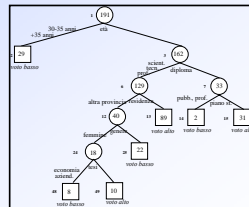
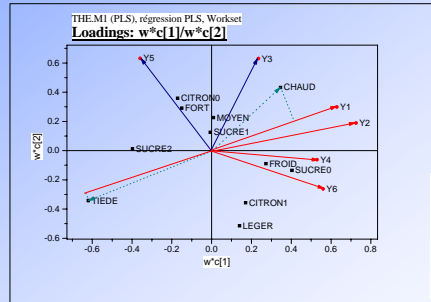
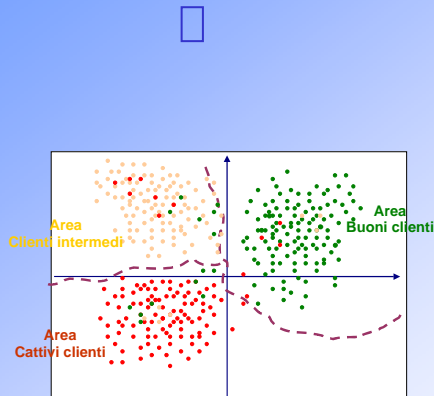


L'Analisi Multidimensionale dei Dati

Una Statistica da vedere

Le Analisi Esplicative



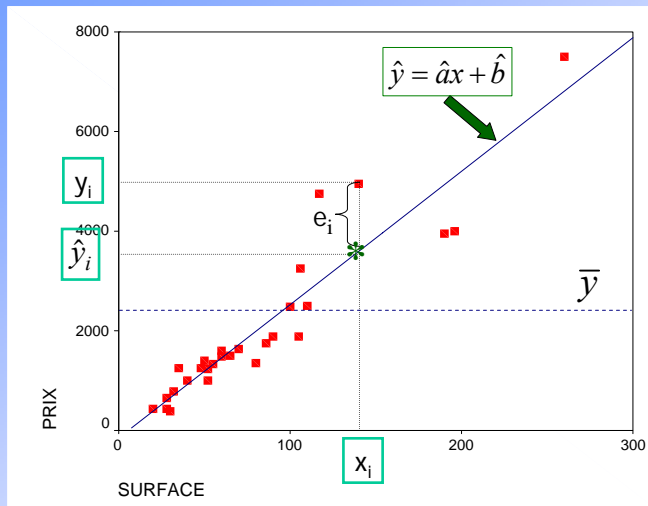
Metodi Esplicativi

Predittori X_1, X_2, \dots, X_k

Variabile di risposta y	Continui	Qualitativi	Misti
Continua	- Regressione Multipla	- Analisi della Varianza	- Analisi della Covarianza
caso Multivariato	- Regressione PLS	- Regressione PLS	- Regressione PLS
Qualitativa	- Regressione Logistica	- Regressione Logistica	- Regressione Logistica
	- Segmentazione	- Segmentazione	- Segmentazione
	- Analisi Discriminante		
caso Multivariato	- PLS-DA	- PLS Logistica	- PLS Logistica
		- PLS-DA	- PLS-DA

La Regressione Lineare Semplice

La retta dei minimi quadrati (OLS)



Si ricercano

\hat{a} et \hat{b}

tali che

$$\sum_{i=1}^n e_i^2$$

sia minima

Valutazione della bontà dei risultati

A) Decomposizione della varianza

$$\underbrace{\sum (y_i - \bar{y})^2}_{\text{Variabilità Totale}} = \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{\text{Variabilità Spiegata}} + \underbrace{\sum e_i^2}_{\text{Variabilità Residua}}$$

$$\text{B) } R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad \text{C) } R = \text{sign}(\hat{a}) \sqrt{R^2}$$

La Regressione Lineare Multipla

- Y
= Variabile di risposta
(o target) numerica
- X_1, \dots, X_k
= Variabili esplicative
numeriche o dummy

	X_1	...	X_k	Y
1	X_{11}	...	X_{k1}	y_1
⋮				
i	X_{1i}	...	X_{ki}	y_i
⋮				
n	X_{1n}	...	X_{kn}	y_n

I dati

Un caso studio: Le Vendite

SEMESTRE	Totale Mercato	Sconti ai grossisti	Prezzo	Budget Ricerca	Investimenti	Pubblicità	Spese Vendite	Totale pubblicità del settore	Vendite
1	398	138	56	12	50	77	229	98	5540
2	369	118	59	9	17	89	177	225	5439
3	268	129	57	29	89	51	166	263	4290
4	484	111	58	13	107	40	258	321	5502
5	394	146	59	13	143	52	209	407	4872
6	332	140	60	11	61	21	180	247	4708
7	336	136	60	25	-30	40	213	328	4627
8	383	104	60	21	-45	32	201	298	4110
9	285	105	63	8	-28	12	176	218	4123
10	277	135	62	11	76	68	175	410	4842
11	456	128	65	22	144	52	253	93	5741
12	355	131	65	24	113	77	208	307	5094
13	364	120	64	14	128	96	195	107	5383
14	320	147	66	15	10	48	154	305	4888
15	311	143	67	22	-25	27	181	60	4033
16	362	145	67	23	117	73	220	239	4942
17	408	131	66	13	120	62	235	141	5313
18	433	124	68	8	122	25	258	291	5140
19	359	106	69	27	71	74	196	414	5397
20	476	138	71	18	4	63	279	206	5149
21	415	148	69	8	47	29	207	80	5151
22	420	136	70	10	8	91	213	429	4989
23	536	111	73	27	128	74	296	273	5927
24	432	152	73	16	-50	16	245	309	4704
25	436	123	73	32	100	43	276	280	5366
26	415	119	75	20	-40	41	211	315	4630
27	462	112	73	15	68	93	283	212	5712
28	429	125	74	11	88	83	218	118	5095
29	517	142	74	27	27	75	307	345	6124
30	328	123	77	20	59	88	211	141	4787
31	418	135	79	35	142	74	270	83	5036
32	515	120	77	23	126	21	328	398	5288
33	412	149	78	36	30	26	258	124	4647
34	455	126	78	22	18	95	233	118	5316
35	554	138	81	20	42	93	324	161	6180
36	441	120	80	16	-22	50	267	405	4801
37	417	120	81	35	148	83	257	111	5512
38	461	132	82	27	-18	91	267	170	5272
39	500	100	83	30	50	90	300	200	.

I dati sulle Vendite

- Variabile di interesse:

$Y = \text{Vendite}$

- Variabili Esplicative:

$X_1 = \text{Mercato Totale}$, $X_2 = \text{Sconti ai grossisti}$, ...

$X_8 = \text{Budget complessivo per la pubblicità del settore}$

- Problema di Previsione:

Prevedere le vendite y_{39} relative al 39esimo semestre con il seguente scenario:

TOT = 500, DIS = 100, PRICE = 83, RES = 30,

INV = 50, ADV = 90, EXP = 300, ADV_sec = 200

Le correlazioni

Correlation Matrix

	Correlation								
	Marché total	Remises aux grossistes	Prix	Budget de recherche	Investissements	Publicité	Frais de ventes	Total publicité de la branche	Ventes
Marché total	1.000	-.069	.549	.164	.144	.200	.903	-.020	.721
Remises aux grossistes	-.069	1.000	.022	.010	-.093	-.120	-.050	-.146	-.084
Prix	.549	.022	1.000	.455	-.058	.255	.625	-.181	.287
Budget de recherche	.164	.010	.455	1.000	.157	.105	.364	-.128	.084
Investissements	.144	-.093	-.058	.157	1.000	.241	.216	-.123	.453
Publicité	.200	-.120	.255	.105	.241	1.000	.134	-.195	.568
Frais de ventes	.903	-.050	.625	.364	.216	.134	1.000	-.022	.637
Total publicité de la branche	-.020	-.146	-.181	-.128	-.123	-.195	-.022	1.000	-.096
Ventes	.721	-.084	.287	.084	.453	.568	.637	-.096	1.000

Il Modello

$$Y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

Fasi dello studio:

- A) Stima dei **coefficienti di regressione** $\beta_0, \beta_1, \dots, \beta_k$.
- B) Stima della **variabilità** del termine residuale ε_i .
- C) Misura della **forza della relazione** tra Y e i predittori.
- D) Analisi dei **residui**.
- E) Significatività della **relazione globale**.
- F) Significatività dell '**effetto marginale**/contributo di ciascun predittore X_j alla spiegazione di Y.
- G) Intervallo di **previsione** al 95% di y_i .
- H) **Selezione automatica** dei migliori predittori.

Stima dei coefficienti di regressione

Notazioni :

- y_i = valore osservato
- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_k x_{ki}$
 - = valore calcolato
 - = **stima di μ_i**
 - = **previsione per y_i**
- $e_i = y_i - \hat{y}_i$ = errore

Metodo dei minimi quadrati (OLS) :

__Ricerchiamo $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ che minimizzano $\sum_{i=1}^n e_i^2$.

Significatività della relazione globale

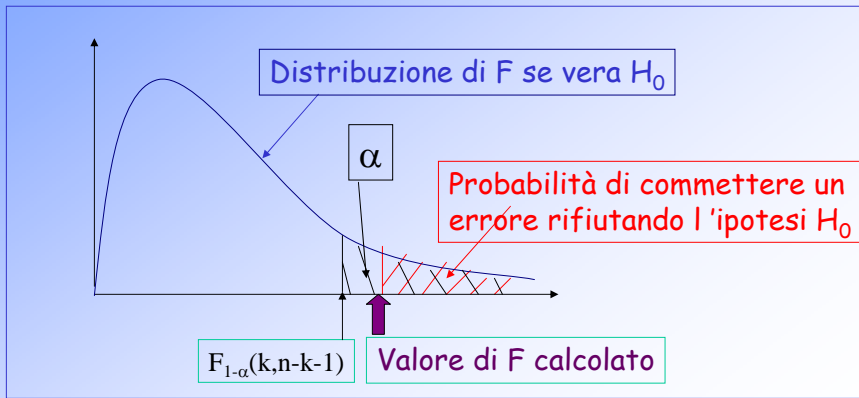
Modello: $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$

Test: $H_0 : \beta_1 = \dots = \beta_k = 0$

$H_1 : \text{almeno un } \beta_j \neq 0$

Statistica di riferimento:

$$F = \frac{\sum (\hat{y}_i - \bar{y})^2 / k}{\sum e_i^2 / (n - k - 1)}$$



Risultati SPSS

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7903373	8	987921.639	15.041	.000 ^a
	Residual	1904830	29	65683.798		
	Total	9808203	37			

a. Predictors: (Constant), Total publicité de la branche, Marché total, Remises aux grossistes, Budget de recherche, Investissements, Publicité, Prix, Frais de ventes

b. Dependent Variable: Ventes

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.898 ^a	.806	.752	256.29

a. Predictors: (Constant), Total publicité de la branche, Marché total, Remises aux grossistes, Budget de recherche, Investissements, Publicité, Prix, Frais de ventes

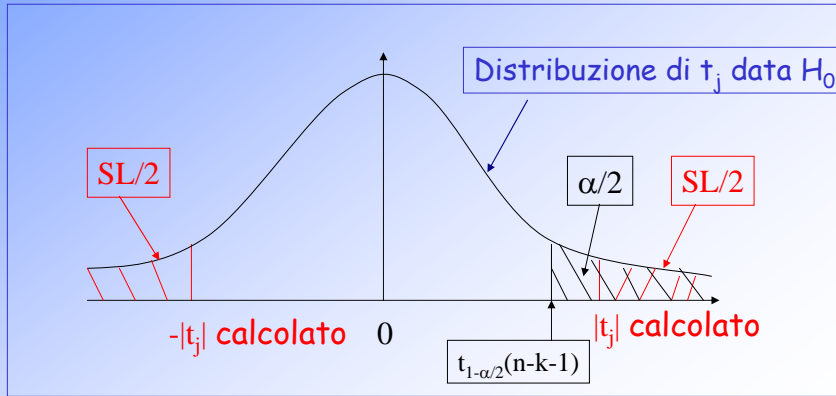
Significatività dell 'effetto marginale

Modello: $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$

Test: $H_0 : \beta_j = 0$
 $H_1 : \beta_j \neq 0$

Statistica di riferimento:

$$t_j = \frac{\hat{\beta}_j}{s_j}$$



Risultati SPSS

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7903373	8	987921.639	15.041	.000 ^a
	Residual	1904830	29	65683.798		
	Total	9808203	37			

a. Predictors: (Constant), Total publicité de la branche, Marché total, Remises aux grossistes, Budget de recherche, Investissements, Publicité, Prix, Frais de ventes
b. Dependent Variable: Ventes

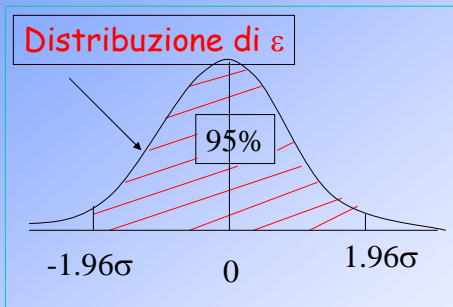
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3129.231	641.355		4.879	.000
	Marché total	4.423	1.588	.605	2.785	.009
	Remises aux grossistes	1.676	3.291	.043	.509	.614
	Prix	-13.526	8.305	-.201	-1.629	.114
	Budget de recherche	-3.410	6.569	-.054	-.519	.608
	Investissements	1.924	.778	.234	2.474	.019
	Publicité	8.547	1.826	.434	4.679	.000
	Frais de ventes	1.497	2.771	.130	.540	.593
	Total publicité de la branche	-2.15E-02	.401	-.005	-.054	.958

a. Dependent Variable: Ventes

Analisi dei Residui

Model : $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_j X_j + \dots + \beta_k X_k + \varepsilon$
 with $\varepsilon \sim N(0, \sigma)$



Un residuo e_i è valutato «troppo importante» se

$$|e_i| > 1.96\hat{\sigma}$$

Risultati SPSS

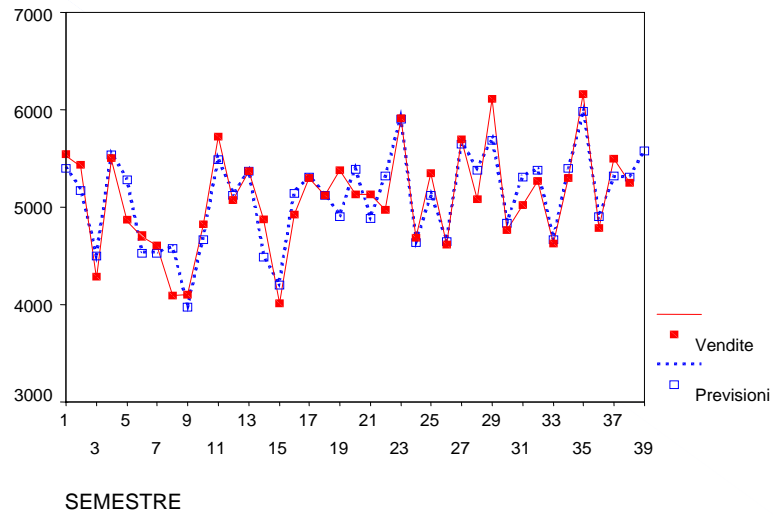
Casewise Diagnostics ^b					
Case Number	Std. Residual	Ventes	Predicted Value	Residual	Status
1	-.477	5540	5417.73	122.27	
2	.995	5439	5184.05	254.95	
3	-.863	4290	4511.08	-221.08	
4	-.205	5502	5554.49	-52.49	
5	-1.663	4872	5298.16	-426.16	
6	.638	4708	4544.43	163.57	
7	.329	4627	4542.62	84.38	
8	-1.896	4110	4595.95	-485.95	
9	-.503	4123	3993.96	129.04	
10	.611	4842	4685.29	156.71	
11	-.921	5741	5504.94	-236.06	
12	-.173	5094	5138.44	-44.44	
13	-.002	5383	5383.53	-.53	
14	1.511	4888	4500.74	387.26	
15	-.713	4033	4215.69	-182.69	
16	-.859	4942	5162.17	-220.17	
17	-.051	5313	5326.12	-13.12	
18	.024	5140	5133.79	6.21	
19	1.849	5397	4923.16	473.84	
20	-.994	5149	5403.76	-254.76	
21	.984	5151	4898.92	252.08	
22	-1.357	4989	5336.90	-347.90	
23	.016	5927	5922.80	4.20	
24	.197	4704	4653.64	50.36	
25	.903	5366	5134.61	231.39	
26	-.134	4630	4664.31	-34.31	
27	.176	5712	5666.85	45.15	
28	-1.192	5095	5400.51	-305.51	
29	1.630	6124	5706.31	417.69	
30	-.266	4787	4855.09	-68.09	
31	-1.127	5036	5324.75	-288.75	
32	-.409	5288	5392.92	-104.92	
33	-.157	4647	4687.18	-40.18	
34	-.390	5316	5415.89	-99.89	
35	.685	6180	6004.57	175.43	
36	-.466	4801	4920.47	-119.47	
37	.685	5512	5336.53	175.47	
38	-.217	5272	5327.66	-55.66	
39	.	.	5593.84	.	M ^a

a. Missing Case

b. Dependent Variable: Ventes

Grafico SPSS

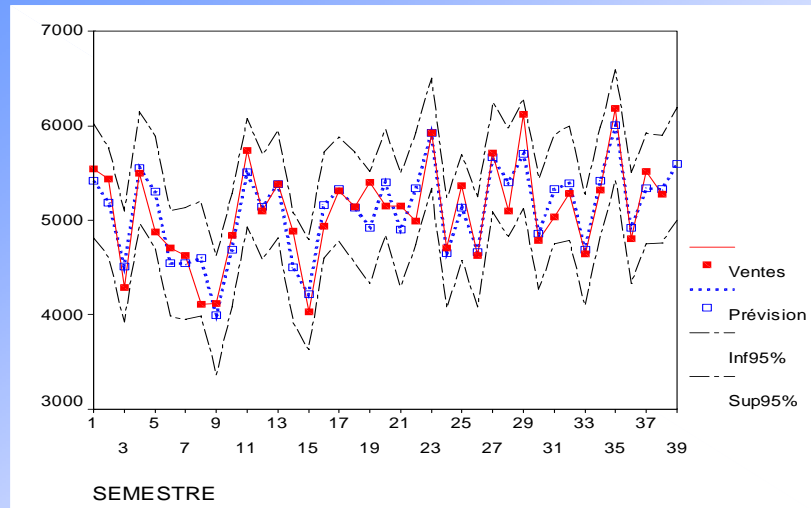
Previsione delle Vendite



Intervalli di Previsione

SEMESTRE	Ventes	Prévision	Inf95%	Sup95%
1	5540	5417.73435	4812.00066	6023.46805
2	5439	5184.05218	4607.34778	5760.75657
3	4290	4511.08482	3924.33140	5097.83823
4	5502	5554.48526	4961.43796	6147.53255
5	4872	5298.16088	4703.45870	5892.86307
6	4708	4544.42945	3983.37627	5105.48263
7	4627	4542.62500	3951.97922	5133.27077
8	4110	4595.95167	3989.66157	5202.24177
9	4123	3993.96113	3367.49333	4620.42894
10	4842	4685.28777	4086.91324	5283.66229
11	5741	5504.94341	4932.75497	6077.13186
12	5094	5138.43957	4584.35964	5692.51951
13	5383	5383.52542	4813.79822	5953.25261
14	4888	4500.73738	3920.03334	5081.44143
15	4033	4215.69091	3632.18367	4799.19814
16	4942	5162.16928	4604.20000	5720.13856
17	5313	5326.11795	4774.67321	5877.56268
18	5140	5133.78522	4548.43835	5719.13209
19	5397	4923.15766	4329.92288	5516.39244
20	5149	5403.76205	4844.35708	5963.16702
21	5151	4898.91836	4300.08237	5497.75434
22	4989	5336.90216	4746.04944	5927.75488
23	5927	5922.79791	5340.98344	6504.61237
24	4704	4653.63757	4069.21332	5238.06182
25	5366	5134.61086	4575.20480	5694.01692
26	4630	4664.30841	4082.12257	5246.49425
27	5712	5666.84752	5085.94736	6247.74769
28	5095	5400.50510	4826.99047	5974.01973
29	6124	5706.30942	5122.68255	6289.93629
30	4787	4855.08960	4267.90602	5442.27318
31	5036	5324.75120	4747.26688	5902.23553
32	5288	5392.92135	4786.23253	5999.61017
33	4647	4687.17699	4095.18697	5279.16702
34	5316	5415.89402	4841.41965	5990.36839
35	6180	6004.56898	5422.36763	6586.77034
36	4801	4920.46554	4334.02785	5506.90323
37	5512	5336.53027	4751.83792	5921.22262
38	5272	5327.66340	4757.83588	5897.49093
39	.	5593.84090	5000.69272	6186.98908

Rappresentazione Intervalli di Previsione



Selezione automatica dei predittori

Regressione Backward Stepwise:

Si inizia dal **modello completo**.

A ciascun passo (stepwise) si rimuove (backward) il predittore X_j che ha il contributo marginale meno significativo

$$|t_j| \text{ minimo}$$

dopo aver constatato che questo contributo è al di sopra di una soglia predeterminata

$$(SL(t_j) > 0.05 \text{ or } 0.1).$$

SPSS: Risultati Regressione Stepwise

Passo 1

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3129.231	641.355		4.879	.000
	Marché total	4.423	1.588	.605	2.785	.003
	Remises aux grossistes	1.676	3.291	.043	.509	.614
	Prix	-13.526	8.305	-.201	-1.629	.114
	Budget de recherche	-3.410	6.569	-.054	-.519	.603
	Investissements	1.924	.778	.234	2.474	.013
	Publicité	8.547	1.826	.434	4.679	.000
	Frais de ventes	1.497	2.771	.130	.540	.593
	Total publicité de la branche	-2.15E-02	.401	-.005	-.054	.953

a. Dependent Variable: Ventes

Variabile rimossa: Pubblicità Totale del Settore

Passo 2

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.898 ^a	.806	.760	251.99

a. Predictors: (Constant), Frais de ventes, Remises aux grossistes, Publicité, Investissements, Budget de recherche, Prix, Marché total

b. Dependent Variable: Ventes

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3115.648	579.517		5.376	.000
	Marché total	4.426	1.561	.606	2.836	.008
	Remises aux grossistes	1.706	3.191	.044	.535	.597
	Prix	-13.445	8.029	-.200	-1.675	.104
	Budget de recherche	-3.392	6.451	-.053	-.526	.603
	Investissements	1.931	.756	.234	2.554	.016
	Publicité	8.558	1.784	.435	4.798	.000
	Frais de ventes	1.482	2.710	.129	.547	.588

a. Dependent Variable: Ventes

Variabile rimossa: Budget per la Ricerca

Passo 3

Model Summary^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.897 ^a	.804	.766	249.04

a. Predictors: (Constant), Frais de ventes, Remises aux grossistes, Publicité, Investissements, Prix, Marché total

b. Dependent Variable: Ventes

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3137.547	571.233		5.493	.000
	Marché total	4.756	1.412	.651	3.368	.002
	Remises aux grossistes	1.705	3.153	.044	.541	.593
	Prix	-14.790	7.521	-.220	-1.966	.058
	Investissements	1.885	.742	.229	2.539	.016
	Publicité	8.519	1.761	.433	4.837	.000
	Frais de ventes	.950	2.484	.082	.382	.705

a. Dependent Variable: Ventes

Variabile rimossa: Spese per le Vendite

Passo 4

Model Summary^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.896 ^a	.803	.772	245.69

a. Predictors: (Constant), Publicité, Remises aux grossistes, Marché total, Investissements, Prix

b. Dependent Variable: Ventes

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3084.009	546.374		5.645	.000
	Marché total	5.222	.704	.714	7.415	.000
	Remises aux grossistes	1.700	3.111	.043	.546	.583
	Prix	-13.467	6.589	-.200	-2.044	.043
	Investissements	1.984	.686	.241	2.893	.007
	Publicité	8.328	1.666	.423	4.998	.000

a. Dependent Variable: Ventes

Variabile rimossa: Sconti ai Grossisti

Passo 5

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.895 ^a	.801	.777	243.07

a. Predictors: (Constant), Publicité, Marché total, Investissements, Prix

b. Dependent Variable: Ventes

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3302.036	369.212		8.943	.000
	Marché total	5.192	.695	.710	7.475	.000
	Prix	-13.172	6.497	-.196	-2.027	.051
	Investissements	1.968	.678	.239	2.903	.007
	Publicité	8.229	1.639	.418	5.021	.000

a. Dependent Variable: Ventes

Tutte le Variabili sono Significative

L'Analisi della Covarianza

Il caso della Consolidated Foods

- **Variabili di risposta:**
 - Vendite settimanali (in unità) del Prodotto 1
 - Quota di mercato del Prodotto 1
- **Variabili esplicative (predittori):**
 - Prezzo del Prodotto 1
 - Prezzo dei concorrenti (prezzo medio ponderato)
 - Promozioni (no, giornali, display, entrambi)
 - Catene di negozi (1, 2, 3)

Il Modello

Regressione Multipla con variabili dummy

VENDITE settimanali del Prodotto 1 =

$$\beta_0 + \beta_1 \text{PrezzoP1} + \beta_2 \text{PrezzoP2-5} + \begin{bmatrix} \beta_3 \text{ se no prom.} \\ \beta_4 \text{ se giornali} \\ \beta_5 \text{ se display} \\ \beta_6 \text{ se entrambe} \end{bmatrix} + \begin{bmatrix} \beta_7 \text{ se catena 1} \\ \beta_8 \text{ se catena 2} \\ \beta_9 \text{ se catena 3} \end{bmatrix} + \varepsilon$$

- A causa della natura (0 - 1) delle variabili dummy, il problema risulta **indeterminato**.
- Pertanto, dobbiamo **eliminare un coefficiente** (scelta arbitraria) per ciascuna variabile qualitativa:

$$\beta_0 + \beta_1 \text{PrezzoP1} + \beta_2 \text{PrezzoP2-5} + \beta_4 (\text{se giornali}) + \beta_5 (\text{se display}) + \beta_6 (\text{se entrambe}) + \beta_7 (\text{se catena 1}) + \beta_8 (\text{se catena 2}) + \varepsilon$$

SPSS: Modello per le Vendite

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.810 ^a	.656	.640	3220.4763

a. Predictors: (Constant), Chaîne 2, Publicité, Pub & PLV, PLV, Chaîne 1, Prix des concurrents, Prix B1

b. Dependent Variable: Ventes B1

Coefficients^a

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	-438.319	3040.634	-.144	.886
	Prix B1	-12055.7	3062.693	-3.936	.000
	Prix des concurrents	12969.760	3999.318	3.243	.001
	Publicité	535.951	1950.622	.275	.784
	PLV	1028.364	593.703	1.732	.085
	Pub & PLV	8816.710	976.036	9.033	.000
	Chaîne 1	5452.256	690.051	7.901	.000
	Chaîne 2	3484.414	977.611	3.564	.000

a. Dependent Variable: Ventes B1

Previsione delle Vendite

VENDITE settimanali del Prodotto 1 =

$$\begin{aligned}
 & -438 + 12056\text{PrezzoP1} + 12970\text{PrezzoP2} - 5 + \\
 & + \begin{bmatrix} 0 \text{ se no prom.} \\ 535 \text{ se giornali} \\ 1028 \text{ se display} \\ 8817 \text{ se entrambe} \end{bmatrix} + \begin{bmatrix} 5452 \text{ se catena 1} \\ 3484 \text{ se catena 2} \\ 0 \text{ se catena 3} \end{bmatrix} \pm 6440
 \end{aligned}$$

Identificazione di outliers

$$(|e_i| > 3\hat{\sigma})$$

Casewise Diagnostics^a

Case Number	Std. Residual	Ventes B1	Predicted Value	Residual
34	3.775	27254.00	15095.838	12158.16
39	5.315	32820.00	15702.573	17117.43
89	-4.471	1203.00	15600.354	-14397.4

a. Dependent Variable: Ventes B1

SPSS: Modello per la Quota di Mercato

Model Summary^a

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.867 ^a	.752	.741	9.083E-02

a. Predictors: (Constant), Chaîne 2, Publicité, Pub & PLV, PLV, Chaîne 1, Prix des concurrents, Prix B1

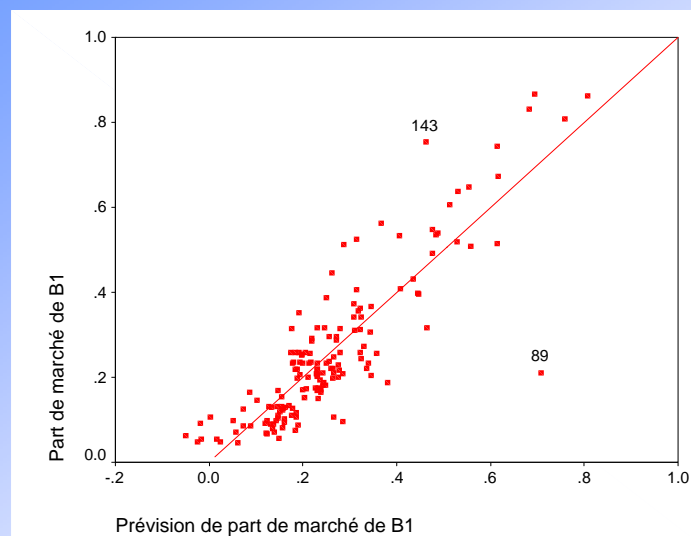
b. Dependent Variable: Part de marché de B1

Coefficients^a

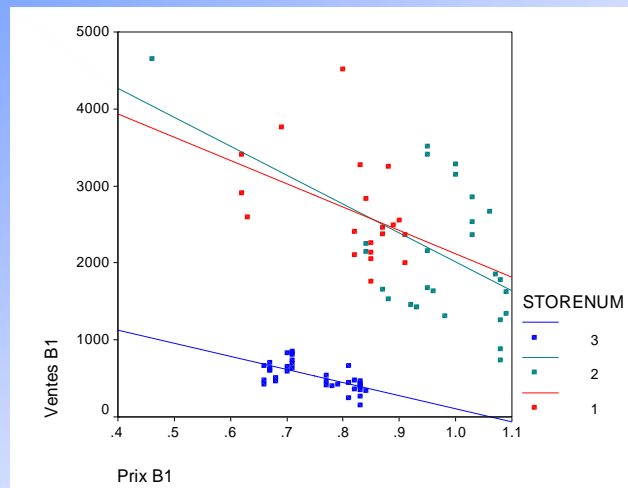
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.53E-02	.086		-.295	.769
	Prix B1	-.644	.086	-.502	-7.451	.000
	Prix des concurrents	1.064	.113	.563	9.431	.000
	Publicité	.117	.055	.090	2.125	.035
	PLV	8.626E-02	.017	.229	5.151	.000
	Pub & PLV	.297	.028	.547	10.800	.000
	Chaîne 1	-8.26E-02	.019	-.219	-4.244	.000
	Chaîne 2	-7.49E-02	.028	-.199	-2.717	.007

a. Dependent Variable: Part de marché de B1

Qualità della Previsione



Studio delle settimane in cui non c'è promozione per il Prodotto 1



Modello con Interazione Prezzo*Catena

VENDITE settimanali del Prodotto 1 =

$$\beta_0 + \beta_1 \text{PrezzoP1} + \beta_2 \text{Catena1} + \beta_3 \text{Catena2} + \\ + \beta_4 \text{PrezzoP1} * \text{Catena1} + \beta_4 \text{PrezzoP1} * \text{Catena2} + \varepsilon$$

SPSS: Modello Completo

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.881 ^a	.775	.761	550.9002

a. Predictors: (Constant), Prix B1 * Chaîne 2, Prix B1 * Chaîne 1, Prix B1, Chaîne 2, Chaîne 1

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1801.135	1056.125		1.705	.092
	Prix B1	-1695.880	1403.874	-.204	-1.208	.231
	Chaîne 1	3351.424	1514.972	1.286	2.212	.030
	Chaîne 2	3966.908	1340.397	1.650	2.960	.004
	Prix B1 * Chaîne 1	-1338.118	1929.801	-.422	-.693	.490
	Prix B1 * Chaîne 2	-2062.656	1638.183	-.842	-1.259	.212

a. Dependent Variable: Ventes B1

SPSS: Selezione - Passo 1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.880 ^a	.774	.762	549.0397

a. Predictors: (Constant), Prix B1 * Chaîne 2, Chaîne 1, Prix B1, Chaîne 2

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2331.854	725.262		3.215	.002
	Prix B1	-2404.031	959.999	-.290	-2.504	.014
	Chaîne 1	2307.281	165.548	.885	13.937	.000
	Chaîne 2	3436.189	1096.666	1.429	3.133	.002
	Prix B1 * Chaîne 2	-1354.506	1276.547	-.553	-1.061	.292

a. Dependent Variable: Ventes B1

SPSS: Selezione - Modello Finale

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.878 ^a	.771	.762	549.4825

a. Predictors: (Constant), Chaîne 2, Chaîne 1, Prix B1

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2905.955	483.360		6.012	.000
	Prix B1	-3170.066	633.276	-.382	-5.006	.000
	Chaîne 1	2357.499	158.767	.905	14.849	.000
	Chaîne 2	2291.724	198.424	.953	11.550	.000

a. Dependent Variable: Ventes B1

La Regressione Multivariata

L'approccio PLS (Proiezione su Strutture Latenti)

	q				p		
	x_1	x_2	...	x_q	y_1	...	y_p
1							
2							
i							
n							

Variabili fortemente correlate

N.ro Osservazioni < Variabili

Dati Mancanti

La Regressione *strictu sensu* non è praticabile

Regressione PLS1 - t_1

Si ricercano quelle **combinazioni lineari** t_h dei predittori che, simultaneamente, **spiegano** la variabile obiettivo y e sono **rappresentativi** dei predittori stessi

1. **Prima** componente PLS t_1 :

$$t_1 = Xw_1 = \sum_j \text{cor}(y, x_j) \times x_j$$

2. **Normalizzazione** del vettore $w_1 = (w_{11}, \dots, w_{1q})$
3. **Regressione** di y su $t_1 = Xw_1$ espresso in termini di X
4. Calcolo dei **residui** y_1 e X_1 delle **regressioni** di y e X su t_1

Regressione PLS1 - $t_2 \dots t_h$

1. **Seconda** componente PLS t_2 :

$$t_2 = X_1 b_2 = \sum_j \text{cov}(y_1, x_{1j}) \times x_{1j}$$

2. **Normalizzazione** del vettore $b_2 = (b_{21}, \dots, b_{2q})$
3. Calcolo di w_2 tale che $t_2 = X_1 b_2 = Xw_2$
4. **Regressione** di y su t_1 e t_2 in termini di X
5. Calcolo dei **residui** y_2 e X_2 delle **regressioni** di y e X su t_1 e t_2

La procedura è **iterata** per le altre componenti

Regressione PLS1: Il caso delle AUTOMOBILI

Dati Completi

Model	Price	Cm3	Power	Speed	Weig.	Length	Width
Honda Civic	83700	1396	90	174	850	369	166
Renault 19	83800	1721	92	180	965	415	169
Fiat Tipo	70100	1580	83	170	970	395	170
:							
Citroën AX Sport	66800	1294	95	184	730	350	160

Dati Incompleti

Model	Price	Cm3	Power	Speed	Weig.	Length	Width
Honda Civic	83700	.	90	174	850	369	166
Renault 19	83800	1721	.	180	965	415	169
Fiat Tipo	70100	1580	83	.	970	395	170
:							
Citroën AX Sport	66800	1294	95	184	730	350	.

Regressione multipla sui dati completi

$$R^2 = 0.847, \quad F = 15.730 \quad \text{Sig.} = 0.0001$$

Coefficients^a

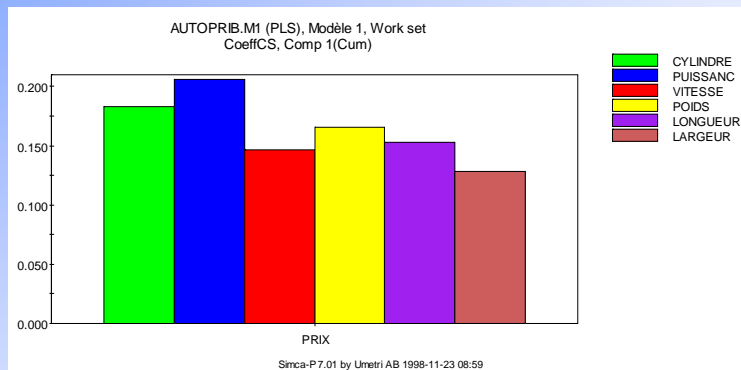
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	12070.406	194786.6		.062	.951
CYLINDRE	-1.936	33.616	-.018	-.058	.955
PUISSANC	1315.906	613.510	.888	2.145	.047
VITESSE	-472.507	740.319	-.207	-.638	.532
POIDS	45.923	100.047	.184	.459	.652
LONGUEUR	209.653	504.152	.151	.416	.683
LARGEUR	-505.429	1501.589	-.067	-.337	.741

a. Dependent Variable: PRIX

Regressione PLS sui dati incompleti

$R^2 = 0.761$

$$\frac{\text{Prix}}{\sigma(\text{Prix})} = 2.18 + 0.183\text{Cylindrée}^* + 0.206\text{Puissance}^* + 0.146\text{Vitesse}^* + 0.165\text{Poids}^* + 0.153\text{Longueur}^* + 0.129\text{Largeur}^*$$



Intervalli di confidenza per i coefficienti

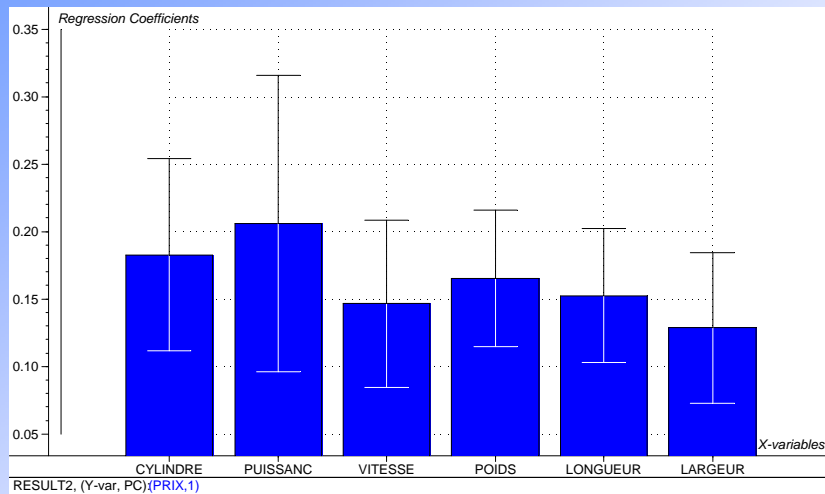
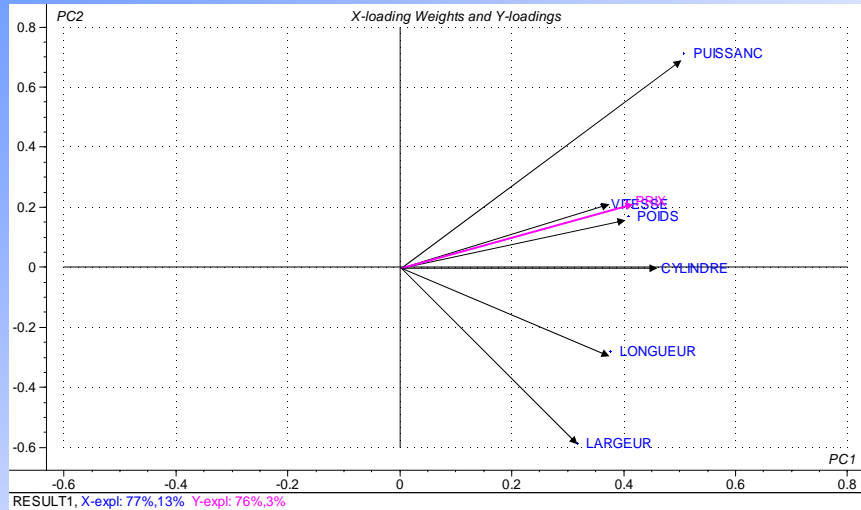
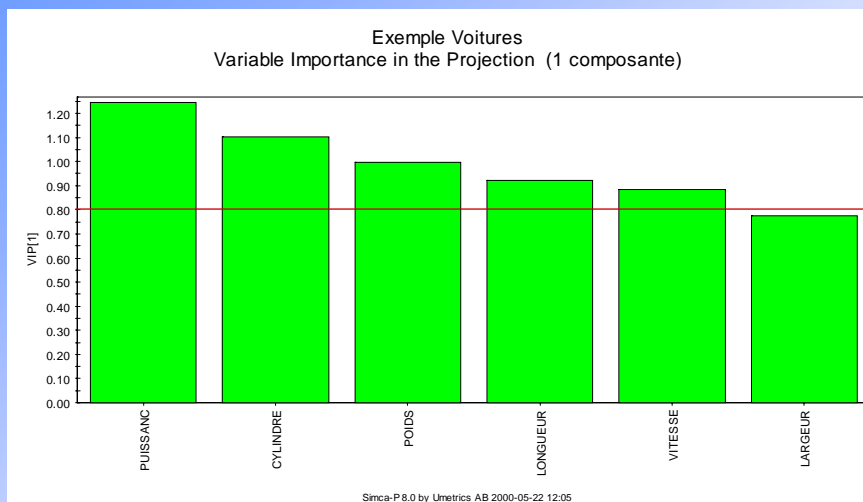


Grafico delle Variabili

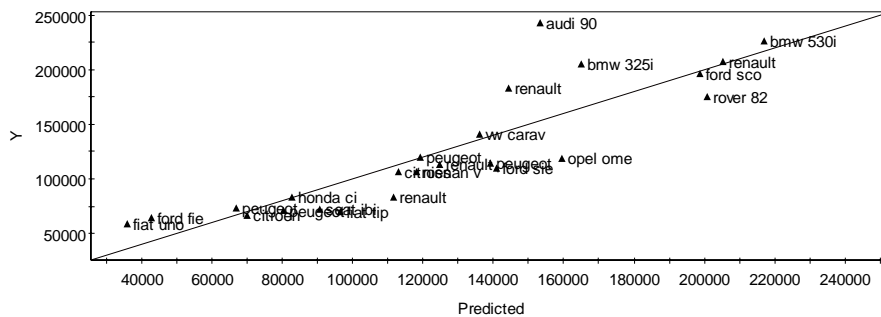


Rilevanza dei predittori - VIP



Valutazione dell'adattamento

AUTOPRIB.M1 (PLS), Modèle 1, Work set
 PRIX, Comp 1(Cum)



RMSEE=28979
 Simca-P7.01 by Umetri AB 1998-11-23 09:40

Previsione del prezzo della FERRARI 512 TR alcuni predittori sono mancanti

Prezzo Reale : 1 197 666 FF
Prezzo Previsto : 390 182 FF

	Ferrari Specifications	Standardized specifications
Displace	4943	6.65
Horsepower	428	8.74
Speed	310	5.15
Weight	1517	1.69
Length	449	0.64
Width	?	?

Regressione PLS2

Il Gusto del Tè

X I Dati

Y

Obs	Temperature	Sugar	Strength	Lemon	Subj1	...	Subj6
1	1	1	1	1	4		5
2	1	2	2	1	2		8
3	1	3	3	2	6		6
⋮							
11	1	2	1	1	1		14
⋮							
18	3	3	1	2	12		15

Temperature	Sugar	Strength	Lemon
1 = Hot	1 = W/out sugar	1 = Strong	1 = With
2 = Medium	2 = 1 tea-spoon	2 = Aver.	2 = W/out
3 = Ice	3 = 2 tea-spoons	3 = Weak	

Risultati della Regressione PLS

- **Modello PLS :**

$$4 \text{ componenti } \mathbf{t}_h = \mathbf{X} \mathbf{w}_h^*$$

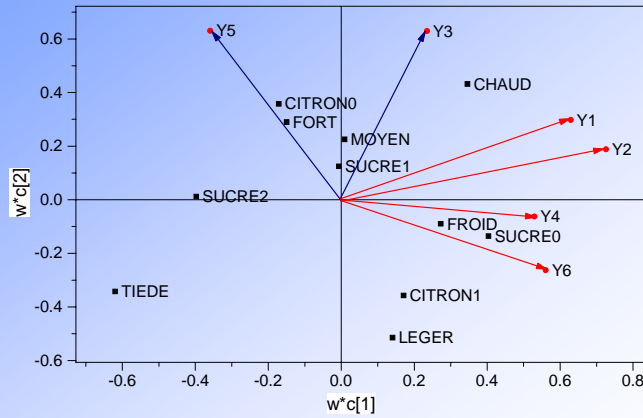
- **Regressione** di \mathbf{Y}_k rispetto a $\mathbf{t}_1, \dots, \mathbf{t}_h$:

$$\mathbf{Y}_k = \mathbf{c}_{1k} \mathbf{t}_1 + \mathbf{c}_{2k} \mathbf{t}_2 + \mathbf{c}_{3k} \mathbf{t}_3 + \mathbf{c}_{4k} \mathbf{t}_4 + \text{residuo}$$

- **Pesi** \mathbf{w}_h^* e \mathbf{c}_h utilizzati per rappresentare graficamente \mathbf{X} e \mathbf{Y} sui piani fattoriali.

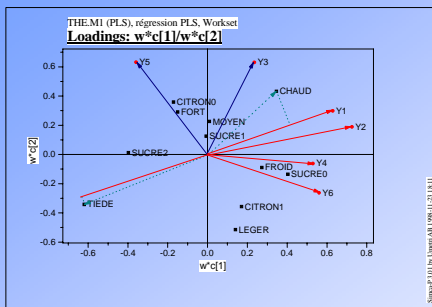
Grafico delle Variabili

THE.M1 (PLS), régression PLS, Workset
Loadings: $w*c[1]/w*c[2]$

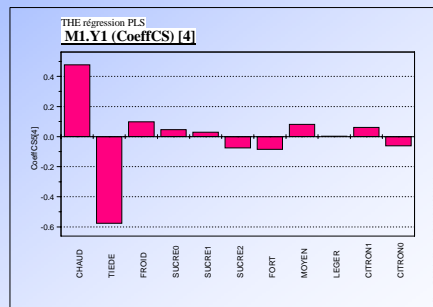


Simca-P 3.01 by Umetrics AB 1998-11-23 18:11

Rappresentazione Grafica del Modello PLS - Giudice 1



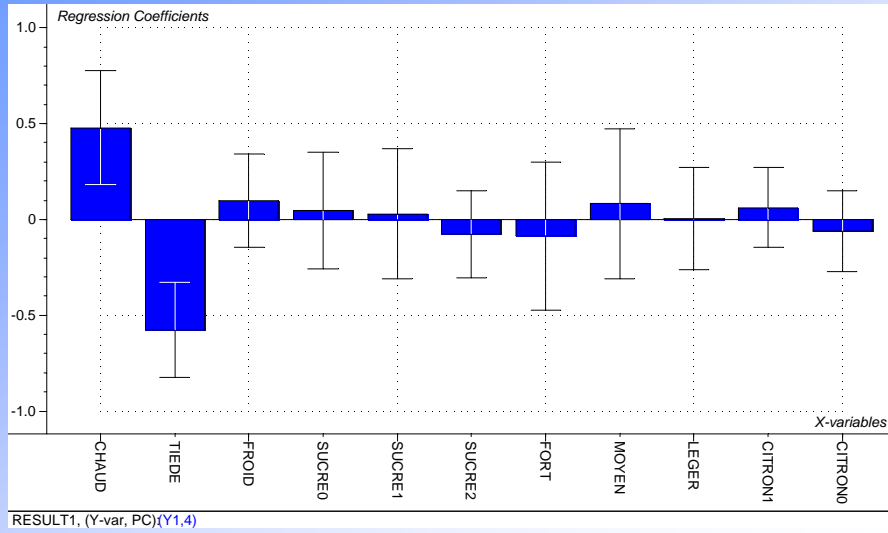
Simca-P 3.01 by Umetrics AB 1998-11-23 18:11



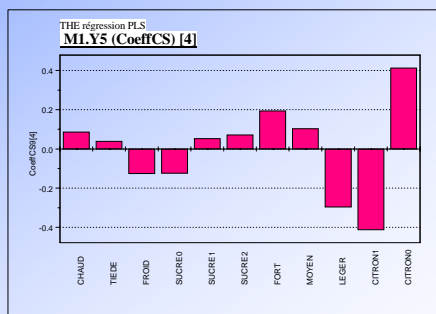
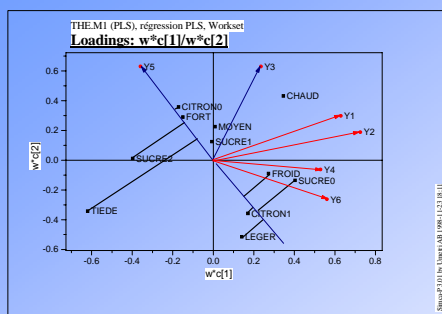
Simca-P 3.01 by Umetrics AB 1998-11-23 18:11

Al Giudice 1 il tè piace caldo e non piace affatto tiepido

Validazione del Modello per il Giudice 1

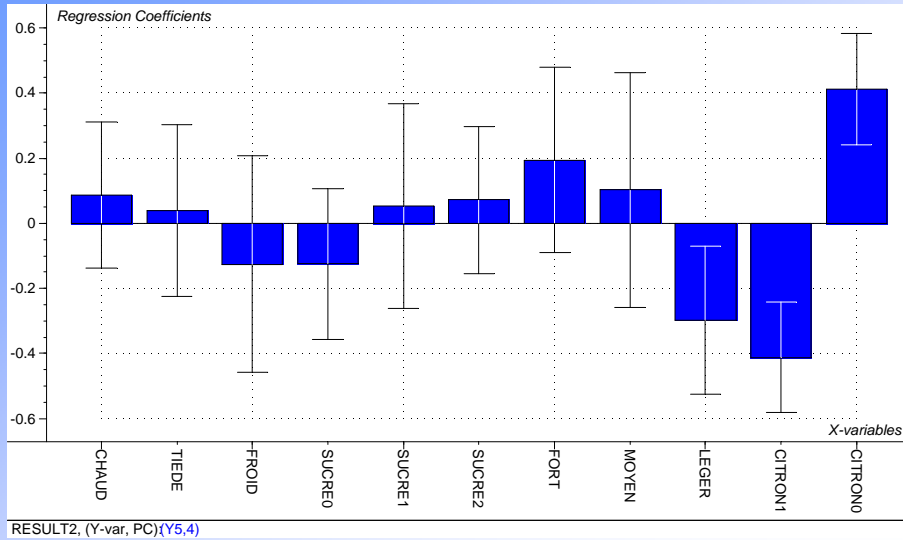


Interpretazione Grafica del Giudice 5



Il **Giudice 5** preferisce il tè **senza limone, forte, zuccherato**; è **indifferente** al tè **tiepido**; non gli piace il tè **freddo, senza zucchero, leggero, con limone**.

Validazione del Modello per il Giudice 5



Rilevanza dei predittori - VIP

