

Regressione multipla: collinearità

Stimiamo la regressione della variabile y sulle variabili x_1 , x_2 e x_3 :

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	53.63724	17.87908	6757.82	<.0001
Error	26	0.06879	0.00265		
Corrected Total	29	53.70603			
Root MSE	0.05144	R-Square	0.9987		
Dependent Mean	1.09993	Adj R-Sq	0.9986		
Coeff Var	4.67632				

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.99969	0.01050	95.19	<.0001
x1	1	0.65585	0.64297	1.02	0.3171
x2	1	0.49429	0.92485	0.53	0.5976
x3	1	0.50969	0.92501	0.55	0.5863

Le tre variabili sembrano tutte non significative.

Elimino x_2 .

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	53.63648	26.81824	10412.1	<.0001
Error	27	0.06954	0.00258		
Corrected Total	29	53.70603			

Root MSE	0.05075	R-Square	0.9987
Dependent Mean	1.09993	Adj R-Sq	0.9986
Coeff Var	4.61404		

Parameter Estimates

Variable	DF	Estimate	Parameter Error	Standard t Value	Pr > t
Intercept	1	0.99784	0.00978	101.99	<.0001
x1	1	0.31226	0.01041	30.01	<.0001
x3	1	1.00405	0.00942	106.63	<.0001

Importante osservare che:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	21.69755	21.69755	18.98	0.0002
Error	28	32.00848	1.14316		
Corrected Total	29	53.70603			

Root MSE	1.06919	R-Square	0.4040
Dependent Mean	1.09993	Adj R-Sq	0.3827
Coeff Var	97.20486		

Parameter Estimates

Variable	DF	Estimate	Parameter Error	Standard t Value	Pr > t
Intercept	1	1.27849	0.19946	6.41	<.0001
x2	1	0.85539	0.19634	4.36	0.0002

La variabile x_2 ha quindi capacità predittiva su y ma questa scompare quando è insieme alle altre due perché ha con queste un forte legame lineare.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	29.65101	14.82550	129412	<.0001
Error	27	0.00309	0.00011456		
Corrected Total	29	29.65410			
Root MSE	0.01070	R-Square	0.9999		
Dependent Mean	-0.20874	Adj R-Sq	0.9999		
Coeff Var	-5.12750				

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-0.00374	0.00206	-1.81	0.0808
x1	1	-0.69512	0.00219	-316.72	<.0001
x3	1	1.00013	0.00199	503.62	<.0001