Résumé du document farm holdings

Nadia GHERNAOUT
Philippine RENAUDIN

1 Introduction

Le but de ce document est de nous donner introduction au techniques de scoring avec l'exemple des exploitations fermières. Il commence par nous expliquer pourquoi à l'époque ils avaient besoin d'utiliser les méthodes de scoring

Le contexte de l'époque se résumait par des réformes politiques sur l'agriculture, l'agrandissement de l'Union Européenne mais aussi des durcissements des contraintes économiques dans le secteur agricole dans les années 80. Tout cela a eu un impact sur les exploitations fermières françaises et s'est résulté par une multiplication des crises agricoles. c'est à partir de ce moment que s'est posé la question d'estimer les risques financiers en agriculture

L'expérience a démontré que les mesures de redressement financier pouvaient être efficaces pourvu que des actions préventives débutaient tôt. Donc il est important d'avoir une méthode pour la détection tôt et rapide des risques financiers en agriculture.

basé sur le concept de la viabilité des exploitations fermières

Donc par la suite nous avons quelques définitions sur la *viabilité* et l' *insolvabilité* :

— une ferme viable peut être définie comme

— une ferme qui assure au fermier un revenu équivalent à celui des autres catégories socio professionnelles

L' *insolvabilité* est définie comme la situation dans laquelle une exploitation agricole n'est pas en mesure d'honorer les obligations générées par la dette existante, à savoir le paiement des intérêts et le paiement des prêts

la méthode du « credit scoring » promet de diagnostiquer de manière préventive les soucis financiers des exploitations.

Puis on nous donne plusieurs critères qui contribueraient à la déstabilisation des fermes d'un point de vue financier : le déclin des prix des produits agricoles l'augmentation des crédits affaiblissement financier dû à : une augmentation des dépenses une baisse du chiffre d'affaire une recrudescence des incidents et des retards de paiement

1.1 Données

Les données concernent 1260 exploitations fermières qui sont réparties en 2 groupes (décrits par la variable DIFF). Le premier groupe rassemble les fermes saines (653) et le second groupe rassemble les fermes défaillantes. La variable à expliquer est donc cette variable DIFF

Pour calculer les risques financiers plusieurs ratios sont présentés on nous définit un ensemble de critères micro économiques qui calculent le degré de faillite des exploitations fermières

— CNTY : code de département
— DIFF : variable à expliquer, est ce qu'il y a déjà eu un incident de paiement (1= sain, 2= risque)
— STATUS : statut légal (1 = propriétaire indépendant, 0= entreprise)
— HECTARE : aire utilisée en hectares
— ToF : index de type de ferme
— OWNLAND : owned land (O= Oui, N= non)
— AGE : l'âge du propriétaire des terres
— HARVEST : année de récolte concernée

EBITDA = "earnings before interest, taxes, depreciation, and amortization" = excédent brut d'exploitation

Capitalisation

$$r1 = \frac{\text{Dette totale}}{\text{Totalité des actifs}} \qquad r2 = \frac{\text{Capitaux propres}}{\text{Capitaux permanents}} \qquad r3 = \frac{\text{Dette de court terme}}{\text{Dette totale}}$$

$$r4 = \frac{\text{Dette de court terme}}{\text{Totalité des actifs}} \qquad r5 = \frac{\text{Dette de long et moyen terme}}{\text{Totalité des actifs}}$$

Poids des dettes

$$r6 = \frac{\text{Dette totale}}{\text{Produit brut}} \qquad r7 = \frac{\text{Dette à long et moyen termes}}{\text{Produit Brut}} \qquad r8 = \frac{\text{Dette de court terme}}{\text{Produit brut}}$$

Liquidité

$$r11 = \frac{\text{Fonds de roulement}}{\text{Produit brut}} \qquad r12 = \frac{\text{Fond de roulement}}{\text{Entrées réelles - frais financiers}} \qquad r14 = \frac{\text{Dette de court terme}}{\text{Actifs en circulation}}$$

Debt servicing

$$r17 = \frac{\text{Frais financiers}}{\text{Dette totale}} \qquad r18 = \frac{\text{Frais financiers}}{\text{Produit brut}}$$

$$r19 = \frac{\text{Frais financiers} + \text{remboursement du capital à long et moyen terme}}{\text{Produit brut}}$$

$$r21 = \frac{\text{Frais financiers}}{\text{EBITDA}}$$

$$r22 = \frac{\text{Frais financiers} + \text{remboursement du capital à long et moyen terme}}{\text{EBITDA}}$$

Capital profitability

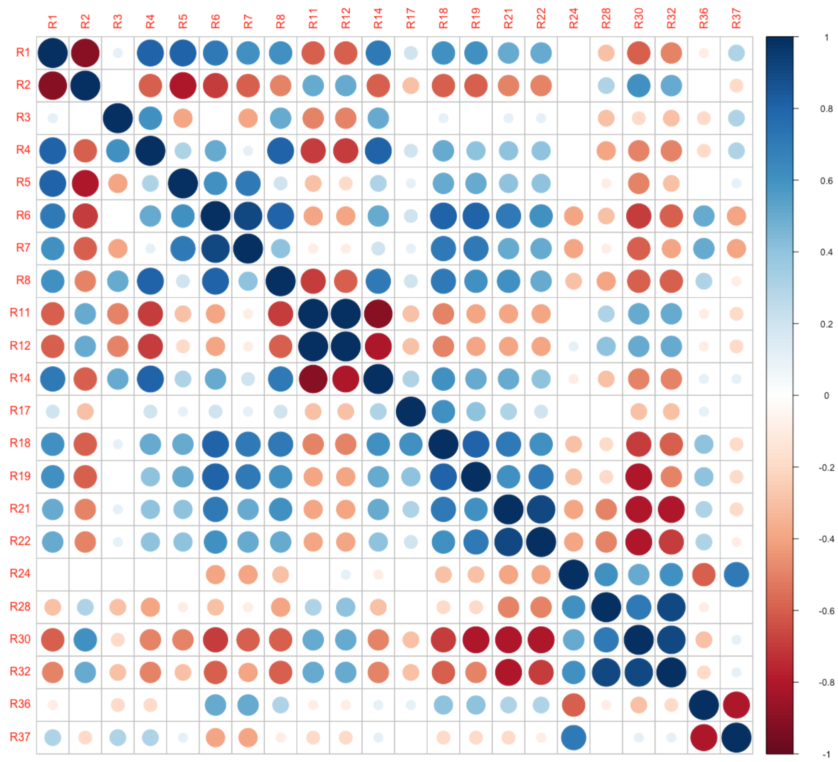
$$r24 = \frac{\text{EBITDA}}{\text{Totalité des actifs}}$$

Earnings

$$r28 = \frac{\text{EBITDA}}{\text{Produit brut}} \qquad r30 = \frac{\text{Revenu disponible}}{\text{Produit brut}} \qquad r32 = \frac{\text{EBITDA - Frais financiers}}{\text{Produit brut}}$$

Productive activity

$$r36 = \frac{\text{Actifs immobilisés}}{\text{Produit brut}} \qquad r37 = \frac{\text{Porduit brut}}{\text{Totalité des actifs}}$$

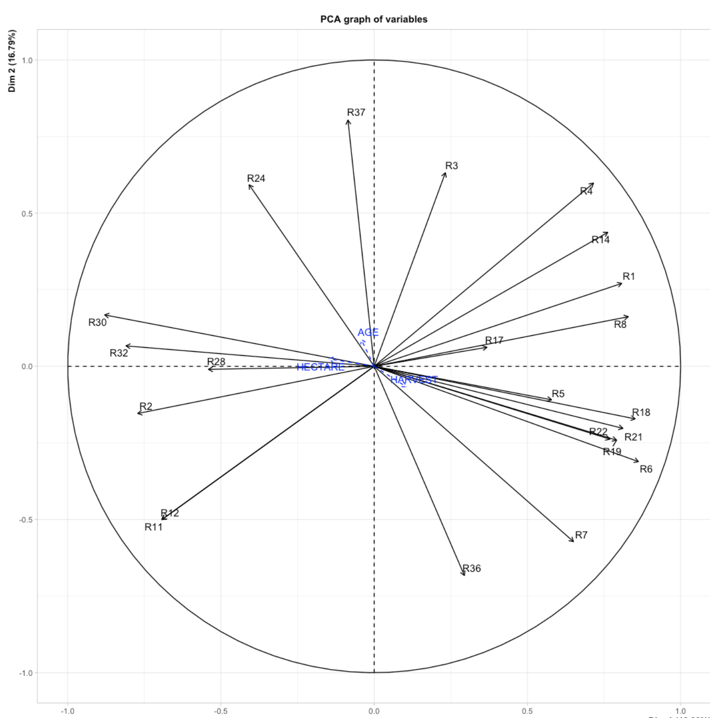


2 ACP

Pour résumer l'information contenue dans ces ratios financiers on fait une Analyse en composante principale.

On peut lire la corrélation entre les différentes variables (ratios) On a également représenté des variables supplémentaires (celles qui ne sont pas des ratios)

```
acp = PCA(farms2, quali.sup = c(1,2,3,5,6), quanti.sup = c(4,7,8))
```



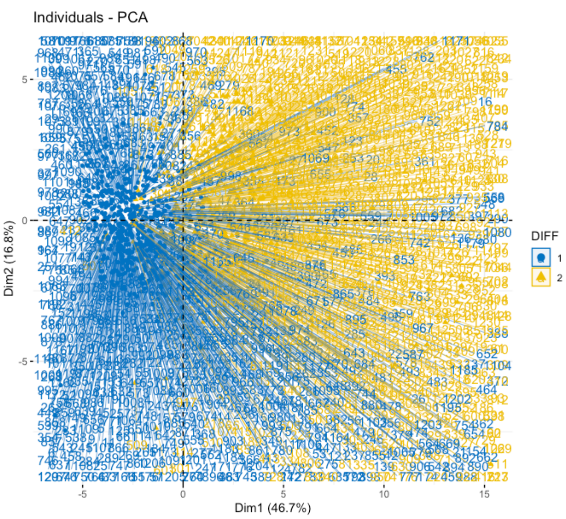
L'axe F1 représente 47% de l'information. Cet axe montre l'opposition entre deux groupes de ratios :

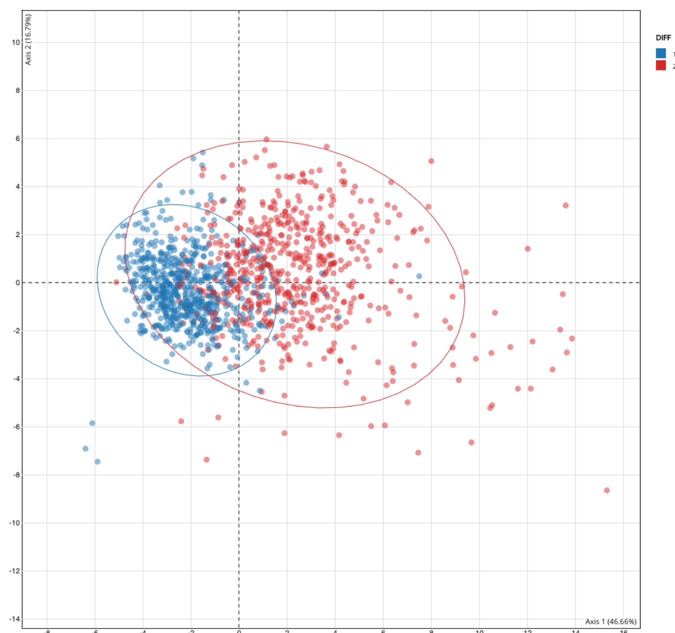
— F1>0 : dette importante en terme de Capitalisation, d'endettement et de liquidité
— F1<0 : importance des gains ? ?

Les autres variables qui décrivent les données sont supplémentaires, leurs projections dans le premier plan factoriel ne peut pas être interprété.

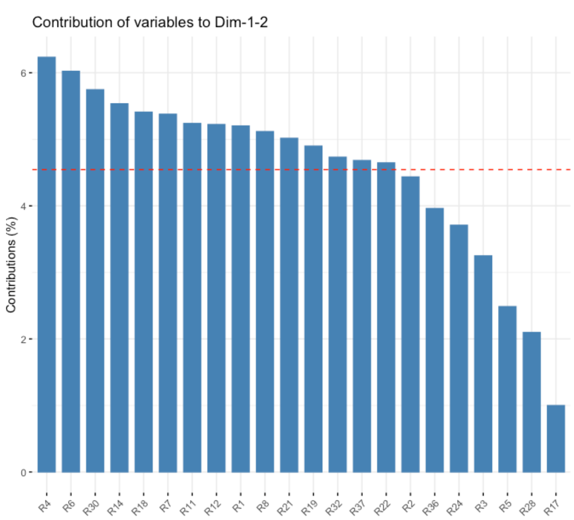
Le deuxième axe F2 représentela relation entre une efficacité productive plus ou moins grande et un montant plus ou moins élevé de dette courant par rapport à la dette totale.

```
fviz_pca_ind(acp,habillage = 2, addEllipses = TRUE, ellipse.type = "confidence",
   palette = "jco",repel = TRUE)  # ce repel=TRUE pose probleme

res <- explor::prepare_results(acp)
explor::PCA_ind_plot(res, xax = 1, yax = 2, ind_sup = FALSE, lab_var = NULL,
                     ind_lab_min_contrib = 0, col_var = "DIFF", labels_size = 9,
                        point_opacity = 0.5,
                     opacity_var = NULL, point_size = 64, ellipses = TRUE,
                        transitions = TRUE,
                     labels_positions = NULL, xlim = c(-9.03, 16.7), ylim = c(-14.
                        4, 11.3))
```



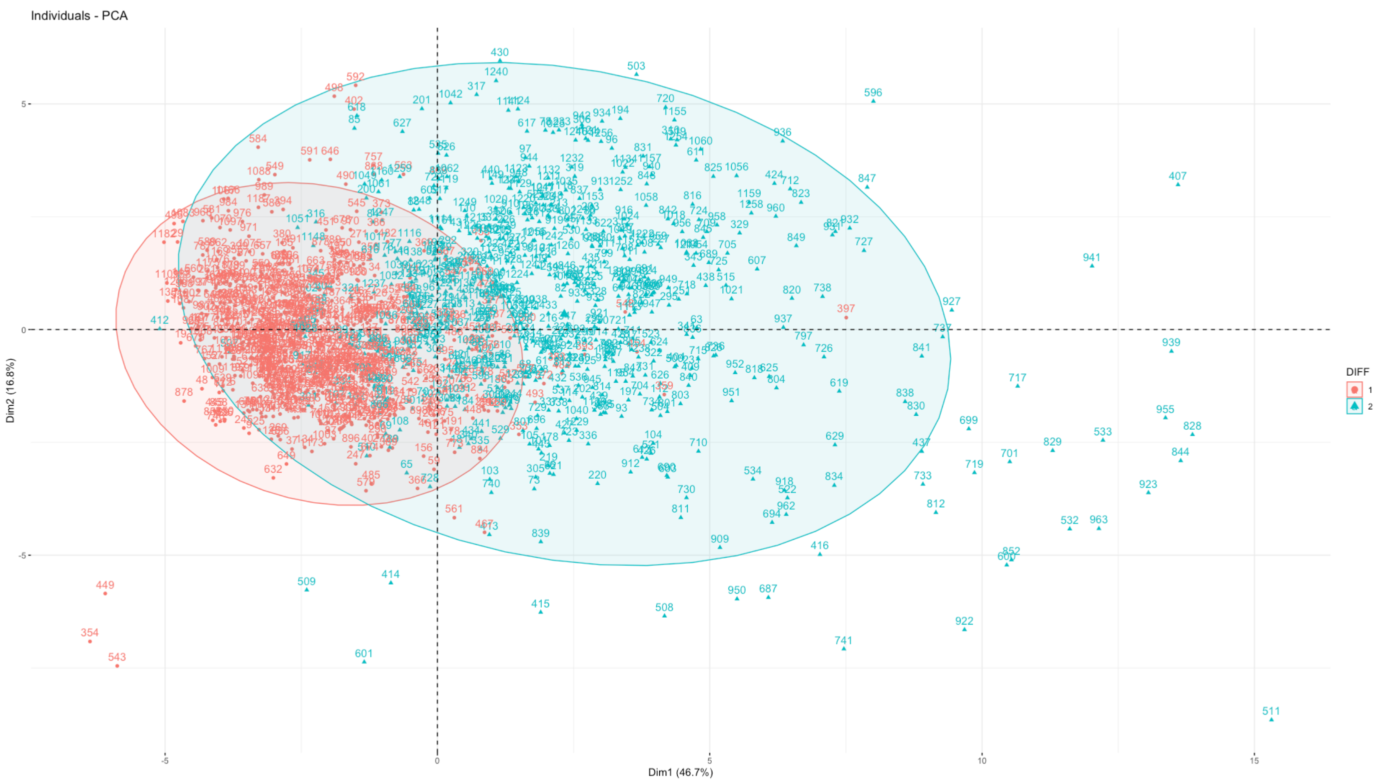



```
fviz_contrib(acp, choice = "var", axes = c(1,2))
```

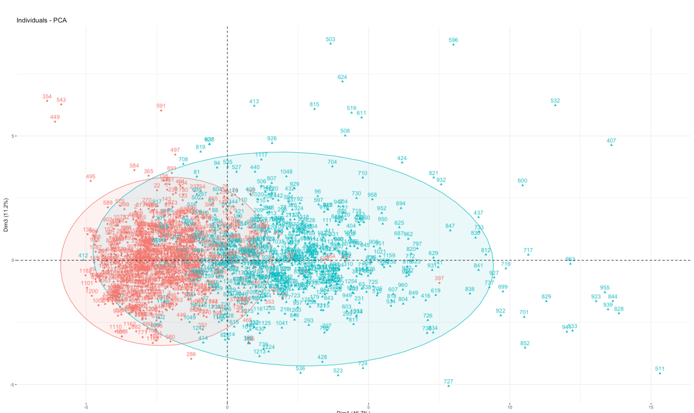


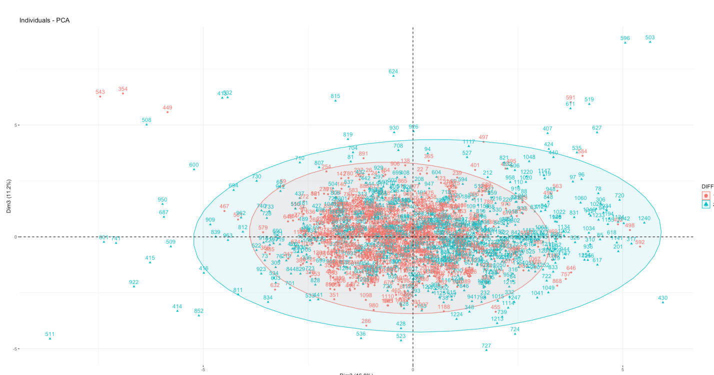
Dans les graphiques ci dessous sont représentés en rouge les fermes saines et en bleu les fermes défaillantes

```
fviz_pca_ind(acp,habillage = 2, addEllipses = TRUE)
```



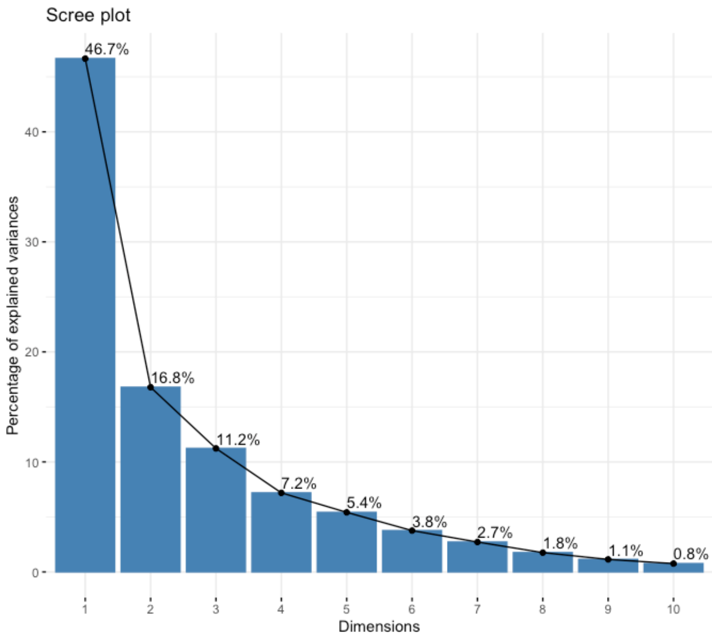

```
fviz_pca_ind(acp,habillage = 2, addEllipses = TRUE, axes = c(1,3))
fviz_pca_ind(acp,habillage = 2, addEllipses = TRUE, axes = c(2,3))
```





On peut tracer l'histogramme des valeurs propres

```
fviz_eig(acp,addlabels = TRUE)
```



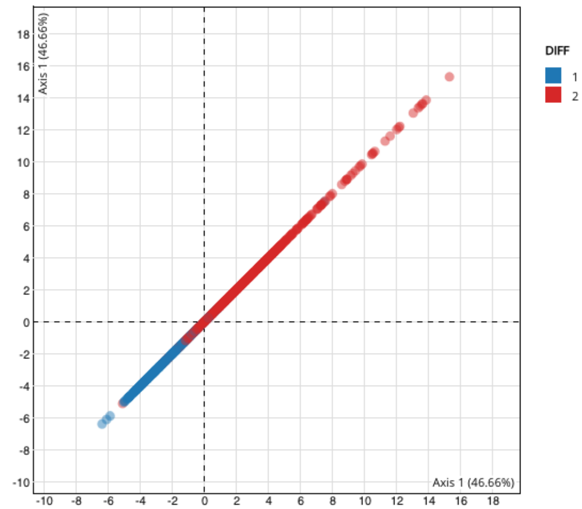

```
acp$eig

  eigenvalue percentage of variance cumulative percentage of variance
comp 1  10.264935316            46.65879689                          46.65880
comp 2   3.692949774            16.78613534                          63.44493
comp 3   2.469801899            11.22637227                          74.67130
comp 4   1.581180290             7.18718314                          81.85849
comp 5   1.191289030             5.41495014                          87.27344
comp 6   0.826545659             3.75702572                          91.03046
comp 7   0.598010331             2.71822878                          93.74869
comp 8   0.386044466             1.75474757                          95.50344
comp 9   0.250359520             1.13799782                          96.64144
comp 10  0.167045846             0.75929930                          97.40074
comp 11  0.137037934             0.62289970                          98.02364
comp 12  0.127529037             0.57967744                          98.60331
comp 13  0.093675957             0.42579981                          99.02911
comp 14  0.058505383             0.26593356                          99.29505
comp 15  0.044486217             0.20221008                          99.49726
comp 16  0.040318956             0.18326798                          99.68053
comp 17  0.022568058             0.10258208                          99.78311
comp 18  0.014627385             0.06648811                          99.84960
comp 19  0.012013792             0.05460815                          99.90420
comp 20  0.010473109             0.04760504                          99.95181
comp 21  0.006359966             0.02890893                          99.98072
comp 22  0.004242073             0.01928215                         100.00000
```

Selon la règle de Kaiser on devrait garder les 5 axes de rang supérieur (règle de la valeur propre supérieure à 1).
Mais les axes F3, F4, F5 sont assez délicats à analyser car sont assez spécifiques à certaines fermes. En effet même si les 5 axes rassemblent plus de 87 % de l'information totale nous allons garder le premier plan factoriel (les deux premiers axes) qui regroupent 63 % de l'inertie totale.

2.1 Premier indicateur : point pivot

```
library(explor)
explor(acp)
```

```
res <- explor::prepare_results(acp)
explor::PCA_ind_plot(res, xax = 1, yax = 1, ind_sup = FALSE, lab_var = NULL,
                     ind_lab_min_contrib = 0, col_var = "DIFF", labels_size = 9, point_opacity = 0.48,
                     opacity_var = NULL, point_size = 64, ellipses = FALSE, transitions = TRUE,
                     labels_positions = NULL, xlim = c(-10.7, 19.7), ylim = c(-10.7, 19.7))
```



Comme on le voit dans la figure des positions des fermes saines et défaillantes, on peut utiliser l'axe F1 comme index de classification. Pour ce faire on calcule la moyenne des individus du groupe des fermes saines pour l'axe F1 puis la même chose avec le groupe des fermes à risque. On fait la moyenne de ces deux moyennes et on obtient un point pivot.

2.2 indicateur : point pivot et axe F1

Une règle géométrique de règle de classification pour une exploitation fermière λ est de la comparer au point pivot obtenu.

```
acp_dim1 = PCA(farms2, quali.sup = c(1,2,3,5,6), quanti.sup = c(4,7,8), ncp = 1)
coordonnees = acp_dim1[["ind"]][["coord"]]
acp_dim1$quali.sup$coord

                Dim.1
CNTY.27 -0.50132116
CNTY.59  0.01207404
CNTY.61  1.10027937
CNTY.76 -0.65770420
DIFF.1  -2.16474485
DIFF.2   2.32879471
0       -0.98807931
1        0.21363877
TOF.1   -0.58660823
TOF.2    0.18611779
TOF.3    0.12327999
TOF.4    0.28008485
TOF.5    0.01429594
TOF.6    3.82298321
0        0.08882875
1       -0.13682493
```

— Ici la moyenne des individus du groupe des fermes saines pour F1 est : $\mu_0 = -2.16474485$
— la moyenne des individus du groupe des fermes défaillantes pour F1 est : $\mu_1 = 2.32879471$
— le point pivot est : $\frac{\mu_0 + \mu_1}{2} = 0.08202495$
— les fermes saines bien classées : 91,7 %
— les fermes défaillantes bien classées : 79,6 %

Avec cette méthode le taux de faux négatifs classés est trop élevé, on va donc changer d'indicateur.

2.3 indicateur r1 : pilot ratio médian

L'auteur du document a décidé d'utiliser le ratio 1 car il est très utilisé en finance (TDR) et est bien corrélé avec le premier axe. Et en France le TDR médian est de 28 % à partir de ce taux on
— les fermes saines bien classées : 66 %
— les fermes défaillantes bien classées : 99 %

2.4 indicateur r1 : pilot ratio troisième quartile

Soit 48 % TDR
— les fermes saines bien classées : 74,1 %
— les fermes défaillantes bien classées : 88,6 %

2.5 autre indicateur : avec le "indebted indicator"

Avec toujours la règle géométrique du point pivot.
— les fermes saines bien classées : 86,4 %
— les fermes défaillantes bien classées : 77,8 %

Ce qui est moins bon que l'indicateur avec l'axe F1.

Même si la première composante principale semble être un facteur discriminant convenable, on va utiliser une autre méthode celle de l'analyse discriminante qui en maximisant le score des fermes bien classées .

3 Analyse discriminante

On cherchera grâce à l'analyse discriminante et aux ratios financiers les plus pertinents à prédire l'appartenance à une ferme à un des groupes. (failing/ healthy). L'analyse discriminante permettra de construire une fonction discriminante en utilisant une règle géométrique ou en utilisant la règle bayésienne.

3.1 Cross Validation : Leave One Out

Procédure de validation croisée. Cela marche sur des petits échantillons. On enlève un individu et on prédit son groupe en fonction du résultat de tous les autres individus.
— les fermes saines bien classées : 92,6 %
— les fermes défaillantes bien classées : 83,7 % (mieux que ACP F1)

On utilise un algo qui choisit les ratios financiers les plus pertinents pour construire la fonction D (discriminante linéaire) pas à pas.

La problématique maintenant est de sélectionner ces ratios financiers et cela en estimant leur **pouvoir discriminant**.

Le critère selectif qu'on va utiliser pour calculer ce pouvoir discriminant est le lambda de Wilks.

3.2 Lambda de Wilks

La valeur du lambda de Wilks Λ varie entre 0 (pouvoir discriminant infini) et 1 (pas de pouvoir discriminant). Des petites valeurs montrent une grande différence entre les groupes et inversement.

$$\Lambda = \frac{1}{1 + \frac{F}{n-2}}$$

La table utilisant le lambda de Wilks nous donne le ratio financier avec le plus forte pouvoir discriminant : r1 (dette totale) d'autres ratios ont aussi une forte capacité discriminante comme r2, r14 , r4, r8 ...

$$T = B + W$$

avec deux groupes on ne peut obtenir qu'une fonction discriminante. Le lambda de Wilks associé à la fonction linéaire D1, est donnée par :

$$\Lambda = \frac{\det(W^*)}{\det(B^*)}$$

avec B^* la matrice de variance covariance inter groupes, et W^* la matrice de variance covariance intra groupes.

Stat de test de Khi deux à 9 ddl ($p(K-1)$) avec $p = 9$ variables et $K = 2$ groupes

p value très faible amène à la conclusion que les scores moyens des deux groupes diffèrent significativement selon la fonction linéaire discriminante. (on rejette l'hypothèse nulle d'égalité des moyennes)

3.3 Autre mesure de pouvoir discriminant

Autre mesure de la capacité discriminante de la fonction linéaire discriminante est la corrélation des coefficients

3.4 Box's M stat

Ne sert à rien car nous n'avons pas de distribution gaussienne.

3.5 Stepwise selection (minimiser lambda de Wilks)

À chaque étape k la variable X_k est sélectionnée avec la plus petite p-value ;
Sélection pas à pas en fonction des corrélations entre variables (comme la fonction `step` ou `bestglm` dans R qui va tester plusieurs combinaisons de variables et va rendre celle qui minimise l'AIC) Par exemple le ratio r1 qui était jusque là toujours gardé a été éliminé dans une étape car on pouvait le reconstituer avec une combinaison linéaire de d'autres ratios.

À partir de cela on obtient la fonction linéaire discriminante D1 (score de fisher) exprimée avec des coefficients standardisés comme combinaison linéaire des variables standardisés discriminantes.

À partir de cette fonction on peut calculer un point pivot c qui est le point médian entre healthy et failing

Pour une ferme i_0

- $D1(i_0) < c$: exploitation fermière saine
- $D1(i0) > c$: exploitation fermière défaillante
- $D1(i0) = c$: random

4 Régression logistique

Le point négatif de l'analyse discriminante est qu'elle nécessite des hypothèses qui ne sont pas toujours vérifiées :

- les variables doivent suivre une loi multinormale
- égalité des matrices de variances covariances des deux groupes

C'est pour cela qu'on utilise la régression logistique qui permet de calculer directement la probabilité qu'un certain évènement apparaît. Les hypothèses à vérifier sont moins nombreuses que pour l'analyse discriminante

On a quelques rappels sur la reglog (paramètres choisit par max de vraisemblance

4.1 Deviance

Les modèles de reglog souvent estimés par maximum de vraisemblance. La notion de déviance est introduite pour faciliter la recherche de ce maximum.

4.2 Stat de Wald

Pas pertinent ici

4.3 Stepwise

Similaire à la méthode utilisée pour l'analyse discriminante. A chaque étape k la variable X_k est sélectionnée avec la plus petite p-value.

Le principe du test pour ajouter une variable supplémentaire est d'utiliser la notion de "déflation de déviance" induite par le ratio de vraisemblance(LR)

$$G_k = -2\log(LR) = -2\log\left[\frac{\text{Vraisemblance sans } X_k}{\text{Vraisemblance avec } X_k}\right]$$

On obtient par cette méthodes les ratios : r1, r32, r14, r17, r36, r12. Les 4 premiers sont les mêmes que ceux troué en analyse discriminante en minimisant le lambda de Wilk.