00 \\ 606.25 & 40640.35 & 311.36 & 29946.90 \\ 5.50 & 311.36 & 3.59 & 246.00 \\ 458.00 & 29946.90 & 246.00 & 22702.00 \end{vmatrix} \qquad (X^t.X)^{-1} = \begin{vmatrix} 2.8701 & -0.02939 & -2.0857 & 0.003475 \\ -0.0293 & 0.00135 & 0.3633 & -0.001587 \\ -2.0857 & 0.03633 & -2.3105 & -0.036310 \\ -0.0034 & -0.00159 & -0.0363 & 0.002461 \end{vmatrix}$$

Luego, la matriz por invertir que constituye el primer término de la matriz solución es:

$$\begin{vmatrix} 489.08 & 69.00 & 3391.73 & 38.90 \\ 69.00 & 10.00 & 606.25 & 5.50 \\ 3391.73 & 606.25 & 40640.35 & 311.36 \\ 38.90 & 5.50 & 31.36 & 3.50 \end{vmatrix}$$

El término inverso de la matriz es:

$$\begin{vmatrix} -\ (Y^t.X.(X^t.X)^{-1}\ .\ X^t.y & Y^t.X^*) \\ X^{*t}.Y & X^{*t}.X^* \end{vmatrix}^{-1} = \begin{vmatrix} -0.28479 & 2.8947 & -0.014529 & -0.088847 \\ 2.89474 & -32.28818 & 0.174840 & 2.93754 \\ -0.014529 & 0.17484 & -0.0010719 & -0.017454 \\ -0.088847 & 2.93753 & -0.017454 & -2.30247 \end{vmatrix}$$

También, dos de los términos no calculados del lado derecho de la matriz son:

$$X^{*t}.y = \begin{vmatrix} 748.98 \\ 44900.30 \\ 411.78 \end{vmatrix} \qquad X^t.y = \begin{vmatrix} 748.98 \\ 44900.39 \\ 411.78 \\ 34245.92 \end{vmatrix}$$

La segunda matriz del lado derecho de la matriz general es igual a:

$$\begin{vmatrix} 5168.57 \\ 748.98 \\ 44900.39 \\ 411.78 \end{vmatrix}$$

Con lo cual se tiene la matriz solución siguiente:

$$\begin{vmatrix} B. \\ A \end{vmatrix} = \begin{vmatrix} 7.1984 \\ -162.1313 \\ 0.5367 \\ 9.1362 \end{vmatrix}$$

Por lo tanto se tiene que la ecuación 2 de la Tabla N.° 2 se expresa como:

$-162.1313 + y_2 + 7.1984.\ y_4 + 0.5367.\ X1 + 9.1362\ X3 = 0$

o Factor de llenado $= y_2 = 162.1313 - 7.1984\ y_4 + 0.5367\ X1 - 9.1362\ X3$

Lo que nos indica que podemos estimar el factor de llenado del cargador de una operación minera superficial conociendo las características de la excavación (y_4) y las variables independientes como altura de banco y fragmentación.

La discrepancia o error es: ∑((Cuadrado de desviaciones para cada observación) / (Tamaño de muestra – Número de coeficientes en la ecuación 1)) = 277.327.

VI. CONCLUSIONES

Son numerosas las variables que intervienen en la performance o productividad de un cargador en una operación minera superficial.

Puede establecerse un orden de importancia de estas variables mediante los diagramas de interrelación o efectuando un análisis de correlación de datos históricos por medio del coeficiente de correlación entre variables.

Tomando en cuenta el principio de Pareto del 80/20, donde el 80% de la exactitud del cálculo de la productividad del cargador puede estimarse con el 20% de las variables más importantes, se acepta preliminarmente las variables tales como tamaño de balde, altura de banco, y fragmentación como variables independientes de gran peso y variables dependientes que operan temporalmente como independientes, como factor de llenado, demoras y características de la excavación como las más influyentes.

En el proceso de cálculo de la flota de equipo para una operación minera superficial, el cargador es el equipo básico y debe elegirse primero para lo cual se necesita toda la información inicial posible de las variables físicas y operativas.

La máxima productividad que puede asignarse a un cargador puede calcularse rápidamente conociendo el ciclo individual del equipo, el tiempo programado de trabajo del equipo por hora, guardia o día, y la capacidad del balde. Este tonelaje se reducirá notablemente a medida que se introduzca una serie de variables restrictivas endógenas y exógenas.

VII. REFERENCIAS BIBLIOGRÁFICAS

1. Baseman R.L. (1967). A generalized classical method of linear estimation of coefficients in a structural equation. Econométrica, v. 25, pp. 77-84.
2. Carrasco Venegas L. (2007). Métodos numéricos aplicados a la Ingeniería. 3.ª ed. Editor R. Figueroa García. Lima. Pp. 144-145.
3. Chapra S.C. and Canale R.P. (2006). Métodos numéricos para ingenieros. 5.ª ed. México, McGraw-Hill Interamericana, pp. 476-492.
4. Scheaffer L.R. and McClave J. T. (1993). Probabilidad y estadística para ingeniería. México, Editorial Iberoamericana S.A., pp. 364-366.