Unidad 3a. Multicolinealidad

Prof. Dra. Eva Romero Ramos

Universidad Europea
LAUREATE INTERNATIONAL UNIVERSITIES

Contenidos

- 1. Multicolinealidad: concepto, causas, consecuencias.
- 2. Ejemplo de análisis de multicolinealidad.
- 3. Medidas de corrección de la multicolinealidad.

1. Multicolinealidad: concepto, causas, consecuencias.

Multicolinealidad

Para analizar adecuadamente los resultados de un modelo de regresión lineal múltiple, es importante estudiar **el grado de relación lineal** existente entre las variables explicativas que lo componen.

Encontramos en este sentido 3 posibles situaciones:

- **Ortogonalidad.**
- **Multicolinealidad perfecta**.
- **Multicolinealidad imperfecta**.

Ortogonalidad

La ortogonalidad surge cuando la relación lineal entre los regresores incluidos en el modelo es nula.

Implica por tanto que no existen relaciones lineales entre los regresores del modelo.

La ortogonalidad no supone el incumplimiento de ninguna de las hipótesis de nuestro modelo. De hecho una de las hipótesis del modelo es que las variables explicativas sean independientes entre sí.

Es difícil encontrar variables en el mundo real que presenten ortogonalidad.

Multicolinealidad Perfecta

Un modelo de regresión lineal presentará multicolinealidad perfecta cuando exista alguna relación lineal exacta entre algunos de los regresores que lo componen.

Considerando la ecuación general del modelo de regresión lineal múltiple:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki} + \varepsilon_i$$

Diremos que el modelo presenta multicolinealidad perfecta si sucede por ejemplo que:

$$X_{4i} = X_{1i} + 2\mathrm{X}_{2\mathrm{i}} \text{ o } X_{3i} = 5 \cdot X_{2i} - 1$$

En ambos casos podemos obtener de forma exacta las observaciones de una variable regresora a partir de una combinación lineal de otra u otras.

Multicolinealidad Perfecta

- La multicolinealidad perfecta supone el incumplimiento de una de las hipótesis en la que se basa el modelo de regresión lineal clásico, que es **la inexistencia de combinaciones lineales exactas** entre los regresores del modelo.
- Esta hipótesis garantiza la existencia de la matriz $(X^tX)^{-1}$, de modo que si no se cumple no se podrán calcular los estimadores de los parámetros del modelo, por no existir el determinante de X^tX.
- En esta situación, Gretl ignora una de las variables que forman el problema de forma automática y estima el modelo sin ella.

Multicolinealidad Perfecta

- La multicolinealidad perfecta surge cuando varias variables explicativas contienen exactamente la misma información, de modo que el modo más sencillo de resolverla es eliminar alguna de las variables explicativas que la generan.
- En la práctica los modelos no suelen presentar multicolinealidad perfecta, ya que las relaciones entre las distintas variables económicas no suele ser exacta.
- El problema suele surgir cuando generamos variables de forma artificial como combinación de otras (variables dummies mal construidas u otro tipo de transformaciones).

Multicolinealidad Imperfecta

Decimos que un modelo de regresión múltiple presenta multicolinealidad imperfecta cuando existen relaciones lineales fuertes entre algunas de sus variables explicativas.

Ejemplo.- Si tratamos de explicar el número de tarjetas de crédito que tienen las familias y utilizamos como variables explicativas:

- El número de miembros de la unidad familiar
- La renta
- El número de vehículos

Fácilmente nos encontraremos con un problema de multicolinealidad fuerte, pues es de esperar que en nuestros datos existan altas correlaciones entre variables como la renta y el número de vehículos, o el número de vehículos y el número de miembros de la familia.

Consecuencias de la Multicolinealidad Imperfecta

- La multicolinealidad imperfecta es un problema muy común en los modelos econométricos que incumple la hipótesis básica de independencia entre las variables explicativas.

- Si un modelo presenta multicolinealidad, **los estimadores de mínimos cuadrados ordinarios siguen siendo los mejores** estimadores que pueden obtenerse y cumpliendo muchas de las propiedades deseadas para un estimador como la **insesgadez** y la **eficiencia**.

Consecuencias de la Multicolinealidad Imperfecta

De la multicolinealidad imperfecta se derivan las siguientes consecuencias:

1) Errores estándar de estimación elevados o varianzas grandes en los estimadores.

2) Inestabilidad de los estimadores ante pequeñas variaciones muestrales. Este problema es consecuencia directa del anterior, ya que si las varianzas de los estimadores son grandes, los estimadores resultan más inestables.

3) Dificultad para interpretar los coeficientes y por tanto sus estimaciones.

   Los coeficientes de regresión (β_i) se interpretan como el cambio que se produce en la variable dependiente (y) ante variaciones de la variable independiente (x_i) de una unidad, siempre que el resto de las variables explicativas permanezca constante.

   Cuando existe multicolinealidad imperfecta es imposible suponer que el resto de las variables permanecen constantes cuando una cambia, ya que si están altamente relacionadas cambios en una implicarán cambios en el resto. Por este motivo los parámetros pierden significado.

2. Ejemplo completo de análisis de multicolinealidad.

Ejemplo completo de análisis de multicolinealidad

El fichero ahorro_familias.gdt recoge datos del ahorro anual de una muestra de familias y otras variables relacionadas, que son:

- Renta: Renta neta anual de la unidad familiar, calculada como la suma de las rentas netas de todos los miembros de la unidad familiar en activo.
- Tamanno: Número total de miembros que conviven en el domicilio familiar.
- GastoViv: Gasto anual en vivienda de la unidad familiar, ya sea como pago de un préstamo para la adquisición de la misma o como alquiler.

Se generará un modelo para explicar el ahorro familiar a partir de la renta familiar, del tamaño y del gasto en vivienda y se analizará si el modelo presenta problemas de multicolinealidad.

Ejemplo completo de análisis de multicolinealidad

gretl: modelo 3

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

```
Modelo 3: MCO, usando las observaciones 1-91
Variable dependiente: Ahorro

             Coeficiente   Desv. Típica  Estadístico t   valor p
  ---------------------------------------------------------------
  const     -10068.8       1201.07          -8.383      8.21e-013 ***
  Renta          0.688736     0.0659644     10.44       5.12e-017 ***
  tamanno       34.8387      92.3558         0.3772     0.7069
  GastoViv       0.395846     0.322199       1.229      0.2225

Media de la vble. dep.  14801.21   D.T. de la vble. dep.   6960.315
Suma de cuad. residuos  42185562   D.T. de la regresión    696.3415
R-cuadrado              0.990325   R-cuadrado corregido    0.989991
F(3, 87)                2968.329   Valor p (de F)          1.78e-87
Log-verosimilitud      -722.7496   Criterio de Akaike      1453.499
Criterio de Schwarz     1463.543   Crit. de Hannan-Quinn   1457.551

Sin considerar la constante, el valor p más alto fue el de la variable 3 (tamanno)
```

Ejemplo completo de análisis de multicolinealidad

Para estudiar la multicolinealidad tenemos las siguiente herramientas:

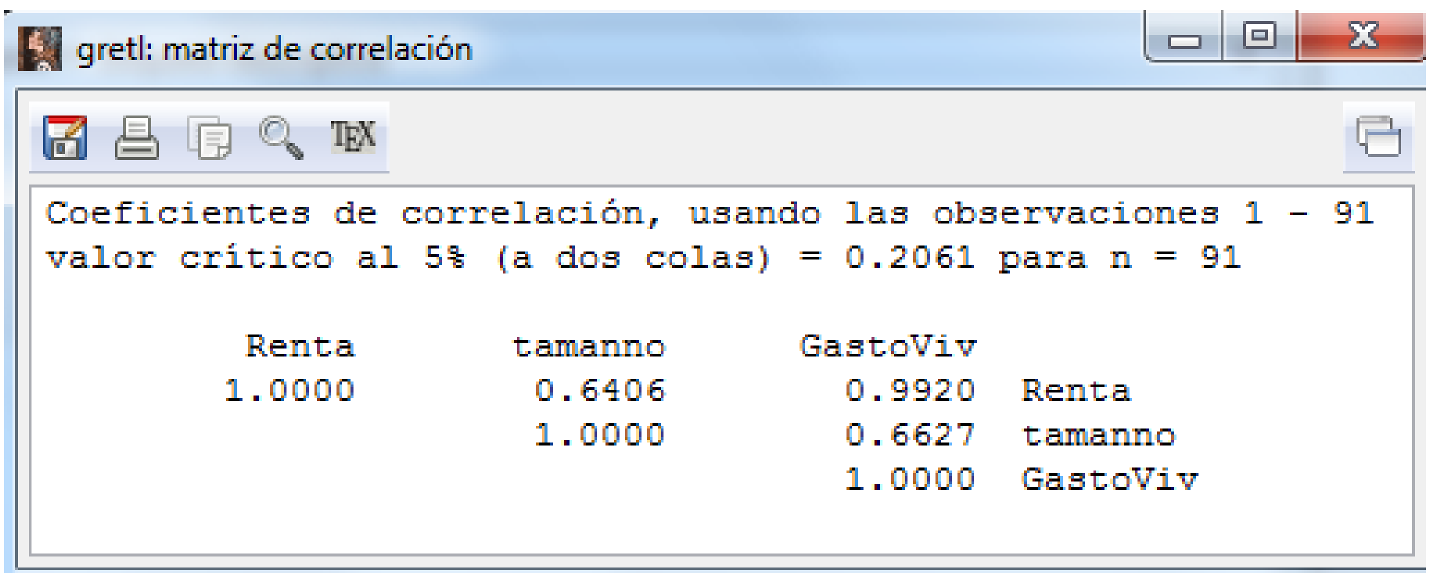
- Observar la matriz de correlaciones entre las variables explicativas y su determinante.

- Analizar la significación individual y conjunta de los regresores.

Matriz de correlaciones de las variable explicativas

- La matriz de correlaciones entre las variables explicativas nos permite observar el grado de relación lineal existente entre cada par de regresores.

- Cuando alguno de los coeficientes de correlación es elevado (próximo a ±1) tendremos un indicio de la existencia de multicolinealidad imperfecta en el modelo.


gretl: matriz de correlación

Coeficientes de correlación, usando las observaciones 1 - 91
valor crítico al 5% (a dos colas) = 0.2061 para n = 91

Renta	tamanno	GastoViv	
1.0000	0.6406	0.9920	Renta
	1.0000	0.6627	tamanno
		1.0000	GastoViv

Determinante de la matriz de correlaciones de las variable explicativas

- En ocasiones, aunque los coeficientes de correlación lineal no presentan grandes correlaciones, existe un problema de multicolinealidad que se debe a la relación entre más de dos variables.

- En estos casos es necesario obtener el determinante de la matriz de correlaciones para detectarla la presencia de multicolinealidad.

- El determinante de la matriz de correlaciones *mide la relación de todas las variables explicativas en conjunto*.

- Cuando la relación entre las variables explicativas es muy alta, el determinante de la matriz de correlaciones toma un valor cercano a cero. En el caso de relaciones lineales perfectas, el determinante tomará valor cero.

Determinante de la matriz de correlaciones de las variable explicativas

Para calcular el determinante de la matriz de correlaciones en Gretl tendremos que hacer uso de la consola de Gretl. Para abrirla tenemos la secuencia:

Herramientas -> consola de Gretl

En la consola de Gretl podemos realizar operaciones con los objetos de nuestro archivo haciendo uso de los comandos de Gretl.

Para obtener el determinante de la matriz de correlaciones de las variables regresoras del modelo, seguiremos los siguientes pasos:

1. Generar una matriz con las variables explicativas.
2. Calcular la matriz de correlaciones de las variables explicativas.
3. Obtener el determinante.

Determinante de la matriz de correlaciones de las variable explicativas

1. Para agrupar las variables en una matriz hacemos:

   matrizvariables = {Renta, tamanno, GastoViv}

2. Una vez que tenemos las variables agrupadas en una matriz, calculamos la matriz de correlaciones:

   matrizcorrelaciones=*mcorr*(matrizvariables)

3. Finalmente obtendremos el determinante con la función det:

   determinante=*det*(matrizcorrelaciones)

Determinante de la matriz de correlaciones de las variable explicativas

- En el ejemplo que nos ocupa sobre el ahorro familiar, el determinante de la matriz de correlaciones de los regresores tiene un valor de ***0,00866012***.

- Como se trata de un valor cercano a cero podemos decir que el modelo presenta un problema de multicolinealidad.

Significatividad individual y conjunta de las variables explicativas

- En presencia de multicolinealidad imperfecta los *errores estándar* de estimación de los parámetros se hacen *anormalmente grandes*.

- En esta situación *los contrastes de significatividad individual tienden a fallar*, aceptando como variables no significativas a variables con capacidad explicativa sobre la variable dependiente.

- Para detectar esta situación tenemos dos opciones:

  - Observar si las variables que aparecen como no significativas en el modelo están altamente correlacionadas con la variable dependiente.

  - Observar si las variables que aparecen como no significativas en el modelo generan por si solas modelos significativos.

Significatividad individual y conjunta de las variables explicativas

- En nuestro modelo si tenemos en cuenta el contraste de significatividad individual diremos que las variables tamaño y gasto en vivienda no son significativas:

	Coeficiente	Desv. Típica	Estadístico t	valor p	
const	−10068.8	1201.07	−8.383	8.21e-013	***
Renta	0.688736	0.0659644	10.44	5.12e-017	***
tamanno	34.8387	92.3558	0.3772	0.7069	
GastoViv	0.395846	0.322199	1.229	0.2225	

Sin embargo, podemos observar que las correlaciones que tienen con la variable dependiente son fuertes.

gretl: matriz de correlación

Coeficientes de correlación, usando las observaciones 1 - 91
valor crítico al 5% (a dos colas) = 0.2061 para n = 91

Renta	tamanno	GastoViv	Ahorro	
1.0000	0.6406	0.9920	0.9950	Renta
	1.0000	0.6627	0.6435	tamanno
		1.0000	0.9889	GastoViv
			1.0000	Ahorro

Internal use

Significatividad individual y conjunta de las variables explicativas

gretl: modelo 1

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

```
Modelo 1: MCO, usando las observaciones 1-91
Variable dependiente: Ahorro

             Coeficiente   Desv. Típica   Estadístico t    valor p
  ---------------------------------------------------------------
  const        3346.03       1549.67          2.159        0.0335    **
  tamanno      4153.08        523.638         7.931        6.01e-012 ***

Media de la vble. dep.  14801.21   D.T. de la vble. dep.   6960.315
Suma de cuad. residuos  2.55e+09   D.T. de la regresión    5357.543
R-cuadrado              0.414103   R-cuadrado corregido    0.407520
F(1, 89)                62.90390   Valor p (de F)          6.01e-12
Log-verosimilitud      -909.4620   Criterio de Akaike      1822.924
Criterio de Schwarz     1827.946   Crit. de Hannan-Quinn   1824.950
```

Por otro lado, si plantemos modelos de regresión simple para explicar el ahorro de las familias a partir de estas variables encontraremos que estos modelos sí son significativos.

gretl: modelo 2

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

```
Modelo 2: MCO, usando las observaciones 1-91
Variable dependiente: Ahorro

             Coeficiente   Desv. Típica  Estadístico t    valor p
  --------------------------------------------------------------
  const      -21724.7       591.493          -36.73       1.47e-055 ***
  GastoViv        3.66356     0.0583118       62.83       1.63e-075 ***

Media de la vble. dep.  14801.21   D.T. de la vble. dep.   6960.315
Suma de cuad. residuos  96141900   D.T. de la regresión    1039.349
R-cuadrado              0.977950   R-cuadrado corregido    0.977702
F(1, 89)                3947.246   Valor p (de F)          1.63e-75
Log-verosimilitud      -760.2301   Criterio de Akaike      1524.460
Criterio de Schwarz     1529.482   Crit. de Hannan-Quinn   1526.486
```

2. Medidas de corrección de la multicolinealidad.

Medidas de corrección de la multicolinealidad imperfecta

- Como la multicolinealidad es un problema muestral, en muchos casos se puede resolver simplemente ampliando la muestra. No obstante, si teníamos acceso a más información deberíamos haberla usado desde el principio.
- Una posible solución para la multicolinealidad imperfecta es la eliminación de alguna de las variables causantes de dicha multicolinealidad. En este caso se puede incurrir en un error de especificación por omisión de una variable relevante.
- Hay que tener en cuenta que, cuanto mayor sea la información compartida por las variables, es decir, cuanto mayor sea el grado de multicolinealidad, menor será el riesgo de cometer un error de especificación por omisión al eliminar una de las variables que la generan.
- Si el objetivo del modelo es principalmente predictivo nos podemos plantear la eliminación de variables para resolver un problema de multicolinealidad, pero si se trata de encontrar los factores que afectan a una variable no deberíamos eliminar ningún factor.

Medidas de corrección de la multicolinealidad imperfecta

- Otra solución práctica a la que se recurre con frecuencia es la transformación de las variables incluidas en el modelo, en un intento de que las variables transformadas presenten correlaciones lineales más bajas.

- Las transformaciones más comúnmente utilizadas son el cálculo de los incrementos de la variables (si se trata de una serie temporal) o relativizarlas con respecto a una variable común (por ejemplo ponerlas en término per cápita).

- En nuestro ejemplo podemos trabajar con las variables en pércapita para tratar de resolver el problema de multicolinealidad.

Ejercicio.- Estimar el nuevo modelo con las variables per capita y estudiar su multicolinealidad.

Medidas de corrección de la multicolinealidad imperfecta

- Otra solución sería utilizar otros métodos de estimación.

- No obstante **si los fines que se persiguen con la construcción del modelo son predictivos, el problema de la multicolinealidad no es tan relevante, ya que no afecta a la capacidad explicativa conjunta de la variables y ni, por tanto, a su capacidad predictiva.**