MA34B - Formulario de Modelos Lineales

9 de julio 2006

Profesor cátedra: Rodrigo Abt B.
Auxiliar: Julio Deride

1) Formulación general del Modelo Lineal:

$$Y = X\beta + \varepsilon$$

En que Y es un vector de $n \times 1$, X es una matriz de $n \times k$, β es un vector de $k \times 1$, y ε es un vector de $n \times 1$. De manera desagregada:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

NOTA: Cualquier fila puede ser representada como:

$$y_i = \mathbf{x}_i^t \beta + \varepsilon_i$$

En que $\mathbf{x}_i$ es el **vector** correspondiente a la fila i-ésima de la matriz X.

2) Estimador MCO:

$$\hat{\beta} = (X^t X)^{-1} X^t Y$$

- El estimador MCO $\hat{\beta}$ para los coeficientes β es aquel que minimiza la suma de los cuadrados de los errores: $\sum \varepsilon_i^2$.

- O de otra forma, corresponde al resultado de proyectar el vector Y sobre el subespacio generado por las columnas de X.

- Para que exista una única solución al problema anterior, la matriz X debe ser de rango completo, esto es, que las columnas de X sean l.i.

- La solución MCO es una solución de tipo geométrica-algebraica

- $\hat{\beta}$ es una combinación lineal de los y_i.

- Además se tienen los siguientes resultados:

$$\hat{Y} = X\hat{\beta}$$
$$Y = X\hat{\beta} + \hat{\varepsilon}$$
$$\text{Por ende: } \hat{\varepsilon} = Y - \hat{Y}$$

- Cuando existe una constante en el modelo, una medida de la çalidad"del ajuste es usar el *coeficiente de determinación*:

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\sum (y_i - \bar{y})^2}$$

- Además si el modelo tiene constante:

$$SST = SSE + SSR$$

  En que $SST = \sum (y_i - \bar{y})^2$, $SSE = \sum (\hat{y}_i - \bar{y})^2$ y $SSR = \sum \hat{\varepsilon}_i^2$.

- Si suponemos que $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$ y $Cov(\varepsilon_i, \varepsilon_j) = 0 \forall i \neq j$, se tendrá además que el estimador MCO es insesgado, con varianza $Var(\beta) = \sigma^2 (X^t X)^{-1}$, y es el mejor estimador entre todos los estimadores insesgados LINEALES en Y (Teorema de Gauss-Markov). Además se puede constuir un estimador insesgado para σ^2:

$$\tilde{\sigma}^2 = \frac{\hat{\varepsilon}^t \hat{\varepsilon}}{n - k}$$

3) Inferencia en el Modelo Lineal:

Suponiendo además que los errores son normales: $\varepsilon \sim N_m(0, \sigma^2 I_n)$, se tiene la base para hacer test de hipótesis. En este caso se tiene que: $\hat{\beta} \sim$

$N_m(\beta, \sigma^2(X^tX)^{-1})$, en cuyo caso se tiene que: $\hat{\beta}_i \sim N(\beta_i, \sigma^2(X^tX)^{-1}_{ii})$, donde $\sigma^2(X^tX)^{-1}_{ii}$ corresponde al elemento i-ésimo de la diagonal de la matriz.

NOTA: Numericamente, el estimador MCO es igual al estimador máximo verosímil de β bajo el supuesto de normalidad de los errores. Además el estimador es eficiente, esto es, alcanza la cota de Cramer-Rao, y es el EIVUM.

a) *Tests sobre un coeficiente del modelo*

- Para hacer tests estadísticos sobre un coeficiente, se utiliza el estadístico t:

$$t_{\beta_i} = \frac{\hat{\beta}_i - \beta_{i_0}}{\tilde{\sigma}\sqrt{(X^tX)^{-1}_{ii}}} \sim t(\alpha/2)_{n-k}$$

Bajo las hipótesis:

$$\begin{array}{ll} H_0: & \beta_i = \beta_{i_0} \\ H_1: & \beta_i \neq \beta_{i_0} \end{array}$$

Se rechaza H_0 si $|t_{\beta_i}| > t(\alpha/2)_{n-k}$. Para testear la significación individual, se toma $\beta_{i_0} = 0$, y se rechaza H_0 con el criterio anterior, o si P-Valor $< \alpha/2$.

NOTA: Este mismo estadístico se utiliza para hacer intervalos de confianza sobre el coeficiente en cuestión.

- Para hacer tests estadísticos sobre combinaciones lineales de coeficientes $a^t\beta$, se tiene que:

$$\begin{array}{l} E(a^t\hat{\beta}) = a^t\beta \\ Var(a^t\hat{\beta}) = a^t\sigma^2(X^tX)^{-1}a \end{array}$$

Por ende, usando el T.C.L.:

$$\frac{N(0,1)}{\sqrt{\dfrac{\chi^2_{n-k}}{n-k}}} = \frac{\dfrac{a^t\hat{\beta} - a^t\beta}{\sigma\sqrt{a^t(X^tX)^{-1}a}}}{\dfrac{(n-k)\tilde{\sigma}^2}{\sigma^2}/(n-k)} = \frac{a^t\hat{\beta} - a^t\beta}{\tilde{\sigma}\sqrt{a^t(X^tX)^{-1}a}} \sim t(\alpha/2)_{n-k}$$

Y se resuelve de manera análoga a los test de un coeficiente.

a) Tests sobre un conjunto de coeficientes

- Si existe constante en el modelo:

$$F = \frac{\frac{SSE}{k-1}}{\frac{SSR}{n-k}} \sim F_{k-1,n-k}$$

Y se rechaza H_0 : *todos los β salvo la constante son 0* si $F > F_{k-1,n-k}$ a un nivel de significación α.

NOTA: El estadístico F se puede construir a partir del R^2 como:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

4) Predicción:

Una predicción corresponde a un valor que toma y usando el modelo bajo una nueva observación x_0:

$$y_{n+1} = \mathbf{x}_0^t \beta + \varepsilon_{n+1}$$

Un predictor natural es $\hat{y}_{n+1} = \mathbf{x}_0^t \hat{\beta}$, donde la esperanza del error de predicción $\hat{y}_{n+1} - y_{n+1}$ es 0, y su varianza es $\sigma^2(1 + x_0^t(X^tX)^{-1}x_0)$.

Bajo la hipótesis de normalidad de los errores:

$$\frac{\hat{y}_{n+1} - y_{n+1}}{\tilde{\sigma}\sqrt{1 + x_0^t(X^tX)^{-1}x_0}} \sim t(\alpha/2)_{n-k}$$

6) Comparación de modelos:

Sean:

$$\begin{aligned}
(1) E(Y) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{k1} x_{k1} \\
(2) E(Y) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{k2} x_{k2}
\end{aligned}$$

$$\frac{SSR2 - SSR1/(k1-k2)}{SSR1/(n-k1)} \sim F_{k1-k2,n-k1}$$

Donde SSR1 y SSR2 corresponden a la suma de los cuadrados de los residuos del modelo original y del modelo con variables omitidas respectivamente, y donde k1 y k2 es la cantidad de coeficientes de cada modelo

6) Comentarios generales:

- El modelo lineal es un modelo **ASOCIATIVO**, por lo que es indispensable conocer el fenómeno estudiado si se quiere hacer una interpretación causal.
- La linealidad de un modelo viene dada por los **COEFICIENTES**, y no por los regresores.
- Siempre hay que graficar los datos para examinar su comportamiento, detectando así puntos extremos, linealidad y relación entre las variables
- Un modelo lineal hace un buen ajuste de los datos si la parte explicada por $X\hat{\beta}$ es "mayor"que la explicada por los residuos. Es por esto mismo, que la calidad de un modelo se parte generalmente estudiando el tamaño y comportamiento de los residuos.
- Muchos de los análisis de sensibilidad, y diagnósticos en los modelos de regresión parten del análisis de la matriz (X^tX), que es una medida de las interacciones entre las variables.
- Es útil examinar la correlación entre todas las variables, regresores y variable dependiente, para detectar problemas de multicolinealidad, y la relación con la variable dependiente (ojalá posteriormente al análisis gráfico).
- Los coeficientes (magnitud y signo) del modelo se interpretan *ceteris paribus*, es decir, manteniendo lo demás constante.
- En un modelo estimado por MCO, los residuos $\hat{\varepsilon}$ son siempre ortogonales a X^t (Por construcción)
- En un modelo con constante, la suma de los residuos es siempre 0.
- Graficar los residuos ayuda a verificar los supuestos del modelo. Si los supuestos son correctos, no debería encontrarse un patrón en los residuos versus las observaciones.

SUERTE. EL PROFESOR