DEA Analyse et Modélisation des Systèmes Biologiques

Introduction au logiciel S-PLUS©

D. Chessel

# 3 - Modèle linéaire

## Résumé

La fiche contient le matériel nécessaire pour une séance de travaux dirigés sur S-PLUS consacrée au modèle linéaire. Elle illustre le cours de J.D. Lebreton, en particulier la régression simple, l'analyse de variance et de covariance et introduit au modèle linéaire multiplicatif.

## Plan

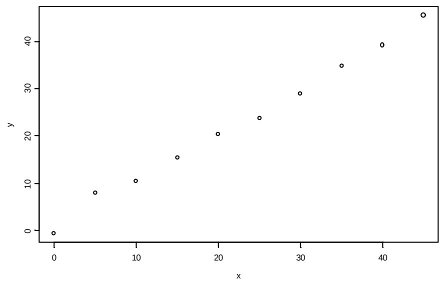
# 1 - Régression simple

Lancer S-PLUS et s'assurer qu'on est bien dans le dossier de travail désiré :

```
Working data will be in D:\Data\DEA1\_Data
```

Implanter le premier exemple proposé par Tomassone R., Charles-Bajard S. & Bellanger L. (1998) Introduction à la planification expérimentale, DEA « Analyse et modélisation des systèmes biologiques »:

```
> y
 [1]  -0.6   7.9  10.5  15.4  20.3  23.8  28.8  34.7  39.1  45.4
> x<-seq(from=0,to=45,by=5)
> x
 [1]  0  5 10 15 20 25 30 35 40 45

> plot(x,y)
```



## lm

```
> ?lm
```

**DESCRIPTION**

**Returns an object of class "lm" or "mlm" that represents a fit of a linear model.**

**USAGE**

```
lm(formula, data=<<see below>>, weights=<<see below>>,
    subset=<<see below>>, na.action=na.fail, method="qr", model=F,
    x=F, y=F, contrasts=NULL, ...)
```

**REQUIRED ARGUMENTS**

**formula a formula object, with the response on the left of a ~ operator, and the terms, separated by + operators, on the right.**

```
> lm(y~x)
Call:
lm(formula = y ~ x)

Coefficients:
 (Intercept)      x
      0.7909 0.9662

Degrees of freedom: 10 total; 8 residual
```

```
Residual standard error: 1.164
```

# *Un modèle linéaire est un objet*

```
> lm1<-lm(y~x)
> lm1
Call:
lm(formula = y ~ x)

Coefficients:
 (Intercept)      x
      0.7909 0.9662

Degrees of freedom: 10 total; 8 residual
Residual standard error: 1.164
```

# *lm1 est de la classe lm*

```
> class(lm1)
[1] "lm"
```

# *La classe lm est une sous-classe de la classe list*

```
> is.list(lm1)
[1] T
```

# *lm1 est une collection de 11 composantes*

```
> length(lm1)
[1] 11
> names(lm1)
 [1] "coefficients"  "residuals"     "fitted.values" "effects"
 [5] "R"             "rank"          "assign"        "df.residual"
 [9] "contrasts"     "terms"         "call"
```

# *Noms et numéros des composantes de lm1*

```
> lm1[[1]]
 (Intercept)      x
      0.7909 0.9662
> lm1$coefficients
 (Intercept)      x
      0.7909 0.9662
> lm1[[2]]
      1     2       3      4      5      6       7       8       9
 -1.391 2.278 0.04727 0.1164 0.1855 -1.145 -0.9764 0.09273 -0.3382
    10
 1.131
> lm1$residuals
      1     2       3      4      5      6       7       8       9
 -1.391 2.278 0.04727 0.1164 0.1855 -1.145 -0.9764 0.09273 -0.3382
    10
 1.131
```

# *Le calcul est possible sur les composantes*

```
> 2*lm1[[1]]
 (Intercept)     x
       1.582 1.932
```

# *Fonctions génériques :* **summary**

```
> summary(y)
 Min. 1st Qu. Median Mean 3rd Qu. Max.
 -0.6    11.7     22 22.5    33.2 45.4

> summary(lm1)
```

```
Call: lm(formula = y ~ x)
Residuals:
   Min     1Q Median    3Q  Max
 -1.39 -0.817   0.07 0.168 2.28

Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept)  0.791  0.684      1.156   0.281
          x  0.966  0.026     37.691   0.000

Residual standard error: 1.16 on 8 degrees of freedom
Multiple R-Squared: 0.994
F-statistic: 1420 on 1 and 8 degrees of freedom, the p-value is 2.69
e-010

Correlation of Coefficients:
  (Intercept)
x -0.843
```

L'ordonnée à l'origine n'est pas significativement non nulle :

```
> lm2<-lm(y~-1+x)
> lm2
Call:
lm(formula = y ~ -1 + x)

Coefficients:
      x
 0.9912

Degrees of freedom: 10 total; 9 residual
Residual standard error: 1.186
> summary(lm2)

Call: lm(formula = y ~ -1 + x)
Residuals:
    Min     1Q Median    3Q  Max
 -0.979 -0.587  0.243 0.574 2.94

Coefficients:
   Value Std. Error t value Pr(>|t|)
x  0.991  0.014     70.560   0.000

Residual standard error: 1.19 on 9 degrees of freedom
Multiple R-Squared: 0.998
F-statistic: 4980 on 1 and 9 degrees of freedom, the p-value is 1.17
e-013
```

# *Fonctions génériques :* plot

```
> ?plot
```

**DESCRIPTION**

**Creates a plot on the current graphics device.**

**This function is generic (see Methods); method functions can be written to handle specific classes of data. Classes which already have methods for this function include:**
**data.frame, design, factor, formula, gam, glm, lm, loess, preplot.gam, preplot.loess, profile, stl, surv.fit, times, tree, tree.sequence.**

**USAGE**

**plot(x, ...)**

**REQUIRED ARGUMENTS**

**x an S-PLUS object.**

```
> ?plot.lm
```

**DESCRIPTION**

**Creates a set of plots suitable for assessing a fitted linear model of class "lm".**

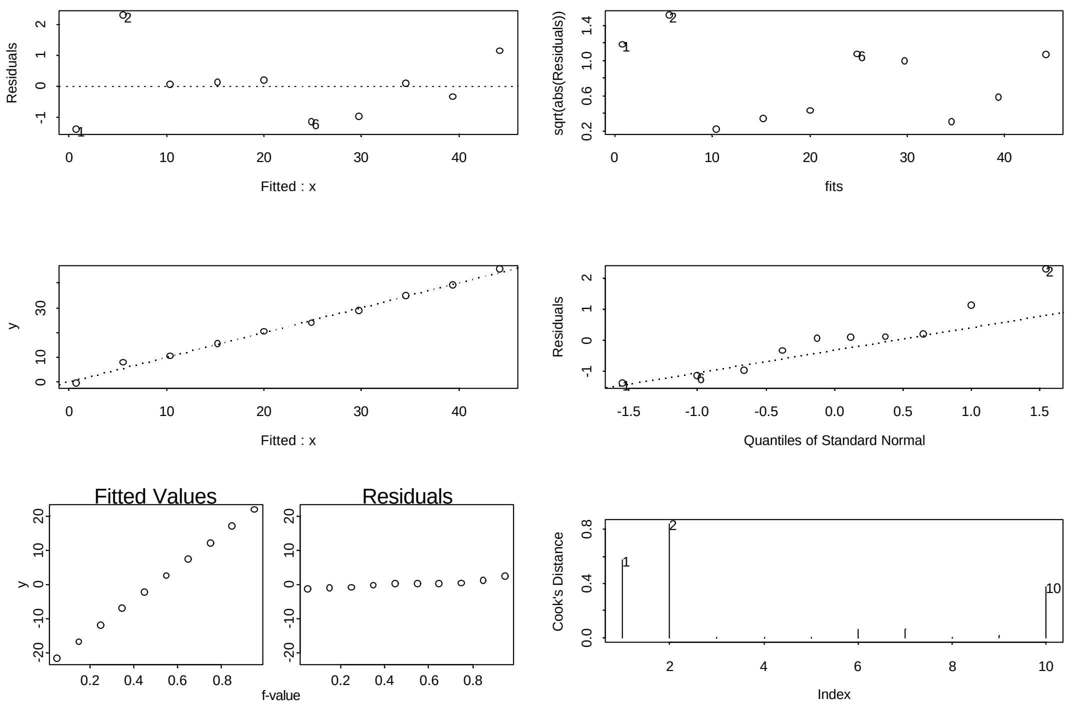
**USAGE**

**plot.lm(lm.obj, residuals = NULL, smooths = F, rugplot = F, id.n = 3, ask = F, ...)**

**REQUIRED ARGUMENTS**

**lm.obj an lm object.**

```
> plot(lm1) Que se passe t'il ?
> plot(lm1,ask=T)
> par(mfrow=c(3,2)) Pourquoi ?
> plot(lm1)
```




*Graphique standard associé à un modèle linéaire*

*1) Résidus en fonction des valeurs prédites*
*2) Racine des valeurs absolues des résidus en fonction des valeurs prédites*
*3) Valeurs observées en fonction des valeurs prédites*
*4) Graphique quantile-quantile normal des résidus (normalité des résidus). N.B. Chacun des graphiques proposés est issu d'une recherche approfondie. Le qq-plot est de Wilk M.B. & Gnanadesikan R. (1968). Probability plotting methods for the analysis of data.* Biometrika*, 55, 1-17 validé par Cleveland W.S. (1994)* The elements of graphing data. *Hobart Press, Summit, New Jersey, p. 143. Les modes de lecture sont décrits dans des ouvrages célèbres comme Tuckey J.W. (1977)* Exploratory data analysis, *Adsison-Wesley, Reading, Massachussets. Ici, les résidus sont sur- dispersés*

*par rapport à une loi normale (cf. du Toit S.H.C., Steyn A.G.W. & Stumpf R.H. (1986)* Graphical Exploratory data analysis, *Springer-Verlag, , New-York, p. 49). Ouvrages classiques : Chambers J.M., Cleveland W.S., Kleiner B. & Tukey P.A. (1983)* Graphical methods for data analysis, *Wadsworth, Belmont, California. Cleveland W.S. (1993)* Visualizing data, *Hobart Press, Summit, New Jersey.*

*5) graphique r-f (r pour residuals, f pour fitted). A gauche, en abscisse le rang des valeurs prédites sur [0,1], en ordonnée les valeurs prédites centrées (fonction de répartition inversée des prédictions). A droite à la même échelle en abscisse le rang des résidus sur [0,1], en ordonnée les valeurs observées des résidus(fonction de répartition inversée des résidus). Le couple permet de comparer l'étendue de la distribution des observations qu'on espère beaucoup plus grande que celle des résidus. Ce graphe exprime le rapport variance expliquée - variance résiduelle.*

*6) Graphe des distances de Cook. Donne pour chacun des points de mesure la distance entre les paramètres estimés par la régression avec et sans ce point. Si l'importance du rôle de chaque point est concentré sur quelques valeurs, la régression n'est pas bonne (prise en compte de points aberrants). Voir Cook, R. D. and Weisberg, S. (1982).* Residuals and Influence in Regression*. Chapman and Hall, New York.*

```
 > cor(x,y)
[1] 0.997
> cor(x,y)*cor(x,y)
[1] 0.994
```

On peut refaire l'expérience :

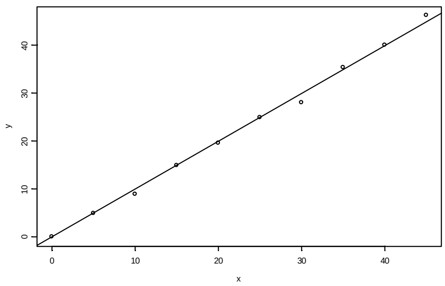
```
> x
 [1]  0  5 10 15 20 25 30 35 40 45
```

# rnorm

```
> e<-rnorm(10)
> e
 [1]  0.008629 -0.038239 -1.016802 -0.132446 -0.360349 -0.033747
 [7] -1.883161  0.336839 -0.000354  1.206677
```

# *Calcul vectoriel*

```
> y<-x+e
> y
 [1]  0.008629  4.961761  8.983198 14.867554 19.639651 24.966253
 [7] 28.116839 35.336839 39.999646 46.206677
> par(mfrow=c(1,1)) (Sinon que se passe t'il ?)
> plot(x,y)
> abline(0,1)
```




# abline

```
> ?abline
> abline(lm(y~x)) est-ce possible ?
> abline(lm(y~-1+x)) est-ce possible ?
```

## 2 -Analyse de variance

Reprendre l'exemple introduit dans la fiche 1 (p. 3).

# lm, anova

La richesse dépend de l'heure :

```
> lm1<-lm(ric~heu.fac)
> anova(lm1)
Analysis of Variance Table

Response: ric

Terms added sequentially (first to last)
            Df Sum of Sq Mean Sq F Value Pr(F)
  heu.fac    1      3071    3071   208.7     0
Residuals 1313     19325      15
```

La richesse dépend de la semaine :

```
> lm2<-lm(ric~heu.fac+sem.fac)
> anova(lm2)
Analysis of Variance Table

Response: ric

Terms added sequentially (first to last)
            Df Sum of Sq Mean Sq F Value Pr(F)
  heu.fac    1      3071    3071   293.8     0
  sem.fac   51      6133     120    11.5     0
Residuals 1262     13192      10
```

La richesse dépend de la station :

```
> lm3<-lm(ric~heu.fac+sem.fac+sta.fac)
> anova(lm3)
Analysis of Variance Table

Response: ric

Terms added sequentially (first to last)
            Df Sum of Sq Mean Sq F Value Pr(F)
  heu.fac    1      3071    3071   396.6     0
  sem.fac   51      6133     120    15.5     0
  sta.fac   13      3519     271    35.0     0
Residuals 1249      9673       8
```

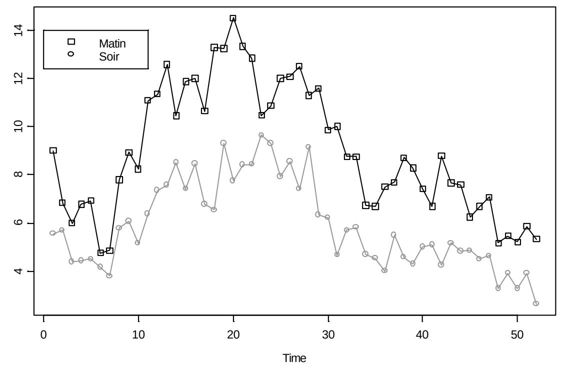
Régression polynomiale :

```
> lm4<-lm(ric~heu.fac+poly(sem,2)+sta.fac)
> anova(lm4)
Analysis of Variance Table

Response: ric

Terms added sequentially (first to last)
               Df Sum of Sq Mean Sq F Value Pr(F)
     heu.fac    1      3071    3071   338.5     0
poly(sem, 2)    2      4025    2012   221.8     0
     sta.fac   13      3523     271    29.9     0
   Residuals 1298     11777       9
> lm5<-lm(ric~heu.fac+poly(sem,3)+sta.fac)
```

```
> anova(lm5)
Analysis of Variance Table

Response: ric

Terms added sequentially (first to last)
               Df Sum of Sq Mean Sq F Value Pr(F)
     heu.fac    1      3071    3071   356.6     0
poly(sem, 3)    3      4614    1538   178.6     0
     sta.fac   13      3541     272    31.6     0
   Residuals 1297     11170       9
> anova(lm4,lm5)
Analysis of Variance Table

Response: ric

                             Terms Resid. Df   RSS    Test Df
1 heu.fac + poly(sem, 2) + sta.fac      1298 11777
2 heu.fac + poly(sem, 3) + sta.fac      1297 11170 1 vs. 2  1
  Sum of Sq F Value     Pr(F)
1
2     607.2    70.5 1.11e-016
```

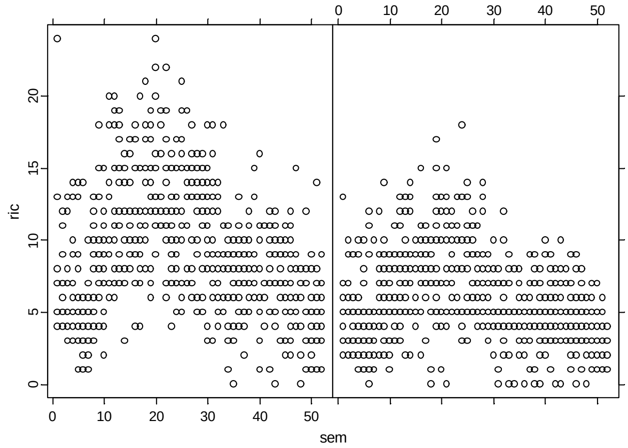
Interactions :

```
> lm6<-lm(ric~heu.fac*poly(sem,3)*sta.fac)
> anova(lm6)
Analysis of Variance Table

Response: ric

Terms added sequentially (first to last)
                               Df Sum of Sq Mean Sq F Value  Pr(F)
                      heu.fac    1      3071    3071   431.3 0.0000
                 poly(sem, 3)    3      4614    1538   216.0 0.0000
                      sta.fac   13      3541     272    38.2 0.0000
         heu.fac:poly(sem, 3)    3       200      67     9.4 0.0000
              heu.fac:sta.fac   13       509      39     5.5 0.0000
         poly(sem, 3):sta.fac   39      1671      43     6.0 0.0000
heu.fac:poly(sem, 3):sta.fac   39       222       6     0.8 0.8047
                    Residuals 1203      8567       7
```

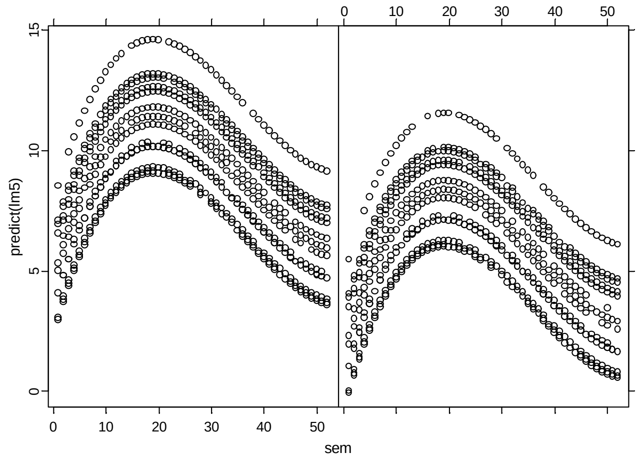



On peut mettre la même courbe pour le matin et le soir (à une constante près).

```
> coplot(ric~sem|heu.fac,show.given = F)
```




```
> coplot(predict(lm5)~sem|heu.fac,show.given = F)
```

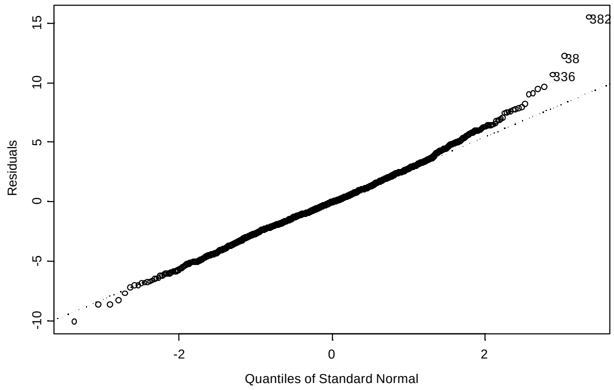



*Modèle additif sans intercation*

| |
|---|
| *Les modèles sont des objets* |

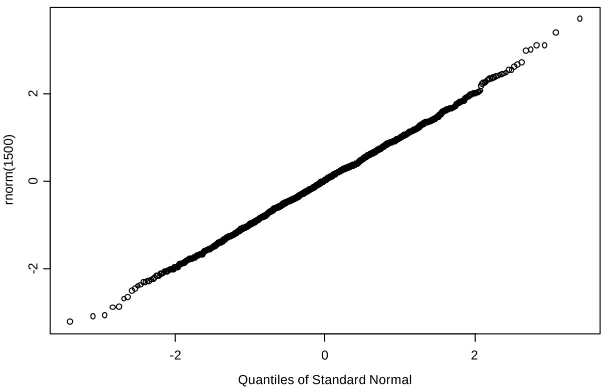
```
> lm5
Call:
lm(formula = ric ~ heu.fac + poly(sem, 3) + sta.fac)

Coefficients:
 (Intercept) heu.fac poly(sem, 3)1 poly(sem, 3)2 poly(sem, 3)3
       7.407  -1.521        -29.19         -56.6         24.66
 sta.fac1 sta.fac2 sta.fac3 sta.fac4 sta.fac5 sta.fac6 sta.fac7
   -1.572   0.3091  -0.2556   0.7351  -0.2176   0.2248   -0.327
 sta.fac8 sta.fac9 sta.fac10 sta.fac11 sta.fac12 sta.fac13
  -0.1317 -0.01319   -0.1864   0.03339    0.1642    0.1017

Degrees of freedom: 1315 total; 1297 residual
Residual standard error: 2.935
> summary(lm5)

Call: lm(formula = ric ~ heu.fac + poly(sem, 3) + sta.fac)
Residuals:
   Min    1Q  Median   3Q  Max
 -10.1 -1.87 -0.0598 1.82 15.5

Coefficients:
                Value Std. Error t value Pr(>|t|)
  (Intercept)   7.407   0.081     91.293   0.000
      heu.fac  -1.521   0.081    -18.775   0.000
poly(sem, 3)1 -29.194   2.940     -9.930   0.000
poly(sem, 3)2 -56.602   2.942    -19.238   0.000
```

```
poly(sem, 3)3  24.658   2.937      8.397   0.000
     sta.fac1  -1.572   0.209     -7.516   0.000
     sta.fac2   0.309   0.124      2.501   0.013
     sta.fac3  -0.256   0.088     -2.908   0.004
     sta.fac4   0.735   0.068     10.775   0.000
...

> plot(lm5,ask=T)

Make a plot selection (or 0 to exit):

1: plot: All
2: plot: Residuals vs Fitted Values
3: plot: Sqrt of abs(Residuals) vs Fitted Values
4: plot: Response vs Fitted Values
5: plot: Normal QQplot of Residuals
6: plot: r-f spread plot
7: plot: Cook's Distances
Selection: 5
```

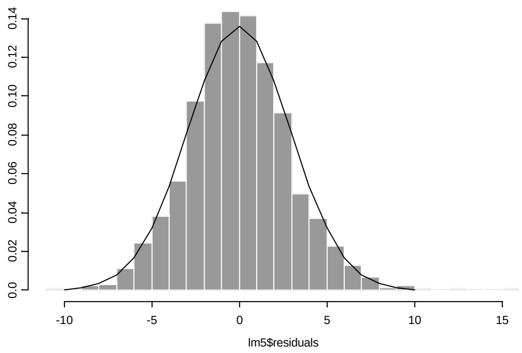



*Graphes quantiles-quantiles : Normalité des résidus*

```
> qqnorm(rnorm(1500))
```




```
> ecrin[382,]
    STA SEM HEU RIC
382   5   1   1  24
> predict(lm5)[382]
   382
 8.522
> ecrin$RIC[382]-predict(lm5)[382]
   382
 15.48 un point aberrant ?
> hist(lm5$residuals,proba=T,nclass=30)
> lines(seq(-10,10,1),dnorm(seq(-10,10,1),mean=0,sd=2.935))
```



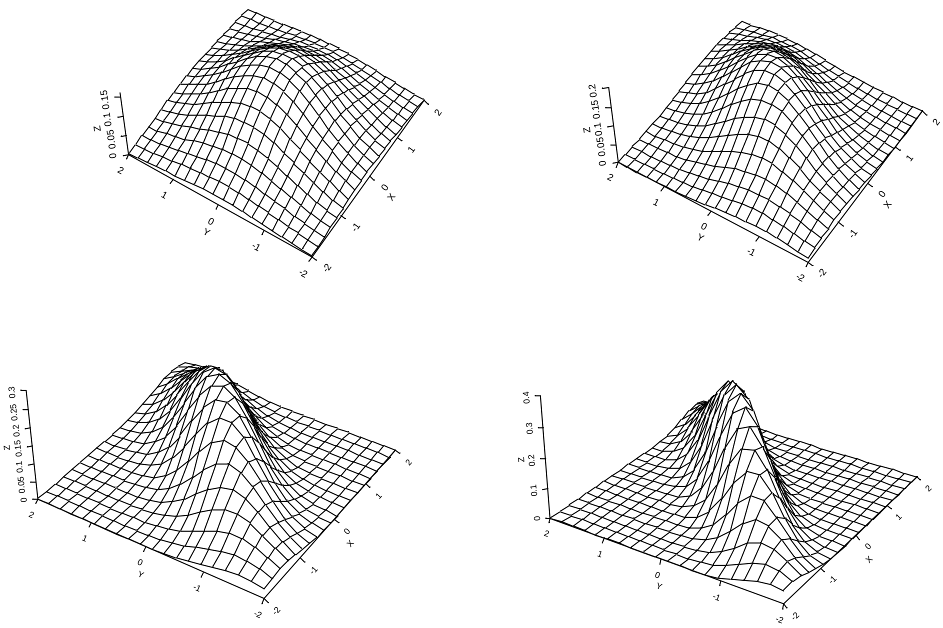

L'approche graphique des modèles statistiques : à consommer sans modération ...

# 3 - Densités de probabilité

## 3.1 - Loi de Gauss bivariée

Pour une corrélation de 0, 0.5, 0.8 et 0.9 :

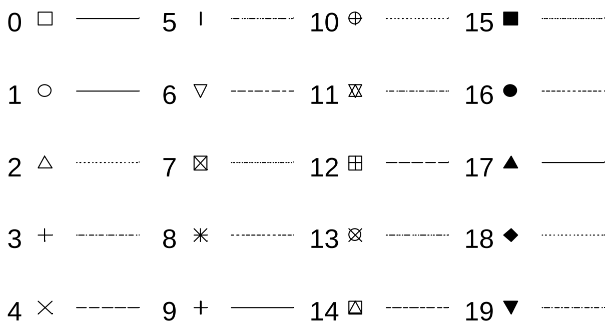
```
x0<-seq(-2,2,le=20)
y0<-seq(-2,2,le=20)
xy0<-expand.grid(x=x0, y=y0)
z0<-dmvnorm(xy0,rho=0.50)
z0<-matrix(z0,nrow=20,ncol=20,byrow=T)
persp(x0,y0,z0,eye=c(-30,-20,2),box = F)
```



## 3.2 - Types de lignes et de caractères :

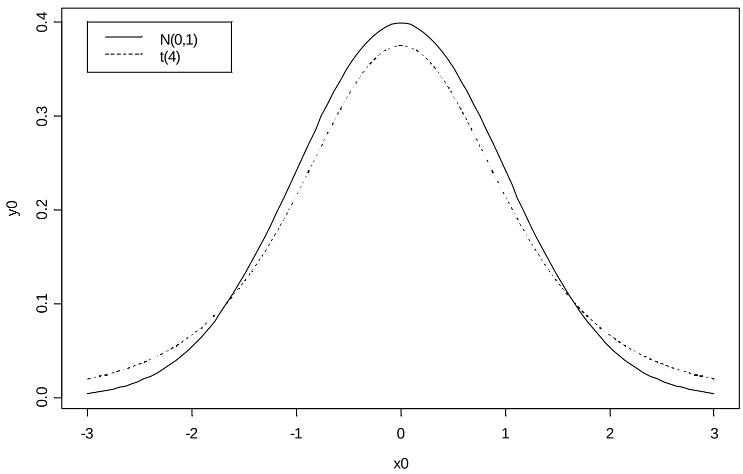
```
par(mai=c(0,0,0,0) ,cex=2)
```

```
plot(0,0,xlim=c(0,20),ylim=c(0,6),type="n",axes=F)
for (i in 1:5) {
   for(j in 1:4) {
         x0<-(j-1)*5+1
         y0<-6-i
         i0<-i+(j-1)*5-1
         text(x0,y0,paste(i0))
         points(x=x0+1,y=y0,pch=i0)
         par(lty=i0)
         segments(x0+2,y0,x0+4,y0)
   }
}
```




## 3.3 - Loi de Student

```
x0<-seq(-3,3,le=100)
y0<-dnorm(x0)
z0<-dt(x0,df=4)
plot(x0,y0,type="n")
lines(x0,y0,lty=1)
lines(x0,z0,lty=8)
legend(-3, 0.4, c("N(0,1)","t(4)"), lty = c(1,8))
```



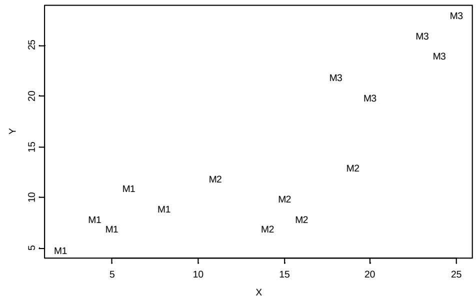

```
> qt(df = 4, c(0.025, 0.975))
[1] -2.776  2.776
> qnorm(c(0.025, 0.975))
[1] -1.96  1.96
```

# 4 - Analyse de covariance

```
> M<-rep(c("M1","M2","M3"),c(5,5,5))
> M
 [1] "M1" "M1" "M1" "M1" "M1" "M2" "M2" "M2" "M2" "M2" "M3" "M3"
[13] "M3" "M3" "M3"
> X
 [1]  2  4  5  8  6 14 16 15 19 11 20 18 23 25 24
> Y
 [1]  5  8  7  9 11  7  8 10 13 12 20 22 26 28 24
```

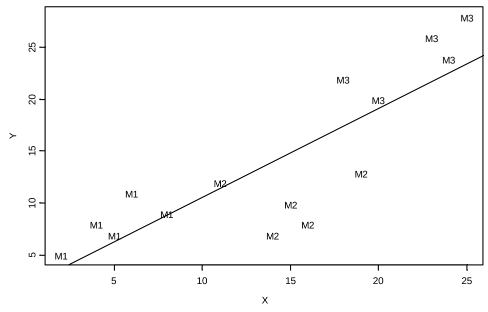
X niveau de départ, Y niveau d'arrivée, M méthode d'enseignement.

```
> covjdl<-cbind.data.frame(X,Y,M)
> names(covjdl)<-c("X","Y","M")
> covjdl
    X  Y  M
 1  2  5 M1
 2  4  8 M1
 3  5  7 M1
...
13 23 26 M3
14 25 28 M3
15 24 24 M3

> plot(X,Y,type="n")
> text(X,Y,M)
```



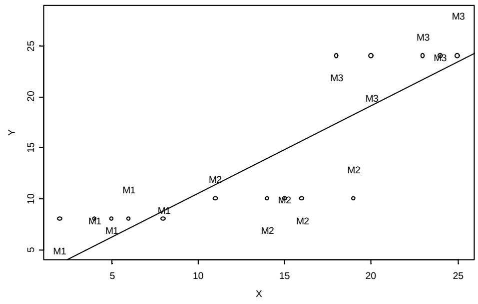
## 4.1 - Une seule droite de régression

```
> lm1<-lm(Y~X)
> anova(lm1)
Analysis of Variance Table

Response: Y

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value      Pr(F)
        X  1       599     599   31.53 0.00008407
Residuals 13       247      19
> abline(lm1)
```




## 4.2 - L'effet du facteur

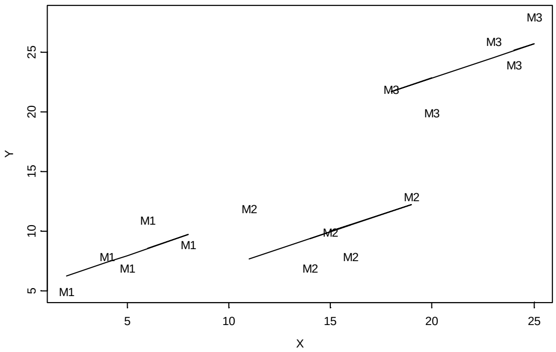
```
> lm2<-lm(Y~M)
> anova(lm2)
Analysis of Variance Table

Response: Y

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value      Pr(F)
        M  2       760   380.0   53.02 1.103e-006
Residuals 12        86     7.2
> points(X,predict(lm2))
```




## 4.3 - Droites parallèles :

```
> lm3<-lm(Y~M+X)
> anova(lm3)
Analysis of Variance Table

Response: Y

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value   Pr(F)
        M  2     760.0   380.0   72.58 0.00000
        X  1      28.4    28.4    5.43 0.03991
Residuals 11      57.6     5.2

> lm4<-lm(Y~X+M)
> anova(lm4)
Analysis of Variance Table

Response: Y

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value     Pr(F)
        X  1     599.0   599.0   114.4 0.0000004
        M  2     189.4    94.7    18.1 0.0003329 Droites non égales
```

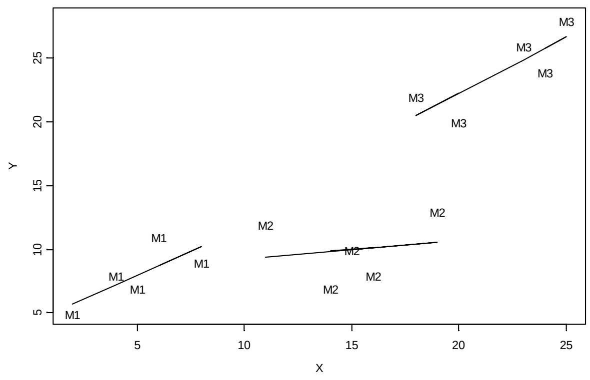
```
Residuals 11      57.6     5.2

> predict(lm3)
     1     2 3     4     5     6     7  8     9    10    11    12
 6.295 7.432 8 9.705 8.568 9.432 10.57 10 12.27 7.727 22.86 21.73
    13   14    15
 24.57 25.7 25.14
> predict(lm4)
     1     2 3     4     5     6     7  8     9    10    11    12
 6.295 7.432 8 9.705 8.568 9.432 10.57 10 12.27 7.727 22.86 21.73
    13   14    15
 24.57 25.7 25.14

> lines(X[M=="M1"],predict(lm3)[M=="M1"])
> lines(X[M=="M2"],predict(lm3)[M=="M2"])
> lines(X[M=="M3"],predict(lm3)[M=="M3"])
```




```
> coefficients(lm(Y[M=="M1"]~X[M=="M1"]))
 (Intercept) X[M == "M1"]
        4.25         0.75
> coefficients(lm(Y[M=="M2"]~X[M=="M2"]))
 (Intercept) X[M == "M2"]
       7.794       0.1471
> coefficients(lm(Y[M=="M3"]~X[M=="M3"]))
 (Intercept) X[M == "M3"]
       4.588       0.8824
```

## 4.4 - Interaction

```
> lm5<-lm(Y~M*X)
> anova(lm5)
Analysis of Variance Table

Response: Y

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value  Pr(F)
        M  2     760.0   380.0   71.93 0.0000
        X  1      28.4    28.4    5.38 0.0456
      M:X  2      10.0     5.0    0.95 0.4220 NON SIGNIFICATIF
Residuals  9      47.5     5.3

> lines(X[M=="M1"],predict(lm5)[M=="M1"])
> lines(X[M=="M2"],predict(lm5)[M=="M2"])
> lines(X[M=="M3"],predict(lm5)[M=="M3"])
```




# 5 - Interaction sans répétition

16 stations météo dans les Rocheuses (4 altitudes x 4 expositions)

```
> roc
     Sud Sommet Nord Vallee
2460  88     83   74     39
2550  61     55   50     33
3000  22      9   17      0
3690 -39    -33  -39    -22

> roc.vec <- roc[, 1]
> for(i in 2:4) roc.vec <- append(roc.vec, roc[, i])
> roc.vec
 [1]  88  61  22 -39  83  55   9 -33  74  50  17 -39  39  33   0 -22

> sta.vec <- names(roc)[rep(1:4, rep(4, le = 4))]
> sta.vec
 [1] "Sud"    "Sud"    "Sud"    "Sud"    "Sommet" "Sommet" "Sommet"
 [8] "Sommet" "Nord"   "Nord"   "Nord"   "Nord"   "Vallee" "Vallee"
[15] "Vallee" "Vallee"
> alt.vec <- row.names(roc)[rep(1:4, 4)]
> alt.vec
 [1] "2460" "2550" "3000" "3690" "2460" "2550" "3000" "3690" "2460"
[10] "2550" "3000" "3690" "2460" "2550" "3000" "3690"
```

## 5.1 - Modèle $y_{ij} = m + a_i + b_j + e_{ij}$

```
> lm0<-lm(roc.vec~sta.vec+alt.vec)
```

**Propriété des plans orthogonaux**

```
> anova(lm0)
Analysis of Variance Table

Response: roc.vec

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value Pr(F)
  sta.vec  3       931     310    1.94 0.194
  alt.vec  3     25162    8387   52.39 0.000
Residuals  9      1441     160

> lm1<-lm(roc.vec~alt.vec+sta.vec)
```

```
> anova(lm1)
Analysis of Variance Table

Response: roc.vec

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value Pr(F)
  alt.vec  3     25162    8387   52.39 0.000
  sta.vec  3       931     310    1.94 0.194
Residuals  9      1441     160

> matrix(predict(lm0),4,4)
       [,1]   [,2]   [,3]    [,4]
[1,]  79.12  74.62  71.63  58.625
[2,]  57.87  53.38  50.38  37.375
[3,]  20.12  15.62  12.63  -0.375
[4,] -25.12 -29.62 -32.63 -45.625

> roc-matrix(predict(lm0),4,4)
        Sud Sommet   Nord  Vallee
2460   8.875  8.375  2.375 -19.625
2550   3.125  1.625 -0.375  -4.375
3000   1.875 -6.625  4.375   0.375
3690 -13.875 -3.375 -6.375  23.625
```

## 5.2 - Modèle $y_{ij} = ma_i b_j + e_{ij}$ [1]

```
> svd(roc)
$d:
[1] 192.914  12.069   7.762   3.524

$v:
        [,1]    [,2]    [,3]    [,4]
[1,] -0.6008  0.4635  0.1312 -0.6380
[2,] -0.5444 -0.4482  0.6340  0.3174
[3,] -0.5119  0.3183 -0.5182  0.6067
[4,] -0.2841 -0.6949 -0.5588 -0.3523

$u:
        [,1]      [,2]     [,3]    [,4]
[1,] -0.7621  0.003526  0.51932  0.3867
[2,] -0.5264 -0.281129 -0.19003 -0.7796
[3,] -0.1390  0.959054 -0.02792 -0.2452
[4,]  0.3505 -0.034198  0.83271 -0.4273

> svd1<-svd(roc)
> svd1$u[,1]
[1] -0.7621 -0.5264 -0.1390  0.3505
> u1<-svd1$u[,1]

> t(svd1$v[,1])
        [,1]    [,2]    [,3]    [,4]
[1,] -0.6008 -0.5444 -0.5119 -0.2841
> v1<-t(svd1$v[,1])
> u1%*%v1
         [,1]     [,2]     [,3]     [,4]
[1,]  0.45783  0.41485  0.39006  0.21648
[2,]  0.31627  0.28659  0.26946  0.14955
[3,]  0.08352  0.07568  0.07116  0.03949
[4,] -0.21055 -0.19078 -0.17938 -0.09956
> svd1$d[1]*u1%*%v1
       [,1]   [,2]   [,3]    [,4]
```

[1] Mandel, J. (1961) Non additivity in two-way analysis of variance. *Journal of the American Statistical Association* : 65,878-888.

```
[1,]  88.32  80.03  75.25  41.762
[2,]  61.01  55.29  51.98  28.850
[3,]  16.11  14.60  13.73   7.618
[4,] -40.62 -36.80 -34.61 -19.206
```

```
> roc-svd1$d[1]*u1%*%v1
         Sud  Sommet   Nord Vallee
2460 -0.3210  2.9693 -1.248 -2.762
2550 -0.0135 -0.2865 -1.983  4.150
3000  5.8880 -5.5996  3.273 -7.618
3690  1.6175  3.8049 -4.394 -2.794
```

**RAPPEL**

```
> roc-matrix(predict(lm0),4,4)
         Sud Sommet   Nord  Vallee
2460   8.875  8.375  2.375 -19.625
2550   3.125  1.625 -0.375  -4.375
3000   1.875 -6.625  4.375   0.375
3690 -13.875 -3.375 -6.375  23.625
```

***Quel est le meilleur modèle ?***

```
> sum((predict(lm0)-roc.vec)^2)/9
[1] 160.1
> sum((as.vector(svd1$d[1]*u1%*%v1)-roc.vec)^2)/9
[1] 24.26
```

```
> qqnorm(predict(lm0)-roc.vec)
> qqline(predict(lm0)-roc.vec)
> qqnorm(as.vector(svd1$d[1]*u1%*%v1)-roc.vec)
> qqline(as.vector(svd1$d[1]*u1%*%v1)-roc.vec)
```