Econometría I (EC402)
Clase #17 - Regresión Lineal Multiple

Prof. Andrés M. Castaño

Ingeniería Comercial
Universidad Católica del Norte
Lunes 04 de Noviembre de 2013

Comentarios generales...

- A veces la variable que intentamos explicar está afectada por más de una variable.

- Modelo de tres variables. 1 dependiente y dos independientes:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \mu_i$$

.

- Se mantienen todos los supuestos.

- Los coeficientes β_2 y β_3 son denominados coeficientes de regresión parcial.

Interpretación de la ecuación de Regresión Múltiple

- $$E(Y_i \mid X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$
- Los efectos de cada parámetro se interpretan ceteris paribus las otras variables.
- Se mantienen todos los supuestos.
- Los coeficientes β_2 y β_3 son denominados coeficientes de regresión parcial.
- Estimadores MCO Y MV

Estimador MCO

- Se diferencia $Min \sum \hat{\mu}_i^2 = \sum (Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i})^2$, se obtienen las ecuaciones normales:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2$$

Estimador MCO

- Resolviendo el sistema de ecuaciones se obtiene:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

Varianzas y errores estándar de los estimadores

- $$Var(\hat{\beta}_1) = (\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2\bar{X}_3 \sum x_{2i}x_{3i}}{\sum_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}) * \sigma^2$$

- $$Var(\hat{\beta}_2) = (\frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}) * \sigma^2$$

- $$Var(\hat{\beta}_3) = (\frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}) * \sigma^2$$

- $$Cov(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{(1 - r_{23}^2)\sqrt{x_{2i}^2}\sqrt{x_{3i}^2}}$$

Coeficiente de determinación múltiple R^2 y el coeficiente de correlación múltiple

- $$\hat{\sigma}^2 = \frac{\sum \mu_i^2}{n-3}$$

- Para obtener el coeficiente de determinación se sigue el siguiente procedimiento:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\mu}_i$$

$$= \hat{Y}_i + \hat{\mu}_i$$

Aplicando las desviaciones a partir de sus medias obtenemos:

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\mu}_i$$

$$y_i = \hat{y}_i + \hat{\mu}_i$$

Si se elevan al cuadrado ambos lados y se suman los valores muéstrales:

$$\sum y_i^2 = \hat{y}_i^2 + \hat{\mu}_i^2 + 2\sum \hat{y}_i \hat{\mu}_i$$

$$\sum y_i^2 = \hat{y}_i^2 + \hat{\mu}_i^2$$

Coeficiente de determinación múltiple R^2 y el coeficiente de correlación múltiple

- Si substituimos el valor de los residuos, obtenemos:

$$\sum y_i^2 = \hat{y}_i^2 + \hat{\mu}_i^2$$

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2}$$

$$R^2 = 1 - \frac{SRC}{STC} = 1 - \frac{\sum \hat{\mu}_i^2}{\sum y_i^2}$$

- Relación entre el R^2 y la varianza de un coeficiente:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}(\frac{1}{1 - R_j^2})$$

Ejemplo: mortalidad infantil respecto al PIB per cápita y la tasa de alfabetización en mujeres. No estandarizada vs estandarizada

$$\mathrm{MI}_i = 263.6416 - 0.0056\,\mathrm{PIBPC}_i - 2.2316\,\mathrm{TAM}_i$$
$$\mathrm{ee} = (11.5932) \quad (0.0019) \quad (0.2099) \qquad R^2 = 0.7077$$
$$\bar{R}^2 = 0.6981$$

$$\mathrm{MI}^* = -0.2026\,\mathrm{PIBPC}_i^* - 0.7639\,\mathrm{TAM}_i^*$$
$$\mathrm{ee} = (0.0713) \quad (0.0713) \qquad r^2 = 0.7077$$

Elección de una forma funcional

- La teoría subyacente sugiere una forma funcional.
- Comparar los coeficientes de pendiente con las elasticidades.
- Los coeficientes de la forma funcional escogida deben satisfacer determinadas expectativas a-priori (signos esperados)
- Comparar signos, significancia y ajuste global.
- Es preferible escoger un modelo con menor ajuste global y coeficientes con signos adecuados y significativos, que el caso contrario.

La regresión simple en el contexto de la regresión múltiple: introducción al sesgo de especificación

- Suponga que el modelo anterior es el verdadero, y que en su lugar usted estimó:

$$MI_i = \alpha_1 + \alpha_2 PIBPC_i + \mu_{1i}$$

- Es $\hat{\alpha}_2$ un estimador insesgado de β_2

$$\begin{aligned} \text{MI}_i &= 263.8635 \; - \; 2.3905\text{TAM}_i \\ \text{ee} &= (21.2249) \quad (0.2133) \qquad\qquad r^2 = 0.6696 \end{aligned}$$

El valor del coeficiente de PIBPC es 0.0114 (error Guajarati), el del parámetro autónomo tampoco es.

La regresión simple en el contexto de la regresión múltiple: introducción al sesgo de especificación

$$\mathrm{MI}_i = 263.8635 - 2.3905\,\mathrm{TAM}_i$$
$$\mathrm{ee} = (21.2249) \quad (0.2133) \qquad r^2 = 0.6696$$

R^2 tradicional vs R^2 ajustado

- $$\bar{R}^2 = 1 - \frac{\frac{\sum \hat{\mu}_i^2}{n-k}}{\frac{y_i^2}{n-1}}$$
- Se debe tener mucho cuidado al comparar dos modelos con la misma variable dependiente via el coeficiente de determinación tradicional. Para estos fines es mejor utilizar el ajustado.
- $$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$
- No decida si su especificación está bien o mal basado en el R cuadrado. Recuerde que el objetivo del análisis de regresión no es obtener un R más alto per se, sino obtener estimadores confiables de los verdaderos coeficientes de regresión. También es importante destacar la importancia de la relevancia teórica.