

Multicolinealidad

Relación de Ejercicios

1. El director de Marketing de cierta empresa láctea desea analizar la posible relación que puede tener el gasto en publicidad, \mathbf{G} , sobre el número de ventas anuales de leche para bebés, \mathbf{V} , realizadas en los últimos 15 años. Con tal objetivo analiza el siguiente modelo de regresión lineal $\mathbf{V} = \alpha + \beta \cdot \mathbf{G} + \mathbf{u}$, donde \mathbf{u} representa a la perturbación aleatoria la cual se supone esférica. ¿Es posible que exista un grado de multicolinealidad aproximada preocupante en este modelo? En caso de existir, ¿qué tipo de multicolinealidad sería? ¿Cómo resolvería el problema?
2. En el modelo de regresión $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{X} + \beta_3 \cdot \mathbf{Z} + \mathbf{u}$ se verifica que $\mathbf{X} = 0.5 \cdot \mathbf{Z}$. ¿Qué parámetros son estimables? ¿Y qué combinaciones lineales de los parámetros?
3. En el modelo de regresión $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{X} + \beta_3 \cdot \mathbf{Z} + \mathbf{u}$ se verifica que $\mathbf{X} = 2 \cdot \mathbf{Z}$. ¿Qué parámetros son estimables si se sabe que $\beta_3 = 1$?
4. Supongamos que en el modelo de regresión $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{X}_2 + \beta_3 \cdot \mathbf{X}_3 + \mathbf{u}$ se dispone de las siguientes matrices de diseño:

$$a) \mathbf{X} = \begin{pmatrix} 1 & 4 & -2 \\ 1 & -2 & 1 \\ 1 & 10 & -5 \\ 1 & 0 & 0 \\ 1 & -1 & 0.5 \end{pmatrix}, \quad b) \mathbf{X} = \begin{pmatrix} 1 & 4 & -2 \\ 1 & 3.9 & 1 \\ 1 & 4.1 & -5 \\ 1 & 4 & 0 \\ 1 & 3.95 & 0.5 \end{pmatrix}, \quad c) \mathbf{X} = \begin{pmatrix} 1 & 4 & 3.85 \\ 1 & 3.9 & 3.71 \\ 1 & 4.1 & 3.9 \\ 1 & 4 & 3.8 \\ 1 & 3.95 & 3.7 \end{pmatrix}.$$

Indique qué tipo de multicolinealidad existe en cada caso, qué parámetros son estimables y cómo mitigaría el problema, si es que es posible.

5. Dado el modelo $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{X}_2 + \beta_3 \cdot \mathbf{X}_3 + \mathbf{u}$ donde se verifica que $2 \cdot \mathbf{X}_2 + \mathbf{X}_3 = 1$, se pide contestar razonadamente a las siguientes cuestiones:
 - a) ¿Es posible obtener una estimación de β_1 , β_2 y β_3 por Mínimos Cuadrados Ordinarios?
 - b) ¿Cuál es la estimación de β_1 y β_2 sabiendo que $\hat{\mathbf{y}} = 3.25 - 1.5 \cdot \mathbf{X}_2$ y $\beta_3 = 0.25$?
6. En el modelo de regresión $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{X} + \beta_3 \cdot \mathbf{Z} + \beta_4 \cdot \mathbf{W} + \mathbf{u}$ se verifica que $\mathbf{X} = \mathbf{Z} - \mathbf{W}$. ¿Qué parámetros son estimables? ¿Y qué combinaciones lineales de los parámetros? ¿Y si la relación fuese $\mathbf{Z} = \mathbf{W} + 2$ o $\mathbf{X} = 5 \cdot \mathbf{W}$?
7. Dados los datos de la Tabla 1 sobre el crédito en Estados Unidos, calcular el Número de Condición especificando la versión normalizada de la matriz de diseño.
8. Para analizar el volumen de consumo textil per cápita, \mathbf{C} , en los años 1923 a 1939 se considera el ingreso per cápita, \mathbf{I} , y el índice de precios de los textiles, \mathbf{P} , obteniéndose la siguiente estimación por Mínimos Cuadrados Ordinarios:

$$\hat{\mathbf{C}} = 130.707 + 1.06171 \cdot \mathbf{I} - 1.383 \cdot \mathbf{P}, \quad R^2 = 0.9513.$$

Teniendo en cuenta que los Factores de Inflación de la Varianza y el Número de Condición son iguales, respectivamente, a 1.033 y 48.953. ¿Existe un grado de multicolinealidad aproximada preocupante en este modelo? En caso de existir, ¿qué tipo de multicolinealidad aproximada sería? ¿Cómo resolvería el problema?

9. En el modelo en el que se explica el reparto de dividendos de una empresa, \mathbf{D} , a partir del endeudamiento a corto plazo de la misma, \mathbf{EC} , del endeudamiento a largo plazo, \mathbf{EL} , y del número de ventas anuales, \mathbf{V} , se sospecha que pueda existir multicolinealidad aproximada preocupante debido a la relación entre las variables \mathbf{EC} y \mathbf{EL} . Por tal motivo se realizan las siguientes regresiones:

- regresión de la variable \mathbf{EC} sobre el resto de variables independientes del modelo, obteniéndose un coeficiente de determinación de 0.990727.
- regresión de la variable \mathbf{EL} sobre el resto de variables independientes del modelo, obteniéndose un coeficiente de determinación de 0.9907107.
- regresión de la variable \mathbf{V} sobre el resto de variables independientes del modelo, obteniéndose un coeficiente de determinación de 0.01864573.

¿Existe multicolinealidad aproximada preocupante en el modelo? En caso afirmativo, ¿cómo la solucionaría?

10. En el modelo $\mathbf{C} = \beta_1 + \beta_2 \cdot \mathbf{R} + \beta_3 \cdot \mathbf{H} + \mathbf{u}$ donde \mathbf{C} es el consumo familiar, \mathbf{R} es la renta familiar y \mathbf{H} el número de hijos, se ha obtenido que el autovalor más grande de $\mathbf{X}^t\mathbf{X}$ convenientemente transformada es 143.08, mientras que el más pequeño es 2.2. ¿Existe multicolinealidad aproximada preocupante en el modelo?

11. Si al modelo del ejercicio anterior se le añade una nueva variable que mida el número de miembros de la familia con trabajo, el autovalor máximo pasa a ser 243.7 y el mínimo a 0.15. ¿Ha cambiado el grado de multicolinealidad aproximada? Indique cuáles son las consecuencias de la presencia de multicolinealidad aproximada grave y cómo resolvería este problema.

12. Suponga que dado el modelo $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{X}_2 + \beta_3 \cdot \mathbf{X}_3 + \beta_4 \cdot \mathbf{X}_4 + \mathbf{u}$ se tiene que la matriz de correlaciones de las variables independientes es:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.95 & 0.3 \\ 0.95 & 1 & 0.2 \\ 0.3 & 0.2 & 1 \end{pmatrix}.$$

Usando el FIV, ¿se puede considerar que la multicolinealidad aproximada existente en dicho modelo es preocupante? ¿Qué tipo de multicolinealidad aproximada sería?

13. Suponga que se ha estimado por Mínimos Cuadrados Ordinarios un modelo de regresión en el que se desea explicar el consumo a partir del ingreso y la riqueza. Detecte la posible presencia de multicolinealidad aproximada grave sabiendo que para dichas variables se tiene la siguiente matriz de correlaciones:

	Consumo	Ingreso	Riqueza
Consumo	1	0.9851	0.985
Ingreso	0.9851	1	0.9644
Riqueza	0.985	0.9644	1

14. Indicar a partir de las siguientes matrices de correlaciones de variables independientes si el grado de multicolinealidad aproximada existente en el modelo correspondiente es preocupante:

$$R = \begin{pmatrix} 1 & -0.8 & 0.8 \\ -0.8 & 1 & -0.75 \\ 0.8 & -0.75 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0.9 & -0.75 \\ 0.9 & 1 & -0.9 \\ -0.75 & -0.9 & 1 \end{pmatrix}.$$

¿Qué tipo de multicolinealidad aproximada sería? Especificar la ecuación del modelo de regresión lineal múltiple.

15. Se ha estimado por Mínimos Cuadrados Ordinarios un modelo que analiza la inflación, \mathbf{I} , en función del desempleo, \mathbf{D} , y el porcentaje de cambio de los salarios, \mathbf{S} , entre los años 1980 y 2019, obteniéndose los siguientes resultados:

$$\hat{\mathbf{I}} = -2.76 + 0.405 \cdot \mathbf{D} + 1.07 \cdot \mathbf{S}, \quad R^2 = 0.814.$$

Analice si la multicolinealidad existente es preocupante teniendo en cuenta que:

$$\hat{\mathbf{D}} = 5.81 + 0.0608 \cdot \mathbf{S}, \quad SCT = 78.7432, \quad \hat{\sigma}^2 = 0.22349.$$

16. Suponga que los autovalores de la matriz $\mathbf{X}^t\mathbf{X}$ (debidamente transformada) asociada al modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ son 0.0015, 0.0687, 1.256 y 3.023. ¿Existe algún problema en el modelo que desaconseje su estimación por MCO?
17. Suponga que en el modelo $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{X} + \beta_3 \cdot \mathbf{Z} + \mathbf{u}$ se verifica que la correlación simple entre \mathbf{X} y \mathbf{Z} es igual a cero y el Número de Condición a 55. ¿Cuánto vale el Factor de Inflación de la Varianza? ¿Existe multicolinealidad aproximada preocupante en el modelo? En caso de existir, ¿de qué tipo de multicolinealidad aproximada se trata? ¿Cómo la mitigaría?
18. Suponga que en el modelo $\mathbf{y}_t = \beta_1 + \beta_2 \cdot \mathbf{X}_t + \beta_3 \cdot \mathbf{Z}_t + \mathbf{u}_t$ se verifica que el Número de Condición calculado considerando la constante y sin considerarla coincide, y es igual a 55. ¿Existe multicolinealidad aproximada preocupante en el modelo? En caso de existir, ¿de qué tipo de multicolinealidad aproximada se trata?
19. Dado el modelo $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{X}_2 + \beta_3 \cdot \mathbf{X}_3 + \beta_4 \cdot \mathbf{X}_4 + \mathbf{u}$, para el cual se dispone de la siguiente información muestral:

$$\mathbf{y} = \begin{pmatrix} 4 \\ -3 \\ 3 \\ 2 \\ -4 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 4 & 2 & 6 \\ 1 & -4 & -2 & -6 \\ 1 & 5 & 2.5 & 7.5 \\ 1 & -3 & -1.5 & -4.5 \\ 1 & -2 & -1 & -3 \end{pmatrix},$$

proporcionar una estimación de β_1 , β_2 , β_3 y β_4 sabiendo que $\hat{\mathbf{y}} = 5 + 2 \cdot \mathbf{X}_2 + 4 \cdot \mathbf{X}_3$.

20. Dadas las matrices de diseño siguientes:

$$a) \mathbf{X} = \begin{pmatrix} 1 & 4 & 1.99 \\ 1 & 3 & 1.51 \\ 1 & -4 & -2.05 \\ 1 & -3 & -1.45 \end{pmatrix}, \quad b) \mathbf{X} = \begin{pmatrix} 1 & 4 & 7 \\ 1 & 3 & -7 \\ 1 & -4 & -13 \\ 1 & -3 & +13 \end{pmatrix},$$

indicar, en cada caso, qué tipo de multicolinealidad existe y si es o no preocupante.

Año	D	C	I	CP	C_p	I_p
1996	3.80510	4.7703	4.8786	808.23	4.815389	4.987343
1997	3.94580	4.7784	5.0510	798.03	4.878906	5.032297
1998	4.05790	4.9348	5.3620	806.12	4.928561	5.461873
1999	4.19130	5.0998	5.5585	865.65	5.164388	5.592996
2000	4.35850	5.2907	5.8425	997.30	5.256441	5.899322
2001	4.54530	5.4335	6.1523	1140.70	5.49978	6.197405
2002	4.81490	5.6194	6.5206	1253.40	5.66251	6.619613
2003	5.12860	5.8318	6.9151	1324.80	5.931252	6.901483
2004	5.61510	6.1258	7.4230	1420.50	6.16048	7.560622
2005	6.22490	6.4386	7.8024	1532.10	6.505051	7.792404
2006	6.78640	6.7394	8.4297	1717.50	6.82464	8.456612
2007	7.49440	6.9104	8.7241	1867.20	6.980757	8.781423
2008	8.39930	7.0993	8.8819	1974.10	7.081545	8.759557
2009	9.39510	7.2953	9.1636	2078.00	7.375371	9.260236
2010	10.68000	7.5614	9.7272	2191.30	7.603095	9.616988
2011	12.07100	7.8036	10.3010	2284.90	7.87653	10.34256
2012	13.44821	8.0441	10.9830	2387.50	8.093913	10.93781

Tabla 1: Datos referentes al crédito en los Estados Unidos

21. En la Tabla 2 se muestran datos sobre volumen de consumo textil per cápita, **C**, ingreso real per cápita con base 1925, **I**, y el precio relativo de los textiles con base 1925, **P**, en Estados Unidos en los años 1923 a 1939. Se pide:
- Estimar por MCO el modelo $\mathbf{C} = \beta_1 + \beta_2 \cdot \mathbf{I} + \beta_3 \cdot \mathbf{P} + \mathbf{u}$.
 - Comprobar que en el modelo anterior el grado de multicolinealidad aproximada existente es preocupante.
 - ¿De qué tipo de multicolinealidad aproximada se trata? ¿Cómo la resolvería?
22. Analizar los conjuntos de datos mostrados en las Tablas 3 y 4 teniendo en cuenta que la variable dependiente es la de la segunda columna.

Año	C	I	P
1923	99.2	96.7	101
1924	99	98.1	100.1
1925	100	100	100
1926	111.6	104.9	90.6
1927	122.2	104.9	86.5
1928	117.6	109.5	89.7
1929	121.1	110.8	90.6
1930	136	112.3	82.8
1931	154.2	109.3	70.1
1932	153.6	105.3	65.4
1933	158.5	101.7	61.3
1934	140.6	95.4	62.5
1935	136.2	96.4	63.6
1936	168	97.6	52.6
1937	154.3	102.4	59.7
1938	149	101.6	59.5
1939	165.5	103.8	61.3

Tabla 2: Datos sobre consumo textil en Estados Unidos

Año	C	I	InA	IA
1936	62.8	43.41	17.1	3.96
1937	65	46.44	18.65	5.48
1938	63.9	44.35	17.09	4.37
1939	67.5	47.82	19.28	4.51
1940	71.3	51.02	23.24	4.88
1941	76.6	58.71	28.11	6.37
1945	86.3	87.69	30.29	8.96
1946	95.7	76.73	28.26	9.76
1947	98.3	75.91	27.91	9.31
1948	100.3	77.62	32.3	9.85
1949	103.2	78.01	31.39	7.21
1950	108.9	83.57	35.61	7.39
1951	108.5	90.59	37.58	7.98
1952	111.4	95.47	35.17	7.42

Tabla 3: Datos referentes al salario mensual en los Estados Unidos

Año	C	IN	PIB	I	TC
1989	118.7	1107.6	1440.1	651.18	25.7
1990	1426	1479	1672.9	902.98	25.59
1991	1730.6	1789.7	1910.4	1073.9	25.52
1992	2010.6	2161.7	2145.7	1131.3	25.4
1993	1397.3	2662.9	2402.8	1266.4	25.32
1994	2710.6	3463.3	2740.6	1460.9	25.15
1995	3203.6	4300.9	3149.9	1762.2	24.92
1996	3643.3	4911.4	3394	1928.2	25.34
1997	4224.7	6060.9	3437.7	1593.2	31.37
1998	4595.9	5472.7	3311	945.97	41.37
1999	4575	5248.3	3334.8	950.61	37.84
2000	4816	4723.7	3628.7	1117.6	40.16
2001	5009.1	4447.9	3776.16	1237.09	44.48
2002	5132	4779.9	3983.53	1297.33	43
2003	5358.1	4954.3	4306.84	1477.48	41.53
2004	5497	5284.3	4794.92	1739.75	40.27
2005	5956.6	5710.3	5182.75	2231.75	40.27

Tabla 4: Datos referentes al depósito de dinero en bancos comerciales de Thailandia