

Trabajo Práctico 5

Multicolinealidad

Contenidos: Multicolinealidad.

Parte I

1. Multicolinealidad perfecta

Considere el siguiente modelo lineal,

$$Y_i = \alpha + \beta X_i + \gamma Z_i + \mu_i \quad i=1, \dots, n$$

en donde μ_i es un término aleatorio no observable, con distribución $N(0,1)$.

- Genere 1000 observaciones aleatorias de las variables u y X . Suponga que las variables tienen distribución $N(0,1)$. Así mismo, genere $Z=c*X$. A partir de las variables anteriores generar la variable Y .¹
- Calcule analíticamente la media y la varianza de cada variable y las covarianzas y los coeficientes de correlación de a pares. Estime sus correspondientes versiones muestrales.
- ¿Es posible aplicar el teorema de Gauss-Markov en este caso?
- Intente estimar empíricamente β y γ por separado. ¿Qué sucede? Justifique.
- Transforme el modelo reemplazando en el modelo original $Z=cX$. ¿Qué combinación de los parámetros sí puede ser estimada por MC? Estime y comente.

2. Multicolinealidad alta pero no perfecta

Considere el siguiente modelo lineal,

$$Y_i = \alpha + \beta X_i + \gamma W_i + \mu_i \quad i=1, \dots, n$$

en donde μ_i es un término aleatorio no observable, con distribución $N(0,1)$.

¹ Suponga un arbitrario para los parámetros, por ejemplo $c=2$, $\alpha=0$ $\beta=2$ y $\gamma=3$.

- Genere 1000 observaciones aleatorias de las variables u y X . Suponga que las variables tienen distribución $N(0,1)$. Así mismo, genere $W=c*X + e$, donde e es también una variable con observaciones aleatorias con distribución $N(0,1)$. A partir de las variables anteriores generar la variable Y .²
- Calcule analíticamente la media y la varianza de cada variable y las covarianzas y los coeficientes de correlación de a pares. Estime sus correspondientes versiones muestrales.
- Intente estimar empíricamente β y γ por separado. Compare su respuesta con el ejercicio 1.c. Analice la significatividad estadística de los estimadores, tanto global como individual. ¿Usted cree que las estimaciones de ambos parámetros son confiables? Justifique.
- ¿El estimador de mínimos cuadrados sigue siendo el de menor varianza entre la familia de estimadores lineales e insesgados?

3. Baja multicolinealidad

- Replique el ejercicio 2.c pero en esta oportunidad asuma un $c=0.01$.
- Compare las respuestas del punto anterior con la del ejercicio 2.c. En especial haga hincapié en los errores estándar y los coeficientes t .
- Vuelva a estimar las covarianzas y los coeficientes de correlación para este caso. Contraste esta respuesta con el ejercicio 2.b.

4. Multicolinealidad perfecta a raíz de correlación global de una variable

Considere el siguiente modelo lineal,

$$Y_i = \alpha + \beta X_i + \gamma P_i + \delta Q_i + \mu_i \quad i=1, \dots, n$$

en donde μ_i es un término aleatorio no observable, con distribución $N(0,1)$.

- Genere 1000 observaciones aleatorias de las variables u , X y P . Suponga que las variables tienen distribución $N(0,1)$. Así mismo, genere $Q=a*X + b*P$. A partir de las variables anteriores generar la variable Y .³
- Calcule analíticamente la media y la varianza de cada variable y las covarianzas y los coeficientes de correlación de a pares. Estime sus correspondientes versiones muestrales.
- ¿Es posible aplicar el teorema de Gauss-Markov en este caso?
- Intente estimar empíricamente β , γ y δ por separado. ¿Qué sucede? Justifique.

5. Correlación no perfecta entre más de dos variables

² Para este ejercicio suponga un $c=0.9$.

³ Suponga un a y b arbitrarios, por ejemplo $a=0.5$ y un $b=0.5$. Adicionalmente utilice $\alpha=0$ $\beta=2$, $\gamma=3$ y $\delta=4$

Considere el siguiente modelo lineal,

$$Y_i = \alpha + \beta X_i + \gamma P_i + \delta Q_i + \mu_i \quad i=1, \dots, n$$

en donde μ_i es un término aleatorio no observable, con distribución $N(0,1)$.

- a) Genere 1000 observaciones aleatorias de las variables Y , X y P . Suponga que las variables Y , X y P tienen distribución $N(0,1)$. Así mismo, genere $Q = aX + bP + e$, donde e es también una variable con observaciones aleatorias con distribución $N(0,1)$. A partir de las variables anteriores generar la variable Y .⁴
- b) Calcule analíticamente la media y la varianza de cada variable y las covarianzas y los coeficientes de correlación de a pares. Estime sus correspondientes versiones muestrales.
- c) Estime la correlación parcial de Q con el resto de las variables. Para esto, estime una regresión donde la variable a explicar sea Q y las explicativas X y P . ¿Qué conclusión puede sacar de los resultados de este inciso en comparación con los resultados del inciso anterior?
- d) Intente estimar empíricamente β , γ y δ por separado. ¿Qué sucede? Justifique. Compare su respuesta con el ejercicio 4.d. Analice la significatividad estadística de los estimadores, tanto global como individual. ¿Usted cree que las estimaciones de ambos parámetros son confiables? Justifique.
- e) ¿El estimador de mínimos cuadrados sigue siendo el de menor varianza entre la familia de estimadores lineales e insesgados?

⁴ Para simplificar, asuma los mismos valores para los coeficientes anteriores.

Parte II

6. Problema empírico: Multicolinealidad en un análisis de empírico de productividad

El objetivo de este ejercicio consiste en introducir el análisis de multicolinealidad a nivel empírico, es decir, a un ejemplo concreto.

A partir del trabajo de Pulido, A. y Rodríguez Vález, J. (2006)⁵, responda las siguientes preguntas:

- a) ¿Cuál es el modelo que intentan estimar los autores?
- b) ¿Qué método de estimación proponen?
- c) ¿Cuál es el problema que encuentran al estimar este modelo? Justifique.
- d) ¿Cómo intentan evaluar la hipótesis de multicolinealidad?
- e) ¿Por qué no les es suficiente con evaluar los coeficientes de correlación entre las variables de a pares? Justifique
- f) ¿Cuál es la conclusión del paper?
- g) En vistas a lo visto a lo largo de la guía y en el trabajo, ¿qué problema cree usted que aumenta más la probabilidad de que un investigador caiga en un problema de multicolinealidad, un escenario en presencia de multicolinealidad perfecta o un escenario donde exista multicolinealidad no perfecta?

⁵ Pulido, A. y Rodríguez Vález, J. (2006): *¿Afecta la multicolinealidad al análisis empírico de la productividad de las infraestructuras regionales?*, Working Paper, Instituto Lawrence R. Klein: Universidad Autónoma de Madrid.