

Universidad Nacional de Ingeniería

7 Diciembre 2007

**Día del Estadístico  
5 Diciembre**

**Diseño y Análisis de  
experimentos Aplicados  
en investigación Agrícola**

**Felipe de Mendiburu**



# Contenido

---

- 1. La investigación y el diseño de experimentos.**
2. Principios básicos de experimentación
3. Uso del R para diseños de experimentos
4. Aplicaciones

## ¿Porque Investigar?

Casos sobre la incidencia de plagas, enfermedades que afectan al cultivo, el efecto del suelo, agua, nutrientes y labores culturales son componentes importantes que de alguna forma intervienen en el desarrollo del cultivo favoreciendo o no la producción.

**Estas situaciones deben ser estudiadas mediante la experimentación.**

## ¿Porqué diseñar el experimento?

**En primer lugar el investigador se formula una serie de preguntas, que espera tener respuesta al iniciar, conducir y culminar el experimento, por ejemplo:**

¿Cómo medir el efecto de estudio?,

¿Cuáles serán las características a analizar?

¿Qué factores deben estudiarse?.

¿Cuántas veces debe realizar experimentos preliminares antes de conducir un experimento formal?.

¿Cuál sería el modelo de estudio para los datos del experimento?.

## Diseño Experimental

El Diseño es una etapa fundamental de la experimentación.

Experimentación comprende toda investigación científica que se realiza por la repetición del mismo.

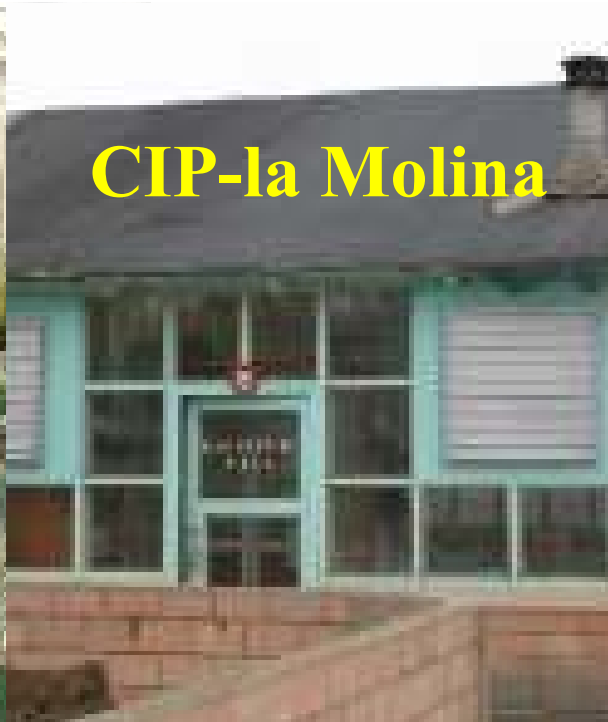
El diseño comprende la forma de aplicar los tratamientos a las unidades experimentales.

**Mediante un modelo estadístico se cuantifica la variación debido a factores controlables y no controlables.**

## Diseños clásicos aplicados en el Campo agrícola e Invernadero

Completo al azar,  
Diseños de bloques,  
Cuadrado latino,  
Parcelas divididas,  
Bloques divididos, etc.

**En campo de agricultores puede planearse los diseños de bloques o diseños con balance incompleto, grupos balanceados, aumentados, Mother & baby, etc .**





# Huancayo







SAN  
RAMON





Preparación del terreno



## Diseño en Bloques







**Barrera**

**Barrera**

**Barrera**



# Rotación de cultivos





## Experimento para el control de rancho





**Experimento para evaluar la resistencia a enfermedades**





**Banco de  
Germoplasma de  
papa nativa**





# Contenido

---

1. La investigación y el diseño de experimentos.
- 2. Principios básicos de experimentación**
3. Uso del R para diseños de experimentos
4. Aplicaciones

¿Qué objetivo debe lograr el investigador al diseñar el experimento?.

Conseguir toda la información relacionada al problema en estudio.

Lograr un diseño simple y eficiente como sea posible.

Optimizar los recursos de tiempo, dinero, personal y material experimental

## **Diseño Experimental**

**El tema del diseño experimental ha recibido siempre mucha atención**

El investigador dispone de nuevos métodos estadísticos para la solución de los problemas.

**Los problemas involucran las condiciones del campo en área experimental y en campo de agricultores, laboratorio e invernadero.**

Mediante el uso de diseños y análisis estadístico apropiado, las decisiones son confiables

## Avance en la investigación

---

Desde sus inicios a la fecha, el desarrollo de nuevas técnicas a permitido mejorar la investigación científica, sin embargo, los principios propuestos por Fisher (1924) sigue teniendo vigencia:

**Los principios básicos del diseño experimental, fueron desarrollados por R. A. Fisher y sus asociados en el Rothamsted Experimental Station en Inglaterra.**



## Los tres principios básicos

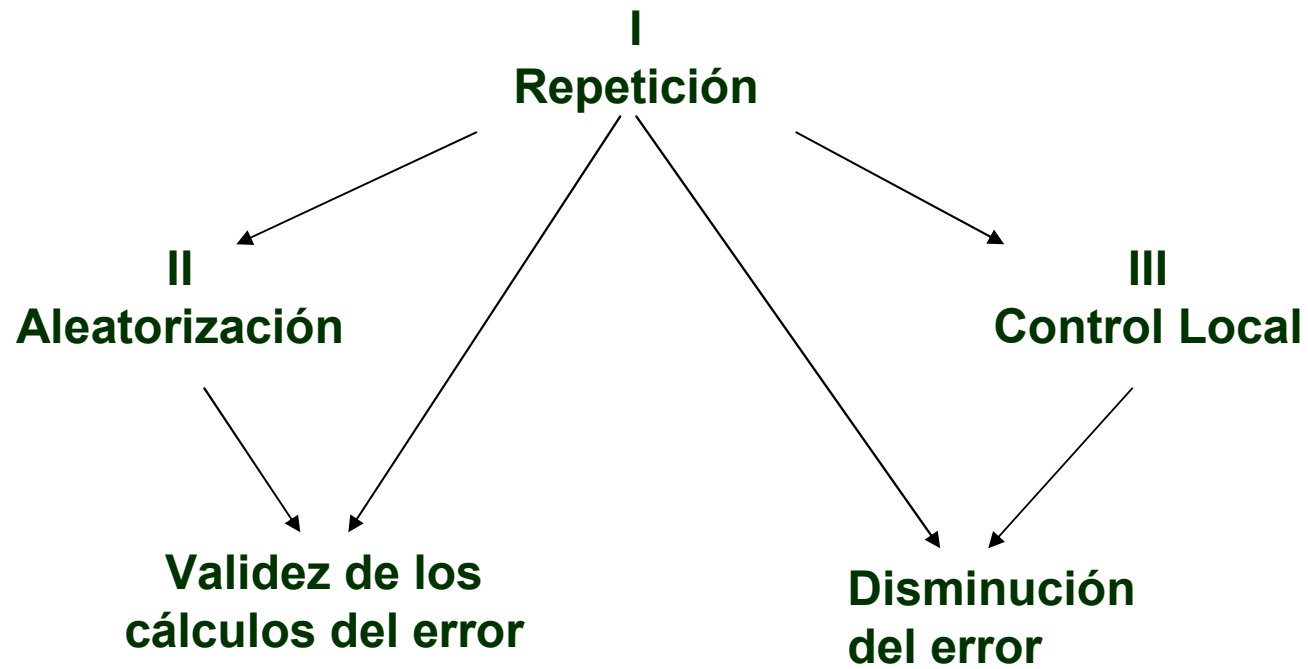
---

- 1. la aleatorización,**
- 2. la repetición y**
- 3. control local.**

---

**Es de vital importancia para el investigador comprender la lógica de estos principios para diseñar experimentos eficaces**

## Fisher ilustra las relaciones generales



# Contenido

---

1. La investigación y el diseño de experimentos.
2. Principios básicos de experimentación
- 3. Uso del R para diseños de experimentos**
4. Aplicaciones



Robert Gentleman



Lenguaje S



Proyecto R



Ross Ihaka

## R & R

R dispone de librerías apropiadas para el diseño y análisis de experimentos:

*Agricolae*: diseños de experimentos

config.design: diseños confundidos

BHH2: Generación de la matriz diseño factorial completo y fraccionado

AlgDesign: Algoritmo de diseños experimentales

multcomp y mvtnorm: Para contrastes ortogonales.

experiment: randomización de tratamientos

# 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* de R.

Soporta un diseño de lattice simple ( [lattice.simple](#)),  
Factorial en un diseño de bloques ([design.ab](#)),  
Alpha diseño ([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)
```

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

Parcelas					Tratamientos				
	[,1]	[,2]	[,3]	[,4]		[,1]	[,2]	[,3]	[,4]
[1,]	101	102	103	104	[1,]	"d w"	"b v"	"a x"	"c y"
[2,]	105	106	107	108	[2,]	"b y"	"d x"	"c v"	"a w"
[3,]	109	110	111	112	[3,]	"a v"	"c w"	"d y"	"b x"
[4,]	113	114	115	116	[4,]	"c x"	"a y"	"b w"	"d v"

No es posible para: 6,10 y pares  $\geq 14$



# Diseño Alfa (01)

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

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

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
0.83333333
<<< 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"

# Contenido



1. La investigación y el diseño de experimentos.
2. Principios básicos de experimentación
3. Uso del R para diseños de experimentos
- 4. Aplicaciones**

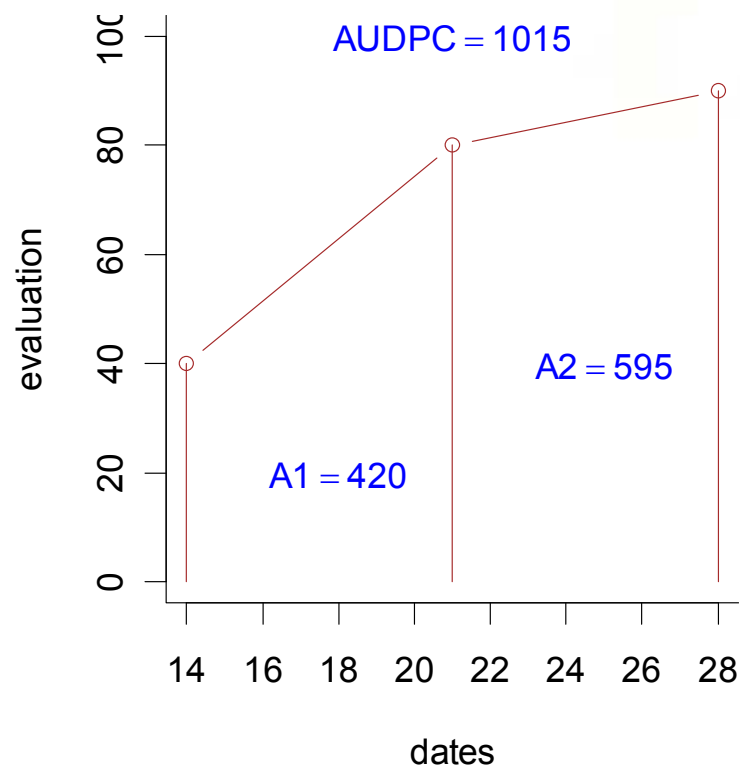
## Midiendo la enfermedad en el tiempo

```
> library(agricolae)
> dates<-c(14,21,28)
> evaluation<-data.frame(E1=40,E2=80,E3=90)
```

```
> absolute<-audpc(evaluation,dates,type="absolute")
> absolute
[1] 1015
>
```

AUDPC = area bajo la curva del progreso de la enfermedad

```
A1 <- 7*(80+40)/2
A2 <- 7*(80+90)/2
A<-A1+A2
```





```

> library(agricolae)
> data(disease)
> dates<-c(7,14,21)
> indice<-audpc(disease[,4:6], dates, type = "relative")
> disease<-data.frame(disease,audpc=indice)
> disease<-data.frame(disease,audpc=indice)

```

## ANTES

```

> disease
  plots rep trt E2 E5 E7 yield
1   101  1  T1  4  9 10   16
2   102  1  T2  3  6  7   11
3   103  1  T3  3  4  6   24
4   104  1  T4  0  0  2   27
5   105  1  T5  0  0  2   28
6   106  1  T6  2  2  4   23
7   107  1  T0  7 14 23   11
8   201  2  T1  6 12 14   14
9   202  2  T2  3  7 14   18
10  203  2  T3  2  3  5   26
...
18  304  3  T4  5 12 15   11
19  305  3  T5  2  9 15   10
20  306  3  T6  3  6 12   11
21  307  3  T0 11 21 35    8

```

## DESPUES

```

> disease
  plots rep trt E2 E5 E7 yield audpc
1   101  1  T1  4  9 10   16 0.0800
2   102  1  T2  3  6  7   11 0.0550
3   103  1  T3  3  4  6   24 0.0425
4   104  1  T4  0  0  2   27 0.0050
5   105  1  T5  0  0  2   28 0.0050
6   106  1  T6  2  2  4   23 0.0250
7   107  1  T0  7 14 23   11 0.1450
8   201  2  T1  6 12 14   14 0.1100
9   202  2  T2  3  7 14   18 0.0775
10  203  2  T3  2  3  5   26 0.0325
...
18  304  3  T4  5 12 15   11 0.1100
19  305  3  T5  2  9 15   10 0.0875
20  306  3  T6  3  6 12   11 0.0675
21  307  3  T0 11 21 35    8 0.2200

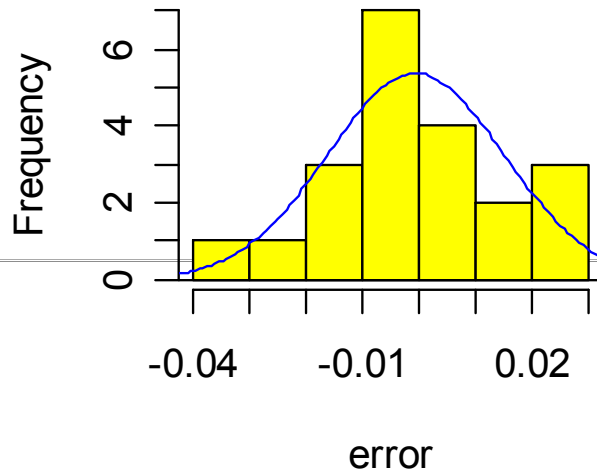
```

```
> disease[,2]<-as.factor(disease[,2])
> attach(disease)
> correlation(yield,audpc)
> modelo<- aov(audpc ~ rep + trt, disease)
> anova(modelo)
> cv.model(modelo)
> error <-residuals(modelo)
```

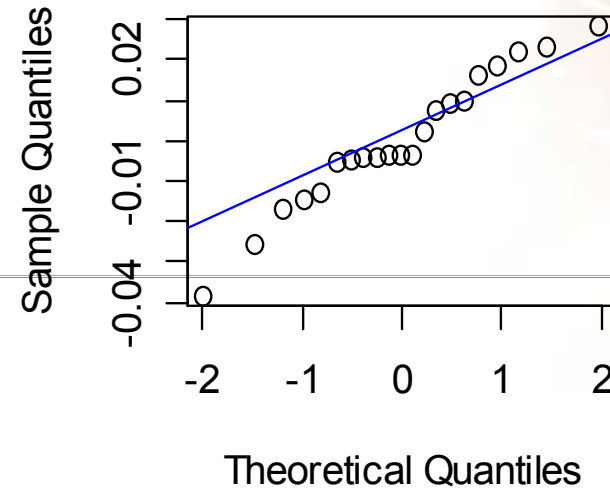
```
shapiro.test(error)
par(mfrow=c(2,2))
normal.freq(hist(error,col="yellow"),col="blue")
qqnorm(error)
qqline(error,col="blue")
modelo<- aov(asin(sqrt(audpc))*180/pi ~ rep + trt, disease)
anova(modelo)
cv.model(modelo)
error <-residuals(modelo)
normal.freq(hist(error,col="yellow"),col="blue")
qqnorm(error)
qqline(error,col="blue")
shapiro.test(error)
compara<-LSD.test(asin(sqrt(audpc))*180/pi, trt,12,9.97)
```

**Sin  
transformación**

**Histogram of error**

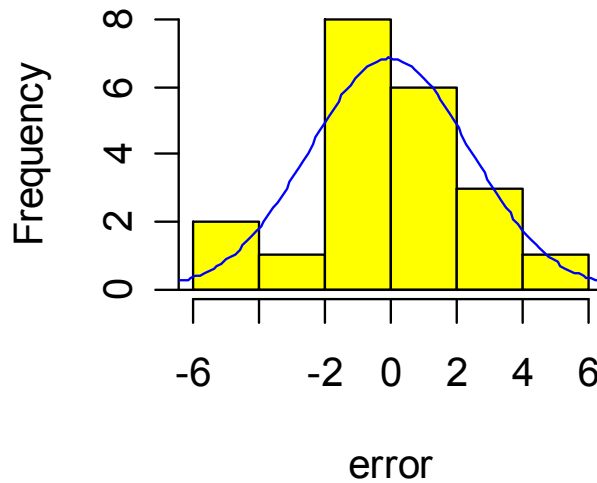


**Normal Q-Q Plot**

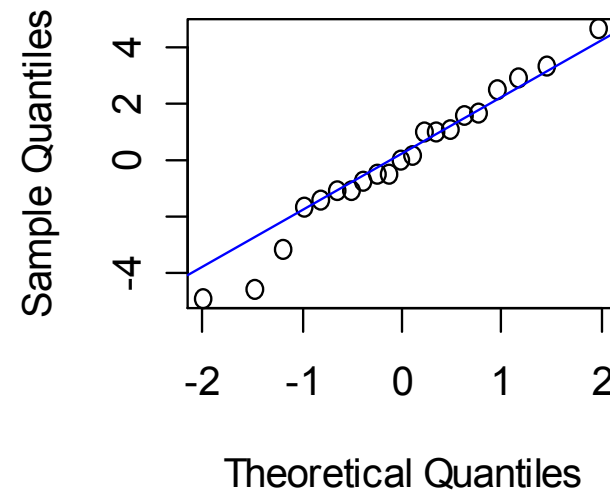


**Con  
transformación  
Angular**

**Histogram of error**



**Normal Q-Q Plot**



```
> shapiro.test(error)
```

Shapiro-Wilk normality test

data: error

W = 0.9685, p-value = 0.6996

```
> shapiro.test(error)
```

Shapiro-Wilk normality test

data: error

W = 0.9775, p-value = 0.8861

```
Response: audpc
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
rep	2	0.012070	0.006035	13.016	0.0009867	***
trt	6	0.041799	0.006966	15.026	5.952e-05	***
Residuals	12	0.005564	0.000464			

---

```
> cv.model(modelo)
```

```
[1] 29.26723
```

```
Response: asin(sqrt(audpc)) * 180/pi
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
rep	2	190.38	95.19	9.5453	0.0033061	**
trt	6	512.16	85.36	8.5595	0.0009077	***
Residuals	12	119.67	9.97			

---

```
> cv.model(modelo)
```

```
[1] 21.68161
```

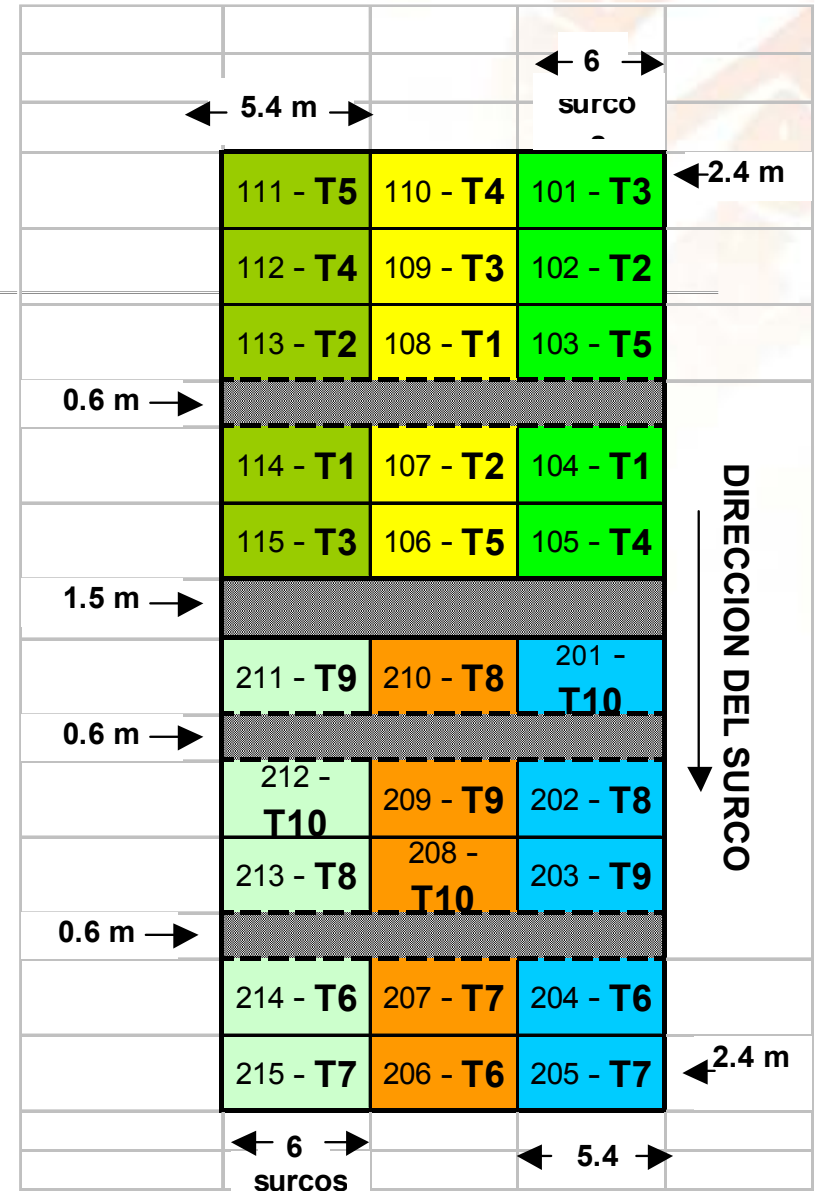
## Cultivo PAPA

Ensayo con Algas: RCBD (3 bloques)  
 2 experimentos (Organico y Quimico )  
 Cada Experimento con 5 tratamientos  
 Diferentes. (28-06-2007) En Estudio

### Tratamientos

T1	Control absoluto	Tratamientos biológicos:
T2	Control 10 ton/ha estiércol	
T3	Alga - 2 ton/ha	
T4	Alga - 5 ton/ha	
T5	Alga - 10 ton/ha	
T6	Control químico 50% □ 70-45-50 NPK	Tratamientos químicos:
T7	Control químico 100% □ 140-90-100 NPK	
T8	Control 10 ton/ha estiércol + 70-45-50 NPK	
T9	Alga - 5 ton/ha + 70-45-50 NPK	
T10	Alga - 10 ton/ha + 70-45-50 NPK	

### Croquis



## Estudio Genotipo - Medio Ambiente (2006)

Localidades: Oxapampa, San Ramon, La Molina

10 plantas en 2.25 m<sup>2</sup> ( 1.25 largo del surco x 2 surcos x 0.9 m que es la separación entre surcos)

3 Localidades, 12 genotipos y 2 bloques por localidad

Localidad	Cip_Number	Rep	Fe
Oxapampa	199024.7	1	13.37
Oxapampa	199024.7	2	15.00
Oxapampa	199027.1	1	12.03
Oxapampa	199027.1	2	11.00
Oxapampa	199035.7	1	10.88
Oxapampa	199035.7	2	13.00
Oxapampa	199062.3	2	14.00
Oxapampa	199062.3	1	12.84
Oxapampa	199069.1	1	10.00
Oxapampa	199069.1	2	9.80



LaMolina	102021.5	2	26.000
LaMolina	102028.7	1	30.226
LaMolina	102028.7	2	27.000

Cultivo: CAMOTE

Variables Observadas:

Follaje: Kg.

Peso Tub: Kg.

Materia seca: %

Almidón: %

Zn, Ca, Mg, Fe : ppm

Proteína: %

Beta Caroteno: ppm

Cortesía: Fredy Gamarra

```
> 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
```

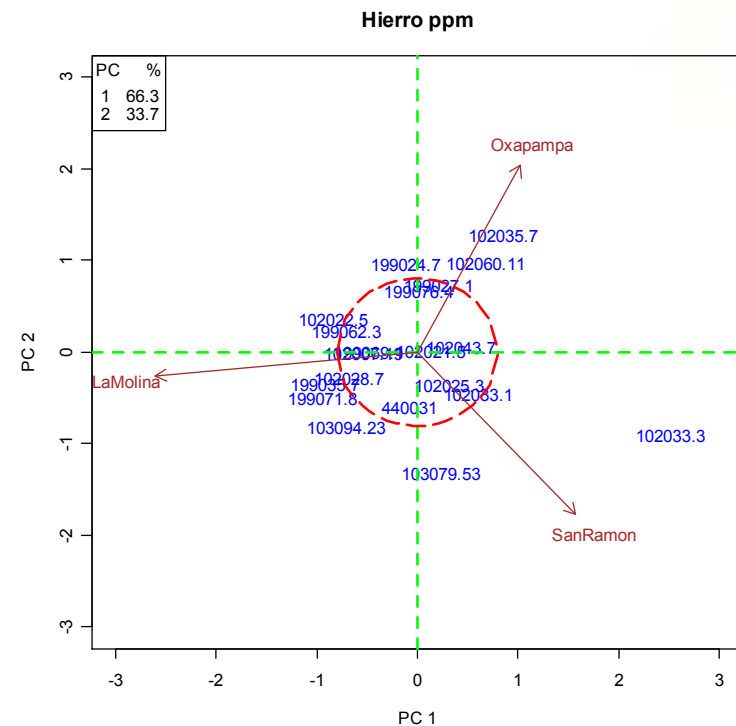
```
$`GENOTYPE IN`
```

```
[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`
```

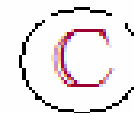
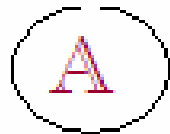
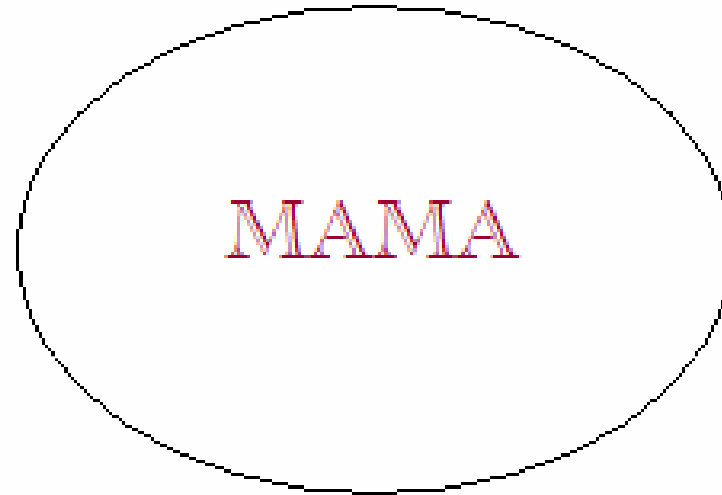
```
[1] "102022.5" "102033.3" "102035.7" "102060.11" "103079.53" "103094.23"
"199024.7" "199035.7" "199071.8"
```

	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





Experimento  
formal con  
repeticiones



BEBES

Experimento en parcelas de  
Agricultores

Sin Repeticion



## **Evaluación de 10 clones de papa en el valle Chillón. 2004. Ensayo CIP.**

Experimento: Diseño mother-baby.

Mother: RCBD con 3 bloques y 3 campos de agricultores.

**Fecha:** Noviembre 2004

**Días después de la siembra:** 133,129,128 y 129

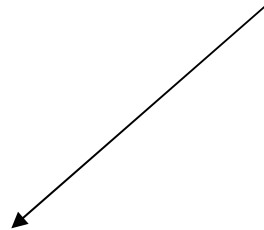
**Evaluación 1:** Peso en kilos del rendimiento de 20 plantas observadas en el campo principal (mama) y en otros tres campos de agricultores (bebes).

**Evaluación 2:** Participación de los agricultores en el campo Mama.

## Evaluacion Participativa

### Parcela MAMA

### Parcelas BEBES



bloque	clon	Tuber_c	Peso_c	Evaluacion
I	C1	65	10.3	49
I	C2	40	3.3	3
I	C3	43	4.2	0
I	C4	35	6.3	15
I	C5	5	0.9	0
I	C6	44	8.5	31
I	C7	21	3.8	1
I	C8	74	8.1	20
I	C9	26	3.7	0
I	C10	20	3.5	1
II	C1	64	10.6	38
II	C2	27	2.1	0

III	C9	15	1.9	1
III	C10	55	11.2	13

bebe	clon	Tuber_c	Peso_c
A	C1	65	13.8
A	C2	96	10.9
A	C3	154	24.9
A	C4	101	15.8
A	C5	85	15.1
A	C6	130	30.3
A	C7	71	12.4
A	C8	153	14.5
A	C9	81	16.6
A	C10	130	18.6
B	C1	75	16.1

C	C7	48	8.7
C	C8	88	10.2
C	C9	109	18.4
C	C10	86	12.2

# Parcela MAMA

```
> library(RODBC)
> library(agricolae)
> canal <- odbcConnectExcel( "mother-baby.xls" )
> exp1 <- sqlFetch(canal,"mother")
> exp2 <- sqlFetch(canal,"babies")
> odbcCloseAll()
> model <- aov(Peso_c ~ bloque + clon , exp1)
> anova(model)
```

## Analysis of Variance Table

Response: Peso\_c

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloque	2	20.909	10.454	2.1240	0.14854
clon	9	177.783	19.754	4.0133	0.00586 **
Residuals	18	88.598	4.922		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05  
'.' 0.1 ' ' 1

```
> cv.model(model)
[1] 34.64
```

Para el test de Waller-Duncan

```
> comparison <- waller.test(Peso_c, clon, 18, 4.922, Fc=4.0133)
```

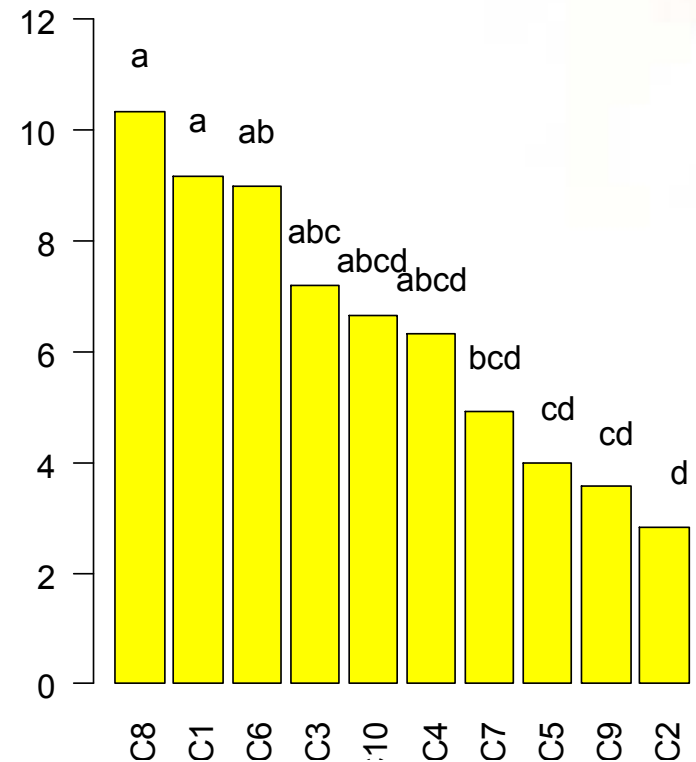
This test minimizes the Bayes risk under additive loss and certain other assumptions.

```
.....  
K ratio                100.0000  
Error Degrees of Freedom 18.0000  
Error Mean Square      4.9220  
F value                4.0133  
Critical Value of Waller 2.2610
```

### Comparison of treatments

#### Groups, Treatments and means

a	C8	10.33333
a	C1	9.166667
ab	C6	9
abc	C3	7.2
abcd	C10	6.666667
abcd	C4	6.333333
bcd	C7	4.933333
cd	C5	4
cd	C9	3.566667
d	C2	2.833333



```
> comparison <- friedman(bebe, clon, Peso_c)
```

Friedman's Test

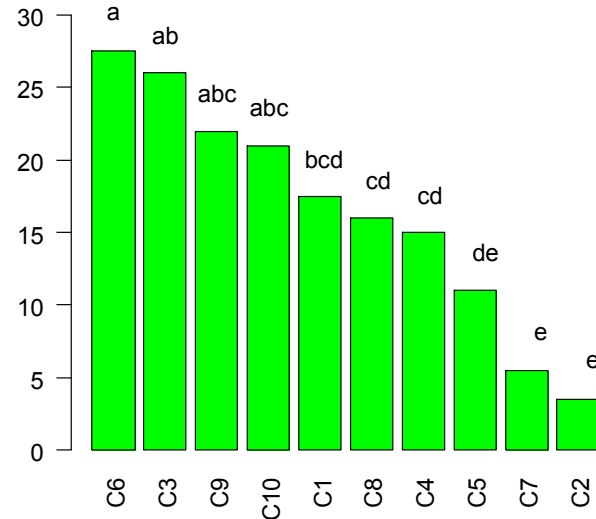
=====

```
Chi-squard: 21.37728      Alpha      : 0.05
Pvalue      : 0.01107658  t-Student  : 2.100922
pFriedman   : 0.0002746419 LSD          : 8.690586
```

Means with the same letter are not significantly different.

Group, Treatment and Sum of the ranks

a	C6	27.5
ab	C3	26
abc	C9	22
abc	C10	21
bcd	C1	17.5
cd	C8	16
cd	C4	15
de	C5	11
e	C7	5.5
e	C2	3.5



# Evaluacion Participativa de los agricultores

```
> comparison <- friedman(bloque, clon, Evaluacion)
```

Friedman's Test

=====

Chi-squard: 18.88889

Pvalue : 0.0261544

pFriedman : 0.003921009

Alpha : 0.05

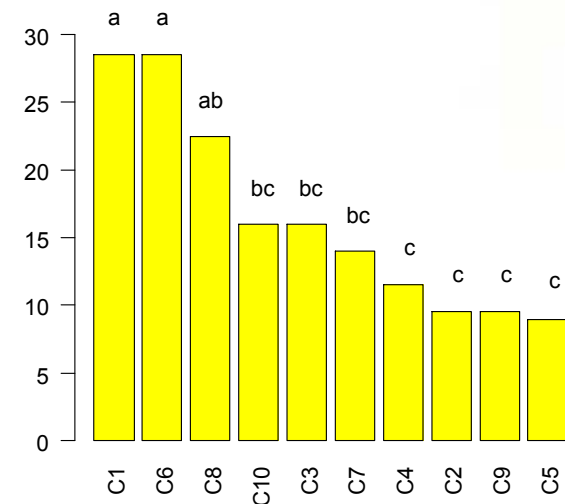
t-Student : 2.100922

LSD : 10.36360

Means with the same letter are not significantly different.

Group, Treatment and Sum of the ranks

a	C1	28.5
a	C6	28.5
ab	C8	22.5
bc	C10	16
bc	C3	16
bc	C7	14
c	C4	11.5
c	C2	9.5
c	C9	9.5
c	C5	9



**Prueba de  
Friedman**



---

*Agricolae*