

DETECCIÓN DE MULTICOLINEALIDAD MEDIANTE MÉTODOS GRÁFICOS

R. Salmerón (romansg@ugr.es) **C.B. García** (cbgarcia@ugr.es)

Dpto. de Métodos Cuantitativos para la Economía y la Empresa
Universidad de Granada

**II Congreso R para Empresa, Docencia e Investigación
Pastrana (Guadalajara), 10 y 11 de noviembre de 2022**

Índice

- 1 Introducción
- 2 Multicolinealidad: tipos y detección
- 3 Ejemplos
- 4 Conclusiones
- 5 Bibliografía
- 6 Agradecimientos

Introducción

Econometría

Econometría es una rama de la Economía que proporciona una base para refinar o refutar conocimiento teórico y conseguir signos y magnitudes de las relaciones de variables que se desean analizar.

Dado el modelo lineal general con n observaciones y p variables independientes:

$$\mathbf{Y} = \beta_1 \cdot \mathbf{X}_1 + \beta_2 \cdot \mathbf{X}_2 + \cdots + \beta_i \cdot \mathbf{X}_i + \cdots + \beta_p \cdot \mathbf{X}_p + \mathbf{u} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

el objetivo anterior se consigue a partir de:

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}_i} = \beta_i.$$

Aplicando Mínimos Cuadrados Ordinarios (MCO), se obtiene la estimación (signo y magnitud) de cada coeficiente mediante:

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

Si las variables que forman \mathbf{X} son linealmente dependientes, entonces no existe $(\mathbf{X}^t \mathbf{X})^{-1}$ (multicolinealidad perfecta).

Introducción

¿Qué ocurre si son **casi** linealmente dependientes? Existe $(\mathbf{X}^t\mathbf{X})^{-1}$, pero...

Ejemplo de Wooldridge (2000) [1]

Variable	Estimación	Desviación típica
Constante	0.4404	0.09556
Tipos de interés a 3 meses	1.00569	0.01343
$R^2 = 0.9787, \hat{\sigma}^2 = 0.2025, F_{2,122} = 5611$		

Estimación por MCO de los tipos de interés a 12 meses en función de los tipos de interés a 3 meses

Correlación entre los tipos de interés a 3 y 12 meses: 0.9893021.

Variable	Estimación	Desviación típica
Constante	0.22471	0.0397
Tipos de interés a 3 meses	-0.62891	0.06582
Tipos de interés a 6 meses	1.59334	0.06394
$R^2 = 0.9965, \hat{\sigma}^2 = 0.03330625, F_{2,122} = 17371.66$		

Estimación por MCO de los tipos de interés a 12 meses en función de los tipos de interés a 3 y 6 meses

Correlación entre los tipos de interés a 3 y 6 meses: 0.9965734 (*ceteris paribus*).

Tipos de Multicolinealidad y Medidas de Detección

Tipos de Multicolinealidad (según Marquardt y Snee (1975) [2])

Multicolinealidad no esencial: relación lineal de las variables independientes con la constante (es sabido que se solventa centrando las variables).

Multicolinealidad esencial: relación lineal entre las variables independientes (excluida la constante).

Medidas de Detección (más usuales)

Multicolinealidad no esencial: coeficiente de variación, índice de Stewart y número de condición.

Multicolinealidad esencial: matriz de correlaciones simples y su determinante, factor de inflación de la varianza, índice de Stewart y número de condición.

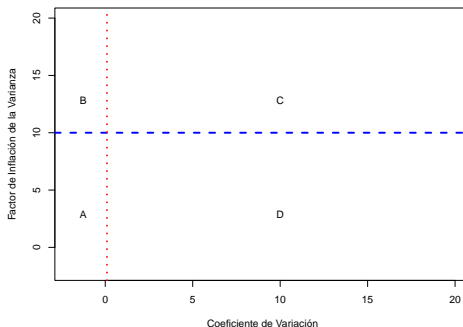
Usarlas de forma conjunta y no aislada

Detección Gráfica

Reglas de Decisión

Coefficiente de Variación: inferior a 0.1002506 indica multicolinealidad aproximada no-esencial preocupante (Salmerón et al. 2019 [3]). **A**

Factor de Inflación de la Varianza: superior a 10 indica multicolinealidad aproximada esencial preocupante (Marquardt y Snee (1975) [2]). **C**



Ejemplo 1

Simulación

set.seed(2022)

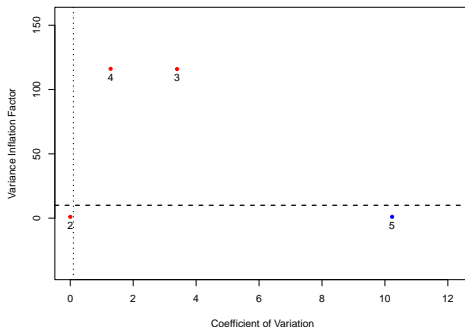
$x_1 = \text{rep}(1, \text{obs})$

$x_2 = \text{rnorm}(100, 5, 0.01) \rightarrow$ muy poca variabilidad: relacionada con la constante

$x_3 = \text{rnorm}(100, 5, 10) \rightarrow$ mucha variabilidad: no relacionada con la constante

$x_4 = x_3 + \text{rnorm}(100, 5, 1) \rightarrow$ relacionada con la tercera variable

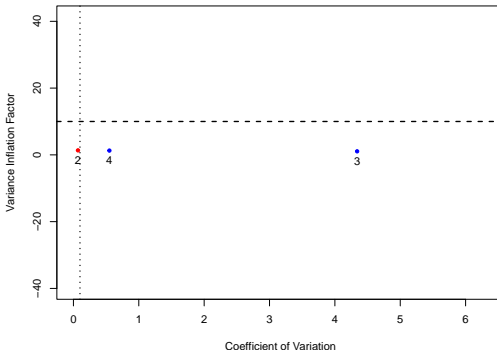
$x_5 = \text{rnorm}(100, -1, 30) \rightarrow$ independiente del resto



Ejemplo 2

Salmerón, R., Rodríguez, A. y García, C.B. (2019) [3]

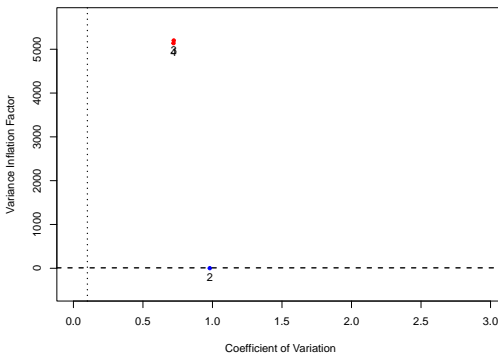
Consideran un modelo financiero en el que el euribor (**E**, %) es analizado a partir del índice de precios de consumo armonizado (**HICP**, %), la balanza de pagos a la cuenta corriente neta (**BC**, millones de euros) y el déficit público a cuentas no financieras netas (**GD**, millones de euros).



Ejemplo 3

Salmerón, R., Rodríguez, A., García, C.B. y García, J. (2020) [4]

En este trabajo se analiza la relación entre el número de empleados (**NE**) de 15 compañías españolas en función de los activos fijos (**FA**, en euros), el ingreso operativo (**OI**, en euros) y las ventas (**S**, en euros).



Principales resultados

Conclusiones

- Mediante el gráfico de dispersión de los coeficientes de variación de las variables independientes frente a sus factores de inflación de la varianza se puede determinar fácilmente
 - 1 si la multicolinealidad aproximada existente es preocupante,
 - 2 en caso de serlo, de qué tipo se trata (esencial o no-esencial) y
 - 3 en este caso, qué variables la provocan.
- En consecuencia, se puede determinar qué técnica de estimación es más adecuada: centrar variables, cresta, alzado, LASSO, etc.

```
install.packages("rvif", repos="http://R-Forge.R-project.org")
```

Bibliografía (por orden de aparición)



Wooldridge, J.M. (2000). *Introductory Econometrics: A modern approach*. South-Western, Cengage Learning.



Marquardt, D. W. and R. Snee (1975). "Ridge regression in practice". *The American Statistician* 29 (1), pp. 3–20.



Salmerón, R., Rodríguez, A. y García, C.B. (2019). "Diagnosis and quantification of the non-essential collinearity". *Computational Statistics*, 35, pp. 647–666.



Salmerón, R., Rodríguez, A., García, C.B. y García, J. (2020). "The VIF and MSE in ridge regression". *Mathematics*, 8 (4), pp. 605.

Agradecimientos

Este trabajo ha sido financiado por:

- el proyecto PP2019-EI-02 de la Universidad de Granada titulado “Redefinición del factor de inflación de la varianza y de sus umbrales” y
- por el proyecto FEDER A-SEJ-496-UGR20 de la Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades titulado “Disrupción en el problema de multicolinealidad: una nueva visión para su diagnóstico y tratamiento”.

DETECCIÓN DE MULTICOLINEALIDAD MEDIANTE MÉTODOS GRÁFICOS

R. Salmerón (romansg@ugr.es) **C.B. García** (cbgarcia@ugr.es)

Dpto. de Métodos Cuantitativos para la Economía y la Empresa
Universidad de Granada

**II Congreso R para Empresa, Docencia e Investigación
Pastrana (Guadalajara), 10 y 11 de noviembre de 2022**