



Econometría

Multicolinealidad

Que es?


- Hay dependencia lineal entre las variables explicativas

- Ejemplo 1

$$X = \begin{pmatrix} 1 & 3 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 8 \\ 1 & 6 & 12 \end{pmatrix}$$

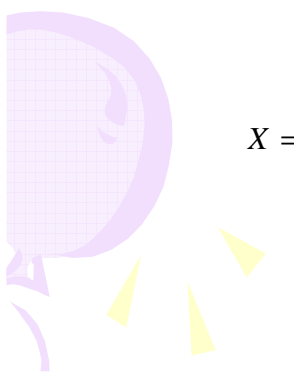
$$X'X = \begin{pmatrix} 4 & 15 & 30 \\ 15 & 65 & 130 \\ 30 & 130 & 260 \end{pmatrix}$$

- Determinante de $X'X$ es cero
- No se puede invertir $X'X$
- No se pueden calcular el MCO



Multicolinealidad aproximada

- Es muy difícil que se de la multicolinealidad exacta.
- Es posible tener multicolinealidad aproximada
- Relación casi lineal entre variables explicativas o regresores.


$$X = \begin{pmatrix} 1 & 3 & 6,01 \\ 1 & 2 & 4 \\ 1 & 4 & 8 \\ 1 & 6 & 12 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 4 & 15 & 30,01 \\ 15 & 65 & 130,03 \\ 30,01 & 130,03 & 260,1201 \end{pmatrix}$$

$$|X'X| = 0,0024$$

Ejemplo de datos con multicolinealidad

	Y (empleo)	X1 (viajeros)	X2 (estancia media)	X3 (ocupación)	X4 (fin de semana)	X5 (internet)
Andalucía	28,4	11.902,5	3,1	52	58	46
Aragón	3,6	1.848,0	2,1	37	44	43
Asturias (Principado de)	2,4	1.088,2	2,3	33	40	49
Balears (Illes)	25,9	6.716,0	7,2	70	71	69
Canarias	27,2	4.875,7	7,8	68	69	58
Cantabria	2,0	933,8	2,4	40	48	39
Castilla y León	6,2	3.647,6	1,7	34	42	30
Castilla-La Mancha	2,8	1.805,1	1,7	31	38	30
Cataluña	23,5	10.771,7	3,4	55	61	45
Comunidad Valenciana	13,4	5.579,7	3,9	60	65	49
Extremadura	2,2	1.000,7	1,7	31	38	26
Galicia	6,3	3.040,5	2,1	33	36	26
Madrid (Comunidad de)	10,7	5.748,9	2,1	51	58	28
Murcia (Región de)	2,0	882,5	3,0	49	55	38
Navarra (Comunidad Foral de)	1,1	557,7	2,0	35	37	42
País Vasco	3,2	1.540,6	1,9	45	53	50
Rioja (La)	0,7	446,2	1,8	43	51	43
Media	9,5	3.669,7	3,0	45,1	50,9	41,9
Error	10,2	3.515,7	1,8	12,7	11,7	11,7

Ejemplo: Matriz de coeficientes de correlación

	X1 (viajeros)	X2 (estancia media)	X3 (ocupación)	X4 (fin de semana)	X5 (internet)
X1 (viajeros)	1	0,4163992	0,60133905	0,60790176	0,2945619
X2 (estancia media)	0,4163992	1	0,86596259	0,79117108	0,76235495
X3 (ocupación)	0,60133905	0,86596259	1	0,9827093	0,73823895
X4 (fin de semana)	0,60790176	0,79117108	0,9827093	1	0,70276862
X5 (internet)	0,2945619	0,76235495	0,73823895	0,70276862	1

determinante

0,001114

Ejemplo : Estimación del modelo completo

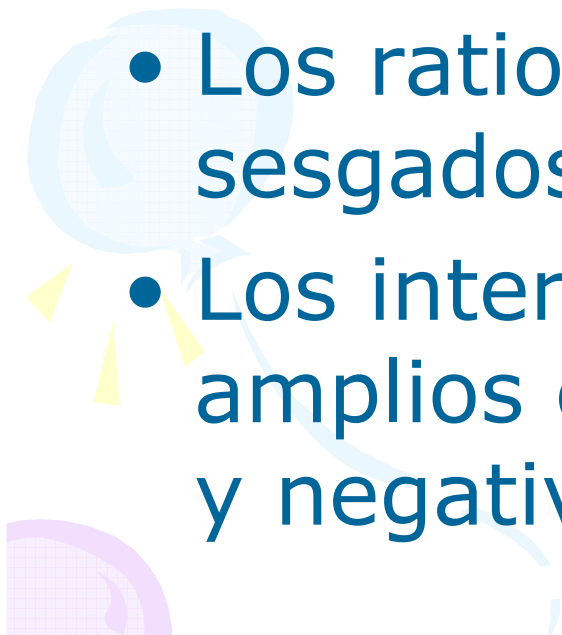
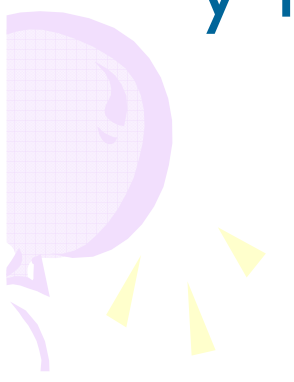
	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>
Intercepción	0	#N/A	#N/A	#N/A	#N/A	#N/A
X1 (viajeros)	0,002164	0,00012128	17,8424421	1,5955E-10	0,00190198	0,00242602
X2 (estancia media)	3,38552621	0,50630968	6,68667087	1,501E-05	2,29171064	4,47934178
X3 (ocupación)	-0,06148436	0,22676924	-0,27113183	0,79054712	-0,55138952	0,4284208
X4 (fin de semana)	-0,10986863	0,17380749	-0,63212829	0,53826193	-0,48535689	0,26561963


Ejemplo: Estimación sin estancias fines de semana

	<i>Coeficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>
Intercepción	0	#N/A	#N/A	#N/A	#N/A	#N/A
X1 (viajeros)	0,00218804	0,00011267	19,4194192	1,6026E-11	0,00194638	0,0024297
X2 (estancia media)	3,6463474	0,28707388	12,7017736	4,4945E-09	3,03063518	4,26205963
X3 (ocupación)	-0,20404825	0,02316191	-8,8096463	4,3819E-07	-0,25372561	-0,15437089

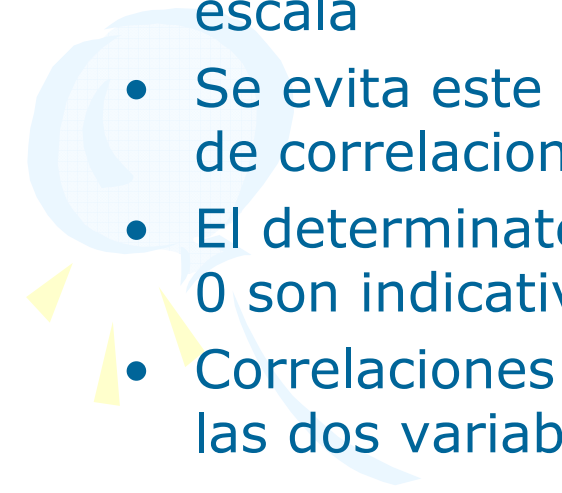
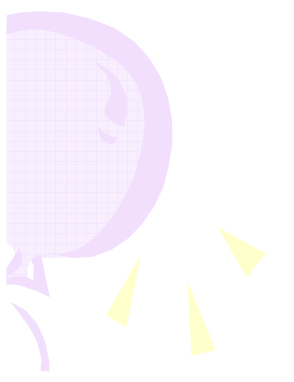


Consecuencias de la multicolinealidad aproximada:

- Las varianzas de los estimadores son muy grandes
 - Los ratios t son pequeños y están sesgados a aceptar la hipótesis nula
 - Los intervalos de confianzas son muy amplios e incluyen valores positivos y negativos
- 
- 




Detección de la multicolinealidad

- La detección de la multicolinealidad alta a partir de la matriz de productos cruzados $(X'X)$, tiene el inconveniente de que el valor de su determinante depende del factor escala
 - Se evita este inconveniente obteniendo la matriz de matriz de correlaciones simples.
 - El determinante de R fluctúa entre 0 y 1, valores próximos a 0 son indicativos de alta correlación
 - Correlaciones simples altas, mayores de 0,8, indicarían que las dos variables explicativas están correlacionadas.
- 
- 

Ejemplo: matriz de coeficientes de correlación

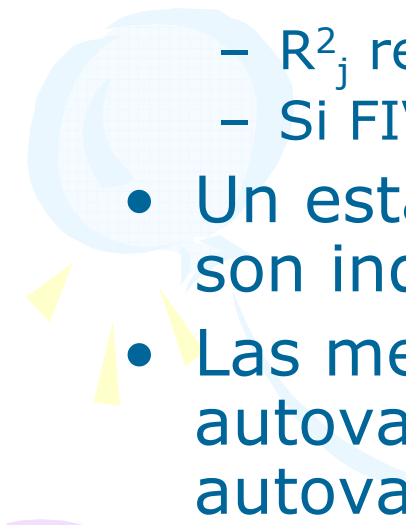

XX						
		X1 (viajeros)	X2 (estancia media)	X3 (ocupación)	X4 (fin de semana)	X5 (internet)
	X1 (viajeros)	426702792	227644,938	3243315,91	3574713,45	2805539,66
	X2 (estancia media)		202,90851	2592,0453	2831,90513	2368,15243
	X3 (ocupación)			37165,7411	41362,8146	33854,5636
	X4 (fin de semana)				46230,0025	37761,0398
	X5 (internet)					31983,5785
R						
		X1 (viajeros)	X2 (estancia media)	X3 (ocupación)	X4 (fin de semana)	X5 (internet)
	X1 (viajeros)	1	0,4163992	0,60133905	0,60790176	0,2945619
	X2 (estancia media)	0,416399204	1	0,86596259	0,79117108	0,76235495
	X3 (ocupación)	0,601339053	0,86596259	1	0,9827093	0,73823895
	X4 (fin de semana)	0,607901762	0,79117108	0,9827093	1	0,70276862
	X5 (internet)	0,2945619	0,76235495	0,73823895	0,70276862	1
	determinante	0,001114				



Detección de la multicolinealidad

- Factor incremento de la varianza (FIV):

$$FIV_j = \frac{1}{1 - R_j^2}$$

- R_j^2 regresión sobre el resto de explicativas
 - Si $FIV > 10$ las variables están altamente correlacionadas
 - Un estadístico F alto y varios estadísticos t bajos, son indicativos de multicolinealidad
 - Las medidas más satisfactorias se basan en los autovalores de la matriz $(X'X)$. La raíz del autovalor más grande entre el autovalor más pequeño, cuando es mayor de 30 denota multicolinealidad.
- 
- 


Ejemplo: Regresión fin de semana resto explicativas

	X4 (fin de semana)	X1 (viajeros)	X2 (estancia)	X3 (ocupación)	X5 (internet)
	0,6	2,3	0,1	0,5	0,4
	-0,6	-0,5	-0,5	-0,6	0,1
	-1,0	-0,7	-0,4	-1,0	0,6
	1,7	0,9	2,3	2,0	2,3
	1,6	0,3	2,7	1,8	1,4
	-0,2	-0,8	-0,3	-0,4	-0,3
	-0,7	0,0	-0,7	-0,8	-1,0
	-1,1	-0,5	-0,7	-1,1	-1,0
	0,8	2,0	0,2	0,8	0,3
	1,2	0,5	0,5	1,2	0,6
	-1,1	-0,8	-0,7	-1,1	-1,3
	-1,3	-0,2	-0,5	-1,0	-1,4
	0,6	0,6	-0,5	0,4	-1,2
	0,3	-0,8	0,0	0,3	-0,3
	-1,2	-0,9	-0,5	-0,8	0,0
	0,2	-0,6	-0,6	0,0	0,7
	0,0	-0,9	-0,6	-0,1	0,1

regresion lineal					
coeficientes	0,01430369	1,19602346	-0,25083066	-0,01108149	
errores cpefic	0,0634376	0,09669553	0,08516657	0,05165734	
R2	0,98020915	0,1560704	#N/A	#N/A	
suma cuadrac	160,967301	13	#N/A	#N/A	
FIV	100,554302				
TOL	0,01979085				



Soluciones a la multicolinealidad

- No hacer nada y asumir que no podemos medir efectos individuales.
 - Obtener más datos (mas "n"), en espera de que la nueva información minore el problema de las correlaciones entre explicativas.
 - Incorporar estimaciones sobre los parámetros procedentes de otros estudios econométricos. $Y_t^* = \beta_1 + \beta_2 X_{2t} + u_t$ $Y_t^* = Y_t - \hat{\alpha}_3 X_{3t}$
 - Eliminar variables explicativas.
- 

Soluciones a la multicolinealidad

- Transformar los datos (con datos de sección cruzada se recomienda utilizar cocientes, con datos de serie temporal se recomienda utilizar primeras diferencias)

$$\frac{Y_i}{X_{3i}} = \beta_1 \frac{1}{X_{3i}} + \beta_2 \frac{X_{2i}}{X_{3i}} + \beta_3 + \frac{u_i}{X_{3i}} \quad \nabla Y_t = Y_t - Y_{t-1} \quad \nabla X_{jt} = X_{jt} - X_{jt-1} \quad \nabla Y_t = \beta_2 \nabla X_{2t} + \beta_3 \nabla X_{3t} + e_t$$

- Usar un estimador cresta:

$$\hat{\beta}_R = (X'X + rI_k)^{-1} X'y$$

Donde r es un numero real

- Usar componentes principales.