

Regressione logistica

*Metodi Quantitativi per Economia,
Finanza e Management*

Esercitazione n°11

Consegna Lavoro di gruppo

- Scadenza per la consegna del lavoro di gruppo e dell'homework facoltativo (per chi ha deciso di farlo):

Lunedì 11 Gennaio 2016

- La consegna va effettuata entro le ore 12 alla Sig.ra Enrica Luezza (Segreteria 4° Piano)
- Il materiale da consegnare consiste in:
 - stampa cartacea della presentazione in Power Point;
 - CD-ROM o chiavetta USB contenente:
 - questionario;
 - base dati in formato Excel;
 - programma SAS e output;
 - presentazione Power Point.

N.B. Il supporto elettronico (CD-ROM/chiavetta USB) non sarà restituito

Regressione logistica - Modello

Modello di regressione logistica

- si vuole modellare la relazione tra una variabile dipendente dicotomica (0-1) e un insieme di regressori che si ritiene influenzino la variabile dipendente
- la variabile dicotomica rappresenta presenza/assenza di un fenomeno (es. abbandono cliente, acquisto prodotto...)
- l'obiettivo è stimare l'equazione

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

dove $\pi := \Pr(Y=1 \mid X)$ è la probabilità che il fenomeno si verifichi

Regressione logistica – Esempio

DATA SET: banca_churn

	NAME	LABEL
1	cliente	Cliente
2	target	Target: abbandono
3	mavere	Numero movimenti avere
4	mdare	Numero movimenti dare
5	utenze	Numero utenze in c/c
6	pprod	Percentuale famiglie prodotti posseduti
7	flag_acc_sti	Accredito stipendio Y/N
8	mesi_bmov	Numero mesi bassa movimentazione ultimo semestre
9	PremiVita	Totale premi ass.ni Vita
10	NumAssDanni	Num ass.ni Danni
11	PremiDanni	Totale premi ass.ni Danni
12	AnzCliente	Anzianità cliente
13	NumAssVita	Num ass.ni Vita
14	eta	Età Cliente

Variabile Dipendente/Variabile

Target:

0: non ha abbandonato la banca

1: ha abbandonato la banca

Obiettivo:

prevedere la probabilità di abbandono a partire da un insieme di variabili indipendenti e capire come queste ultime influenzano l'esito della variabile target

Regressione logistica – Esempio

Qual è la percentuale di clienti che ha abbandonato la banca?

```
proc freq data= corso.Banca_churn ;  
  table target;  
run;
```

Target: abbandono				
target	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	31281	81.97	31281	81.97
1	6882	18.03	38163	100.00

PROC LOGISTIC – Sintassi

Modello di regressione logistica – k regressori

```
proc logistic data= dataset descending;  
  model variabile dipendente=  
    regressore_1  
    .  
    .  
    regressore_k  
/stb selection=stepwise slentry=... slstay=...;  
run;
```

ordina la variabile dipendente

elencare solo le variabili quantitative o dummy

OPTIONS:

- **selection=stepwise** applica la procedura stepwise per la selezione dei regressori
- **slentry=...** livello di significatività richiesto per il test statistico affinché il singolo regressore possa entrare nel modello
- **slstay=...** livello di significatività richiesto per il test statistico affinché il singolo regressore non sia rimosso dal modello
- **stb** coefficienti standardizzati

PROC LOGISTIC– Esempio

```
proc logistic data= corso.Banca_churn descending;  
model target=  
mesi_bmov pprod utenze mdare maverere flag_acc_sti  
eta PremiVita PremiDanni NumAssVita NumAssDanni  
AnzCliente  
/ selection=stepwise  
  slentry=0.05  
  slstay=0.05  
  stb;  
run;
```

Regressori Selezionati

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	0.3771	0.0445	71.7005	<.0001	
mesi_bmov	1	0.4456	0.00847	2764.9054	<.0001	0.3856
pprod	1	-5.3232	0.1986	718.4034	<.0001	-0.3603
utenze	1	-0.0584	0.0134	19.1126	<.0001	-0.0596
mdare	1	-0.0441	0.00317	193.4787	<.0001	-0.2414
mavere	1	-0.2835	0.0145	384.2701	<.0001	-0.4671
flag_acc_sti	1	-0.8580	0.0496	298.7248	<.0001	-0.2163

Valutazione bontà del modello

Valutazione bontà del modello (output della PROC LOGISTIC)

- 1. Percentuale di Concordant** → valuta la capacità del modello di stimare la probabilità che il fenomeno si verifichi (quanto più la percentuale è alta tanto migliore è il modello)
- 2. Test di significatività congiunta dei coefficienti (Likelihood ratio test/score test/Wald test)** → OK p-value inferiori al livello di significatività fissato
 - equivalenti al test F nella regressione lineare (valuta la capacità esplicativa del modello)
- 3. Test di significatività dei singoli coefficienti (Wald Chi_square test)** → OK p-value inferiori al livello di significatività fissato
 - equivalente al test t nella regressione lineare (valuta la significatività dei singoli coefficienti = la rilevanza dei corrispondenti regressori nella spiegazione della variabile dipendente)

Percentuale di Concordant (1/2)

Si definiscono PAIRS il numero di coppie di osservazioni (i, h con $i \neq h$) che in un caso hanno $Y=1$ e nell'altro $Y=0$.

La coppia di osservazioni (i, h con $i \neq h$) per la quale $Y_i = 1$ e $Y_h = 0$ è:

- concordante se $\hat{\pi}_i > \hat{\pi}_h$
- tied se $\hat{\pi}_i = \hat{\pi}_h$
- discordante se $\hat{\pi}_i < \hat{\pi}_h$

dove:

- π_i : probabilità stimata dal modello che $Y=1$ in corrispondenza dell'osservazione i
- π_h : probabilità stimata dal modello che $Y=1$ in corrispondenza dell'osservazione h

Tanto maggiore è il numero dei CONCORDANT (e quindi tanto minore è il numero dei DISCORDANT), tanto più il modello rappresenterà adeguatamente il fenomeno indagato.

Percentuale di Concordant (2/2)

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	86.8	Somers' D	0.738
Percent Discordant	13.0	Gamma	0.740
Percent Tied	0.2	Tau-a	0.218
Pairs	215275842	c	0.869



Percentuale di Concordant → valuta la capacità del modello di stimare la probabilità che il fenomeno si verifichi (quanto più la percentuale è alta tanto migliore è il modello)

Test di significatività congiunta dei coefficienti

Test di significatività congiunta dei coefficienti

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11618.7876	6	<.0001
Score	11214.2350	6	<.0001
Wald	6923.4937	6	<.0001

- **Likelihood ratio test**
- **Score test**
- **Wald test**

$$H_0 : \beta_1 = \dots = \beta_p = 0$$
$$H_1 : \text{almeno un } \beta_j \neq 0$$

(equivalenti al test F nella regressione lineare: valuta la capacità esplicativa del modello)

Test di significatività dei singoli coefficienti

Test di significatività per i singoli coefficienti

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	0.3771	0.0445	71.7005	<.0001	
mesi_bmov	1	0.4456	0.00847	2764.9054	<.0001	0.3856
pprod	1	-5.3232	0.1986	718.4034	<.0001	-0.3603
utenze	1	-0.0584	0.0134	19.1126	<.0001	-0.0596
mdare	1	-0.0441	0.00317	193.4787	<.0001	-0.2414
mavere	1	-0.2835	0.0145	384.2701	<.0001	-0.4671
flag_acc_sti	1	-0.8580	0.0496	298.7248	<.0001	-0.2163

- **Wald Chi_square test**

valuta la significatività dei singoli coefficienti, ossia la rilevanza dei corrispondenti regressori nella spiegazione della variabile dipendente (equivalente al test t nella regressione lineare)

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

Multicollinearità

Per valutare la presenza di multicollinearità tra i regressori, si usa la PROC REG con opzione VIF → specificare solo i regressori significativi

```
proc reg data= corso.Banca_churn ;  
model target= mesi_bmov pprod utenze  
mdare maverere flag_acc_sti  
/ vif;  
run;  
quit;
```

Si leggono solo i
valori del VIF



Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	Intercept	1	0.34349	0.00473	72.55	<.0001	0
mesi_bmov	Numero mesi bassa ... azione ultimo semest	1	0.09278	0.00113	82.19	<.0001	1.14794
pprod	Percentuale famiglie prodott	1	-0.52980	0.01649	-32.13	<.0001	1.49788
utenze	Numero utenze in c/	1	-0.00350	0.00110	-3.18	0.0015	1.51868
mdare	Numero movimenti d	1	-0.00303	0.00020598	-14.70	<.0001	1.52716
maverere	Numero movimenti avere	1	-0.00542	0.00061848	-8.76	<.0001	1.24908
flag_acc_sti	Accredito stipendio Y/N	1	-0.07997	0.00388	-20.60	<.0001	1.15148

Multicollinearità

Per risolvere il problema della multicollinearità, è necessario ricorrere ad una delle seguenti azioni:

- rimuovere le variabili indipendenti affette da multicollinearità;
- mantenere nel modello una sola variabile tra quelle indipendenti affette da multicollinearità;
- eseguire una analisi fattoriale su TUTTE le variabili indipendenti di partenza (l'esito della stepwise potrebbe essere stato influenzato dalla presenza di multicollinearità);



**Esempio di risoluzione
multicollinearità**

```
proc factor data=corso.banca_churn fuzz=0.35  
plots=scree(unpack);
```

```
var mesi_bmov pprod utenze mdare maverere flag_acc_sti eta  
PremiVita PremiDanni NumAssVita NumAssDanni AnzCliente;
```

```
quit;
```

Multicollinearità – risoluzione (1/6)

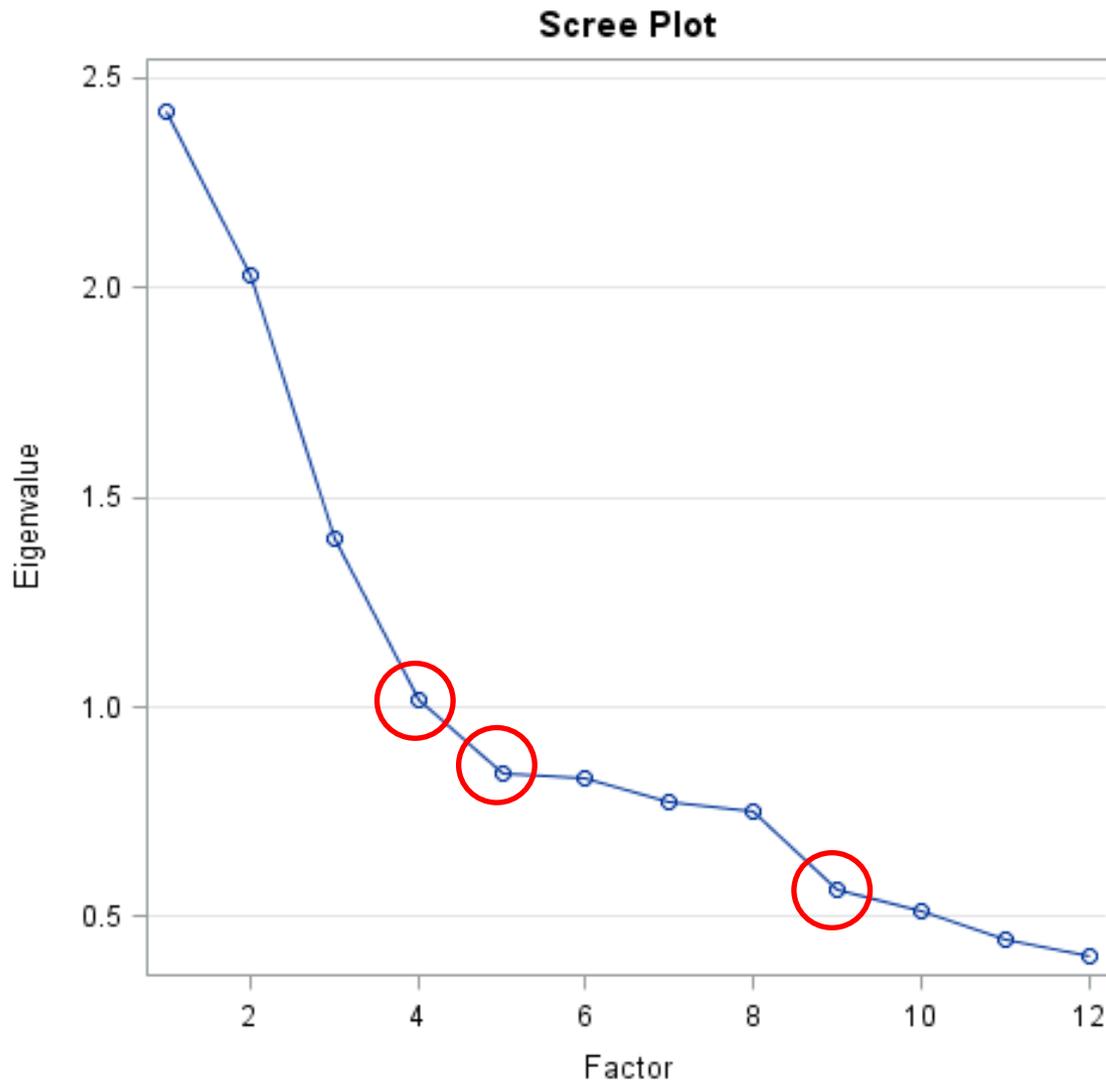
Eigenvalues of the Correlation Matrix: Total = 12 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.42220853	0.39120158	0.2019	0.2019
2	2.03100695	0.62710545	0.1693	0.3711
3	1.40390150	0.38905827	0.1170	0.4881
4	1.01484323	0.17509536	0.0846	0.5727
5	0.83974787	0.00657223	0.0700	0.6426
6	0.83317564	0.05996999	0.0694	0.7121
7	0.77320565	0.01940992	0.0644	0.7765
8	0.75379573	0.18773859	0.0628	0.8393
9	0.56605714	0.05468876	0.0472	0.8865
10	0.51136838	0.06629888	0.0426	0.9291
11	0.44506950	0.03944963	0.0371	0.9662
12	0.40561987		0.0338	1.0000

La regola degli autovalori > 1 suggerisce di prendere in considerazione **4 fattori**

Tale soluzione spiega il 57% della varianza originaria

%varianza spiegata $>60\%$

Multicollinearità – risoluzione (2/6)



Lo scree plot mostra un gomito accennato in corrispondenza del 4 fattore, e due ulteriori gomiti in corrispondenza del fattore 5 e del fattore 9.

- 4 fattori → già selezionata
- 9 fattori → n° fattori elevato rispetto a n° variabili originarie
- **5 fattori → % varianza originaria spiegata = 64 %, adeguata**

**N°fattori = circa 1/3
variabili originali →
circa 4 fattori**

Multicollinearità – risoluzione (3/6)

Confronto tra le comunalità delle soluzioni a 4 e a 5 fattori.

```
proc factor data=corso.banca_churn fuzz=0.35 N=5;  
var elenco variabili ;  
quit;
```

	4 fattori	5 fattori	6 fattori
mesi_bmov	0,29	0,30	0,86
pprod	0,58	0,58	0,58
utenze	0,56	0,56	0,57
mdare	0,74	0,74	0,74
mavere	0,41	0,43	0,63
flag_acc_sti	0,86	0,86	0,90
eta	0,41	0,86	0,89
PremiVita	0,49	0,81	0,83
PremiDanni	0,63	0,64	0,64
NumAssVita	0,61	0,61	0,61
NumAssDanni	0,73	0,73	0,73
AnzCliente	0,57	0,57	0,57

La variabile «mesi_bmov» non risulta adeguatamente spiegata dalle due soluzioni → analizziamo anche la soluzione a 6 fattori (71% varianza spiegata)

```
proc factor data=corso.banca_churn  
fuzz=0.35 N=6;  
var elenco variabili;  
quit;
```

La soluzione a 6 fattori spiega adeguatamente la variabile «mesi_bmov», ma porterebbe ad un numero di componenti principali troppo elevato rispetto alle variabili originarie.

Decidiamo di proseguire con un tentativo di **interpretazione per la soluzione a 5 fattori.**

Multicollinearità – risoluzione (4/6)

Rotazione dei fattori con il metodo Varimax ed interpretazione.

```
proc factor data=corso.banca_churn fuzz=0.35 N=5 rotate=varimax reorder;  
var mesi_bmov pprod utenze mdare maverere flag_acc_sti eta PremiVita  
PremiDanni NumAssVita NumAssDanni AnzCliente;  
quit;
```

Rotated Factor Pattern		Factor1	Factor2	Factor3	Factor4	Factor5
mdare	Numero movimenti dare	0.80207
pprod	Percentuale famiglie prodotti posseduti	0.71554
utenze	Numero utenze in c/c	0.71275
maverere	Numero movimenti avere	0.63351
mesi_bmov	Numero mesi bassa movimentazione ultimo semestre	-0.49963
NumAssDanni	Num ass.ni Danni	.	0.85400	.	.	.
PremiDanni	Totale premi ass.ni Danni	.	0.79945	.	.	.
NumAssVita	Num ass.ni Vita	.	0.76150	.	.	.
PremiVita	Totale premi ass.ni Vita	.	.	0.89820	.	.
AnzCliente	Anzianità cliente	.	.	-0.61999	.	0.41521
flag_acc_sti	Accredito stipendio Y/N	.	.	.	0.91009	.
eta	Età Cliente	0.91236

Ipotesi di
intepretazione:

**Movimentazioni
conto corrente**

**Prodotti
assicurativi**

???

Multicollinearità – risoluzione (5/6)

L'interpretazione della soluzione selezionata non è soddisfacente...

NOTA BENE!!!

In ambito di risoluzione della multicollinearità, l'analisi fattoriale non ha l'obiettivo primario di sintetizzare un elevato numero di variabili correlate tra loro.



L'esigenza di parsimonia nella selezione dei fattori diventa meno stringente!

COSA FARE?

Nell'ottica di sanare il problema della multicollinearità: se l'interpretazione della soluzione ottimale, suggerita dai criteri pratici e dal confronto delle comunalità, non è convincente, possiamo provare ad ammettere un numero più elevato di fattori, purchè vi sia un guadagno in termini di interpretabilità.

Multicollinearità – risoluzione (6/6)

Proviamo a rieseguire lo step di interpretazione aumentando di volta in volta il numero di fattori considerati (nell'esempio: 6 fattori, 7 fattori, ecc). Ci arrestiamo quando la soluzione analizzata fornisce una interpretazione soddisfacente.

```
proc factor data=corso.banca_churn fuzz=0.35 N=8 rotate=varimax reorder
out=corso.factors;
var elenco variabili ;
quit;
```

		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Interpretazione
NumAssDanni	Num ass.ni Danni	0,8688	,	,	,	,	,	,	,	Prodotti assicurativi
PremiDanni	Totale premi ass.ni Danni	0,8313	,	,	,	,	,	,	,	
NumAssVita	Num ass.ni Vita	0,7043	,	,	,	,	,	,	,	
utenze	Numero utenze in c/c	,	0,8604	,	,	,	,	,	,	Prodotti posseduti
pprod	Percentuale famiglie prodotti posseduti	,	0,7989	,	,	,	,	,	,	
mdare	Numero movimenti dare	,	0,5928	0,4969	,	,	,	,	,	
mavere	Numero movimenti avere	,	,	0,9441	,	,	,	,	,	Movimentazioni conto corrente
flag_acc_sti	Accredito stipendio Y/N	,	,	,	0,9389	,	,	,	,	Accredito Stipendio
eta	Età Cliente	,	,	,	,	0,9657	,	,	,	Età cliente
PremiVita	Totale premi ass.ni Vita	,	,	,	,	,	0,9635	,	,	Premi Vita
mesi_bmov	Numero mesi bassa movimentazione ultimo semestre	,	,	,	,	,	,	0,9712	,	Mesi bassa movimentazione
AnzCliente	Anzianità cliente	,	,	,	,	,	,	,	0,9554	Anzianità cliente

N.B.: la variabile «mdare» ha correlazioni simili con Factor 2 e Factor3 → è opportuno tener conto del significato della variabile anche nell'interpretazione di Factor3!

Stima modello

Stima del modello considerando i fattori estratti come variabili indipendenti.

```
proc logistic data= corso.factors descending;  
model target= Factor1 Factor2 Factor3 Factor4 Factor5 Factor6 Factor7  
Factor8 / selection=stepwise slentry=0.05 slstay=0.05 stb;  
run;
```

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	86.6	Somers' D	0.733
Percent Discordant	13.2	Gamma	0.735
Percent Tied	0.2	Tau-a	0.217
Pairs	215275842	c	0.867

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11467.6219	4	<.0001
Score	11086.2778	4	<.0001
Wald	6836.6442	4	<.0001

Stima modello

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-2.3736	0.0250	8984.1258	<.0001	
Factor2	1	-1.2579	0.0230	2983.8190	<.0001	-0.6935
Factor3	1	-1.1528	0.0355	1055.3886	<.0001	-0.6356
Factor4	1	-0.4792	0.0227	446.4109	<.0001	-0.2642
Factor7	1	0.9121	0.0131	4861.7200	<.0001	0.5029

Factor2 → prodotti posseduti

Factor3 → movimentazioni conto corrente

Factor4 → flag accredito stipendio

Factor7 → numero mesi bassa movimentazione ultimo semestre

Multicollinearità

Verifica della presenza di multicollinearità per il nuovo modello stimato (solo i regressori significativi).

```
proc reg data= corso.factors ;  
model target= Factor2 Factor3 Factor4 Factor7/ vif;  
run;  
quit;
```

Si leggono solo i
valori del VIF

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	Intercept	1	0.18033	0.00166	108.78	<.0001	0
Factor2		1	0.00000	0.00000	0.0000	<.0001	1.00000
Factor3		1	0.00000	0.00000	0.0000	<.0001	1.00000
Factor4		1	-0.04733	0.00166	-28.51	<.0001	1.00000
Factor7		1	0.15998	0.00166	96.50	<.0001	1.00000

Interpretazione dei Coefficienti (1/2)

Importanza dei regressori

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-2.3736	0.0250	8984.1258	<.0001	
Factor2	1	-1.2579	0.0230	2983.8190	<.0001	-0.6935
Factor3	1	-1.1528	0.0355	1055.3886	<.0001	-0.6356
Factor4	1	-0.4792	0.0227	446.4109	<.0001	-0.2642
Factor7	1	0.9121	0.0131	4861.7200	<.0001	0.5029

Si ordinano i regressori in modo decrescente rispetto al valore assoluto del coefficiente standardizzato.

Il fattore 2 (prodotti posseduti) è il regressore maggiormente influente nel modello. Seguono in termini di importanza il fattore 3, il fattore 7 e il fattore 4.

Interpretazione dei Coefficienti (2/2)

Analisi del segno dei coefficienti

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-2.3736	0.0250	8984.1258	<.0001	
Factor2	1	-1.2579	0.0230	2983.8190	<.0001	-0.6935
Factor3	1	-1.1528	0.0355	1055.3886	<.0001	-0.6356
Factor4	1	-0.4792	0.0227	446.4109	<.0001	-0.2642
Factor7	1	0.9121	0.0131	4861.7200	<.0001	0.5029

- Più è elevato il numero di prodotti posseduti, più diminuisce la probabilità di abbandono (coeff. std. Factor2= -0.6935 segno negativo)
- Più è alta la movimentazione di C/C, più diminuisce la probabilità di abbandono (coeff. std. Factor3= -0.6365 segno negativo)
- La presenza dell'accredito dello stipendio in C/C diminuisce la probabilità di abbandono (coeff. std. Factor4= - 0.2642 segno negativo)
- Più aumentano i mesi di bassa movimentazione nell'ultimo semestre, più aumenta la probabilità di abbandono (coeff. std. Factor7= 0.5029 segno positivo)

Regressione logistica – Passi da fare

- 1) Individuare la variabile oggetto di analisi (variabile dipendente dicotomica (0/1)) e i potenziali regressori (variabili quantitative o dummy).
- 2) Stimare un modello di regressione logistica utilizzando il metodo di selezione automatica STEPWISE per selezionare le variabili.
- 3) Valutare:
 - I. la bontà del modello (*percentuale di Concordant*);
 - II. la significatività congiunta dei coefficienti (*Likelihood ratio test/Score test/Wald test*);
 - III. la significatività dei singoli coefficienti stimati (*Wald Chi-square test*).

Regressione logistica – Passi da fare

- 4) Valutare la presenza di multicollinearità tra i regressori (utilizzare la PROC REG con opzione VIF).
- 5) Nel caso di multicollinearità, provvedere alla risoluzione del problema tramite una delle seguenti opzioni:
 - rimuovere le variabili indipendenti affette da multicollinearità;
 - mantenere nel modello una sola variabile tra quelle indipendenti affette da multicollinearità;
 - analisi fattoriale su tutte le variabili indipendenti di partenza.
- 6) Rieseguire gli step 2-3-4-5 fino ad individuare il modello finale.
- 7) Interpretare i coefficienti standardizzati:
 - I. stabilire tra i regressori un ordine di importanza nella spiegazione della variabile target;
 - II. valutare la direzione dell'impatto di ogni regressore sulla variabile target, tramite analisi del segno dei coefficienti.

Rischio di Credito

Nel rischio di credito la tecnica della regressione logistica viene utilizzata per prevedere se, nell'ambito di un'operazione creditizia, un debitore sarà in grado di restituire o meno il debito al suo creditore

Modelli di PD (Probability of Default)

→ Evento target da prevedere: Default / Insolvenza del debitore

Step:

1. Predisposizione di una base dati in cui, in corrispondenza di ogni debitore a cui è stato concesso un credito, vengono raccolte alcune informazioni:
 - Variabile «target» dicotomica:
 - vale 1 se il debitore è risultato insolvente (non è stato in grado di assolvere all'obbligo di restituzione del debito) nell'ultimo anno del prestito
 - vale 0 altrimenti

Rischio di Credito

- Variabili indipendenti legate a: caratteristiche o comportamenti del cliente, caratteristiche del prestito o del bene acquistato tramite il prestito...
2. Stima di un modello di regressione logistica in cui si stima la probabilità che la variabile target assuma valore 1 (default/ insolvenza) in funzione delle variabili indipendenti
 3. Utilizzo dell'equazione stimata per prevedere l'insolvenza dei debitori nell'arco dell'anno successivo

Homework

Case study: Società finanziaria che concede prestiti finalizzati all'acquisto di veicoli

Materiale fornito:

- Tabella SAS contenente: una variabile identificativa del contratto di finanziamento, la variabile dipendente dicotomica e un elenco di variabili indipendenti
- File Excel con elenco delle variabili contenute nella tabella e relativa descrizione

Obiettivo:

Stimare un modello di regressione logistica utile a prevedere la probabilità di insolvenza in funzione delle variabili indipendenti e capire quali variabili e in che modo la influenzano

Da consegnare (unitamente al materiale relativo al lavoro di gruppo e in corrispondenza della stessa scadenza):

- Presentazione Power Point (sia formato cartaceo che elettronico) con la descrizione degli step di analisi e dei risultati ottenuti
- Codice SAS utilizzato e output HTML