Ejemplo: Queremos analizar la relación existente entre el grado de estrés de los trabajadores Y, medido a partir del tamaño de la empresa en que trabajan, X_1, el número de años que llevan en el puesto de trabajo, X_2, el salario anual percibido, X_3 y la edad del trabajador, X_4. Se dispone de las observaciones siguientes:

X_1	X_2	X_3	X_4	Y
812	15	30	38	101
334	8	20	52	60
377	5	20	27	10
303	10	54	36	27
505	13	52	34	89
401	4	27	45	60
177	6	26	50	16
598	9	52	60	184
412	16	34	44	34
127	2	28	39	17
601	8	42	41	78
297	11	84	58	141
205	4	31	51	11
603	5	38	63	104
484	8	41	30	76

a) Factores de incremento de la varianza:
La matriz de correlaciones entre las variables explicativas y su inversa son:

$$R = \begin{pmatrix} 1.0000 & 0.5011 & 0.0884 & -0.0189 \\ 0.5011 & 1.0000 & 0.3842 & -0.1135 \\ 0.0884 & 0.3842 & 1.0000 & 0.2608 \\ -0.0189 & -0.1135 & 0.2608 & 1.0000 \end{pmatrix} \quad R^{-1} = \begin{pmatrix} 1.3703 & -0.7804 & 0.2093 & -0.1172 \\ -0.7804 & 1.6891 & -0.6718 & 0.3522 \\ 0.2093 & -0.6718 & 1.3503 & -0.4245 \\ -0.1172 & 0.3522 & -0.4245 & 1.1485 \end{pmatrix}$$

Los factores de incremento (o de inflación) de la varianza son:

$FIV_{(1)}=1.3703$, $FIV_{(2)}=1.6791$, $FIV_{(3)}=1.3503$, $FIV_{(4)}=1.1485$,.

Observemos que ninguno de ellos es mayor que 10.

b) Índice de condicionamniento:
Los valores propios de la matriz **R** son: $\lambda_1=0.3547$, $\lambda_2= 0.7609$, $\lambda_3= 1.2067$, $\lambda_4= 1.6777$. Por tanto,

$$cond(R) = \sqrt{\frac{1.6777}{0.3547}} = 1.8444 < 10 \Rightarrow R \text{ está bien definida}$$

Conclusión: Nada indica que vayamos a tener problemas de multicolinealidad.

Multiple Regression Analysis

Dependent variable: Y

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-126,505	32,2811	-3,91887	0,0029
X1	0,176293	0,0400949	4,39691	0,0013
X2	**-1,56295**	**2,01205**	**-0,776793**	**0,4553**
X3	1,57454	0,445674	3,53293	0,0054
X4	1,62929	0,628717	2,59144	0,0269

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	30873,5	4	7718,37	13,37	0,0005
Residual	5774,93	10	577,493		
Total (Corr.)	36648,4	14			

R-squared = 84,2423 percent
R-squared (adjusted for d.f.) = 77,9393 percent

Multiple Regression Analysis

Dependent variable: Y

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-132,338	30,8244	-4,2933	0,0013
X1	0,160318	0,0337924	4,7442	0,0006
X3	1,42055	0,3919	3,62478	0,0040
X4	1,75274	0,597231	2,93479	0,0136

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	30525,0	3	10175,0	18,28	0,0001
Residual	6123,4	11	556,672		
Total (Corr.)	36648,4	14			

R-squared = 83,2915 percent
R-squared (adjusted for d.f.) = 78,7346 percent

Comparación de modelos:

Modelo	R-squared	R-squared (adjusted)
y = -126.505 + 0,176293 X_1 -1,56295 X_2 + 1,57454 X_3 +1,62929 X_4	**84,2423**	77,9393
y = -132.338 + 0.160318 X_1 + 1.42055 X_3 +1.75274 X_4	83,2915	**78,7346**

Si nos fijamos solamente en el coeficiente de determinación (sin corregir) seleccionaríamos un modelo erróneo, puesto que la variable X_2 es no significativa.

Observemos que al eliminar del modelo a la variable X_2, los efectos de las variables explicativas sobre la variable respuesta han disminuido levemente, excepto el efecto de X_4 (con quien X_2 estaba correlacionada negativamente). Por lo tanto, la introducción de la variable X_2 no significativa en el modelo provocaba una **sobreestimación** de los efectos de las restantes variables sobre la respuesta.

Hallar las predicciones del grado de estrés de los trabajadores para los valores siguientes de las variables explicativas:

individuo	X_1	X_2	X_3	X_4
1	302	9	44	42
2	351	8	65	62
3	381	9	52	53

¿Extrapolación oculta?

Debemos comprobar si $x_h' (X'X)^{-1} x_h \leq v_{max}$

donde $v_{max}=max(diag(H))=max(diag(X (X' X)^{-1} X'))$ y $X = [1 \ X_1 \ X_3 \ X_4]$.

$max(diag(H))=max($ 0.4316, 0.2334, 0.2887, 0.2450, 0.2264, 0.1048, 0.2438, 0.2894, 0.0721, 0.2663, 0.1460, **0.6647**, 0.1997, 0.3711, 0.2170 $)$

v_{max}**=0.6647**

individuo	$x_h' (X'X)^{-1} x_h$ (***)	*Predicción*
1	0.1121	52.1972
2	0.3651	124.9393
3	0.1407	95.5069

(***) Atención: ¿quién es x_h?