



User: Modèle discriminatoire de risque de crédit

```

1 .
   name: <unnamed>
   log:  D:\stata 15\Stata_15.0x64\Stata_15.0x64\Projet risque de crédit.sm
> cl
   log type:  smcl
   opened on: 31 May 2023, 20:21:57

2 . import excel "D:\données projets.xlsx", sheet("Feuil5") firstrow

3 . drop R10

4 . do "C:\Users\GIGACYTE\AppData\Local\Temp\STD1ac_000000.tmp"

5 . /* Ce projet consiste à mettre en place un modèle capable de discriminer les
> entreprises qui, souhaitent obtenir un crédit, en fonction de leurs probabilité
> s
> de défaut. Pour ce faire, nous partons d'une base de données qui, contient le
> s
> données de plisieurs entreprises qui sont jugées soit saine soit défaillante.
> C'est à partir de données qui, sont des ratios que l'on construire un modèle
> à
> même de prédire la probabilité de défaut et de distinguer les bons candidats au
>
> crédit des mauvais.
>
> Notre modèle de base est le suivant:
>
>  $Z = a_0 + a_1R_1 + a_2R_2 + a_3R_3 + a_4R_4 + a_5R_5 + a_6R_6 + a_7R_7 + a_8R_8 + a_9R_9 + a_{10}R_{10}$ 
> 0
>
> Z est le score
>  $a_0, \dots, a_{10}$  sont les paramètres du modèle
> R1 est le ratio dettes à long et moyen terme sur capitaux permanents, attendu
> (+).
> R2 est le ratio dette à long et moyen terme sur capitaux propres, attendu (+)
> .
> R3 est le ratio capitaux propres sur total passif, attendu (-).
> R4 est le ratio excédent brut d'exploitation sur chiffre d'affaire, attendu (
> -).
> R5 est le ratiocharges financières sur valeur ajoutée, attendu (+).
> R6 est le ratio charges financières sur dettes à long et moyen terme, attendu
> (+).
> R7 est le ratio cash-flows sur dette à long et moyen terme, attendu (-).
> R8 est le ratio disponibilités et assimilés sur actif courant, attendu (-).
> R9 est le ratio fournisseur sur total passif courant, attendu (+).
> */

6 .
7 . * Nous allons importer les données
8 . import excel "D:\données projets.xlsx", sheet("Feuil5") firstrow
   no; data in memory would be lost
   r(4);

   end of do-file

   r(4);

9 . do "C:\Users\GIGACYTE\AppData\Local\Temp\STD1ac_000000.tmp"

10 . * Les trois prumières lignes de la base
11 . list in 1/3

```

1.	Entrep~s E115	R1 .14743739	R2 .17583632	R3 .85827565	R4 .26643244	R5 .06354754
	R6 .11455915	R7 1.19248	R8 .06357074	R9 .40671265	Etat Saine	

2.	Entrep~s E116	R1 .81906781	R2 -.8519641	R3 .54944047	R4 .09365011	R5 .1241651
	R6 .0294069	R7 .23802834	R8 .00906947	R9 .27328842	Etat Saine	

3.	Entrep~s E117	R1 .23063395	R2 .31629682	R3 .52933829	R4 .10248399	R5 .23589674
	R6 .30431641	R7 1.1350412	R8 .01018617	R9 .11604783	Etat Saine	

12 . * Les trois dernières lignes

13 . list in -5/-1

181.	Entrep~s E286	R1 0	R2 0	R3 .01109615	R4 .08698562	R5 0
	R6 .15665217	R7 .96672851	R8 .13375388	R9 0	Etat Défaillante	

182.	Entrep~s E287	R1 .84096504	R2 5.4789514	R3 .11533746	R4 .19442183	R5 .11541796
	R6 .04493092	R7 .20452224	R8 .01222889	R9 .29503896	Etat Défaillante	

183.	Entrep~s E289	R1 0	R2 0	R3 .94832545	R4 .1130093	R5 0
	R6 .15665217	R7 .96672851	R8 0	R9 0	Etat Défaillante	

184.	Entrep~s E292	R1 0	R2 0	R3 .79909467	R4 .20470249	R5 0
	R6 .15665217	R7 .96672851	R8 0	R9 0	Etat Défaillante	

185.	Entrep~s E295	R1 .35849743	R2 .68687231	R3 .47973184	R4 .1326009	R5 0
	R6 0	R7 1.0318996	R8 0	R9 .2631403	Etat Défaillante	

14 .

```
15 . * Description des variables de la base de données
16 . describe
```

Contains data

```
  obs:      185
  vars:      11
  size:    16,465
```

variable name	storage type	display format	value label	variable label
Entreprises	str5	%9s		Entreprises
R1	double	%10.0g		R1
R2	double	%10.0g		R2
R3	double	%10.0g		R3
R4	double	%10.0g		R4
R5	double	%10.0g		R5
R6	double	%10.0g		R6
R7	double	%10.0g		R7
R8	double	%10.0g		R8
R9	double	%10.0g		R9
Etat	str12	%12s		Etat

Sorted by:

Note: Dataset has changed since last saved.

```
17 .
    end of do-file
```

```
18 . do "C:\Users\GIGACYTE\AppData\Local\Temp\STD1ac_000000.tmp"
```

```
19 . * Les valeurs manquantes
20 . misstable summarize
    (variables nonmissing or string)
```

```
21 .
22 . /* Nous allons créer une nouvelle variable que nous nommerons score. Elle pre
> ndra
> comme valeur 1 si l'entreprise est défaillante puisque c'est la probabilité de
> défaut qui est modélisée ici, et la modalité 1 si l'entreprise est saine.
> */
23 . tabulate Etat, gen (etat)
```

Etat	Freq.	Percent	Cum.
Défaillante	55	29.73	29.73
Saine	130	70.27	100.00
Total	185	100.00	

```
24 . gen score = etat1
```

```
25 . replace score = 0 if missing(score)
    (0 real changes made)
```

```
26 .
    end of do-file
```

```
27 . do "C:\Users\GIGACYTE\AppData\Local\Temp\STD1ac_000000.tmp"
```

```

28 . /* Notre base est constituée au total de 185 entreprises dont environs 30% dé
> faillantes
> et 70% saines.
> */
29 .
end of do-file

30 . do "C:\Users\GIGACYTE\AppData\Local\Temp\STD1ac_000000.tmp"

31 . /*Estimation des paramètres
> Pour estimer les paramètres de ce modèles, nous allons appliquer une régresss
> ion
> logistique.
> */
32 . logit score R1 R2 R3 R4 R5 R6 R8 R9

```

```

Iteration 0: log likelihood = -112.58302
Iteration 1: log likelihood = -97.587063
Iteration 2: log likelihood = -95.75366
Iteration 3: log likelihood = -95.420045
Iteration 4: log likelihood = -95.418422
Iteration 5: log likelihood = -95.418421

```

```

Logistic regression                                Number of obs    =      185
                                                    LR chi2(8)       =      34.33
                                                    Prob > chi2      =      0.0000
Log likelihood = -95.418421                    Pseudo R2       =      0.1525

```

score	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
R1	-.3623865	.4265608	-0.85	0.396	-1.19843	.4736574
R2	.0643425	.1178295	0.55	0.585	-.1665992	.2952841
R3	-1.34115	.7541991	-1.78	0.075	-2.819353	.1370527
R4	-1.12238	1.495541	-0.75	0.453	-4.053587	1.808827
R5	.7263621	.5726099	1.27	0.205	-.3959326	1.848657
R6	-2.668184	1.06233	-2.51	0.012	-4.750312	-.5860564
R8	-2.092274	2.502263	-0.84	0.403	-6.996619	2.812071
R9	-1.232423	.6446372	-1.91	0.056	-2.495889	.0310422
_cons	1.041625	.6114839	1.70	0.088	-.156861	2.240112

```

33 . /* Le tableau ci-dessus donne le résultat de la régression logistique. La
> colonne p>|z| indique les probabilités associées au test significativité des
> variaboes du modèle au seuil de 95%. Une probilité inférieure à 5% signifie
>
> que la variable est signicative donc peut expliquer la probabilité de défaut.
> À
> l'inverse, une probilité supérieure à 5% indique une non significativité.
>
> Dans le but de conserver uniquement les variables pertinentes, nous allons
> procéder à une progressive élimination des variables non déterminantes.
> */
34 .
35 . logit score R1 R3 R5 R6 R8 R9

```

```

Iteration 0: log likelihood = -112.58302
Iteration 1: log likelihood = -98.428405
Iteration 2: log likelihood = -96.448217
Iteration 3: log likelihood = -96.155143
Iteration 4: log likelihood = -96.15352
Iteration 5: log likelihood = -96.153519

```

```

Logistic regression                                Number of obs    =      185
                                                    LR chi2(6)       =      32.86
                                                    Prob > chi2      =      0.0000
Log likelihood = -96.153519                    Pseudo R2       =      0.1459

```

score	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
R1	-.1444409	.3450561	-0.42	0.676	-.8207384	.5318565
R3	-1.597365	.7179256	-2.22	0.026	-3.004473	-.1902568
R5	.7052851	.5038078	1.40	0.162	-.2821601	1.69273
R6	-2.658356	1.030246	-2.58	0.010	-4.6776	-.6391109
R8	-2.543702	2.509448	-1.01	0.311	-7.46213	2.374727
R9	-1.168671	.6227074	-1.88	0.061	-2.389155	.0518134
_cons	1.002687	.5402031	1.86	0.063	-.0560913	2.061466

```

36 .
37 . * Éliminons R1 et R8
38 . logit score R3 R5 R6 R9

```

```

Iteration 0:  log likelihood = -112.58302
Iteration 1:  log likelihood = -98.552514
Iteration 2:  log likelihood = -97.033532
Iteration 3:  log likelihood = -96.944582
Iteration 4:  log likelihood = -96.944511
Iteration 5:  log likelihood = -96.944511

```

```

Logistic regression              Number of obs   =      185
                                LR chi2(4)        =      31.28
                                Prob > chi2         =      0.0000
Log likelihood = -96.944511      Pseudo R2      =      0.1389

```

score	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
R3	-1.574666	.7066905	-2.23	0.026	-2.959754	-.189578
R5	.6347045	.4323372	1.47	0.142	-.2126608	1.48207
R6	-2.462433	.9136241	-2.70	0.007	-4.253103	-.6717629
R9	-1.152539	.603337	-1.91	0.056	-2.335058	.0299795
_cons	.7915202	.4693578	1.69	0.092	-.1284041	1.711445

```

39 .
40 . *Enfin, éliminons R5
41 . logit score R3 R6 R9

```

```

Iteration 0:  log likelihood = -112.58302
Iteration 1:  log likelihood = -100.40677
Iteration 2:  log likelihood = -98.857087
Iteration 3:  log likelihood = -98.803405
Iteration 4:  log likelihood = -98.803243
Iteration 5:  log likelihood = -98.803243

```

```

Logistic regression              Number of obs   =      185
                                LR chi2(3)        =      27.56
                                Prob > chi2         =      0.0000
Log likelihood = -98.803243      Pseudo R2      =      0.1224

```

score	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
R3	-1.713393	.6981169	-2.45	0.014	-3.081677	-.3451088
R6	-2.286326	.8505377	-2.69	0.007	-3.95335	-.6193031
R9	-1.271505	.596722	-2.13	0.033	-2.441058	-.1019513
_cons	.9451304	.4579427	2.06	0.039	.0475791	1.842682

```

42 .
43 . /* Les ratios R3, R6 et R9 sont significatives pour expliquer le modèle.
>
> Le modèle logit est aussi globalement significatif car la probabilité associée
> au test du rapport de vraisemblance (0) est inférieure à 5%.
>
> Les signes négatifs des coefficients associés à ces ratios révèlent qu'ils ont
> un effet négatif sur la probabilité de défaut des entreprises. En d'autres termes
> la probabilité de faire défaut diminue lorsque ces ratios augmentent.
>
> Le pseudo R2 indique que 12,24 % des fluctuations de la probabilité d'abonnement
> sont expliquées par les ratios R3, R6 et R9.
>
> */
44 .
45 . *Calcul des effets marginaux
46 . mfx

```

```

Marginal effects after logit
      y = Pr(score) (predict)
      = .23764536

```

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
R3	-.3104154	.12579	-2.47	0.014	-.556962 - .063869	.501413
R6	-.4142138	.13011	-3.18	0.001	-.669216 - .159211	.323011
R9	-.2303586	.1069	-2.15	0.031	-.439877 - .02084	.403567

```

47 . /* Interprétation des effets marginaux
> Si R3 augmente de 100%, la probabilité de défaut diminue de 31%.
> Si R6 augmente de 100%, la probabilité de défaut diminue de 41%.
> Si R9 augmente de 100%, la probabilité de défaut diminue de 23%.
> */
48 .
49 . /* Test de Hosmer-Lemeshow
> H0: L'ajustement au modèle est bon
> H1: L'ajustement au modèle n'est pas bon
> */
50 . estat gof

```

Logistic model for score, goodness-of-fit test

```

      number of observations =      185
      number of covariate patterns =      179
      Pearson chi2(175) =      154.54
      Prob > chi2 =      0.8651

```

```

51 .
52 . /* La probabilité associée au test (86,51%) est supérieure à 5% donc le modèle
> est
> logit est validé; il est compatible avec les données.
> */
53 .
54 . *Calcul du pourcentage de prédictions correctes
55 . lstat

```

Logistic model for score

Classified	True		Total
	D	~D	
+	13	8	21
-	42	122	164
Total	55	130	185

Classified + if predicted $\Pr(D) \geq .5$
True D defined as score $\neq 0$

Sensitivity	$\Pr(+ D)$	23.64%
Specificity	$\Pr(- \sim D)$	93.85%
Positive predictive value	$\Pr(D +)$	61.90%
Negative predictive value	$\Pr(\sim D -)$	74.39%
False + rate for true $\sim D$	$\Pr(+ \sim D)$	6.15%
False - rate for true D	$\Pr(- D)$	76.36%
False + rate for classified +	$\Pr(\sim D +)$	38.10%
False - rate for classified -	$\Pr(D -)$	25.61%
Correctly classified		72.97%

```

56 . /* Le pourcentage de prédictions correctes du modèle est de 72,97%.
> Soit un taux d'erreur relativement élevé, d'environ 27%.
> */
57 .
58 . * Pouvoir discriminant du modèle
59 . lroc

Logistic model for score

number of observations =      185
area under ROC curve   =    0.6968

60 . /* L'aire sous la courbe ROC vaut 0,6968 soit près de 70%.
> Le pouvoir de prédiction du modèle est relativement bon.
> */
61 .
62 . * Prédire les probabilités de défaut
63 . predict prob, pr

64 . * Afficher les probabilité de défaut des cinq premières entreprises
65 . list prob in 1/5

```

	prob
1.	.2134104
2.	.3986728
3.	.3089237
4.	.6199628
5.	.2908348

```

66 . * Celles des cinq dernières
67 . list prob in -5/-1

```

	prob
181.	.6382924
182.	.567005
183.	.2615589
184.	.313848
185.	.4473403

```

68 .
69 . * Calcul du coefficient des coefficients de cote
70 . gen ci = prob/(1-prob)

71 . * Afficher le coefficient de la 92e entreprise
72 . list ci in 92/92

```

	ci
92.	1.104342

```

73 . * Elle a presque autant de chance d'être défailante que d'être saine.
74 .
75 .
76 .
77 .
78 .
79 .
80 .
    end of do-file

81 . log close
    name: <unnamed>
    log: D:\stata 15\Stata_15.0x64\Stata_15.0x64\Projet risque de crédit.sm
> cl
    log type: smcl
    closed on: 31 May 2023, 20:51:53

```
