CENTRO INTERNACIONAL DE LA PAPA CIP

IOWA State University. August, 2007

Agricolae – a free statistical library for agricultural research

Felipe de Mendiburu, Reinhard Simon

Description

Agricolae Version 1.0-3
contains a total 68 statistical routines and 33 data sets.

Agricolae is a statistical library for agricultural research with the goal of supporting developing countries. Focuses on statistical tools used in the breeding program of the International Potato Center for its main commodity crops, potato and sweetpotato. Thus, Agricolae supports a variety of field trial designs, including incomplete block designs techniques, genetic designs, stability analysis, AMMI with biplot and triplot analysis, multiple comparisons of treatments Other functions include the construction of consensus clusters, optimal size and shape of experimental field plots.

Agricoale was developed using R and is available via the CRAN, repository at http://www.r-project.org.

Planning of field experiment

Randomize and field book

Alpha design, Graeco, latin square, CRD, RCBD , BIB.

The planning of field experiments is one of the main areas of Agricolae.

It supports simple lattice design (lattice.simple),
Factorial a block design (design.ab),
Alpha design (alpha.design),
Balanced Incomplete Block Design (design.bib),
Randomized complete block design (design.rcbd).
Complete randomized design (design.crd),
Graeco-latin square design (design.graeco),
Latin square design (design.lsd).

Planning of field experiment

Greaco latin

args: trt1, trt2, number = 1, seed = 0, kinds = "Super-Duper"

```
> T1<-c("a","b","c","d")
> T2<-c("v","w","x","y")
> Plan <-  design.graeco(T1,T2,number=101)
```

Plots

```
     [,1] [,2] [,3] [,4]
[1,]  101  102  103  104
[2,]  105  106  107  108
[3,]  109  110  111  112
[4,]  113  114  115  116
```

Treatments

```
     [,1]  [,2]  [,3]  [,4]
[1,] "d w" "b v" "a x" "c y"
[2,] "b y" "d x" "c v" "a w"
[3,] "a v" "c w" "d y" "b x"
[4,] "c x" "a y" "b w" "d v"
```

It’s not possible to construct: 6,10 and pair >= 14

Planning of field experiment

Alpha design

(trt, k, r, number = 1, seed = 0, kinds = "Super-Duper“)

```
> Trt <- letters[1:12]
> plan<-design.alpha(trt,k=3, r=2, number=101)
```

```
alpha design (0,1) - Serie  I
Parameters Alpha design
=======================
treatmeans : 12
Block size : 3
Blocks     : 4
Replication: 2

Efficiency factor

(E ) 0.6470588

<<< Book >>>
```

Field Book

```
> plan
     plots cols block trt replication
1     101     1      1   j           1
2     102     2      1   h           1
3     103     3      1   c           1
4     104     1      2   d           1
...
23    123     2      8   h           2
24    124     3      8   e           2
```

Comparison of multiple treatments
Test:

LSD, HSD, Waller, Durbin, Kruskal Wallis, Friedman, Waerden

Test parametrics:

LSD: Least significant difference and Adjust P-values
HSD: Honestly significant difference Tukey.
Waller: Bayesian t-values for multiple comparisons

Test Non parametrics

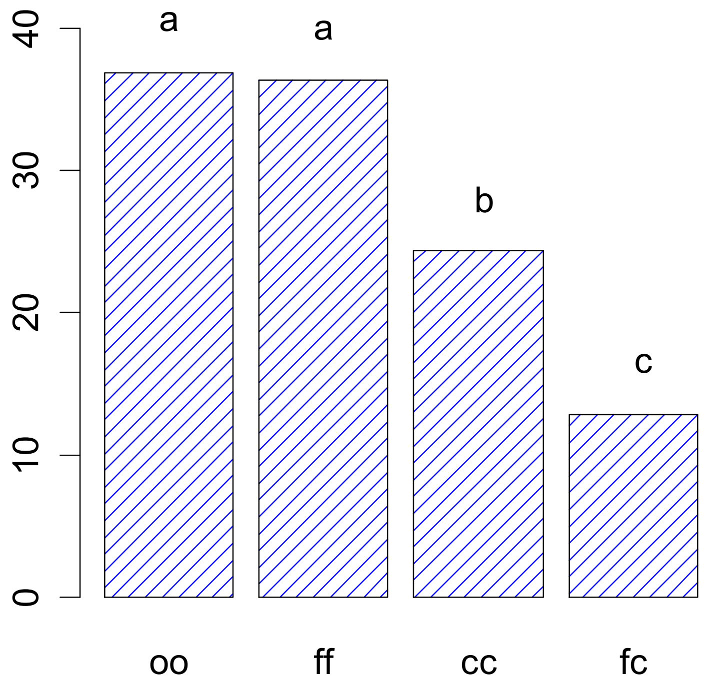
Kruskal Wallis: Complete randomized design
Friedman: Randomized complete block design
Durbin: Balanced Incomplete Block Design
Waerden: The van der Waerden (Normal Scores)

Comparison of multiple treatments

Waller-Duncan

(y, trt, DFerror, MSerror, Fc, K = 100, group = TRUE, main = NULL)

```
> attach(sweetpotato)
> model<-aov(yield~virus)
> comparison <- waller.test(yield, virus, DFerror=8, MSerror=22.49,Fc=17.345)
> bar.group(comparison,horiz=FALSE,ylim=c(0,45),density=10,col=“blue")
```



```
Critical Value of Waller   2.236
Minimum Significant Difference
8.658066

Means with the same letter are not
significantly different.

Groups, Treatments and means

a         oo        36.9
a         ff        36.33333
 b        cc        24.4
  c       fc        12.86667
```

Comparison of multiple treatments

LSD

(y, trt, DFerror, MSerror, alpha = 0.05, p.adj = c("none", "holm", hochberg", "bonferroni", "BH", "BY", "fdr"), group = TRUE, main = NULL))

```
> comparison <- LSD.test(yield, virus, DFerror=8, MSerror=22.49,
p.adj=“bonferroni”)
```

group = TRUE

```
LSD t Test for yield
P value adjustment method: bonferroni
                              ......
Alpha                       0.050000
Error Degrees of Freedom    8.000000
Error Mean Square          22.490000
Critical Value of t         3.478879

Least Significant Difference 13.47065

Means with the same letter are not
significantly different.

Groups, Treatments and means

a        oo      36.9
a        ff      36.33333
ab       cc      24.4
 b       fc      12.86667
```

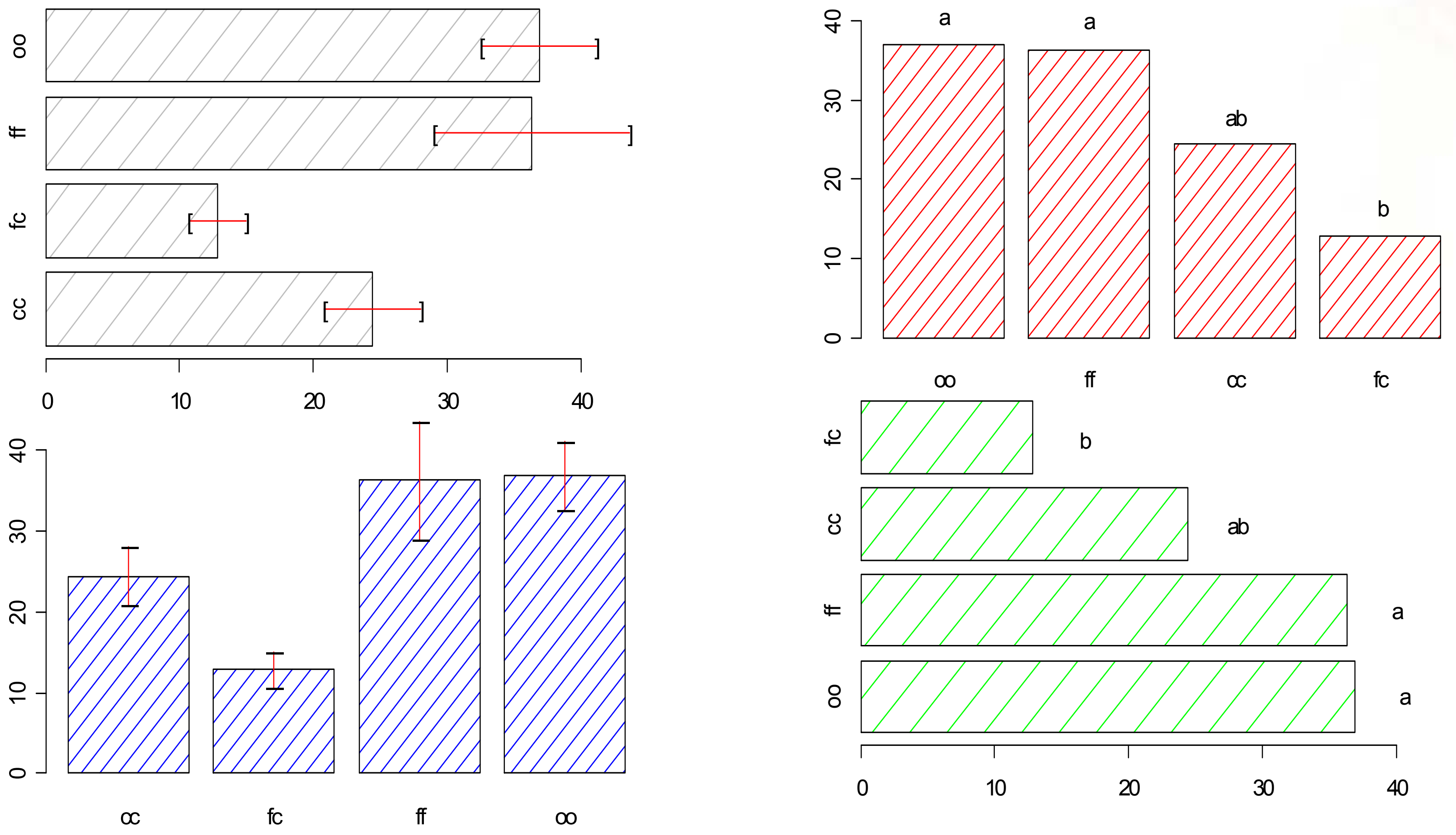
group = FALSE

```
Treatment Means

   virus    yield  std.err replication

1    cc 24.40000 2.084067          3
2    fc 12.86667 1.246774          3
3    ff 36.33333 4.233727          3
4    oo 36.90000 2.482606          3

Comparison between treatments means

  tr.i tr.j       diff pvalue
1    1    2 11.5333333 0.1056
2    1    3 11.9333333 0.0900
3    1    4 12.5000000 0.0720
4    2    3 23.4666667 0.0024
5    2    4 24.0333333 0.0012
6    3    4  0.5666667 1.0000
```

Comparison of multiple treatments

Graphics.

bar.err & bar.group



Stability analysis

AMMI, stability.par, stability.nonpar

AMMI: Additive Main Effects and Multiplicative Interaction models are widely used to analyze main effects and genotype by environment (GEN, ENV) interactions in multilocation variety trials. Furthermore, this function generates biplot and triplot graphs as well as principal component analysis.

stability.par: SHUKLA'S STABILITY VARIANCE AND KANG'S. This procedure calculates the stability variations as well as the statistics of selection for the yield and the stability

stability.nonpar: A method based on the statistical ranges of the study variable per environment for the stability analysis

Stability analysis

AMMI

```
(ENV, GEN, REP, Y, MSE=0, number=TRUE,graph="biplot",..)
```

```
> model<- AMMI(ltrv[,2], ltrv[,1], ltrv[,3], ltrv[,5],
xlim=c(-3,3),ylim=c(-4,4), graph="biplot",number=FALSE)

ANALYSIS AMMI:  ltrv[, 5]
Class level information

ENV:  Ayac LM-02 SR-02 Hyo-02 LM-03 SR-03
GEN:  102.18 104.22 121.31 141.28 157.26 163.9 221.19 233.11 235.6
241.2 255.7 314.12 317.6 319.20 320.16 342.15 346.2 351.26 364.21
402.7 405.2 406.12 427.7 450.3 506.2 Canchan Desiree Unica
REP:  1 2 3

Number of observations:  504

model Y: ltrv[, 5]  ~ ENV + REP%in%ENV + GEN + ENV:GEN
Random effect REP%in%ENV
```

Stability analysis

AMMI

`(ENV, GEN, REP, Y, MSE=0, number=TRUE,graph="biplot",..)`

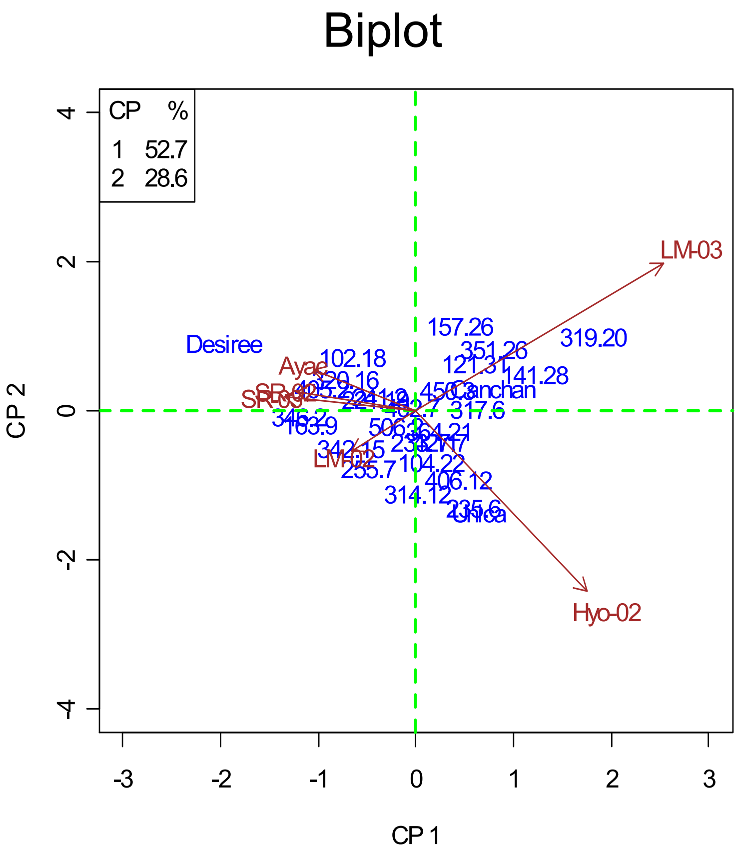
```
Analysis of Variance Table
Response: Y
           Df Sum Sq Mean Sq  F value    Pr(>F)
ENV         5 9607.4  1921.5 284.6352 4.957e-12 ***
REP(ENV)   12   81.0     6.8   2.7313   0.00154 **
GEN        27 1367.4    50.6  20.4904 < 2.2e-16 ***
ENV:GEN   135 1764.8    13.1   5.2891 < 2.2e-16 ***
Residuals 324  800.8     2.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coeff var       Mean ltrv[, 5]
20.07525         7.831188

Analysis
    percent  acum Df    Sum.Sq   Mean.Sq F.value   Pr.F
CP1    52.7  52.7 31 929.89935 29.996753   12.14 0.0000
CP2    28.6  81.3 29 503.95903 17.377898    7.03 0.0000
```

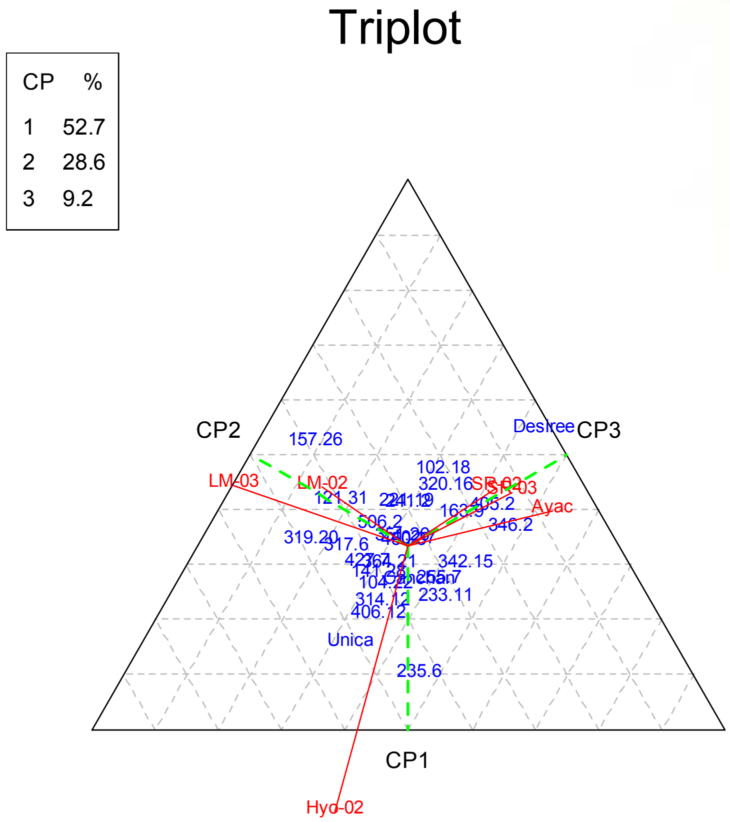
... More ...

Stability analysis

AMMI

(ENV, GEN, REP, Y, MSE=0, number=TRUE,graph="biplot",..)





Stability analysis

AMMI.contour

`(model, distance, shape, ...)`

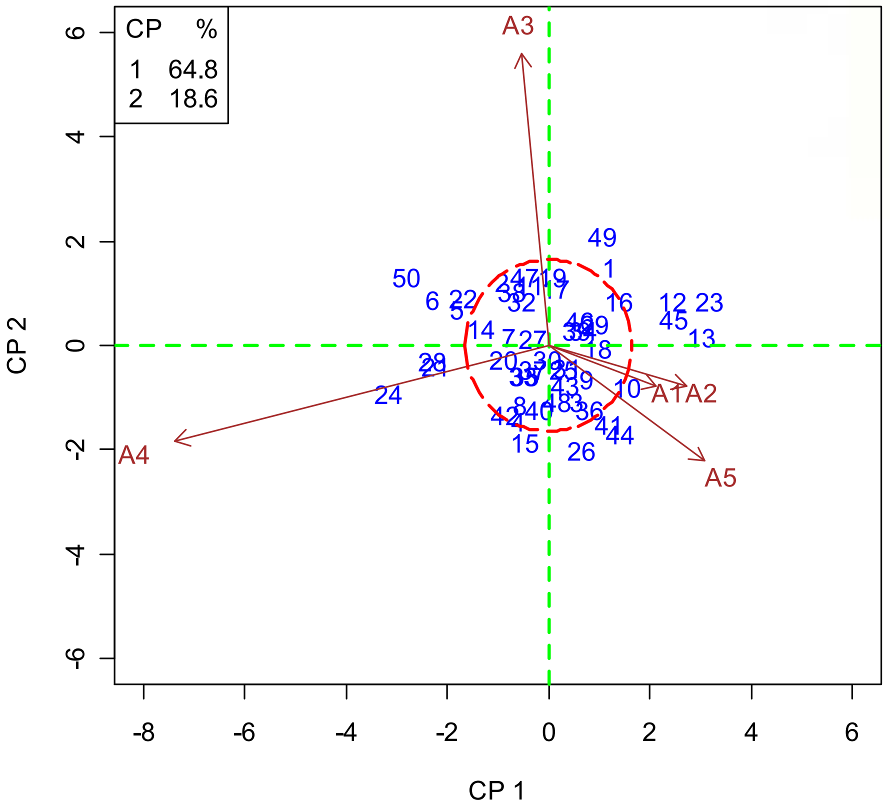
Limit, radio: 1.645421
Genotype in: 31
Genotype out: 19

GENOTYPE IN:

"10" "11" "12" "15" "16" "17" "19"
"21" "24" "25" "26" "27" "31" "33"
"35“ [16] "36" "37" "38" "39" "4"
"40" "41" "42" "43" "44" "45" "47"
"48" "50" "6" "7"

GENOTYPE OUT:

"1" "13" "14" "18" "2" "20" "22"
"23" "28" "29" "3" "30" "32" "34"
"46“ "49" "5" "8" "9"



Stability analysis

AMMI

`(ENV,GEN,REP, Y, MSE=0,number=TRUE,graph="biplot",..)`

Input data:

a) complete or missing value. Experiments in localities under randomized complete block design.

Or

b) Only means and missing value. Estimation variance of error and replication:

MSE = variance error = Mean square error

Rep = constant = Harmonic Mean (r1, r2,.., rk)

Stability analysis (parametric)

stability.par

```
(data, rep, MSerror, alpha = 0.1, main = NULL, cova = F,
name.cov = NULL, file.cov = 0)
```

```
> stability.par(data, rep=4, MSerror=1.8, alpha=0.1, main="Genotype")
```

```
INTERACTIVE PROGRAM FOR CALCULATING SHUKLA'S STABILITY VARIANCE AND
KANG'S
                        YIELD - STABILITY (YSi) STATISTICS
 Genotype
 Environmental index  - covariate
 Analysis of Variance
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Source          d.f.   Sum of Squares  Mean Squares         F
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 TOTAL            155     2121.2544
 GENOTYPES         12      101.0877          8.4240         3.31 *
 ENVIRONMENTS      11     1684.3067        153.1188        85.07 **
 INTERACTION      132      335.8600          2.5444         1.41 **
 HETEROGENEITY     12       34.7256          2.8938         1.15 ns
 RESIDUAL         120      301.1344          2.5095         1.39 *
 POOLED ERROR     432                        1.8000
 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Stability analysis (parametric)

stability.par

```
(data, rep, MSerror, alpha = 0.1, main = NULL, cova =
F, name.cov = NULL, file.cov = 0)
```

```
Simultaneous selection for yield and stability  (++)

   Genotype     Yield Rank Adj.rank Adjusted Stab.var Stab.rating YSi ...
1         A 7.383333   11        1       12 2.134311           0  12   +
2         B 6.783333    2       -1        1 1.672824           0   1
3         C 7.250000    9        1       10 0.805606           0  10   +
4         D 6.783333    2       -1        1 2.919766          -2  -1
5         E 7.075000    7       -1        6 1.604036           0   6   +
6         F 6.916667    6       -1        5 3.924945          -2   3
7         G 7.808333   12        2       14 4.043485          -2  12   +
8         H 7.908333   13        2       15 2.899022          -2  13   +
9         I 7.275000   10        1       11 4.251970          -2   9   +
10        J 7.083333    8       -1        7 1.853320           0   7   +
11        K 6.433333    1       -2       -1 2.167039           0  -1
12        L 6.891667    5       -1        4 1.692631           0   4
13        M 6.791667    4       -1        3 3.108168          -2   1

 Yield Mean: 7.10641
 YS    Mean: 5.846154
 LSD (0.05): 0.4514298
 - - - - - - - - - - -
 +   selected genotype
 ++  Reference: Kang, M. S. 1993. Simultaneous selection for yield
 and stability: Consequences for growers. Agron. J. 85:754-757
```

Stability analysis (Non-parametric)

Haynes K G, Lambert D H, Christ B J, Weingartner D P, Douches D S, Backlund J E, Fry W and Stevenson W. 1998. Phenotypic stability of resistance to late blight in potato clones evaluated at eight sites in the United States American Journal Potato Research 75, pag 211-217.

```
Stability.nonpar(data, variable=NULL, ranking = FALSE)
```

```
> haynes
        clone  FL   MI   ME ...
1    A84118-3 284 1113 1053 ...
2   AO80432-1 254  690 1112 ...
3   AO84275-3 395 1089 1090 ...
4  AWN86514-2 136  296  374 ...
5     B0692-4  87  653  412 ...
6     B0718-3 130  126  329 ...
...   ... ...
```

```
stability.nonpar(haynes,"YIELD",ranking=TRUE)
```

```
 Nonparametric Method for Stability Analysis
 -------------------------------------------

 Estimation and test of nonparametric measures
 Variable: YIELD
 Ranking...
            FL MI ME MN ND   NY PA WI
 A84118-3    7 11 11 14  8 14.0 12 11
 AO80432-1   6  9 13 13 12 12.0 15 14
 AO84275-3  10 10 12  8  9  7.0 11 12
 AWN86514-2  3  3  3  1  3  3.0  2  1
 B0692-4     1  8  4  3  2  2.0  1  3
 B0718-3     2  1  2  2  4  4.0  3  4
 ...
```

Stability analysis Non-parametric

Haynes K G, Lambert D H, Christ B J, Weingartner D P, Douches D S, Backlund J E, Fry W and Stevenson W. 1998. Phenotypic stability of resistance to late blight in potato clones evaluated at eight sites in the United States American Journal Potato Research 75, pag 211-217.

```
Stability.nonpar(data, variable=NULL, ranking = FALSE)
```

```
Statistics...
                Mean Rank   s1   Z1    s2   Z2
A84118-3      741.62   13 4.82 0.22 16.70 0.34
AO80432-1     734.38   12 6.21 0.73 26.57 0.47
AO84275-3     635.88    9 6.20 0.70 28.53 0.87
...
Sum of Z1:  20.08986
Sum of Z2:  25.84532
```

Test...

The Z-statistics are measures of stability. The test for the significance of the sum of Z1 or Z2 are compared to a Chi-Square value of chi.sum. individual Z1 or Z2 are compared to a Chi-square value of chi.ind.

```
MEAN          es1     es2     vs1       vs2      chi.ind   chi.sum
561.4609  5.3125  21.25 1.111905 60.75223  8.733011  26.29623
```

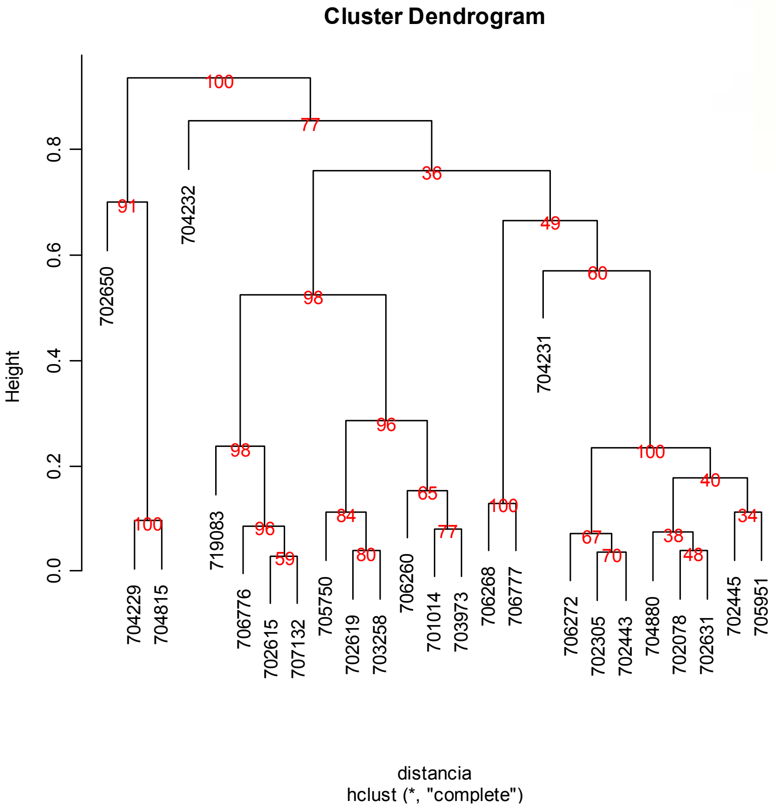
Consensus cluster

Methods distance and clustering of R, functions dist() and hclust().

```
(data, distance = c("binary", ..), method = c("complete", ..),
nboot = 500, duplicate = TRUE, cex.text = 1, col.text = "red",
...)
```

```
output<-consensus( pamCIP,distance="binary", method="complete", nboot=500)
```

```
Duplicates: 18
New data   : 25 Records
Consensus hclust
Method distance: binary
Method cluster : complete
rows and cols  : 25 107
n-boostrap     : 500


Run time       : 16.281 secs
```



Consensus cluster

Methods distance and clustering of R, functions dist() and hclust().

```
(data, distance = c("binary", ..), method = c("complete", ..),
nboot = 500, duplicate = TRUE, cex.text = 1, col.text = "red",
...)
```

OUTPUT

```
> names(output)
[1] "table.dend" "dendrogram" "duplicates"
```

to reproduce dendrogram

```
dend<-output$dendrogram
data<-output$table.dend
plot(dend)
text(data[,3],data[,4],data[,5],col="blue",cex=1)
```

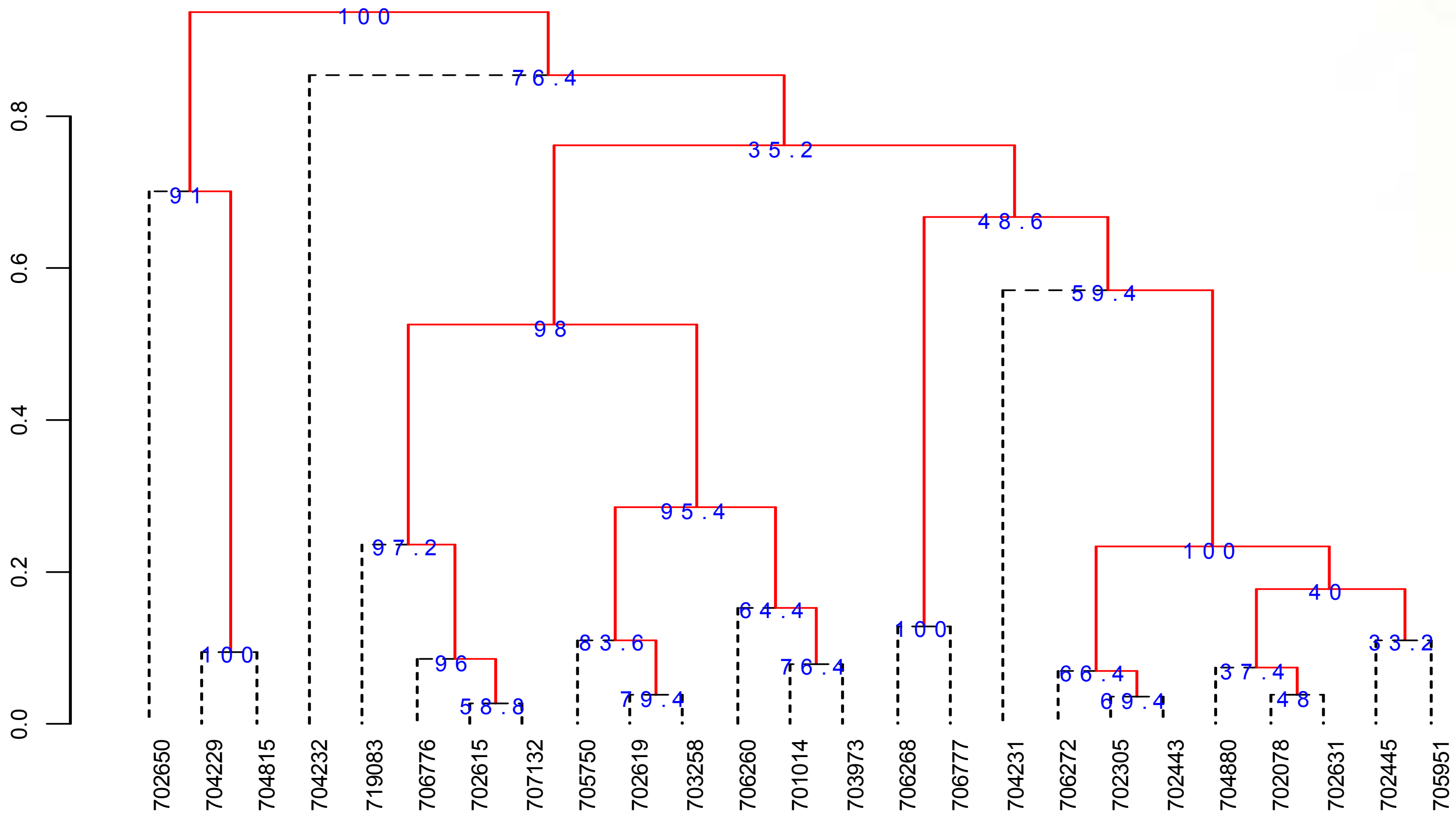
classical dendrogram

```
dend<-as.dendrogram(output$dendrogram)
plot(dend,type="r",edgePar = list(lty=1:2, col=2:1))
text(data[,3],data[,4],data[,5],col="blue",cex=1)
```

Consensus cluster

Methods distance and clustering of R, functions dist() and hclust().

```
(data, distance = c("binary", ..), method = c("complete", ..),
nboot = 500, duplicate = TRUE, cex.text = 1, col.text = "red",
...)
```



Consensus cluster

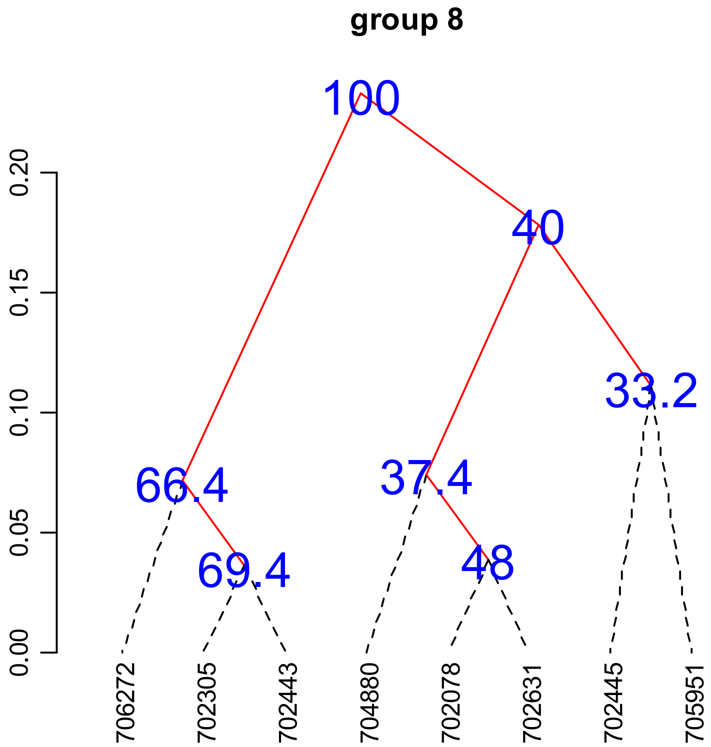
Input: output consensus

hcut()

```
(consensus, h, group, col.text = "blue", cex.text = 1, ...)
```

```
hcut(output,h=0.4,group=8,type="t",edgePar = list(lty=1:2,
col=2:1),main="group 8‶ ,col.text="blue",cex.text=2)
```

numbers	
1	1
2	2
3	1
4	4
5	6
6	2
7	1
8	8



Soil uniformity

```
Index.smith(data, ...)
```

```
table<-index.smith(rice, type="l",lty=4, lwd=3,
main="Relationship between CV\n per unit area and plot
size",col="red")
```

Smith's index of soil heterogeneity is used primarily to derive optimum plot size. The index gives a single value as a quantitative measure of soil heterogeneity in an area. The coefficient of variance is used to determine plot size and shape

```
> table

$model

lm(formula = CV ~
I(log(x)))

Coefficients:
(Intercept)     I(log(x))
    12.4782       -0.7009
```

```
$uniformity

      Size Width Length plots       Vx   CV

 [1,]    1     1      1   648 9044.539 13.0
 [2,]    2     1      2   324 7816.068 12.1
 [3,]    2     2      1   324 7831.232 12.1
 [4,]    3     1      3   216 7347.975 11.7
 [5,]    3     3      1   216 7355.216 11.7
 ...
[40,]  162     9     18     4 4009.765  8.6
```

Soil uniformity

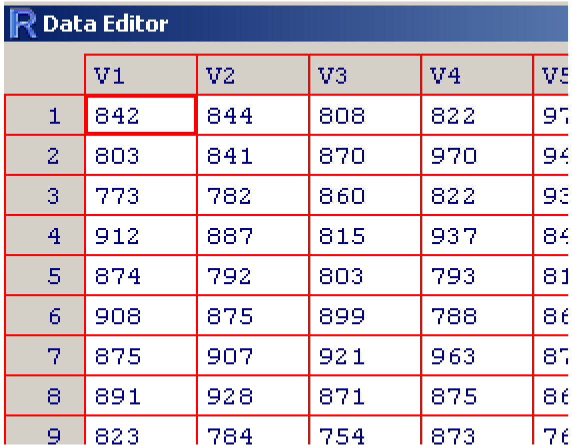
```
Index.smith(data, ...)

 table<-index.smith(rice, type="l",lty=4, lwd=3,
main="Relationship between CV\n per unit area and plot
 size",col="red")
```

```
predict(table$model, new=data.frame(x=30))
```

```
[1] 10.09436
```

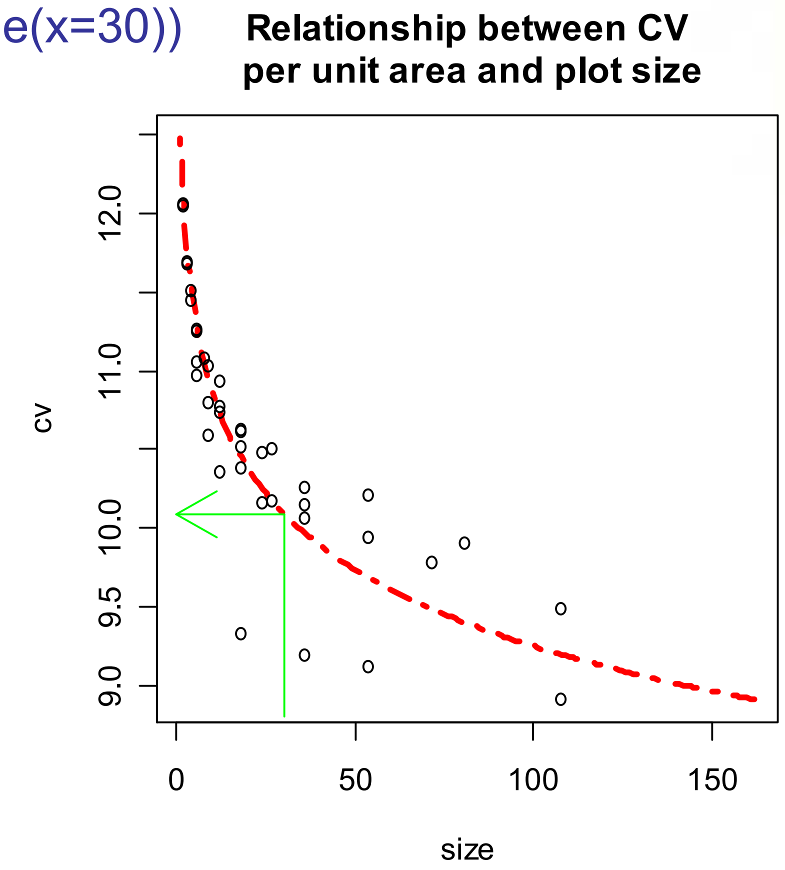
If plot size = 30 unit ^2
then CV = 10 %

rice



Data Editor

	V1	V2	V3	V4	V5
1	842	844	808	822	97
2	803	841	870	970	94
3	773	782	860	822	93
4	912	887	815	937	84
5	874	792	803	793	81
6	908	875	899	788	86
7	875	907	921	963	87
8	891	928	871	875	86
9	823	784	754	873	76



Relationship between CV
per unit area and plot size

Other functions and data sets

Genetic design: north carolina design, line x tester.
Biodiversity index and confidence interval.
Descriptive statistical: cross tabulations,...
Model: simulation and resampling.

Data sets main in package 'agricolae':

ComasOxapampa	Data AUDPC Comas - Oxapampa
Glycoalkaloids	Data Glycoalkaloids
RioChillon	Data and analysis Mother and baby trials
clay	Data of Ralstonia population in clay soil
disease	Data evaluation of the disease overtime
huasahuasi	Data of yield in Huasahuasi
melon	Data of yield of melon in a Latin square experiment
natives	Data of native potato
pamCIP	Data Potato Wild
paracsho	Data of Paracsho biodiversity
ralstonia	Data of population bacterial Wilt: AUDPC
soil	Data of soil analysis for 13 localities
sweetpotato	Data of sweetpotato yield
trees	Data of species trees. Pucallpa
wilt	Data of Bacterial Wilt (AUDPC) and soil

Agricolae Version 1.0-4

Please note that there is a new version of the agricolae on the link below

http://tarwi.lamolina.edu.pe/~fmendiburu