

## PRUEBAS PARA ESTABLECER LA VARIACION ESTACIONAL EN LOS PRECIOS DE HORTALIZAS EN UN MERCADO DE MAYOREO EN LA CIUDAD DE MEXICO

Por Dean Schreiner<sup>1</sup>, Donald K. Hotchkiss<sup>2</sup>, Ovidio Zambrano y Sergio Flores

Centro de Economía Agrícola y Centro de Estadística y Cálculo, Colegio de Postgraduados, Chapingo, Méx.

### *Sinopsis*

La variación de precios no se debe totalmente a los resultados de las ocurrencias del azar que afectan a la oferta y la demanda sino que es, en parte, el resultado de factores cíclicos y reacciones específicas del mercado. La variación estacional en los precios de hortalizas, refleja tendencias de estacionalidad en factores que afectan su oferta y demanda. El objetivo del presente estudio es construir un modelo, para probar si el precio anual promedio es una buena base para estimar cualquier mes del año. Además, se usa el modelo para determinar las desviaciones mensuales del precio medio anual y, la estimación del error resultante, se utiliza para probar las diferencias significativas de precios entre estaciones. Se incluye un extenso apéndice de carácter estadístico, el cual describe diversos métodos para imponer restricciones a la matriz de insumos del modelo de regresión de mínimos cuadrados, con el fin de obtener estimadores de efectos mensuales (o anuales). Se describen técnicas para hacer pruebas de hipótesis adecuadas, así como para interpretar la respuesta de la regresión. Se presenta una aplicación del modelo y se dan varias pruebas, usando precios de dos hortalizas para un mercado de mayoreo específico en la Ciudad de México. Se ha mostrado la estacionalidad de precios para ambos productos, indicando que la media de los precios anuales no es un buen estimador de precios mensuales específicos.

### *Summary*

Price variation is not due completely to the results of random occurrences affecting supply and demand but is partially the result of cyclical factors and specific market reaction phenomenon. Seasonal variation in vegetable crop prices reflect seasonality trends in factors affecting their supply and demand. It is the objective of this study to construct a model testing whether the average annual price is a good estimator for any one month. Further, the model is used to determine the expected monthly deviation from the overall mean price and the resulting error estimate is used to test for significant price differences between seasons. An extensive statistical appendix is included which describes different methods of imposing restrictions on the input matrix of the least squares regression model in order to obtain estimates of month (year) effects. Techniques for making appropriate tests of hypotheses and interpreting the regression output are described.

An application of the model is presented and various tests are given using prices of two vegetables commodities for a specific wholesale market in Mexico City. Price seasonality has been shown for both commodities indicating the mean annual price is not a good estimator of specific monthly prices.

### *Introducción*

#### *La importancia del conocimiento de las fluctuaciones estacionales del precio*

La variación de precio en un año, de 50% a 200% para productos hortícolas en los mercados de mayoreo en la Ciudad de México, no es un fenómeno extra-

1. *Dirección actual:* Assistant Professor, Department of Agricultural Economics, Oklahoma State University, Stillwater, Oklahoma.
2. Associate Professor, Department of Statistics, Iowa State University, Ames, Iowa.

ordinario. Tal variación tiene suma importancia para los productores y para los planificadores agrícolas. Para tomar las decisiones óptimas, es indispensable considerar los precios de los productos, los cuales se deben conocer con certeza. Por consiguiente, se deben clasificar las fuentes de variación de los precios y, de ser posible, analizar las consecuencias que se han previsto para cada una, con el fin de proporcionar al productor o planificador un solo precio esperado.

La determinación de los precios se debe esencialmente al equilibrio, en el mercado, entre oferta y demanda. De aquí que la variación de los factores que influyen las condiciones de oferta o demanda sean causa de fluctuaciones en un mercado previamente en equilibrio, y de las consiguientes oscilaciones de los precios.

La variación del tiempo meteorológico tiene una gran influencia sobre las condiciones de la oferta y, por lo tanto, es una fuente indirecta de fluctuaciones de los precios.

La tendencia de los precios a mostrar una fluctuación o distribución según las estaciones, es una fuente de variación que debe ser prevista y cuantificada. Algunos factores que afectan a la oferta o la demanda pueden distribuirse conforme a las estaciones, y por tanto causan una fluctuación de temporada en los precios. Este tipo de fluctuaciones de los precios es distinta de las variaciones aleatorias en la oferta o la demanda. También es diferente de las variaciones de precio debidas a fenómenos de reacción en el mercado, tales como el efecto de telaraña (4), donde las decisiones se basan en las experiencias de los productores. La trayectoria de reacciones de este tipo sobre los precios de equilibrio en el mercado, no necesariamente ocurren de manera estacional, especialmente en el caso de una situación de doble cultivo.

#### *Objetivos del presente estudio*

Un objetivo de este estudio es el de analizar, en un marco de decisión, por qué las variaciones estacionales de los precios pueden ser económicamente óptimas. El análisis teórico se presenta en la siguiente sección; sin embargo no se hacen aplicaciones específicas a las condiciones de la agricultura en México.

Pero el objetivo principal de este estudio es desarrollar y aplicar un método de pruebas de significancia de la estacionalidad en los precios de mayoreo de las hortalizas en un mercado. La aplicación del método se da para los productos de un mercado mayorista específico de la Ciudad de México\*.

#### *Razones económicas para las fluctuaciones estacionales*

Se presentan dos formas de exponer las fluctuaciones estacionales de precio de las hortalizas: 1) un caso se presenta suponiendo la demanda estable\*\* durante todo el año, y la oferta inestable, pero representada por dos niveles distintos en dos

\* En una próxima publicación del Centro de Economía Agrícola se mencionan análisis y aplicaciones más detalladas para otros productos.

\*\* Para una definición de demanda u oferta estable e inestable, ver Shepherd, páginas 143-149 (4).

diferentes períodos. 2) el segundo caso también supone la demanda estable, pero estable únicamente dentro de estaciones distintas del año, y también la oferta inestable. Podrían describirse muchos otros casos y combinaciones; pero, para propósitos de ilustración de las razones económicas para las variaciones estacionales de precio, basta con los dos casos mencionados.

*Caso 1. La demanda estable y dos niveles distintos de oferta inestable*

Las suposiciones para este análisis son: la maximización de las ganancias de los productores, y la competencia perfecta. Se presume que la inestabilidad de la oferta se debe a factores incontrolables, con los cuales se tienen que enfrentar los productores, tales como los cambios climáticos o pérdidas inusitadas por el ataque de insectos y enfermedades.

En consecuencia, los productores tienen prevista una respuesta de la oferta para cada nivel de precio de los productos; pero los rendimientos de la producción real fluctúan conforme a los dichos factores incontrolables.

Para simplificar más el análisis, supondremos que sólo hay un recurso variable y que la misma función de producción existe en ambos períodos del año para transformar el recurso variable en el producto deseado. Además, se supone que el producto es perecedero, y de ahí que el almacenamiento pueda transferir la producción de un período al siguiente. Estas suposiciones se hacen exclusivamente para simplificar la presentación del problema, y se pueden cambiar fácilmente, de acuerdo a condiciones más reales.

Las ganancias que le deja al agricultor la producción de las hortalizas específicas en los diversos períodos de tiempo, se pueden expresar con una ecuación que contiene la diferencia entre los ingresos obtenidos de la producción, y los costos totales de producción, tanto variables como fijos. La referida ecuación de las ganancias puede plantearse así:

$$\Pi = P_1q_1 + P_2q_2 - Pr_1 \cdot R_1 - Pr_2 \cdot R_2 - C_1 - C_2 \quad (1)$$

donde:

- $P_1$  = precio del producto en el período 1,
- $P_2$  = precio del producto en el período 2,
- $q_1$  = cantidad producida en el período 1,
- $q_2$  = cantidad producida en el período 2,
- $Pr_1$  = precio del recurso variable en el período 1,
- $Pr_2$  = precio del recurso variable en el período 2,
- $R_1$  = cantidad del recurso variable usado en el período 1,
- $R_2$  = cantidad del recurso variable usado en el período 2,
- $C_1$  = costos fijos en el período 1, y
- $C_2$  = costos fijos en el período 2.

Las ganancias máximas de cada productor necesitan ajustarse a las siguientes condiciones:

$$\frac{\partial \pi}{\partial R_1} = 0 = P_1 \frac{\partial q_1}{\partial R_1} - Pr_1 \quad (2)$$

y

$$\frac{\partial \pi}{\partial R_2} = 0 = P_2 \frac{\partial q_2}{\partial R_2} - Pr_2 \quad (3)$$

$$P_1 = \frac{Pr_1}{\frac{\partial q_1}{\partial R_1}} = f_1 (Pr_1, q_1) \quad (4)$$

y

$$P_2 = \frac{Pr_2}{\frac{\partial q_2}{\partial R_2}} = f_2 (Pr_2, q_2) \quad (5)$$

donde la ecuación (4) representa la función de la oferta para la hortaliza en el período uno, y la ecuación (5) representa la oferta en el período dos (ver Heady y Dillon, 3).

En qué condiciones podemos esperar una oferta diferente de un período a otro, cuando los precios del producto son los mismos en ambos períodos? Si observamos las ecuaciones (4) y (5), podemos ver que cuando  $P_1$  es igual a  $P_2$ , los diferentes precios de los recursos ( $Pr_1$  y  $Pr_2$ ) requerirán diferentes niveles de producción para mantener la igualdad. Esto es, si  $Pr_1 > Pr_2$ , entonces...

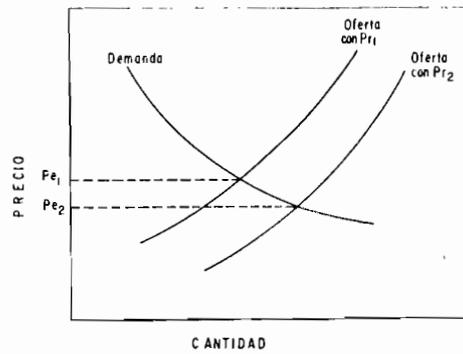
$$\frac{\partial q_1}{\partial R_1} > \frac{\partial q_2}{\partial R_2}, \text{ donde } \frac{\partial q_1}{\partial R_1} \text{ es el producto marginal físico de } q, \text{ usando el}$$

curso  $R$  en el período uno y  $\frac{\partial q}{\partial R}$  representa el mismo fenómeno en el período dos.

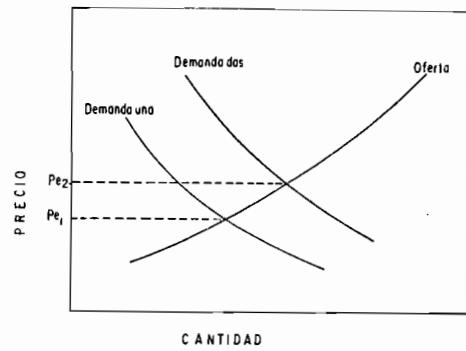
Suponiendo que los productores estén operando dentro de los límites racionales de la función de producción (es decir, el producto marginal físico decreciente) para un recurso variable y los demás fijos (2, páginas 90-96), un producto marginal físico mayor en el período uno significa menos producción en el período uno, que en el período dos. Por consiguiente, un costo superior de recursos en el período uno significa una curva de oferta colocada más hacia la izquierda en el período uno, que en el período dos.

El Caso 1 se puede representar gráficamente como en la Figura 1, donde los diferentes costos de los recursos nos dan curvas de oferta diferentes. Suponiendo curvas de demanda con igual pendiente para ambas estaciones, el precio de equilibrio esperado es diferente para cada estación, originando la variación por temporada.

Como ejemplo, consideremos a la tierra como el recurso variable en la producción de hortalizas. El costo de esta producción puede tomarse como el costo de



**Figura 1.** Precios de equilibrio estacionalmente diferentes, debido a costos de los recursos estacionalmente diferentes.



**Figura 2.** Precios de equilibrio estacionalmente diferentes, debido a funciones de demanda estacionalmente diferentes.

oportunidad de la tierra, o sea si se usara para obtener algún producto alternativo. Si el costo de oportunidad de la tierra es diferente en la estación uno, que en la estación dos la distribución de la tierra para la obtención de nuestra producción, será distinta.

Una fuente distinta de variación en la oferta puede deberse al hecho de que la función productiva en el período uno es diferente de la función productiva en el período dos. En consecuencia los productos marginales físicos son distintos en el período uno y en el período dos, y dan como resultado curvas de oferta diferentes. Este caso puede darse si hay diferencias climáticas entre los dos períodos.

*Caso 2. Oferta inestable y dos niveles distintos de demanda estable.*

El Caso 2 puede explicarse gráficamente con la Figura 2, con demandas estacionales diferentes para el producto que se está considerando. Las funciones diferentes de la demanda indican que, para el mismo nivel de precios, la demanda del consumidor es mayor durante un período del año, que en el otro. Un ejemplo puede ser la demanda de sandía en el verano, en comparación con la del invierno. La hipótesis es que la demanda de sandía es mayor durante los períodos cálidos, que durante los períodos fríos. Los diferentes precios de equilibrio existen para cualquier curva de oferta que no sea la de tipo perfectamente elástico. Para condiciones rígidas, como las comprendidas en el Caso 1 (es decir, un recurso variable y los otros fijos), la curva de la oferta se inclina necesariamente hacia arriba, dando diversos precios de equilibrio para los dos períodos.

*Pruebas para la estacionalidad de los precios*

*Los datos*

Se toman como ejemplo los datos mensuales de precios de mayoreo de una gran cadena de supermercados, para los productos chile verde serrano y ejote. Los precios se miden al nivel de mayoreo de una cadena de supermercados, antes de la dis-

tribución a sus diversos revendedores. Los precios cubren el período comprendido entre enero de 1959 y 1965, o sea un período de 7 años completos.

### El modelo

Los análisis de estos datos fueron efectuados por medio de un análisis de regresión por mínimos cuadrados, en el cual se impuso un conjunto de restricciones sobre la matriz de insumos merced al uso de variables falsas. El resultado del modelo de regresión, obtenido mediante la computadora, se usó para la evaluación total de estos datos.

En el Apéndice de este artículo se incluye una explicación de la técnica general de mínimos cuadrados en la cual se usan variables falsas. Se describen conjuntos de restricciones usados frecuentemente, así como se dan ejemplos, para ayudar en la utilización e interpretación de la solución típica de la regresión.

El conjunto de restricciones impuesto al modelo usado en este estudio, ajusta esencialmente todos los datos a un nivel de referencia de 1965. Consiguientemente, cada efecto "anual" representado por los respectivos valores de  $b$ , es una desviación de ese año, con respecto al año base 1965. Un diferente conjunto de restricciones se usan también para comparar cada mes con el promedio total de 1965. El Cuadro 1 da las constantes (valores de  $b$ ) para los años y meses correspondientes, y el error estándar de los respectivos valores de  $b$ .

### Resultados

Los resultados del Cuadro 1 se muestran en la Figura 3 en la parte (a) para chile verde serrano, y en la parte (b), para ejote. La distribución mensual se presenta en el eje horizontal. Los precios, ajustados al nivel de 1965, según explica el modelo, están representados en el lado izquierdo del eje vertical. El precio promedio, tal como está representado por  $b_0$  en el Cuadro 1, incluye la media mensual del año para el nivel base 1965 y lo representa la línea horizontal gruesa. Las desviaciones mensuales de precio con respecto al precio medio de 1965 repre-

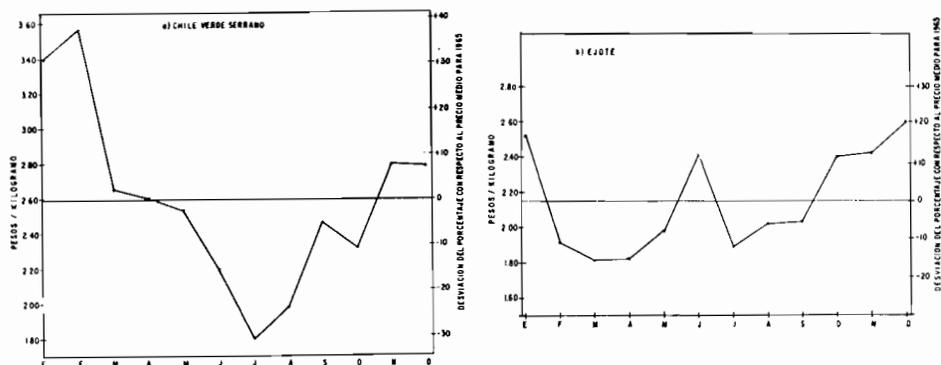


Figura 3. Precios mensuales estimados, al nivel de precios del año de 1965, para dos productos hortícolas en un mercado de mayoreo de la Ciudad de México, y las desviaciones mensuales de precios estimados con respecto al precio medio total. (1959-65).

CUADRO 1

Valores de regresión para el modelo de estacionalidad, explicando las fluctuaciones de precio para productos hortícolas, en un mercado de la Ciudad de México<sup>a</sup>

(a) Chile Verde Serrano					
Efecto anual		Valor del coeficiente	Efecto mensual		Valor del coeficiente
1959	b <sub>1</sub>	-0.067	Enero <sup>c</sup>		0.812*
1960	b <sub>2</sub>	-0.974*	Febrero	b <sub>7</sub>	0.974**
1961	b <sub>3</sub>	-0.781*	Marzo	b <sub>8</sub>	0.062
1962	b <sub>4</sub>	0.108	Abril	b <sub>9</sub>	0.014
1963	b <sub>5</sub>	-1.034**	Mayo	b <sub>10</sub>	-0.056
1964	b <sub>6</sub>	0.543	Junio	b <sub>11</sub>	-0.395
			Julio	b <sub>12</sub>	-0.784*
Error estándar		0.378	Agosto	b <sub>13</sub>	-0.611#
			Septiembre	b <sub>14</sub>	-0.131
Ordenada al origen <sup>b</sup>		2.586	Octubre	b <sub>15</sub>	-0.276
R <sup>2</sup>		0.457	Noviembre	b <sub>16</sub>	0.201
			Diciembre	b <sub>17</sub>	0.191
			Error estándar		0.335
(b) Ejote					
Efecto anual		Valor del coeficiente	Efecto mensual		Valor del coeficiente
1959	b <sub>1</sub>	-0.587**	Enero <sup>c</sup>		0.366*
1960	b <sub>2</sub>	-0.638**	Febrero	b <sub>7</sub>	-0.240
1961	b <sub>3</sub>	-0.353**	Marzo	b <sub>8</sub>	-0.339*
1962	b <sub>4</sub>	-0.741**	Abril	b <sub>9</sub>	-0.332*
1963	b <sub>5</sub>	-0.299#	Mayo	b <sub>10</sub>	-0.169
1964	b <sub>6</sub>	-0.151	Junio	b <sub>11</sub>	-0.264#
			Julio	b <sub>12</sub>	-0.262#
Error estándar		0.165	Agosto	b <sub>13</sub>	-0.132
			Septiembre	b <sub>14</sub>	-0.124
Ordenada al origen <sup>b</sup>		2.150	Octubre	b <sub>15</sub>	0.250#
R <sup>2</sup>		0.525	Noviembre	b <sub>16</sub>	0.274#
			Diciembre	b <sub>17</sub>	0.443**
			Error estándar		0.147

a Ver el apéndice para una presentación formal del modelo

b La ordenada al origen incluye el efecto del nivel de precio para 1965

c El efecto de enero es la suma negativa de todos los otros efectos mensuales

# Significancia del coeficiente de regresión al nivel de 10%

\* Significancia del coeficiente de regresión al nivel de 5%

\*\* Significancia del coeficiente de regresión al nivel de 1%

sentadas por los valores  $b$  del modelo, se han dibujado y relacionado con los meses adyacentes, para representar el ciclo por estaciones.

Como ejemplo, el precio medio total del chile verde serrano es de \$2.586 por kilogramo, ajustado al nivel total del precio para 1965. El valor de  $b$  para el mes de febrero, que es de \$0.974, indica la desviación de precio esperada, conforme se estimó a partir de 7 años de observaciones, con respecto al precio medio total de \$2.586 por kilo, nuevamente sobre la base de niveles totales de precio para 1965. Al añadir la diferencia prevista para el efecto de \$0.94 del mes de febrero, al nivel de precio para 1965, de \$2.586 por kilo, obtenemos el precio esperado para febrero, que es de \$3.560. Este valor queda expresado en la Figura 3 (a).

Las diferencias en porcentaje al nivel medio de precios para 1965, se han representado en el eje derecho de las ordenadas. Por consiguiente, podemos observar que el máximo de precios del ciclo para chile serrano, que se dio en el mes de febrero, está aproximadamente 38% sobre la medida mensual de precios del año. La parte inferior del ciclo tuvo lugar durante el mes de julio y está 30% por abajo de la media de precio en el año.

La variación estacional de precio para los dos productos puede apreciarse en la Figura 3, (a) y (b). El chile verde serrano muestra una fluctuación de un ciclo, con los precios más altos durante los meses de enero y febrero y los precios más bajos durante julio y agosto. El ejote tiene dos ciclos dentro del año, con períodos de precio bajo durante marzo y abril, y luego durante julio. Los períodos de alto precio ocurren durante diciembre y enero y luego durante junio.

Las pruebas de desviaciones mensuales significativas en los precios indican que febrero es significativamente diferente del efecto mensual promedio, al nivel de 1%; enero y julio son significativamente diferentes del efecto mensual promedio, al nivel de 5%, y agosto al nivel de 10% para chile verde serrano. Para el ejote el precio de diciembre es significativamente diferente del efecto medio mensual al nivel de 1%; los de enero, marzo y abril al nivel de 5%; y junio, julio, octubre y noviembre al nivel de 10%. Estos resultados indican que, a ciertos niveles de confianza, la media de precio del año no es un buen estimador del precio efectivo para algunos meses.

Una prueba más interesante consiste en determinar si hay diferencias significativas entre los períodos de bajo precio y los períodos de alto precio, que indiquen la existencia de fluctuaciones estacionales definidas.

Al usar el mes de febrero como período de alto precio y el mes de julio como período de bajo precio para chile verde serrano, vemos que el valor  $t$  es igual a 3.56, que indica que la diferencia entre los períodos alto y bajo es altamente significativa\*. En la misma forma, podemos decir que, para el ejote, marzo es diferente de junio; junio es diferente de julio; julio es diferente de noviembre y; noviembre es diferente de marzo, a un nivel altamente significativo.

#### *Conclusiones*

Se ha demostrado que el precio anual promedio de dos hortalizas en un mercado de mayoreo específico de la Ciudad de México, no es un buen estimador del

\* Las pruebas que se exponen en este análisis se describen en el Apéndice.

precio esperado para algunos meses del año. Además se ha encontrado que existe una fluctuación de un ciclo para el precio de chile verde serrano, con una diferencia significativa (al nivel de 1%) entre el mes alto de febrero y el mes bajo de julio. Hay un doble ciclo para el precio de ejote, con precios altos en marzo y julio, significativamente diferentes de los meses de precios bajos que son junio y diciembre.

Estos resultados indican que las decisiones de los productores basadas en el precio promedio del año previsto, pueden caer en error apreciable al aplicarlas a meses específicos para estas hortalizas. Para chile verde serrano, la parte alta del ciclo de precios, que vienen durante el mes de febrero, muestra que la desviación esperada es de 38% por encima del precio anual promedio. La parte baja del ciclo de precios, que viene durante el mes de julio, tiene una desviación esperada de 30% por abajo del precio anual promedio.

El propósito del presente estudio no fue recomendar la eliminación de la estacionalidad de precios. De hecho, según se explicó en el cuerpo de este artículo, existen causas racionales de origen económico para las diferencias estacionales en los precios de hortalizas. Por consiguiente, el propósito de este trabajo fue reconocer que la estacionalidad de precios existe y puede ser pronosticada mediante el uso de un modelo estadístico por una regresión por mínimos cuadrados.

#### Referencias citadas

1. FOOTE, RICHARD S. Y FOX, KARL A. *Analytical tools for measuring demand*. Agriculture Handbook No. 64. United States Department of Agriculture, Washington, D. C. January 1954.
2. HEADY, EARL O. (1965). *Economics of agricultural production and resource use*. Englewood Cliffs, New Jersey, Prentice-Hall, Inc.
3. HEADY, EARL O. Y DILLON, JOHN L. (1961) *Agricultural production functions*. Amcs. Iowa. The Iowa State University Press.
4. SHEPHERD, GEOFFREY S. (1957). *Agricultural price analysis*. Amcs, Iowa. The Iowa State University Press.
5. WAUGH, FREDERICK V. *Demand and price analysis*. Technical Bulletin No. 1316. United States Department of Agriculture, Washington, D. C. November 1964.

## A P E N D I C E

### ESTABLECIMIENTO DE RESTRICCIONES EN EL MODELO DE REGRESIÓN DERIVADO POR EL METODO DE MINIMOS CUADRADOS

Por Donald K. Hotchkiss

Comúnmente se usa el modelo de regresión para situaciones en las cuales encontramos una variable dependiente (precio), que está relacionada con una segunda variable que es continua en su respuesta (suministro), y en la cual el interés se encuentra principalmente en la naturaleza lineal o posiblemente cuadrática de la relación. En este caso, la variable independiente queda incluida en la matriz X, en las unidades de la medida (lineal), en el cuadrado de la medida (cuadrática). Los coeficientes de regresión (b) que resultan, se usan para interpretar la relación que existe entre ambas variables.

Sin embargo, las variables independientes frecuentemente pueden y deben ser consideradas en grupos cualitativos, en los cuales una relación lineal entre los grupos, no tendría sentido o bien podría no dar una comprensión completa de la relación. (Los años en el presente estudio, expresados como una relación lineal de un número de meses desde 1959, no tienen ningún significado si no se espera ninguna tendencia). Cuando se requiere una clasificación cualitativa de los datos, los valores de la matrix  $X$  constituyen una serie de variables simuladas que toman en cuenta los grados de libertad de la variable. Cualquier conjunto de variables linealmente independientes a  $-1$  (donde  $a$  es el número de clasificaciones de la variable cualitativa), que no tienen correlación de 1.0 entre cualquiera de dos grupos dentro del conjunto, explicará la misma suma de cuadrados para un grupo de datos dados. Sin embargo, algunos conjuntos de restricciones resultan en valores de  $b$ , los cuales tienen una interpretación más útil y, por lo tanto, se usan más frecuentemente.

*La matriz del diseño*

La descripción del modelo es la siguiente:

$$Y_{ij} = \mu + \alpha_i + \lambda_j + \epsilon_{ij} \quad \begin{array}{l} i = 1, \dots, 7 \\ j = 1, \dots, 12 \end{array}$$

donde:

- $Y_{ij}$  = respuesta para el  $j$ -ésimo mes del  $i$ -ésimo año
- $\alpha_i$  = contribución del  $i$ -ésimo año
- $\lambda_j$  = contribución del  $j$ -ésimo mes
- $\epsilon_{ij}$  = componente aleatoria o variación del error.

Este modelo podría también describirse como

$$Y_{ij} = \beta_0 X_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_7 X_7 + \lambda_1 X_8 + \dots + \lambda_{12} X_{19} + \epsilon_{ij}$$

Donde  $\alpha_i$  y  $\lambda_i$  se definen como se hizo previamente, entonces:

$X_0 = 1$  para cada observación

$X_i = 1$  si el efecto del  $i$ -ésimo año está presente (o el  $j$ -ésimo mes)

$X_i = 0$  si el efecto del  $i$ -ésimo año no está presente (o del  $j$ -ésimo mes)

Para ejemplificar el modelo, emplearemos tres años con dos observaciones en cada uno:

$$Y_{ij} = \beta_0 X_0 + \sum \alpha_i X_i + \epsilon_{ij} \quad \begin{array}{l} \text{donde, } i = 1, 2, 3 \\ j = 1, 2 \end{array}$$

o para el ejemplo,

$$Y_{ij} = \beta_0 X_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \epsilon_{ij}$$

En notación matricial:  $Y = X\beta + \epsilon$ :

$$Y = \begin{array}{c} X \\ X_0 \quad X_1 \quad X_2 \quad X_3 \end{array} \beta + \epsilon$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix}$$

Al multiplicar este conjunto de matrices se obtienen los datos en la forma del modelo original.

La solución por el método de mínimos cuadrados para encontrar los coeficientes de regresión consiste en encontrar un conjunto de estimadores para los parámetros que minimicen la suma de los cuadrados de las desviaciones con respecto a la línea de regresión. De la solución por el método de mínimos cuadrados resulta una serie de ecuaciones (ecuaciones normales), que se resuelven en forma simultánea para dar los estimadores de los parámetros en la matriz  $\beta$ .

Usando notación matricial las ecuaciones normales quedarían como sigue:

$$X'X\beta = X'Y$$

El valor para la matriz  $X'X$  de nuestro ejemplo es:

$$X'X = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

Decimos que la matriz  $X'X$  es singular porque su determinante es cero, por lo tanto, no existe una solución que sea única para  $\beta$ . Tal vez una explicación más sencilla de la singularidad de la matriz es la siguiente: tenemos 4 incógnitas (parámetros  $\beta_0, \alpha_1, \alpha_2, \alpha_3$ ) y sólo tenemos 3 ecuaciones independientes. Esta situación puede observarse en la matriz  $X'X$ , en donde las hileras, 2, 3 y 4 sumadas dan la hilera 1, o bien sumando la suma de las columnas 2, 3 y 4 de la matriz del diseño nos dan la columna 1, o sea que existe una dependencia entre la primera columna y el conjunto de las tres columnas restantes. Consecuentemente debe imponerse una restricción (una relación lineal entre los parámetros) para tener un conjunto de ecuaciones solubles. Al establecer esta restricción, no afectará la solución final, ya que estamos principalmente interesados en una evaluación relativa de los efectos del año.

#### *La matriz del diseño (con restricciones)*

Al imponer un conjunto de restricciones lineales sobre los parámetros en la matriz del diseño, esencialmente se reparametriza el modelo original, lo cual, nos conducirá a una solución única para cada restricción en particular. Existe un número infinito de formas para reparametrizar el modelo original, sin embargo, la selección del conjunto de restricciones, depende de la facilidad que haya en la construcción de la matriz restringida  $X$ , y es más importante aún la utilidad directa de los resultados de la regresión para interpretar los datos.

Para ilustrar el modelo reparametrizado, en contraste con el modelo original, usaremos tres conjuntos de restricciones comúnmente utilizadas. Además de discutir el mecanismo para crear la matriz restringida  $X$ , se analizará la utilidad de cada conjunto de restricciones para la interpretación de un conjunto de datos.

*Restricción (1)*

$$\Sigma \alpha_i = 0; \text{ ó } \alpha_1 + \alpha_2 + \alpha_3 = 0.$$

Esto puede expresarse así:

$$\alpha_3 = -\alpha_1 - \alpha_2$$

De nuestro modelo original:

$$Y_{ij} = \beta_0 X_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \epsilon_{ij},$$

podemos expresar el modelo reparametrizado como

$$Y_{ij} = \lambda_0 X_0 + \lambda_1 X_1 + \lambda_2 X_2 + \epsilon_{ij}$$

donde,

$$X_0 = 1 \text{ para cada observación.}$$

$$X_i = 1 \text{ si el efecto del } i\text{-ésimo año está presente.}$$

$$X_i = 0 \text{ si no está presente el efecto del } i\text{-ésimo año.}$$

Usando la restricción anterior el efecto del año 3 puede expresarse en términos del efecto del año 1 y del efecto del año 2, (para año 3,  $X_1 = -1$ ,  $X_2 = -1$ ). Por lo tanto, para el año 1, tenemos:

$$Y_{ij} = \lambda_0(1) + \lambda_1(1) + \lambda_2(0) + \epsilon_{ij} = \beta_0 + \alpha_1 + \epsilon_{ij}$$

y para el año 3 tenemos,

$$Y_{3j} = \lambda_0(1) + \lambda_1(-1) + \lambda_2(-1) + \epsilon_{3j} = \lambda_0 - \lambda_1 - \lambda_2 + \epsilon_{3j}$$

que es equivalente a,  $Y_{3j} = \beta_0 + \alpha_3 + \epsilon_{3j}$

Así, el nuevo modelo con sus restricciones es,

$$Y = \begin{matrix} & & X_{(1)}^* & & \beta & + & \epsilon \\ & & X_0 & X_1 & X_2 & & \\ \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} & = & \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} & \begin{bmatrix} \beta_0 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} & \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix} \end{matrix}$$

$$\text{y, } [X_{(1)}^*]' [X_{(1)}^*] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 4 \end{bmatrix}$$

donde  $X_{(1)}^*$  es la matriz del diseño con restricciones. La matriz  $[X_{(1)}^*]' [X_{(1)}^*]$  no es singular, y por lo tanto, tiene una solución única de las ecuaciones normales para los parámetros  $\beta_0$ ,  $\lambda_1$  y  $\lambda_2$ . El lector notará que ninguna combinación de las columnas en  $X_{(1)}^*$ , o conjunto de hileras en  $[X_{(1)}^*]' [X_{(1)}^*]$ , sumadas dan alguna combinación de columnas o hileras.

*Restricción (2)*

Una de las  $\alpha_i =$  cero, o más específicamente nuestro ejemplo  $\alpha_3 = 0$ .

En esta forma definimos ahora un nuevo conjunto de parámetros ( $\gamma_i$ ):

$$\gamma_0 = \beta_0 + \alpha_3$$

$$\gamma_1 = \alpha_1 - \alpha_3$$

$$\gamma_2 = \alpha_2 - \alpha_3$$

el modelo reparametrizado es ahora:

$$Y_{ij} = \gamma_0 X_0 + \gamma_1 X_1 + \gamma_2 X_2 + \epsilon_{ij},$$

o en términos del modelo original:

$$Y_{ij} = (\beta_0 + \alpha_3) X_0 + (\alpha_1 - \alpha_3) X_1 + (\alpha_2 - \alpha_3) X_2 + \epsilon_{ij},$$

donde:

$$X_0 = 1 \text{ para cada observación.}$$

$$X_i = 1 \text{ si el efecto del } i\text{-ésimo año está presente.}$$

$$X_i = 0 \text{ si el efecto del } i\text{-ésimo año no está presente.}$$

Así, para el efecto del año 1 tenemos:

$$Y_{1j} = (\beta_0 + \alpha_3)(1) + (\alpha_1 - \alpha_3)(1) + (\alpha_2 - \alpha_3)(0) + \epsilon_{1j} = \beta_0 + \alpha_1 + \epsilon_{1j}$$

y para el año 3:

$$Y_{3j} = (\beta_0 + \alpha_3)(1) + (\alpha_1 - \alpha_3)(0) + (\alpha_2 - \alpha_3)(0) + \epsilon_{3j} = \beta_0 + \alpha_3 + \epsilon_{3j}$$

La matriz  $[X_{(2)}^*]$  y  $[X_{(2)}^*]'[X_{(2)}^*]$  es:

$$X_{(2)}^* = \begin{matrix} & X_0 & X_1 & X_2 \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} & [X_{(2)}^*]'[X_{(2)}^*] = \begin{bmatrix} 6 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \end{matrix}$$

Nuevamente tenemos una matriz no singular y podemos obtener una solución única para los parámetros de las ecuaciones normales donde estos están expresados en la matriz  $\beta$ :

$$\beta = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix}$$

*Restricción (3)*

Otra restricción muy útil, es un conjunto de contrastes ortogonales, los cuales, pueden representar una serie de comparaciones susceptibles de interpretación entre el conjunto de tratamiento (años). Como ilustración podemos comparar el año 1 con el año 3 (contraste 1) y el año 2 con el promedio de los años 1 y 3 (contraste 2).

Para este conjunto de restricciones los nuevos parámetros son:

$$\begin{aligned} \xi_1 &= -\alpha_1 + \alpha_3 \\ \xi_2 &= \alpha_1 - 2\alpha_2 + \alpha_3 \end{aligned}$$

o en términos de  $\alpha_i$ :

$$\begin{aligned} \alpha_1 &= -\xi_1 + \xi_2 \\ \alpha_2 &= -2\xi_2 \\ \alpha_3 &= \xi_1 + \xi_2 \end{aligned}$$

Por lo tanto el modelo reparametrizado es:

$$Y_{ij} = \beta_0 X_0 + \{(-1)\xi_1 + (1)\xi_2\} X_1 + (-2)\xi_2 X_2 + \{(1)\xi_1 + (1)\xi_2\} X_3 + \epsilon_{ij}$$

ó

$$Y_{ij} = \beta_0 X_0 + \xi_1 (X_3 - X_1) + \xi_2 (X_1 - 2X_2 - X_3) + \epsilon_{ij} = \\ = \beta_0 Z_0 + \xi_1 Z_1 + \xi_2 Z_2 + \epsilon_{ij}$$

En esta última expresión tenemos que:

$$Z_0 = X_0, Z_1 = X_3 - X_1, \text{ y } Z_2 = X_1 - 2X_2 + X_3$$

y definimos:

$X_0 = 1$  para cada observación.

$X_i = 1$  si el efecto del  $i$ -ésimo año está presente.

$X_i = 0$  si el efecto del  $i$ -ésimo año no está presente.

Los valores de  $Z$  por los años 1, 2 y 3 se presentan a continuación:

Año	$Z_0$	$Z_1$	$Z_2$
1	1	-1	1
2	1	0	-2
3	1	1	1

Por lo tanto, para el año 1:

$$Y_{1j} = \beta_0(1) + \xi_1(-1) + \xi_2(1) + \epsilon_{1j} = \beta_0 + \alpha_1 + \epsilon_{1j}$$

para el año 2:

$$Y_{2j} = \beta_0(1) + \xi_1(0) + \xi_2(-2) + \epsilon_{2j} = \beta_0 + \alpha_2 + \epsilon_{2j}$$

para el año 3:

$$Y_{3j} = \beta_0(1) + \xi_1(1) + \xi_2(1) + \epsilon_{3j} = \beta_0 + \alpha_3 + \epsilon_{3j}$$

La matriz con restricciones ( $Z$ ), la matriz de parámetros ( $\beta$ ) y  $Z'Z$  son:

$$Y = Z \beta + \epsilon$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \begin{bmatrix} Z_0 & Z_1 & Z_2 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix}$$

$$\text{y } Z'Z = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 12 \end{bmatrix}$$

Tenemos de nuevo una matriz no singular y una solución única para los parámetros  $\beta_0$ ,  $\xi_1$  y  $\xi_2$ .

#### Interpretación de resultados

La mecánica de la inversión de una matriz no se discutirá en este trabajo, ya que esta operación puede realizarse con una computadora de acuerdo a las instrucciones de un programa. Sin embargo, el investigador práctico debe entender el resultado de las regresiones y la manera de hacer comparaciones específicas. Por otra parte, la información de esta sección puede guiar al investigador a seleccionar un conjunto de restricciones.

La salida de un programa típico de regresión incluirá la varianza ( $s_{y.x}^2$  o  $s_{y.x}$ ), el coeficiente de correlación múltiple ( $R^2$ ), los coeficientes de regresión respectivos ( $b_0, b_1, \dots, b_k$ ), el error estándar de los coeficientes de regresión ( $s_{b_i}$ ) y los valores de  $t$  que prueban las hipótesis generales sobre los parámetros.

$$H_0 : \beta_{y \text{ sobre } 1, 2, \dots, i-1, i+1, \dots, k} = 0.$$

*¿Qué significa esa información y cómo puede usarse?*

La prueba de la efectividad del modelo mediante  $R^2$  o  $F$ . Puede usarse una prueba de  $F$  formada con el cuadrado medio de regresión sobre el cuadrado medio del error para determinar el modelo que explica alguna variabilidad más allá de la variación aleatoria.

$$F = \frac{\text{Cuadrado medio de regresión}}{\text{Cuadrado medio del error}}$$

Existen tablas en los libros de texto de Métodos Estadísticos para hacer la misma prueba usando  $\sqrt{R^2} = R$  como una prueba rápida y directa para determinar el nivel de significación del modelo de regresión.

En el Cuadro A-1 se presenta el análisis de varianza para chile verde de las ecuaciones de regresión usadas en este trabajo. Los pasos 2 y 3 en este cuadro dan los cuadrados medios que se usarán para la prueba de  $F$ .

$$F = \frac{2.790}{0.855} = 3.26$$

La  $F$  calculada (3.26) se compara con el valor de la tabla con 17 y 66 grados de libertad.

El valor para  $R^2$  en este estudio (ver Cuadro 1), fue,  $R^2 = 0.457$  ó  $R = 0.676$ , lo cual es compara con los valores de la tabla para  $R$  ( $R_{(1,0.5)} = 0.235$  y  $R_{(1,0.1)} = 0.307$ ), teniendo 66 grados de libertad.

*Prueba de la importancia de un subconjunto de parámetros en el modelo*

Para determinar la importancia de un subconjunto de parámetros, se ajusta un modelo reducido de regresión en el cual se ha omitido el subconjunto de parámetros bajo prueba. La diferencia en la suma de cuadrados de regresión para el modelo completo y el modelo reducido, es una medida de la importancia del subconjunto de parámetros en la explicación de la variabilidad de las observaciones originales. Puede hacerse una prueba de significación expresando esta diferencia como un cuadrado medio y calculando la  $F$ , usando el cuadrado medio del error del modelo original.

Esto puede demostrarse usando los datos de chile verde (Cuadro A-1) y haciendo la prueba para ver si existen efectos de año. El paso 5 del Cuadro A-1, da la variabilidad y los precios del chile verde explicado por el factor de años, ajustado por el efecto de meses. El valor de  $F$  para esta prueba es:

$$F = \frac{4.489}{0.855} = 5.25$$

que se compara con los valores de la tabla con 6 y 66 grados de libertad.

## CUADRO A-1

*Análisis de la varianza de precio para chile verde obtenido por medio del análisis de regresión de los precios de chile verde sobre los efectos de años y meses usando la matriz X, con restricciones*

Pasos*	Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrado de la media	F
1	Total	83	103.862		
2	R( $b_1, \dots, b_6, b_7, \dots, b_{17}$ )	17	47.428	.855	3.26
3	Residual	66	56.434	2.790	
4	R( $b_7, \dots, b_{17}$ )	11	20.495		
5	Efecto de años ajustados a meses (Paso 2 — Paso 4)	6	26.933	4.489	5.25

\* En los pasos 2 y 3 se considera el modelo completo (años y meses) y en la línea 4, se considera el modelo reducido con la variable en meses únicamente.

*Interpretación de los valores de "t" y elección de restricción*

La hipótesis probada con los valores de t en el análisis de regresión ( $H_0: \beta_i = 0$ ) varía de acuerdo al tipo de la restricción impuesta a la matriz original X (Cuadro A-2).

En el estudio de chile verde, por ejemplo, hemos utilizado las restricciones "año 1965 = 0" ( $\alpha_7 = 0$ ) y "la suma de los efectos de meses = 0" ( $\sum \lambda_i = 0$  ó  $\lambda_{12} = -(\lambda_1 + \dots + \lambda_{11})$ ). En consecuencia cada valor de t que prueba los valores de b para años, contrasta el año respectivo con la media de 1965. Cada valor de t correspondiente a los efectos del mes prueba la significancia de la desviación de mes.

El Cuadro 1 de este artículo presenta los valores de b para el estudio del chile verde. El valor de t para 1964 ( $b_6 = 0.543$ ) es:

## CUADRO A-2

*Hipótesis de nulidad probadas con los valores de t en el análisis de regresión bajo las diferentes restricciones en el problema hipotético (grupos de 3 años)*

Restricciones		Hipótesis de nulidad ( $H_0$ )*	
Núm.	Definición	con restricción	modelo original
(1)	$\sum \alpha_i = 0$	$\lambda_i = 0$	$\alpha_i - \alpha_3 = 0$
(2)	$\alpha_3 = 0$	$\gamma_i = 0$	$\alpha_i = 0$
(3)	Contrastes		
	(a) $\alpha_3 - \alpha_1 = 0$	$\xi_1 = 0$	$\alpha_3 - \alpha_1 = 0$
	(b) $\alpha_3 - 2\alpha_2 + \alpha_1 = 0$	$\xi_2 = 0$	$\alpha_3 - 2\alpha_2 + \alpha_1 = 0$

\* La primera columna expresa la hipótesis de nulidad en términos de los parámetros resultantes después de imponer las restricciones. La segunda columna expresa la hipótesis de nulidad en términos del modelo original (efecto de años).

$$t = \frac{.543}{.378} = 1.43,$$

el cual no es significativo cuando se compara con la  $t$  de las tablas con 66 grados de libertad. Esta prueba indica, que no existe evidencia que 1964 difiere de 1965 después de considerar los ajustes para otros años y otros meses.

El valor para febrero ( $b_7 = 0.974^{**}$ ) es significativo indicando que las desviaciones en precio con respecto a la media general se deben al efecto del mes de febrero, el cual aparentemente no es debido al efecto aleatorio únicamente.

La selección del tipo de restricción, en consecuencia, depende de los objetivos específicos del estudio. Por ejemplo, si una clasificación del estudio tiene un tratamiento testigo, la desviación para el testigo podría establecerse igual a cero, ( $\alpha_T = 0$ ), por lo tanto, todos los demás tratamientos en tal clasificación podrían compararse con el testigo. El conjunto de restricciones del tipo 3, por otra parte, se puede utilizar para establecer contrastes específicos entre las medias (no necesariamente ortogonales), que están de acuerdo con los objetivos del estudio.

Es conveniente que cuando el total de grados de libertad para una clasificación, están incluidos en la matriz restringida, la suma de cuadrados para esta clasificación, será la misma, independientemente de la combinación de restricciones impuestas, siempre y cuando, ningún subconjunto tenga una correlación perfecta con cualquier otro subconjunto.

Las interacciones de dos factores cualitativos son frecuentemente de interés. Para crear columnas en la matriz restringida para interacciones entre dos factores, simplemente se multiplica cada columna de un factor,  $(a-1)$  columnas, por cada columna del otro factor,  $(b-1)$  columnas, creando  $(a-1)(b-1)$  nuevas columnas para los grados de libertad de la interacción. La discusión anterior para la interpretación de pruebas de hipótesis con  $F$  y  $t$  se aplica para el efecto de interacción del modelo.

*Prueba entre los valores b*

Con frecuencia, una prueba de particular interés puede no formar parte del análisis de regresión. Por ejemplo, supóngase que se quiere comparar el año 1 y el año 2 en el ejemplo hipotético. ( $\alpha_1$  vs  $\alpha_2$  ó  $H_0 : \alpha_1 - \alpha_2 = 0$ ).

Bajo la restricción (1)

$\alpha_1$  está estimado por  $\lambda_1$   
 $\alpha_2$  está estimado por  $\lambda_2$

de ahí que  $t = \frac{\hat{\lambda}_1 - \hat{\lambda}_2}{s_{\lambda_1 - \lambda_2}}$

donde  $s_{\lambda_1 - \lambda_2} = \sqrt{s_{y,x}^2 (C_{11} + C_{22} - 2C_{12})}$

los valores  $C_{ij}$  se obtienen de la matriz invertida la cual es parte de la rutina de cálculo de análisis de regresión.

Bajo la restricción (2)  $\gamma_1 = \alpha_1 - \alpha_3$  y  $\gamma_2 = \alpha_2 - \alpha_3$

por lo tanto:  $\gamma_1 - \gamma_2 = (\alpha_1 - \alpha_3) - (\alpha_2 - \alpha_3) = \alpha_1 - \alpha_2$

$$y, t = \frac{\gamma_1 - \gamma_2}{s_{\gamma_1 - \gamma_2}}$$

donde  $s_{\gamma_1 - \gamma_2} = \sqrt{s_{y,x}^2 (C_{11} + C_{22} - 2C_{12})}$

usando el inverso  $(C_{ij})$  para este juego de restricciones.

$$\text{Bajo la restricción (3)} \quad \begin{aligned} \xi_1 &= \alpha_1 - \alpha_3 \\ \xi_2 &= \alpha_1 - 2\alpha_2 + \alpha_3 \end{aligned}$$

$$\xi_1 + \xi_2 = \alpha_1 - \alpha_3 + \alpha_1 - 2\alpha_2 + \alpha_3 = 2\alpha_1 - 2\alpha_2$$

Por lo tanto:  $1/2 (\xi_1 + \xi_2) = \alpha_1 - \alpha_2$

para probar la hipótesis  $H_0: \alpha_1 - \alpha_2 = 0$  usamos  $t = \frac{1/2(\xi_1 + \xi_2)}{s_{1/2(\xi_1 + \xi_2)}}$

donde  $s_{1/2(\xi_1 + \xi_2)} = \sqrt{s_{y,x}^2 (1/2)^2 (C_{11} + C_{22} + 2C_{12})}$

Cada una de estas pruebas de  $t$  se compara con los valores de las tablas usando los grados de libertad de  $s_{y,x}$  a partir del modelo de regresión.

En la discusión, se han hecho comparaciones entre dos meses, por ejemplo marzo y agosto en el estudio de chile verde.

Del Cuadro 1 podemos obtener los coeficientes de regresión para febrero y julio ( $b_7 = 0.974$ ,  $b_{12} = -0.784$ ). El Cuadro A-1 da  $s_{y,x} = 0.855$  y de la matriz invertida encontramos que

$$\begin{aligned} C_{7,7} &= 0.1310 \\ C_{12,12} &= 0.1310 \\ C_{7,12} &= -0.0119 \end{aligned}$$

$$\text{así, } t = \frac{0.974 - (-.784)}{\sqrt{0.855 \{ .1310 + .1310 + 2(-.0119) \}}} = 3.56$$

este valor de  $t$  se compara con el valor de las tablas y con 66 grados de libertad.

### Resumen

a) Una variable cualitativa (con  $a$  niveles) puede considerar el análisis de regresión usando un conjunto de  $(a-1)$  variables simuladas. Estas variables simuladas deben basarse en las restricciones impuestas a la matriz del diseño.

b) Se han presentado tres tipos de restricciones en este Apéndice, así como la discusión e interpretación del análisis de regresión.

c) El problema en el presente trabajo pudo haberse resuelto usando la rutina del análisis de la varianza. Sin embargo, cuando existen combinaciones de variables de clasificación con una o más variables cualitativas, el enfoque con el modelo de regresiones derivado por el método de los cuadrados mínimos permite que sean aplicables los métodos delineados en este Apéndice.