Pontificia Universidad Catolica del Peru

24 Abril 2008

Presentación de la librería “agricolae” de R para la investigación agrícola

Felipe de Mendiburu

---

http://www.cipotato.org

Welcome to the International Potato Center

Peruvian potato project wins global contest

LIMA - The Peruvian potato project T'ikapapa has been chosen as the winner of The World Challenge 2007, a global competition promoted by the BBC and Newsweek magazine, in association with Shell.

See Award Ceremony Video (WMV 50 MB)

See video Potato Goldmine (WMV 34 MB)

Read more Versión en Español

President launches the International Year of the Potato in Peru

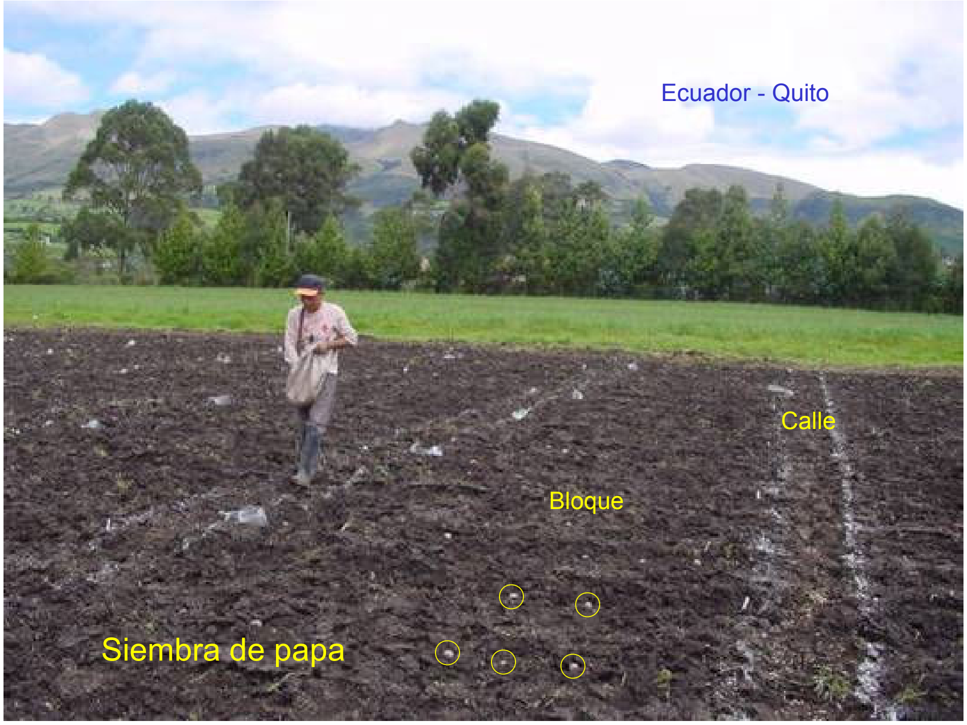
On Thursday 22 November, the President of Peru, Alan García Pérez, officially launched the International Year of the Potato in the country, highlighting the contribution that the potato makes to the world.

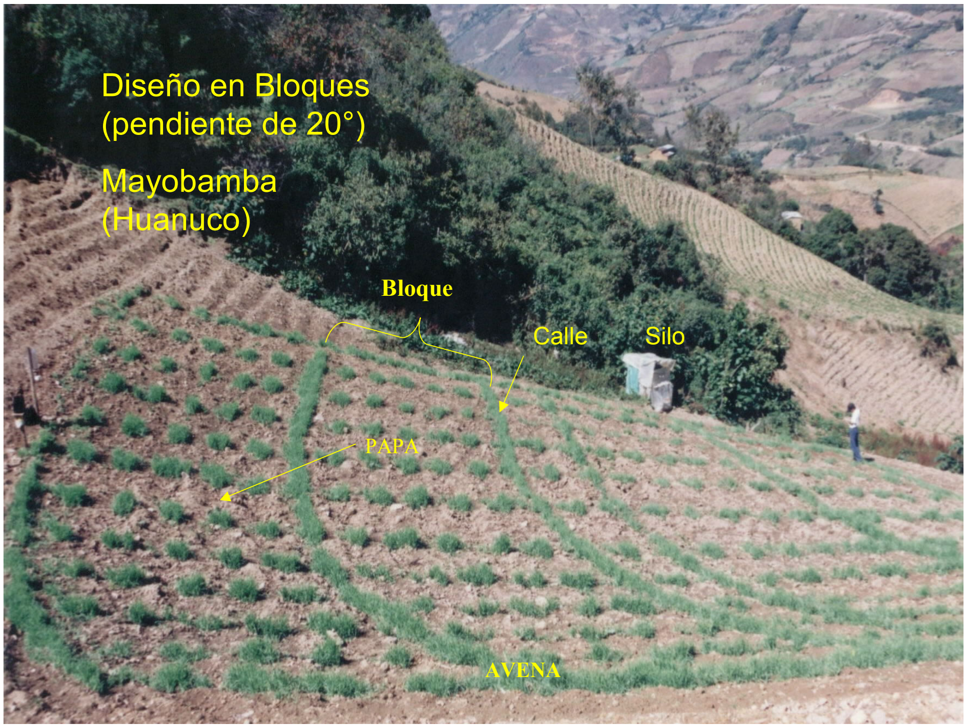
Av. La Molina 1895, La Molina, Lima, Peru. Postal address: Apartado 1558, Lima 12, Peru. Tel +51 1 3496017 webmaster-cip@cgiar.org

El CIP es un centro de investigacion con orientacion a cultivos de raices y tuberculos como la papa, el camote y otros cultivos andinos.

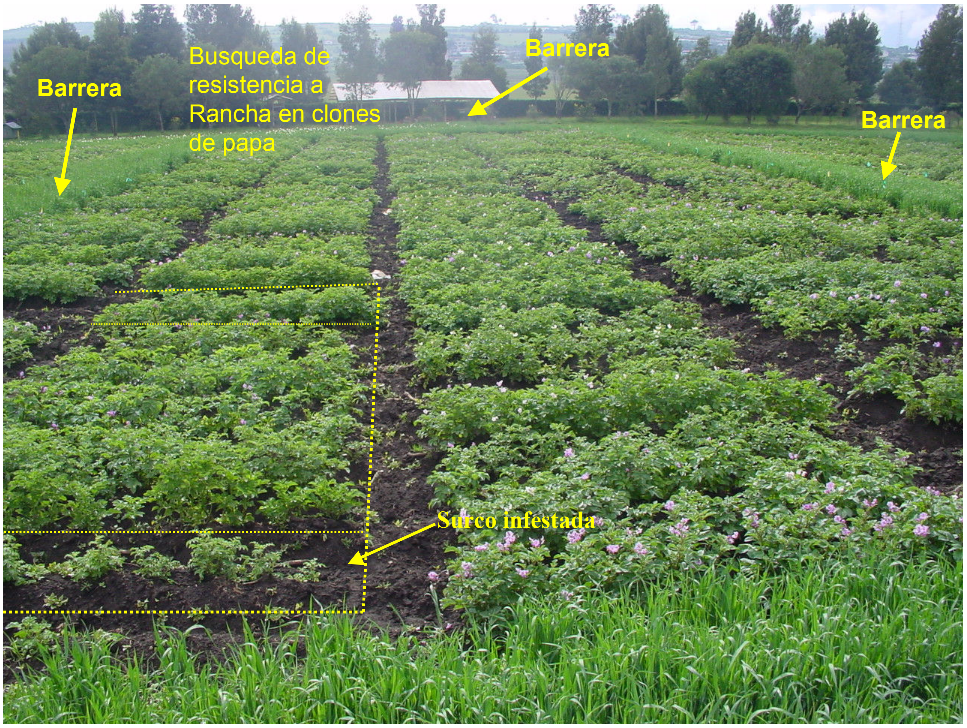
Los estudios usan mayormente los diseños de experimentos para analisis comparativo,

Son muchas las areas que realizan estas actividades y usan los diseños y analisis estadistico: Recursos Geneticos, Entomologia, Virologia, Patologia, Biodiversidad, Mejoramiento genetico de plantas y otras. Areas importantes que motivaron la cracion de una libreria de apoyo a los investigadores del CIP.


Siembra de papa


Diseño en Bloques (pendiente de 20°)

Mayobamba (Huanuco)



Inicialmente fue papa y se determino la marchitez inicial

Rotación de cultivos en Carhuaz para marchitez bacteriana (2006): Pedro Aley.

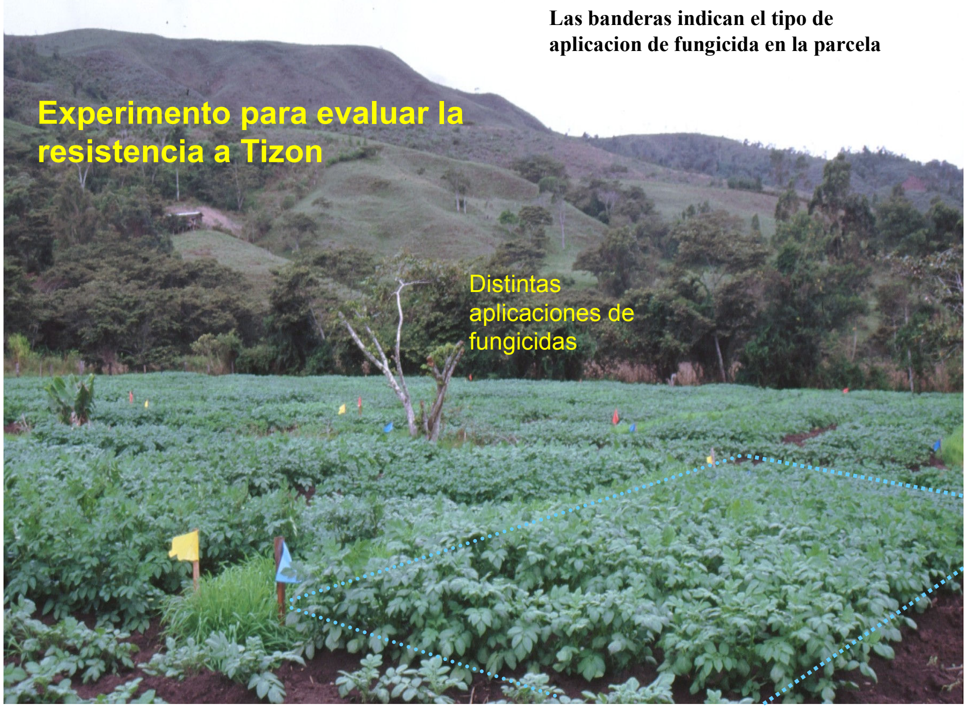
Reduccion de la marchitez bacteriana de la papa mediante rotacion de cultivos

Experimento para el control de rancha

Strip Plot: 4 variedades y 5 tratamientos con 4 rep

Las banderas indican el tipo de aplicacion de fungicida en la parcela

Experimento para evaluar la resistencia a Tizon



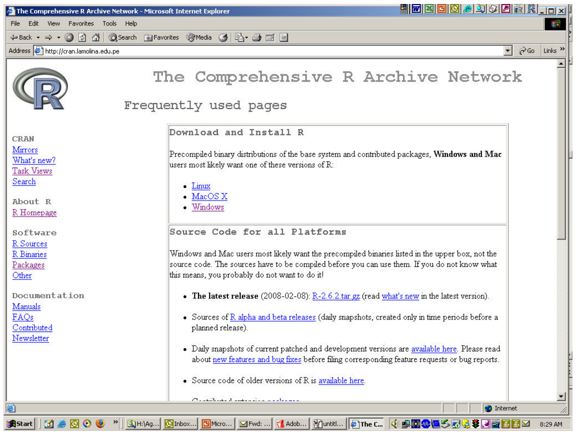
Parcelas de Observacion: Banco de Germoplasma de papa nativa (6000 clones)

Objetivo: Caracterizacion y mantenimientos del material genetico

Analisis: Cluster y dendrogramas

R en PERU

http://cran.lamolina.edu.pe



The Comprehensive R Archive Network

Frequently used pages

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed
Newsletter

Download and Install R

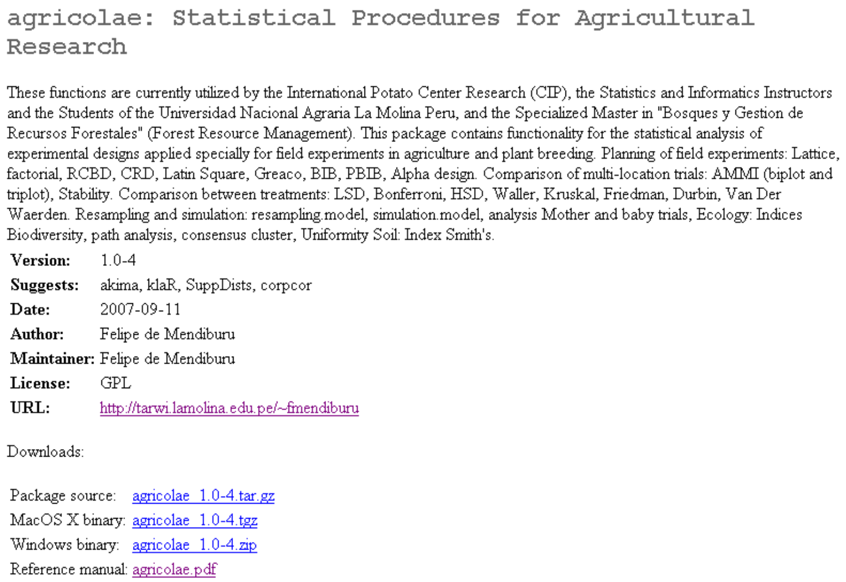
Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Linux
- MacOS X
- Windows

Source Code for all Platforms

Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- **The latest release** (2008-02-08): R-2.6.2.tar.gz (read what's new in the latest version).
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.



agricolae: Statistical Procedures for Agricultural Research

These functions are currently utilized by the International Potato Center Research (CIP), the Statistics and Informatics Instructors and the Students of the Universidad Nacional Agraria La Molina Peru, and the Specialized Master in "Bosques y Gestion de Recursos Forestales" (Forest Resource Management). This package contains functionality for the statistical analysis of experimental designs applied specially for field experiments in agriculture and plant breeding. Planning of field experiments: Lattice, factorial, RCBD, CRD, Latin Square, Greaco, BIB, PBIB, Alpha design. Comparison of multi-location trials: AMMI (biplot and triplot), Stability. Comparison between treatments: LSD, Bonferroni, HSD, Waller, Kruskal, Friedman, Durbin, Van Der Waerden. Resampling and simulation: resampling.model, simulation.model, analysis Mother and baby trials, Ecology: Indices Biodiversity, path analysis, consensus cluster, Uniformity Soil: Index Smith's.

Version:	1.0-4
Suggests:	akima, klaR, SuppDists, corpcor
Date:	2007-09-11
Author:	Felipe de Mendiburu
Maintainer:	Felipe de Mendiburu
License:	GPL
URL:	http://tarwi.lamolina.edu.pe/~fmendiburu

Downloads:

Package source:	agricolae_1.0-4.tar.gz
MacOS X binary:	agricolae_1.0-4.tgz
Windows binary:	agricolae_1.0-4.zip
Reference manual:	agricolae.pdf

Caracteristicas de Agricolae

1. Version 1.0 - Diciembre 10, del 2006
2. Version 1.0.4 Setiembre 11, del 2007 tiene un tamaño de 550 kb
3. Version 1.0.5 en actualizacion, 33 dataset y 69 funciones

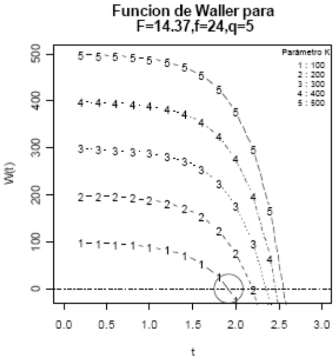
Aplicaciones

1. Manipulacion de datos
2. Estadistica descriptiva
3. Diseño de experimentos
4. Comparación multiple de tratamientos
5. Pruebas no paramétricas
6. Diseños Geneticos
7. Análisis de Estabilidad, biodiversidad
8. Pruebas de uniformidad para tamaño y forma de parcela
9. Consensus en cluster

Funcion de Waller-Duncan

$$W(t) = k - \frac{\int_1^{\infty} \sqrt{x-1}\, g(x) h(tb(x)) d(x)}{\int_1^{\infty} \sqrt{x-1}\, g(x) h(-tb(x)) d(x)}$$

$$g(x) = x^{\left(-(q+3)/2\right)} (f+q^*F/x)^{\left(-(f+q-1)/2\right)} \qquad b(x) = \sqrt{\frac{(q+f)(x-1)}{f*x+q*F}} \qquad b(x) = \sqrt{\frac{(q+f)(x-1)}{f*x+q*F}}$$



Ecuación a resolver es:

W(t) = 0

> waller(K=100, q=5, f=24, Fc=14.37)

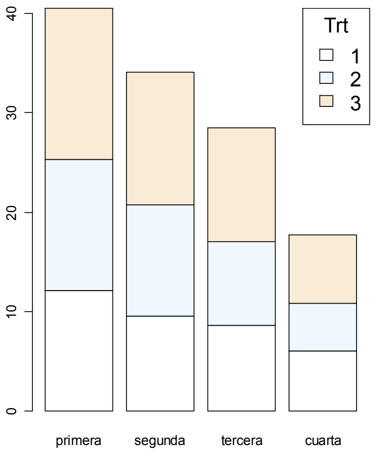
[1] 1.919

> TAPPLY.STAT(……)

```
tapply.stat(papa[,2],papa[,3:6],mean)
tapply.stat(papa[,2],papa[,3:6],function(x) sd(x)*100/mean(x))
```

```
> papa[,3:6] <-round(papa[,3:6],1)
> papa
   bloque tratamiento primera segunda tercera cuarta
1       1           1    10.0    11.2    11.3    8.3
2       2           1    13.0    12.1     8.8    7.7
3       3           1    17.6    10.7    10.5    5.1
4       4           1     9.3     7.9     6.6    5.1
5       5           1    10.7     5.9     6.0    4.1
6       1           2    12.4    13.3    11.6    7.5
7       2           2    20.3    14.1     9.6    4.7
8       3           2    19.0    12.6     8.2    3.9
9       4           2     6.5     7.2     6.6    3.8
10      5           2     7.7     8.6     5.9    4.1
11      1           3    12.1     9.8    11.5    9.8
12      2           3    15.0    15.5    13.0    7.7
13      3           3    18.4    13.4     9.4    6.4
14      4           3    15.4    13.0    11.2    6.3
15      5           3    15.1    15.2    12.0    4.2
```

```
> tapply.stat(papa[,2],papa[,3:6],mean)
  papa[, 2] primera segunda tercera cuarta
1         1   12.12    9.56    8.64   6.06
2         2   13.18   11.16    8.38   4.80
3         3   15.20   13.38   11.42   6.88
```



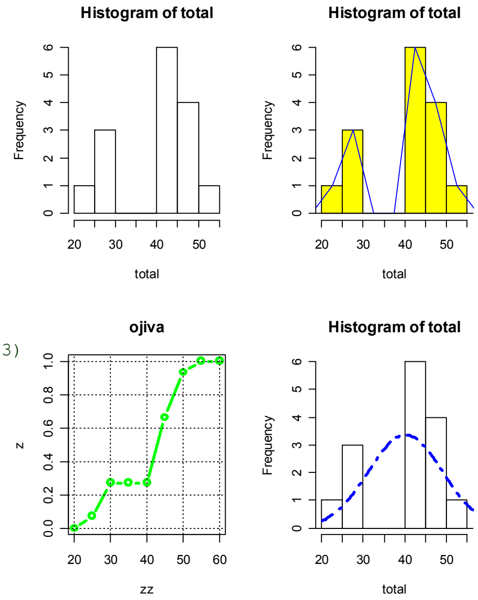
```
> barplot(as.matrix(medias[,-1]), col=colors()[1:3])

> legend("topright",as.character(medias[,1]),fill=colors()[1:3],cex=1.5,title="Trt")
```

Estadistica descriptiva

```
X<-cbind(c(1,1,1,1))
total <- as.matrix(papa[,3:6])%*%X
par(mfrow=c(2,2))
h<-hist(total)
plot(h,col="yellow")
polygon.freq(h,col="blue")
```

```
 ojiva.freq(h,type="b",col="green",
 main="ojiva",lwd=3)
 stat.freq(h)
 plot(h)
 normal.freq(h,col="blue",lty=4,lwd=3)
```

```
> round(table.freq(h),2)
  Inf Sup   MC fi  fri Fi  Fri
   20  25 22.5  1 0.07  1 0.07
   25  30 27.5  3 0.20  4 0.27
   30  35 32.5  0 0.00  4 0.27
   35  40 37.5  0 0.00  4 0.27
   40  45 42.5  6 0.40 10 0.67
   45  50 47.5  4 0.27 14 0.93
   50  55 52.5  1 0.07 15 1.00
```



Diseño de experimentos
aleatorización y libro de campo

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

La planificación de experimentos de campo es una de las principales tareas de la librería Agricolae.

- Soporta un diseño de latice simple (lattice.simple),
- Factorial en un diseño de bloques (design.ab),
- Diseño alfa (alpha.design),
- Bloques incompleto balanceado (design.bib),
- Bloques completos al azar (design.rcbd).
- Completamente al azar (design.crd),
- Greco latino (design.graeco),
- Cuadrado latino (design.lsd).

Diseño Greco Latino

```
> T1<-c("a","b","c","d");
> T2<-c("v","w","x","y");
> Plan <-  design.graeco(T1,T2,number=101)
> t(matrix(paste(Plan$T1,Plan$T2),c(4,4)))
```

---

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

Parcelas

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

Tratamientos

```
      [,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"
```

No es posible para: 6,10 y pares >= 14

Diseño Alfa (01)

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

```
> trt <- letters[1:12]
> plan<-design.alpha(trt,k=3, r=3, number=101, seed=55)
```

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

Efficiency factor

(E ) 0.7096774

<<< Book >>>
```

Libro de campo

```
> plan
  plots cols block trt replication
1     1    1     1   c           1
2     2    2     1   a           1
3     3    3     1   d           1
4     4    1     2   j           1
...
35   35    2    12   g           3
36   36    3    12   d           3
```

Diseño alfa con semilla = 55

```
> fields<-plan[,4]
> dim(fields)<-c(k,s,r)
> for (i in 1:r) print(t(fields[,,i]))

> plan$design   # ver 1.0-5
```

I

```
     [,1] [,2] [,3]
[1,] c    a    d
[2,] j    f    k
[3,] i    l    g
[4,] b    h    e
```

II

```
     [,1] [,2] [,3]
[1,] j    c    b
[2,] k    g    h
[3,] a    i    f
[4,] l    d    e
```

III

```
     [,1] [,2] [,3]
[1,] j    a    e
[2,] l    f    h
[3,] c    k    i
[4,] b    g    d
```

Diseño de bloques incompleto balanceado

```
> trt<-c("A","B","C","D","E")
> k<-3
> bib <-design.bib(trt,k,number=101,kinds ="Super-Duper")
```

```
Parameters BIB
==============
Lambda     : 3
treatmeans : 5
Block size : 3
Blocks     : 10
Replication: 6
Efficiency factor
0.8333333
<<< Book >>>
```

```
> field <-as.character(bib[,3])
> t(matrix(field,c(3,5)))

      [,1] [,2] [,3]
 [1,] "C"  "D"  "E"
 [2,] "E"  "B"  "D"
 [3,] "E"  "C"  "B"
 [4,] "B"  "A"  "D"
 [5,] "A"  "E"  "C"
 [6,] "B"  "C"  "D"
 [7,] "A"  "C"  "B"
 [8,] "C"  "A"  "D"
 [9,] "B"  "E"  "A"
[10,] "D"  "E"  "A"
```

Diseño de bloques completos y cuadrado latino

```
> trt<-c("A","B","C","D")
```

```
> rcbd <-design.rcbd(trt,5,45, kinds= "Super-Duper")
> plan <-as.character(rcbd[,3])
> dim(plan) <-c(4,5)
> print(t(plan))
```

Bloques

```
•       [,1] [,2] [,3] [,4]
•  [1,] "C"  "B"  "D"  "A"
•  [2,] "B"  "D"  "C"  "A"
•  [3,] "A"  "B"  "C"  "D"
•  [4,] "B"  "A"  "D"  "C"
•  [5,] "D"  "C"  "A"  "B"
```

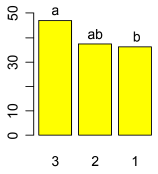
```
> lsd <-design.lsd(trt,45,kinds="Super-Duper")
> plan <-as.character(lsd[,4])
> dim(plan) <-c(4,4)
> print(t(plan))
```

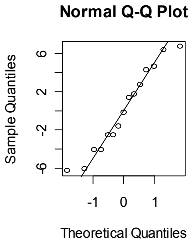
Latino
- Filas
- Columnas

```
     [,1] [,2] [,3] [,4]
[1,] "C"  "A"  "B"  "D"
[2,] "D"  "B"  "C"  "A"
[3,] "A"  "C"  "D"  "B"
[4,] "B"  "D"  "A"  "C"
```

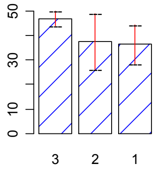
Comparación multiple

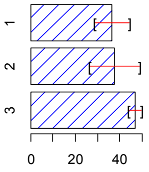
```
> modelo <- aov(primera+segunda+tercera+cuarta ~ bloque + tratamiento,papa)
> anova(modelo)
> cv.model(modelo)

> attach(papa)

> comparacion<- HSD.test(primera +
segunda + tercera +
cuarta,tratamiento, 8, 33.08)
```





```
> bar.group(comparacion, ylim=c(0,55), col="yellow")
> error<-residuals(modelo)
> shapiro.test(error)
> qqnorm(error)
> qqline(error)

> bar.err(comparacion,
ylim=c(0,55),col="blue",density=4)
```





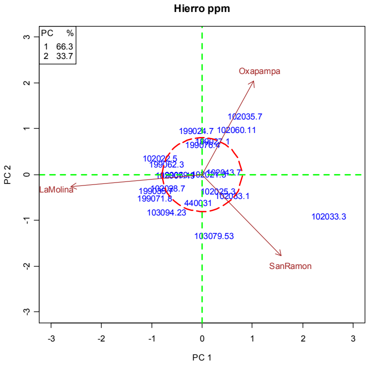
```
> bar.err(comparacion, xlim=c(0,55),col="blue",density=8,horiz=T)
```

AMMI : Estabilidad

```
> model <- AMMI(Localidad, Cip_Number,Rep, Fe,xlim=c(-3,3),ylim=c(-3,3)
,number=F, main="Hierro ppm")

ANALYSIS AMMI:  Fe

Number of observations:  120

model Y: Fe  ~ ENV + REP%in%ENV + GEN + ENV:GEN
Random effect REP%in%ENV


 Analysis of Variance Table

 Response: Y
           Df  Sum Sq Mean Sq  F value    Pr(>F)
 ENV        2 1273.82  636.91 169.3027  0.000823 ***
 REP(ENV)   3   11.29    3.76   1.6092  0.197280
 GEN       19  703.11   37.01  15.8293 3.143e-16 ***

 ENV:GEN   38  490.51   12.91   5.5215 4.463e-09 ***
 Residuals 57  133.26    2.34
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Coeff var       Mean Fe
 9.914176         15.42227

 Analysis
     percent  acum Df   Sum.Sq   Mean.Sq F.value Pr.F
 CP1    66.3  66.3 20 325.4555 16.272773    6.96    0
 CP2    33.7 100.0 18 165.0577  9.169872    3.92    0
 CP3     0.0 100.0 16   0.0000  0.000000    0.00    1
```

```
> AMMI.contour(model,distance=0.3,shape=20,col="red",lwd=2,lty=5)

Limit, radio: 0.8026402
Genotype  in: 11
Genotype out: 9

$`GENOTYPE IN`
 [1] "102007.19" "102021.5"  "102025.3"  "102028.7"  "102033.1"  "102043.7"
 [7] "199027.1"  "199062.3"  "199069.1"  "199076.4"  "440031"

$`GENOTYPE OUT`
[1] "102022.5"  "102033.3"  "102035.7"  "102060.11" "103079.53" "103094.23"
"199024.7"  "199035.7"  "199071.8"
```

```
          distance
   102007.19         0.5363391
   102021.5          0.1360669
   102022.5          0.9136192
   102025.3          0.4729085
   102028.7          0.7337491
   ....
   199069.1          0.4897033
   199071.8          1.0525786
   199076.4          0.6850364
   440031            0.5925231
```



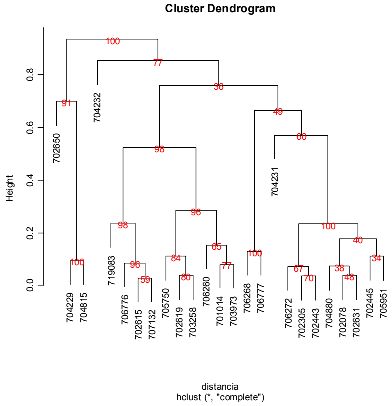
Consensus cluster

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",
...)
```

```
data(pamCIP);   rownames(pamCIP)<-substr(rownames(pamCIP),1,6)
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)
```

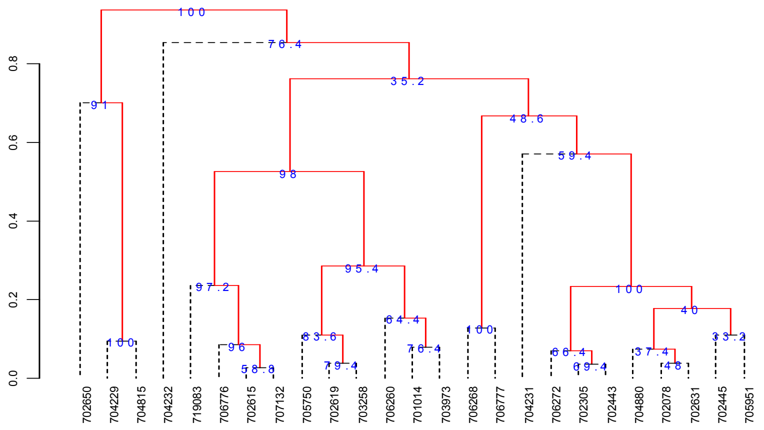
classical dendrogram

```
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

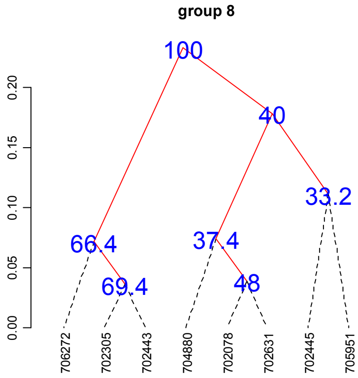
hcut()

```
(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

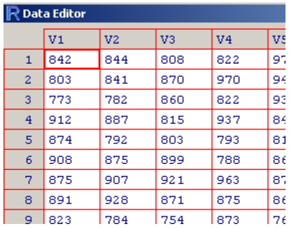
```
Index.smith(data, ...)
```

```
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
```

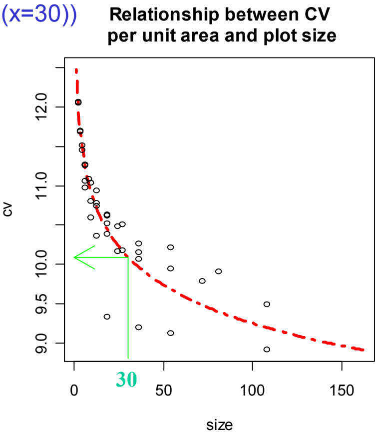
If plot size = 30 unit ^2
then CV = 10 %

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



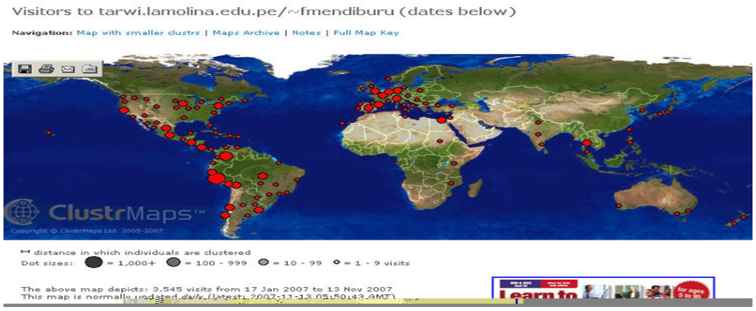
Other functions and data sets

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
```

2007

Enero
-
Diciembre



Visitors to tarwi.lamolina.edu.pe/~fmendiburu (dates below)

The above map depicts: 3,545 visits from 17 Jan 2007 to 13 Nov 2007
This map is normally updated daily (latest: 2007-11-13 05:50:43 GMT)

2008

Enero
-
Abril.

Prof. Dr. Christine Müller
Fachbereich Mathematik/Informatik
Universität Kassel

Lineare Modelle und Versuchsplanung WS 2007/2008
Linear Models and Experimental Design WS 2007/2008

http://www.mathematik.uni-kassel.de/~cmueller/Lehre/Versuchsplanung_WS0708.html

•**Content** The course has three parts: the lecture on Tuesday, which provides the methods, the tutorial and practice on Tuesday, where the methods are trained at data sets, and a second lecture for mathematicians at Thursday, where the mathematical foundations are given.
The aim of the methodical parts of the course is that students should be able to analyse and design complex surveys and experiments with the free statistical software R. In particular the R package “agricolae” is used but also other R packages for statistical analysis and for designing experiments are mentioned. At first an introduction into the statistical software R is given and foundations of statistical testing are repeated at the example of the two-sample t-test. Then the analysis of variance (ANOVA) of one-way, two-way and multi-way layouts are presented and corresponding design questions as completely randomised block designs, balanced incomplete block designs, and split-block designs are considered. In a second part, regression experiments are studied, where linear and polynomial regression, multiple regression, the analysis of covariance and corresponding design considerations are treated. At last the multivariate analysis of variance (MANOVA) is presented.

Agricolae Version 1.0-5

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