

REGRESION LINEAL MULTIPLE

Regresión Lineal Múltiple

En muchos problemas existen dos o más variables que están relacionadas y puede ser importante modelar y explorar esta relación.

Ejemplo

Un médico se encuentra investigando la relación del peso en kilogramos de un grupo de personas, con respecto a su estatura, talla de cintura y talla de cadera. Los resultados de las tallas y Pesos se muestran en la siguiente tabla,

Estudiantes	y=Peso	x1=Estatura	x2= talla de cintura	x3=talla de cadera
1	71.5	1.7	90	94
2	79.1	1.85	77.5	100
3	66.3	1.66	79	89.5
4	89.1	1.73	88	99.5
5	58.2	1.61	60	84
6	62.1	1.63	74	83
7	89.3	1.83	87	97
8	54.6	1.68	68	95
9	93.7	1.75	95.5	99
10	54.3	1.55	70	83
11	62	1.87	63.9	81.2
12	87.7	1.76	91	97
13	88.5	1.73	81.5	105
14	60	1.65	73	98
15	68.2	1.68	70	100
16	83.2	1.6	92	105
17	50.5	1.56	64	86.5
18	80.8	1.8	77	97
19	79.6	1.7	91	96.5

MODELOS DE REGRESION MULTIPLE

HIPOTESIS DE LOS PARAMETROS DE LOS MODELOS DE REGRESION

HIPOTESIS Y MODELOS DE REGRESION SIGNIFICATIVOS

PROCEDIMIENTOS DE SELECCION DEL MEJOR MODELO DE REGRESION

PROCEDIMIENTO DE TODAS LAS VARIABLES

PROCEDIMIENTO DE FORWARD

PROCEDIMIENTO DE BACKWARD

PROCEDIMIENTO CON TODAS LAS VARIABLES Y CONSTANTE EN EL MODELO

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	1004,93	461,681	2,17669	0,0575
x1	743,475	631,491	1,17733	0,2693
x2	-7,28268	3,93546	-1,85053	0,0973
x3	-30,5972	7,95476	-3,8464	0,0039
x1*x2	4,28139	3,14885	1,35967	0,2070
x1*x3	1,80546	3,35509	0,538124	0,6035
x2*x3	-0,00714679	0,0261337	-0,27347	0,7907
x1^2	-338,479	199,289	-1,69843	0,1237
x2^2	0,00882581	0,0149837	0,589026	0,5703
x3^2	0,154036	0,042554	3,61976	0,0056

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	3384,51	9	376,057	21,18	0,0001
Residual	159,77	9	17,7522		
Total (Corr.)	3544,28	18			

R-squared = **95,4922** percent

$$y = 1004,93 + 743,475*x1 - 7,28268*x2 - 30,5972*x3 + 4,28139*x1*x2 + 1,80546*x1*x3 - 0,00714679*x2*x3 - 338,479*x1^2 + 0,00882581*x2^2 + 0,154036*x3^2$$

PROCEDIMIENTO FORWARD O AGREGANDO VARIABLES Y CONSTANTE EN EL MODELO

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	-30,8142	14,722	-2,09306	0,0526
x1*x3	0,522194	0,101719	5,13367	0,0001
x2^2	0,00318549	0,000958252	3,32427	0,0043

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	2856,15	2	1428,08	33,20	0,0000
Residual	688,132	16	43,0083		
Total (Corr.)	3544,28	18			

R-squared = **80,5847** percent

$$y = -30,8142 + 0,522194*x1*x3 + 0,00318549*x2^2$$

PROCEDIMIENTO BACKWARD O QUITANDO VARIABLES Y CON CONSTANTE EN EL MODELO

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	1543,13	313,78	4,91787	0,0003
x2	-9,38895	2,41274	-3,89141	0,0019
x3	-26,918	5,47716	-4,91459	0,0003
x1*x2	5,92271	1,4268	4,15105	0,0011
x1^2	-110,015	30,5564	-3,6004	0,0032
x3^2	0,148969	0,0293869	5,06923	0,0002

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	3304,96	5	660,992	35,90	0,0000
Residual	239,323	13	18,4095		
Total (Corr.)	3544,28	18			

R-squared = 93,2476 percent

$$y = 1543,13 - 9,38895*x2 - 26,918*x3 + 5,92271*x1*x2 - 110,015*x1^2 + 0,148969*x3^2$$

PROCEDIMIENTO CON TODAS LAS VARIABLES SIN CONSTANTE EN EL MODELO

		<i>Standard</i>	<i>T</i>		
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>	
x1	1682,85	540,353	311,435	0,0110	
x2	-611,628	456,976	-133,843	0,2104	
x3	-269,929	911,946	-295,992	0,0143	
x1*x2	316,006	364,101	0,867908	0,4058	
x1*x3	155,764	393,021	0,396324	0,7002	
x2*x3	0,00486325	0,0299406	0,16243	0,8742	
x1^2	-581,081	193,639	-300,085	0,0133	
x2^2	0,00567577	0,0174802	0,324697	0,7521	
x3^2	0,13164	0,0483973	271,998	0,0216	
Analysis of Variance					
<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	103343,	9	11482,6	470,83	0,0000
Residual	243,88	10	24,388		
Total	103587,	19			
R-squared = 99,7646 percent					

$$y = 1682,85*x1 - 6,11628*x2 - 26,9929*x3 + 3,16006*x1*x2 + 1,55764*x1*x3 + 0,00486325*x2*x3 - 581,081*x1^2 + 0,00567577*x2^2 + 0,13164*x3^2$$

PROCEDIMIENTO FORWARD O AGREGANDO VARIABLES Y SIN CONSTANTE EN EL MODELO

		<i>Standard</i>	<i>T</i>		
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>	
x2	-0,855989	0,402539	-212,648	0,0494	
x1*x2	0,765108	0,226763	337,404	0,0039	
x3^2	0,00424294	0,00110217	384,961	0,0014	
Analysis of Variance					
<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	102885,	3	34295,1	781,75	0,0000
Residual	701,91	16	438,694		
Total	103587,	19			
R-squared = 99,3224 percent					

$$y = -0,855989*x2 + 0,765108*x1*x2 + 0,00424294*x3^2$$

PROCEDIMIENTO BACKWARD SIN CONSTANTE EN EL MODELO

<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
x1	1716,75	359,759	477,194	0,0004
x2	-720,215	213,634	-337,126	0,0050
x3	-267,406	560,543	-477,047	0,0004
x1*x2	460,115	125,606	366,316	0,0029
x1^2	-580,832	124,862	-46,518	0,0005
x3^2	0,146435	0,0297519	492,188	0,0003

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	103338,	6	17223,1	899,98	0,0000
Residual	248,783	13	191,372		
Total	103587,	19			

R-squared = **99,7598** percent

$$y = 1716,75*x1 - 7,20215*x2 - 26,7406*x3 + 4,60115*x1*x2 - 580,832*x1^2 + 0,146435*x3^2$$

¿CUAL ES EL MEJOR MODELO DE REGRESION?

PROCEDIMIENTO DE TODAS LAS VARIABLES CON CONSTANTE

Hay términos en el modelo que no son significativos para un alfa de 0.05

Según el ANOVA el modelo es significativo para un alfa de 0.05

R cuadrado de 95.49% es muy bueno

PROCEDIMIENTO DE FORWARD CON CONSTANTE

Los términos en el modelo que son significativos para un alfa de 0.05

Según el ANOVA el modelo es significativo para un alfa de 0.05

R cuadrado de 80.58% es muy bueno

PROCEDIMIENTO DE BACKWARD CON CONSTANTE

Los términos en el modelo que son significativos para un alfa de 0.05

Según el ANOVA el modelo es significativo para un alfa de 0.05

R cuadrado de 80.58% es muy bueno

¿CUAL ES EL MEJOR MODELO DE REGRESION?

PROCEDIMIENTO DE TODAS LAS VARIABLES SIN CONSTANTE

Hay términos en el modelo que no son significativos para un alfa de 0.05

Según el ANOVA el modelo es significativo para un alfa de 0.05

R cuadrado de 99.76% es muy bueno

PROCEDIMIENTO DE FORWARD SIN CONSTANTE

Los términos en el modelo que son significativos para un alfa de 0.05

Según el ANOVA el modelo es significativo para un alfa de 0.05

R cuadrado de 99.32% es muy bueno

PROCEDIMIENTO DE BACKWARD SIN CONSTANTE

Los términos en el modelo que son significativos para un alfa de 0.05

Según el ANOVA el modelo es significativo para un alfa de 0.05

R cuadrado de 99.75% es muy bueno

PROCEDIMIENTO BACKWARD SIN CONSTANTE EN EL MODELO, ES EL MEJOR MODELO.

<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>	
x1	1716,75	359,759	477,194	0,0004	
x2	-720,215	213,634	-337,126	0,0050	
x3	-267,406	560,543	-477,047	0,0004	
x1*x2	460,115	125,606	366,316	0,0029	
x1^2	-580,832	124,862	-46,518	0,0005	
x3^2	0,146435	0,0297519	492,188	0,0003	
Analysis of Variance					
<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	103338,	6	17223,1	899,98	0,0000
Residual	248,783	13	191,372		
Total	103587,	19			
R-squared = 99,7598 percent					

$$y = 1716,75*x1 - 7,20215*x2 - 26,7406*x3 + 4,60115*x1*x2 - 580,832*x1^2 + 0,146435*x3^2$$

	<i>Observed</i>	<i>Fitted</i>	<i>Std. Error</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
<i>Row</i>	<i>Value</i>	<i>Value</i>	<i>CL for Forecast</i>	<i>CL for Forecast</i>	<i>CL for Forecast</i>	<i>CL for Mean</i>	<i>CL for Mean</i>
1	71,5	759,383	468,862	658,091	860,675	722,938	795,828
2	79,1	799,054	499,384	691,169	90,694	74,702	851,088
3	66,3	633,882	463,584	533,731	734,034	600,737	667,028
4	89,1	873,512	466,525	772,726	974,299	838,494	908,531
5	58,2	577,711	51,488	466,478	688,945	51,905	636,372
6	62,1	664,436	515,229	553,127	775,744	605,632	72,324
7	89,3	864,369	493,394	757,777	970,961	815,074	913,665
8	54,6	61,91	475,363	516,404	721,796	578,916	659,284
9	93,7	945,689	499,304	83,782	105,356	89,369	997,687
10	54,3	499,102	492,267	392,753	60,545	450,335	547,868
11	62,0	629,726	594,412	501,311	758,142	542,785	716,667
12	87,7	877,902	478,158	774,602	981,202	836,199	919,606
13	88,5	848,985	568,309	726,209	971,761	770,613	927,357
14	60,0	655,595	473,825	553,231	757,959	616,268	694,922
15	68,2	720,372	472,599	618,273	822,471	68,174	759,005
16	83,2	812,481	578,643	687,472	937,489	730,655	894,306
17	50,5	456,615	511,506	346,111	567,119	399,348	513,882
18	80,8	753,789	473,056	651,591	855,987	714,896	792,682
19	79,6	794,465	466,776	693,624	895,307	759,292	829,639

Predicciones del mejor modelo de regresión múltiple.