Maestría en Ciencia y Análisis de Datos- Universidad Mayor de San Andrés

Modelos lineales y modelos lineales generalizados

Rolando Gonzales Martinez, PhD

Fellow postdoctoral Marie Skłodowska-Curie

Universidad de Groningen (Países Bajos)

Investigador (researcher)

Iniciativa de Pobreza y Desarrollo Humano de la Universidad de Oxford (UK)

Contenido del curso

(2) Diagnóstico y Evaluación de Modelos Lineales

- Diagnóstico de residuos.
- Multicolinealidad.
- Otras patologías que pueden surgir en los modelos lineales.
- Métricas de evaluación de ajuste.
- Laboratorio: Evaluación de modelos lineales

Teorema Gauss-Markov

En un modelo lineal de regresión múltiple (MLRM), si se cumplen las siguientes condiciones:

$$y = X\beta + \epsilon$$
$$E(\epsilon) = 0$$
$$Var(\epsilon) = \sigma^2 I$$
$$Cov(\epsilon_i, \epsilon_j) = 0$$
$$Cov(X, \epsilon) = 0$$

1. **Linealidad**: El modelo es lineal en los parámetros
2. Los errores tienen media cero
3. **Homocedasticidad**: Los errores tienen varianza constante
4. **No autocorrelación**: Los errores no están correlacionados entre sí
5. **Independencia entre X y ϵ**: Los errores no están correlacionados con las variables explicativas

Entonces, el estimador de mínimos cuadrados ordinarios es el mejor estimador lineal insesgado (MELI, BLUE). Esto significa que:

- **Insesgado**: $E(\hat{\beta}) = \beta$
- **Varianza mínima**: Entre todos los estimadores lineales insesgados, el estimador MCO tiene la varianza más baja.

Normalidad de los residuos

- La normalidad de los residuos no es parte del Teorema Gauss-Markov, por lo que no es necesaria para que los estimadores MCO/OLS sean MELI/BLUE.
- Sin embargo la normalidad de residuos es relevante para la docimasia de hipótesis y para la construcción de intervalos de confianza precisos.
- Si los residuos son normales los estimadores MCO coinciden con los estimadores máximo verosímiles y son eficientes en el sentido de tener varianza mínima.

Diagnóstico de residuos

Diferentes pruebas estadísticas pueden aplicarse a los residuos para testear normalidad y los supuestos del teorema Gauss-Markov respecto a los residuos de la regresión:

	Test	Hipótesis nula
Linealidad	Harvey-Collier	Linealidad
Normalidad	Jarque-Bera	Normalidad
	D’Agostino,and Pearson	Normalidad
Homocedasticidad (heteroscedasticidad)	Breusch–Pagan	Homocedasticidad
	Goldfeld-Quandt	Homocedasticidad
	White	Homocedasticidad
	Engle	Homocedasticidad

Diagnóstico de residuos

	Test	Hipótesis nula
Correlación en los residuos (autocorrelación, correlación serial de primer orden y de órdenes superiores)	Durbin-Watson	No correlación (+/-) de primer orden
	Breusch-Godfrey	No (auto)correlación (+/-) de primer orden y órdenes superiores
	Ljung-Box	No (auto)correlación (+/-) de primer orden y órdenes superiores
Independencia entre X y ϵ (exogeneidad)	Hausman	Los errores no están correlacionados con los regresores

Test Jarque-Bera de normalidad

- El estadígrafo de Jarque-Bera se distribuye asintóticamente como una distribución chi-cuadrado con dos grados de libertad
- La hipótesis nula es una hipótesis conjunta de que la asimetría y el exceso de curtosis son nulos (asimetría = 0 y curtosis = 3).

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$$

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}$$

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

Test omnibus D'Agostino-Pearson de normalidad

$$z_1 = \sqrt{\frac{(n+1)(n+3)}{6(n-2)}} \cdot S$$

$$z_2 = \frac{(n-2)}{24} \cdot \left(\frac{(n+3)(n+5)}{(n-1)(n+1)} \right) \cdot (K-3)$$

$$DP = z_1^2 + z_2^2$$

$$DP \sim \chi^2_{(2)}$$

Linealidad: test Harvey-Collier

$$y_j = x_j'\beta + u_j, \qquad j = 1, \ldots, n,$$

$$\tilde{u}_j = \frac{y_j - x_j' b_{j-1}}{(1 + x_j'(X_{j-1}' X_{j-1})^{-1} x_j)^{1/2}}, \qquad j = k+1, \ldots, n,$$

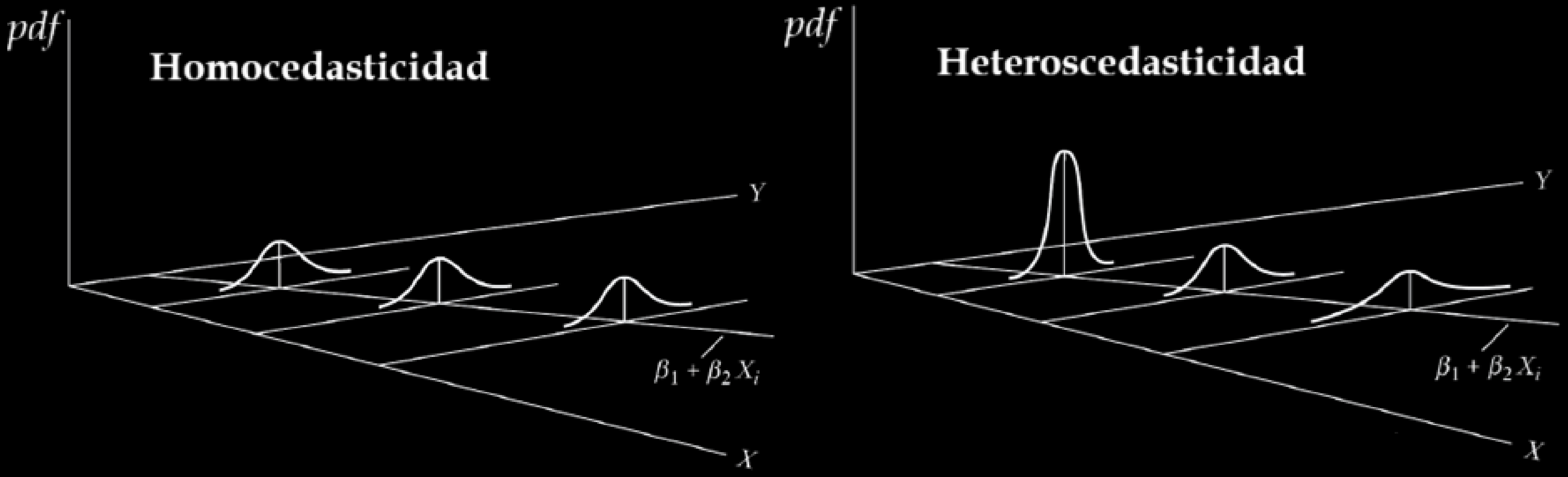
$$\psi = \left[(n-k-1)^{-1} \sum_{j=k+1}^{n} (\tilde{u}_j - \bar{\tilde{u}})^2 \right]^{-1/2} (n-k)^{-1/2} \sum_{j=k+1}^{n} \tilde{u}_j$$

$$\psi \sim t(n - k - 1)$$

Harvey, A. C., & Collier, P. (1977). Testing for functional misspecification in regression analysis. Journal of Econometrics, 6(1), 103-119.

Homocedasticidad y Heterocedasticidad

$$Y_i = \beta_1 + \beta_2 X_i$$



Afecta eficiencia y validez de la inferencia

Homocedasticidad: Test Breusch–Pagan

$$y_i = X_i\beta + \varepsilon_i, \quad i = 1, \ldots, n$$

$$g_i = \hat{\varepsilon}_i^2/\hat{\sigma}^2, \quad \hat{\sigma}^2 = \sum \hat{\varepsilon}_i^2/n$$

$$g_i = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k + v$$

$$LM = \frac{n \cdot R^2}{2}$$

$$F = \frac{\left(SSR_{\text{reducido}} - SSR_{\text{completo}}\right)/k}{SSR_{\text{completo}}/(n-k-1)}$$

Homocedasticidad: Test Goldfeld-Quandt

Ordenar los datos y dividir en 2 grupos:

Para el primer grupo:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \epsilon_i$$

Para el segundo grupo:

$$y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj} + \epsilon_j$$

$$RSS_1 = \sum_{i \in \text{Grupo 1}} \hat{\epsilon}_i^2$$
$$RSS_2 = \sum_{j \in \text{Grupo 2}} \hat{\epsilon}_j^2$$

$$F = \frac{\frac{RSS_2}{df_2}}{\frac{RSS_1}{df_1}}$$

Homocedasticidad: Test de White

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon$$

$$\hat{\epsilon}^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k + \alpha_{k+1} X_1^2 + \alpha_{k+2} X_2^2 + \ldots + \alpha_{2k} X_k^2 + \alpha_{2k+1} X_1 X_2 + \ldots + v$$

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

$$\hat{\epsilon}^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1^2 + \alpha_4 X_2^2 + \alpha_5 X_1 X_2 + v$$

$$LM = n \cdot R^2 \sim \chi^2_{(d)}$$

Homocedasticidad: Test ARCH de Engle

$$Y_t = X'_t b_0 + \varepsilon_t \qquad t = 1, \ldots, T$$

$$\varepsilon_t = \xi_t h_t^{1/2}$$

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2$$

$$e_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + v_t$$

$$\mathrm{H_0}\colon \alpha_j = 0 \text{ for } j = 1, \ldots, q$$

$$\mathrm{H_A}\colon \alpha_j \geqslant 0,\ j = 1, \ldots, q$$

Heterocesdasticidad: soluciones

- Transformaciones de variables, WLS
- En el contexto de series de tiempo: modelizar la heteroscedasticidad dinámica
- Matriz varianza-covarianza robusta:
    - HC0: Matriz varianza-covarianza robusta clásica Huber-White. No ajusta los grados de libertad. Adecuada para muestras grandes.
    - HC1: Ajuste de MacKinnon y White (1985) para muestras pequeñas.
    - HC2, HC3: Ajustes alternativos para muestras más pequeñas.

Heterocesdasticidad: WLS

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \epsilon_i^2 = \min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} w_i \epsilon_i^2 = \min_{\beta_0,\beta_1} \sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2$$

$$W = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix}$$

$$y = X\beta + \epsilon$$

$$S(\beta) = (y - X\beta)^T W (y - X\beta)$$

$$\frac{\partial S(\beta)}{\partial \beta} = -2X^T W (y - X\beta) = 0$$

$$\beta = (X^T W X)^{-1} X^T W y$$

Heterocesdasticidad: matrices robustas

$$\mathrm{HC0} = X'X^{-1}X'\mathrm{diag}(u_i^2)XX'X^{-1}$$

$$\mathrm{HC1} = \left(\frac{n}{n-k}\right)\mathrm{HC0}$$

$$\mathrm{HC2} = X'X^{-1}X'\mathrm{diag}\left(\frac{u_i^2}{1-h_i}\right)XX'X^{-1}$$

$$H = X(X'X)^{-1}X'$$

$$\mathrm{HC3} = X'X^{-1}X'\mathrm{diag}\left(\frac{u_i^2}{(1-h_i)^2}\right)XX'X^{-1}$$

Ejemplo: matriz Huber-White (estimador sandwich)

$$X = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{pmatrix} \qquad \hat{\epsilon} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$$

$$\hat{D} = \mathrm{diag}(\hat{\epsilon}_1^2, \hat{\epsilon}_2^2, \hat{\epsilon}_3^2) = \mathrm{diag}(4, 1, 1)$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 \\ x_{11} & x_{21} & x_{31} \end{pmatrix} \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{pmatrix} = \begin{pmatrix} 3 & \sum x_{i1} \\ \sum x_{i1} & \sum x_{i1}^2 \end{pmatrix}$$

$$X^T \hat{D} X = \begin{pmatrix} 1 & 1 & 1 \\ x_{11} & x_{21} & x_{31} \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{pmatrix} = \begin{pmatrix} 4 & 4x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{pmatrix} \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{pmatrix}$$

$$\hat{\Sigma}_{\text{robusta}} = (X^T X)^{-1} (X^T \hat{D} X)(X^T X)^{-1}$$

Heterocesdasticidad: modelos de la familia (G)ARCH

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

$$Y_t = \nabla^d X_t$$

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

$$\nabla^d X_t = \phi_1 \nabla^d X_{t-1} + \phi_2 \nabla^d X_{t-2} + \cdots +$$
$$\phi_p \nabla^d X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_q \epsilon_{t-q}^2$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

Correlación de los residuos: Estadígrafo Durbin-Watson

$$e_t = \rho e_{t-1} + \nu_t$$

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

DW varía entre 0 y 4:

- DW = 2 indica que no hay autocorrelación en los residuos.
- DW < 2 sugiere autocorrelación positiva.
- DW > 2 sugiere autocorrelación negativa.

Valores Críticos dL y dU:

- Si DW < dL: hay evidencia de autocorrelación positiva.
- Si dL ≤ DW ≤ dU: la prueba es inconclusa.
- Si DW > dU: no hay evidencia de autocorrelación positiva.
- Para la autocorrelación negativa, se compara DW con 4 - dU y 4 - dL.

Correlación de los residuos: Estadígrafo Durbin-Watson

- El estadígrafo DW es sesgado (subestima autocorrelacion) en el contexto de modelos AR(I)MA
- El estadígrafo H se distribuye asintóticamente como una distribución gaussiana estándar

$$e_t = \rho e_{t-1} + \nu_t$$

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

$$H = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{n}{1 - n\rho^2}}$$

Correlación de los residuos: test Breusch-Godfrey

Correlación de primer orden (p=1) y de ordenes superiores (p > 1):

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \epsilon_t$$

$$\hat{e}_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \cdots + \alpha_k x_{kt} + \rho_1 \hat{e}_{t-1} + \cdots + \rho_p \hat{e}_{t-p}$$

$$LM = n\, R^2$$

$$nR^2 \sim \chi^2_p$$

Correlación de los residuos: test Breusch-Godfrey

Correlación de primer orden (p=1) y de ordenes superiores (p > 1) hasta el rezago *m*:

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{n}(e_t - \bar{e})(e_{t-l} - \bar{e})}{\sum_{t=1}^{n}(e_t - \bar{e})^2}$$

$$Q = n(n+2)\sum_{l=1}^{m}\frac{\hat{\rho}_l^2}{n-l}$$

$$Q \sim \chi^2_{(m)}$$

Exogeneidad: Test de Hausman

$$y = X\beta + \epsilon$$

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y \qquad \mathrm{Var}(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}$$

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y \qquad \mathrm{Var}(\hat{\beta}_{IV}) = \sigma^2(X'P_Z X)^{-1}$$

$$H = (\hat{\beta}_{IV} - \hat{\beta}_{OLS})' \left[\mathrm{Var}(\hat{\beta}_{IV}) - \mathrm{Var}(\hat{\beta}_{OLS})\right]^{-1} (\hat{\beta}_{IV} - \hat{\beta}_{OLS})$$

$$H \sim \chi^2_{(k)}$$

- Nula: el estimador OLS (consistente) es preferible al estimador IV (porque se cumple el supuesto de exogeneidad
- Alternativa: el estimador IV es preferible porque el estimador OLS es inconsistente (porque no existe exogeneidad)

Multicolinealidad

- La multicolinealidad se refiere a la situación en la que dos o más variables explicativas en un modelo de regresión están altamente correlacionadas entre sí.
- El determinante de la matriz $X'X$ proporciona información sobre la dependencia lineal entre las variables explicativas: si $\det(X'X)$ tiende a cero, existe evidencia de multicolinealidad, y si $\det(X'X) = 0$ la multicolinealidad es perfecta.
- El número de condición de Belsley (B), el cociente entre el mayor y el menor valor singular de X, también ofrece información sobre el grado de multicolinealidad, siendo ésta severa cuando $B > 30$. Los valores singulares son las raíces cuadradas no negativas de los autovalores de la matriz $X'X$.

Multicolinealidad

Una manera práctica de evaluar y lidiar con la multicolinealidad es emplear los factores de inflación de varianza (variance inflation factors, VIF):

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2}$$

$$X_1 = \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_k X_k + c_0 + e$$

- VIF ≥ 10: multicolinealidad severa
- VIF ≥ 5: multicolinealidad elevada
- VIF < 5: multicolinealidad débil a moderada