Análisis de Regresión Multivariado

1. Introducción
2. Metodología Estadística
   2.1 Formulación del modelo
   2.2 Estimación de parámetros
   2.3 Prueba de hipótesis lineal general
   2.4 Pruebas estadísticas
3. Ejemplo de aplicación

1. Introducción

El análisis de regresión multivariado (ARM) es la generalización en el campo multivariado del análisis de regresión lineal múltiple. El modelo de un ARM representa la relación funcional de un grupo de variables dependientes (Que deben estar correlacionadas) en función de un conjunto de variables independientes o predictoras.

Entonces, todo el proceso de estimación y de inferencia estadística aplicado al análisis de regresión múltiple se puede generalizar para el ARM.

El propósito de usar el ARM pueden ser para fines explicativos o predictivos, a partir de la estimación de los parámetros del modelo y en base de una muestra.

2.1 Pruebas Estadísticas Preliminares

Con el Software R se hará la prueba de normalidad p-variada de Shapiro para las variables dependientes.

H0: Las p variables tienen distribución normal p-variada

2.2 Pruebas Estadísticas Preliminares

Prueba de esfericidad de Bartlett. Hacerla usando el Análisis Factorial. Lo que se quiere es determinar que las variables dependientes estén correlacionadas.

3.1 Formulación del modelo

1. Modelo de regresión lineal multivariado

Para un modelo de regresión lineal multivariado donde se tiene “q” variables dependientes , “p” variables independientes y considerando un conjunto de datos de “n”observaciones se expresa como:

$$\underset{(n\times q)}{Y} = \underset{(n\times(p+1))}{X}\ \underset{((p+1)\times q)}{\beta} + \underset{(n\times q)}{\varepsilon}$$

Y En una matriz de (n×q) de variables dependientes.

X Es la matriz de (n×(p+1)) de variables independientes

β Es la matriz de ((p+1)xq) de parámetros desconocidos del modelo.

ε Es la matriz de (n×q) que representa los errores del modelo

3.2 Estimación de Parámetros

El estimador mínimo cuadrático para la matriz de coeficientes de regresión será:

$$\hat{\beta}_{(p+1)xq} = (\mathrm{X}'\,\mathrm{X})^{-1} X\,'Y$$

Se puede observar que la estimación del modelo del ARM corresponde a p modelos de regresión múltiple; esto es, cada modelo estimado tiene las mismas variables independientes pero diferentes variable dependiente. Sin embargo, no serán las mismas estimaciones individuales.

3.3 Prueba de hipótesis lineal general

- **Formulación de hipótesis**

$$\mathrm{H}_0 : CBM = K$$

$$\mathrm{H}_1 : CBM \neq K$$

donde las matrices: $C_{(g\times q)}$ $M_{(p\times r)}$ $K_{(g\times r)}$

son conocidas.

3.4 Prueba de hipótesis lineal general

- **Cuadro del análisis de variancia**

Fuentes de Variación	G.L	Matrices de Suma de Cuadrados Y Productos
Debido a H_0	q	$H = (C\hat{\beta}M - K)'(C(X'X)C')^{-1}(C\hat{\beta}M - K)$
Residual	$n-q$	$E = (YM - X\hat{\beta}M)'(YM - X\hat{\beta}M)$
Total	n	$T = H + E$

3.5 Pruebas estadísticas

Existen cuatro criterios que son los más usados como pruebas estadísticas, para probar la hipótesis Ho. Estos criterios se basan en el cálculo de los autovalores de la matriz HE^{-1}.

Algunas pruebas se han hallado la distribución exacta y otras son aproximaciones. Así mismo, se han desarrollado buenas aproximaciones con convergencia asintótica y robustas a la distribución F.

3.5 Pruebas estadísticas

1) Criterio del Máximo Autovalor (Roy` s)

Se basa en el máximo autovalor de HE^{-1}

$$R = \frac{l_1}{1 + l_1}$$

2) Criterio de la Traza (Lawley y Hotelling)

$$T = Tr\,(HE^{-1}) = \sum_{i=1}^{p} l_i$$

3.5 Pruebas estadísticas

3) Criterio de Bartlett – Nanda – Pillai

$$V = \sum_{i=1}^{p} \frac{l_i}{1+l_i} = Tr\left[H(E+H)^{-1}\right]$$

4) Criterio de Razón de verosimilitud (Lambda de Wilk` s)

$$W = \frac{|E|}{|T|} = \frac{|E|}{|H+E|} = \prod_{i=1}^{p} (1+l_i)^{-1}$$

4. Ejemplo de aplicación

- Se hizo un estudio cinético de la deshidratación osmótica y el contenido de vitamina C en piña por medio de una solución osmótica. Para lo cual se estudia la Temperatura en °C como variable X1 y la presión en mb como variable X2, sobre la determinación de peso en humedad en % (Y1) y el contenido de vitamina C en mg/muestra (Y2), en 60 minutos. Se utilizaron 8 muestras experimentales de piña.
- Se quiere determinar si el peso en humedad y el contenido de vitamina C dependen de la temperatura y la presión.
- Ver el archivo **Estudio de piña**.

4. Ejemplo de Aplicación

- Una investigadora ha recolectado datos de las siguientes variables dependientes: Contenido de colesterol (Y1), Contenido de triglicéridos (Y2) y Contenido de glucosa (Y3) de 30 personas. También recolectó datos de las siguientes variables independientes: cuántos gramos de carne roja (X1), pescado (X2), productos diarios (X3) y chocolate (X4) consumen a la semana.
- Las variables Y1, Y2 y Y3 se expresan en mg/dL.
- Se quiere determinar si el contenido de colesterol, triglicéridos y glucosa dependen del consumo de carne roja, pescado, productos diarios y chocolate.
- Ver el archivo **Efecto de algunos alimentos.**