Corporate Risk - Scores & Rules with R

Let us consider the data `finance` available at: `http://alexandrelourme.free.fr/scoring_data_finance`.

Introduction

1. Load the data into a R session.

```
finance <- read.table('http://alexandrelourme.free.fr/scoring_data_finance',
header=TRUE,sep=',',dec='.')
names(finance)=c('Year','Health','EBTA','VATS','QR','APTS')
attach(finance)
```

2. What is the sample size ?

```
nrow(finance)

## [1] 889
```

3. How many descriptors are there ?

```
ncol(finance)

## [1] 6
```

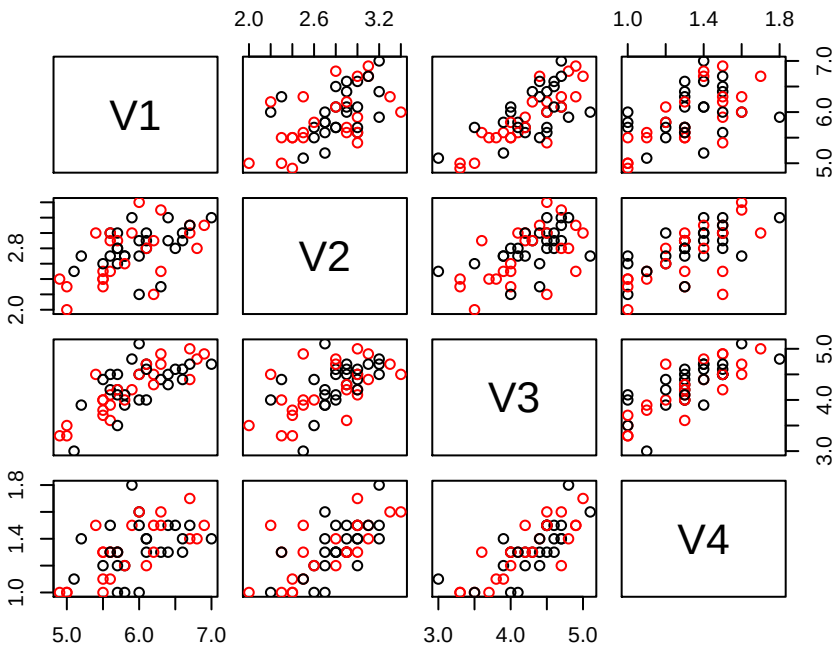
4. Please, give a more simple name to each variable:

original	Year	Health	EBITDA.Total.Assets	Value.Added.Total.Sales	Quick.Ratio	Accounts.Payable.Total.Sales
new name	Year	Health	EBTA	VATS	QR	APTS

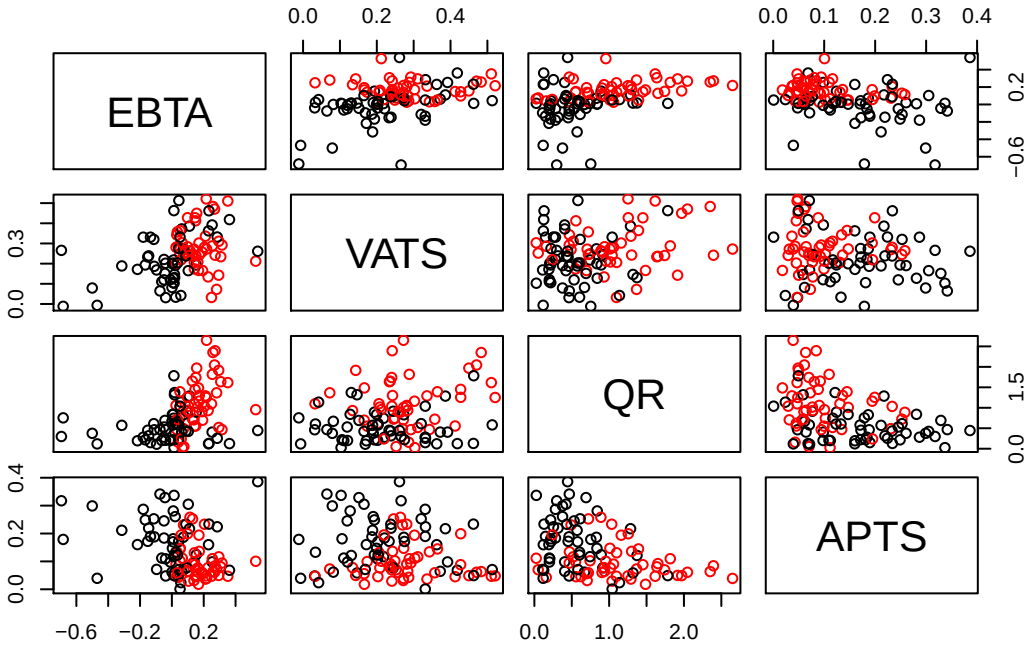
```
names(finance)=c('Year','Health','EBTA','VATS','QR','APTS')
```

5. Plot the data into all canonical planes.

```
plot(finance[,3:6])
```



```
a=sample(1:889)[1:100]
plot(finance[a,3:6],col=Health[a])
```



6. How many firms in 2002 ?

```
select02=(Year=='2002')
sum(select02)

## [1] 428
```

7. How many healthy firms in 2002 ?

```
health02=(Health=='healthy')&(Year=='2002')
sum(health02)

## [1] 216
```

8. Are there more healthy firms in 2003 than in 2002 ?

```
health02=(Health=='healthy')&(Year=='2002')
health03=(Health=='healthy')&(Year=='2003')
sum(health02)

## [1] 216

sum(health03)

## [1] 241
```

9. What is the ratio of healthy firms in 2002 ?

```
sum(health02)/sum(Year=='2002')
```

```
## [1] 0.5046729
```

```
sum(health03)/sum(Year=='2003')
```

```
## [1] 0.5227766
```

10. About the ratio of healthy firms: is it greater in 2002 or 2003 ?

A one-dimensional study

Here, we consider only healthy and bankruptcy firms of year 2002 described by EBITDA.

11. What is the mean of EBITDA in 2002 ?

```
mean(EBTA[Year=='2002'])
```

```
## [1] 0.06473636
```

12. Is EBITDA greater in average for healthy firms than for bankruptcy firms ?

```
mean(EBTA[(Year=='2002')&(Health=='healthy')])
```

```
## [1] 0.1661564
```

```
mean(EBTA[(Year=='2002')&(Health=='bankruptcy')])
```

```
## [1] -0.03859726
```

13. What is the variance of EBITDA for healthy firms in 2002 ?

```
var(EBTA[(Year=='2002')&(Health=='healthy')])
```

```
## [1] 0.01188665
```

```
var(EBTA[(Year=='2002')&(Health=='bankruptcy')])
```

```
## [1] 0.02993546
```

14. Pease, complete Table 1.

	EBITDA		
	weight	mean	variance
healthy	$w_h =$	$m_h =$	$v_h =$
bankruptcy	$w_b =$	$m_b =$	$v_b =$

Table 1

```
wh=sum((Year=='2002')&(Health=='healthy'))/sum(Year=='2002')
wb=sum((Year=='2002')&(Health=='bankruptcy'))/sum(Year=='2002')
mh=mean(EBTA[(Year=='2002')&(Health=='healthy')])
mb=mean(EBTA[(Year=='2002')&(Health=='bankruptcy')])
vh=var(EBTA[(Year=='2002')&(Health=='healthy')])
vb=var(EBTA[(Year=='2002')&(Health=='bankruptcy')])
wh;wb;mh;mb;vh;vb
```

```
## [1] 0.5046729
## [1] 0.4953271
## [1] 0.1661564
## [1] -0.03859726
## [1] 0.01188665
## [1] 0.02993546
```

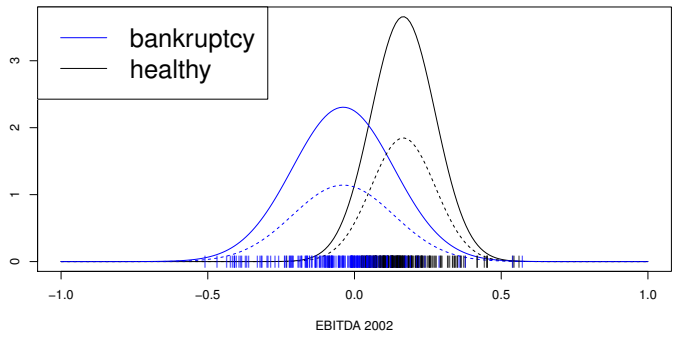


Fig. 1: *Gaussian densities of bankruptcy/healthy EBITDA values (plain line) and weighted densities (dashed line).*

15. Plot the Gaussian densities of bankruptcy/healthy EBITDA values as on Figure 1.

```
dh <- function(x){dnorm(x,mh,sqrt(vh))}
db <- function(x){dnorm(x,mb,sqrt(vb))}

xx=seq(from=-1,to=1,by=0.01)
yyh=dh(xx)
yyb=db(xx)

plot(xx,yyh,'l',xlab='EBITDA 2002',ylab='')
points(xx,yyb,'l',col='blue')

yyh=wh*dh(xx)
yyb=wb*db(xx)

points(xx,yyh,'l',lty=2)
points(xx,yyb,'l',lty=2,col='blue')

colist <- NULL
colist[Health[Year==2002]=='healthy']='black'
colist[Health[Year==2002]=='bankruptcy']='blue'

points(EBTA,rep(0,length(EBTA)),pch='|',lwd=0.01,col=colist)

legend('topleft',c('bankruptcy','healthy'),
       #lwd=c(1,1),
       lty=c(1,1),
       col=c('blue','black'),
cex=2
)
```

16. Beyond which value of a MAP score do you think a firm is healthy ?

```
score <- function(ebitda){wh*dh(ebitda)-wb*db(ebitda)}
uniroot(score,lower=-0.5,upper=0.5)

## $root
## [1] 0.04656009
##
## $f.root
## [1] -2.370593e-05
##
## $iter
## [1] 9
##
## $init.it
## [1] NA
##
## $estim.prec
## [1] 6.103516e-05
```

The first firm in Year 2003 is called Brown & Sons

17. Would you estimate Brown & Sons is healthy ?

```
Brown_ebitda=finance[Year=='2003',3][1]
score(Brown_ebitda)

## [1] 0.7426383
```

18. Your estimate about Brown & Sons, is it right ?

```
Brown_health=finance[Year=='2003',2][1]
Brown_health

## [1] bankruptcy
## Levels: bankruptcy healthy
```

We consider the MAP as an allocation rule learnt on EBITDA in 2002 and tested on EBITDA in 2003.

19. Draw the contingency table: true class $\times$ estimated class for the firms in 2003.

```
score03=sapply(finance[Year=="2003",3],score)
estimated_health03=(score03>=0)
true_health03=(finance[Year=="2003",2]=='healthy')
head(estimated_health03)

## [1]  TRUE FALSE FALSE FALSE FALSE FALSE

head(true_health03)

## [1] FALSE FALSE FALSE FALSE FALSE FALSE

table(true_health03,estimated_health03)

##              estimated_health03
## true_health03 FALSE TRUE
##         FALSE   144   76
##         TRUE     51  190
```

20. Would you claim Quick Ratio is better than EBITDA for estimating the health of firms in 2003 ?

A plain dimensional study

We aim at estimating the health of firms in 2003 by taking information on the financial features of firms in 2002.

21. Compute the centres μ_h, μ_b of the two classes (healthy/bankruptcy) in plain dimension.

```
muh=colMeans(finance[Year=='2002'&Health=='healthy',3:6])
mub=colMeans(finance[Year=='2002'&Health=='bankruptcy',3:6])
muh
```

```
##      EBTA      VATS        QR      APTS
## 0.1661564 0.2748861 1.0660966 0.1078791
```

```
mub
```

```
##        EBTA        VATS          QR        APTS
## -0.03859726  0.20694656  0.60885557  0.17741156
```

22. Is Brown & Sons closer to μ_h or μ_b ?

```
Brown=finance[Year=='2003',3:6][1,]
Brown
```

```
##        EBTA    VATS      QR    APTS
## 429 0.10286 0.23877 0.88864 0.24117
```

```
Brown=finance[Year=='2003',3:6][1,]
Brown_muh=sqrt(sum((Brown-muh)^2))
Brown_muh
```

```
## [1] 0.2335982
```

```
Brown_mub=sqrt(sum((Brown-mub)^2))
Brown_mub
```

```
## [1] 0.3215079
```

23. Would you say Brown & Sons is healthy, or not ?

24. Compare your estimation with the true status of Brown & Sons.

```
Brown_health=finance[Year=='2003',2][1]
Brown_health
```

```
## [1] bankruptcy
## Levels: bankruptcy healthy
```

John generalizes the previous rule: one firm is considered as healthy if this firm is closer to μ_h than to μ_b.

25. What is the score associated to John's rule ?

```
John_score <- function(ratios){
  u=as.matrix(ratios-muh)
  v=as.matrix(ratios-mub)
  u%*%t(u)-v%*%t(v)}
John_score(finance[Year=="2003",3:6][16,])
```

```
##           444
## 444 0.4301496
```

26. Determine the classes of the firms in 2003 according to John's rule.

```
for (i in 1:length(score03)){score03[i]=John_score(finance[Year=="2003",3:6][i,])}
estimated_health03=(score03<0)
true_health03=(finance[Year=="2003",2]=='healthy')
head(estimated_health03)

## [1]  TRUE FALSE FALSE FALSE FALSE FALSE

head(true_health03)

## [1] FALSE FALSE FALSE FALSE FALSE FALSE

table(true_health03,estimated_health03)

##              estimated_health03
## true_health03 FALSE TRUE
##         FALSE   172   48
##         TRUE     94  147
```

27. Determine John's contingency table: true class $\times$ estimated class.

28. How many firms are wrongly allocated by John in 2003 ?

29. Compute and interpret John's false positive rate.

Sam considers within each class of firm: healthy/bankruptcy, the vector of financial ratios: EBITDA.Total.Assets, Value.Added.Total.Sales, Quick.Ratio, Accounts.Payable.Total.Sales is Gaussian.

30. Compute the covariance matrices: Σ_h, Σ_b.

```
Sigmah=cov(finance[Year=='2002'&Health=='healthy',3:6])
Sigmab=cov(finance[Year=='2002'&Health=='bankruptcy',3:6])
```

Sam considers as a score the condititional probability to belong to the healthy class and one firm is estimated as healthy if the score is greater or equal to $\tau = 0,5$.

31. Brown & Sons is it healthy or not, according to Sam ?

```
library(mvtnorm)
wh*dmvnorm(Brown,muh,Sigmah)/(wh*dmvnorm(Brown,muh,Sigmah)+wb*dmvnorm(Brown,mub,Sigmab))

##       429
## 0.4284869
```

32. Compare Sam's estimation with the true class of Brown & Sons.

```
Brown_health

## [1] bankruptcy
## Levels: bankruptcy healthy
```

33. Determine Sams's contingency table: true class $\times$ estimated class.

```
Sam_score <- function(ratios){
wh*dmvnorm(ratios,muh,Sigmah)/(wh*dmvnorm(ratios,muh,Sigmah)+wb*dmvnorm(ratios,mub,Sigmab))
}
for (i in 1:length(score03)){score03[i]=Sam_score(finance[Year=='2003',3:6][i,])}
estimated_health03=(score03>=0.5)
true_health03=(finance[Year=="2003",2]=='healthy')
head(estimated_health03)
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE
```

```
head(true_health03)
```

```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE
```

```
table(true_health03,estimated_health03)
```

```
##              estimated_health03
## true_health03 FALSE TRUE
##         FALSE   159   61
##         TRUE     56  185
```

34. How many firms are wrongly allocated by Sam in 2003 ?

35. Compute Sam's false positive rate.

Sam changes the threshold τ so as to change the decision about Brown & Sons.

36. Which value of τ changes the decision about Brown & Sons.

```
Sam_score(Brown)
```

```
##       429
## 0.4284869
```

```
tau_Brown=Sam_score(Brown)
```

37. What are the estimated classes in 2003 obtained with the new threshold ?

```
estimated_health03_Brown_threshold=(score03>=tau_Brown)
```

38. How many firms are allocated differently with the new threshold ?

```
table(estimated_health03_Brown_threshold,estimated_health03)
```

```
##                                   estimated_health03
## estimated_health03_Brown_threshold FALSE TRUE
##                              FALSE   183    0
##                              TRUE     32  246
```

39. Please, update the contingency table: true class $\times$ estimated class.

```
##              estimated_health03_Brown_threshold
## true_health03 FALSE TRUE
##         FALSE   142   78
##         TRUE     41  200
```

40. How many firms are wrongly allocated with the new threshold ?