

# *Planeando y diseñando un experimento agrícola*

Sharon Nielsen

Traducido por Sabela Muñoz

Estadística

sabela.munozsanta@adelaide.edu.au





Este taller es parte del programa de formación de *Statistics for Australian Grains Industry* en la Universidad de Adelaida, que está cofinanciado por la entidad *Australian Grains Research and Development Corporation*, proyecto UA00164. El material debe utilizarse estrictamente para fines de formación y queda prohibida su distribución y modificación.

# Índice general

Índice general	2
<b>1 Planeando y diseñando un experimento agrícola</b>	<b>5</b>
1. Definiciones	6
1.1. Población y Muestra	6
1.2. Tratamientos	6
1.3. Unidades experimentales y observacionales	6
1.4. Réplicas	7
1.4.1. Pseudoréplicas	7
1.5. Bloques	8
1.6. Aleatorización	8
1.6.1. Variabilidad y la media	8
1.6.2. Aleatorización	9
1.7. Factores confundidos	10
1.8. Heterogeneidad	10
1.9. Factores y niveles	10
1.10. Efectos principales e interacciones	11
<b>2 Diseños</b>	<b>15</b>
1.11. Generando números aleatorios	16
1.12. Diseños totalmente aleatorizados	16
1.12.1. Características del diseño	16
1.12.2. Uso apropiado	16
1.12.3. Creando este tipo de diseños	16
1.12.4. Comprobación del diseño	17
1.12.5. archivo EXCEL	19
1.13. Diseño en bloques completos aleatorizados	20
1.13.1. Características del diseño	20
1.13.2. Uso apropiado	20
1.13.3. Creando el diseño	20
1.13.4. Comprobación del diseño	21
1.13.5. archivo EXCEL	21
1.14. Cuadrado latino	22
1.14.1. Características del diseño	22
1.14.2. Uso apropiado	22
1.14.3. Creando el diseño	22
1.14.4. Comprobando el diseño	23
1.14.5. Archivo EXCEL	23
1.15. Estructura de los tratamientos	25
1.15.1. Estructura de tratamiento con factores cruzados	25
1.15.2. Estructura de tratamiento con factores anidados	26
1.15.3. Creando el diseño	26
1.15.4. Comprobación del diseño	27
1.15.5. Archivo EXCEL	28
1.15.6. Comprobación del diseño	29
1.15.7. Archivo EXCEL	30
1.15.8. Comprobación del diseño	31

1.15.9.	Archivo EXCEL . . . . .	32
1.16.	Diseño de parcelas divididas . . . . .	32
1.16.1.	Características del diseño . . . . .	32
1.16.2.	Uso apropiado . . . . .	33
1.16.3.	Creando el diseño . . . . .	33
1.16.4.	Comprobación del diseño . . . . .	34
1.16.5.	Archivo EXCEL . . . . .	35
2.	Grados de libertad del error . . . . .	36
2.1.	Diseño totalmente aleatorizado . . . . .	36
2.2.	Diseño en bloques completos aleatorizados . . . . .	36
2.3.	Parcelas divididas . . . . .	37
3.	Diseños más complejos . . . . .	39
3.0.1.	Diseño de parcelas en banda o de bloques divididos . . . . .	39
3.0.2.	<i>p</i> -rep . . . . .	44
Referencias	. . . . .	47



## Capítulo 1

# Planeando y diseñando un experimento agrícola

Este curso está en gran parte basado en el contenido del libro *Statistical Methods in Biology: Design and Analysis of Experiments and Regression* (Welham, Gezan, Clark, y Mead, 2014), así que en vez de citar dicho libro continuamente, preferimos referirnos a él al principio de estas notas.

El rigor científico significa implementar los más altos estándares y las mejores prácticas al método científico y aplicarlos a nuestra investigación. Se trata de descubrir la verdad. El rigor científico requiere minimizar los sesgos en la selección de individuos y en el análisis de datos. Es desde esta perspectiva que vamos a presentar este taller.

Hofseth (2018) se refiere a rigor científico como *“Las reglas del rigor científico pueden diferir de estudio a estudio... pero generalmente incluyen... réplicas en cada experimento... aleatorización... métodos estadísticos robustos y precisos, diseños de experimentos apropiados y exactos... modelos adecuados y autenticados”*.

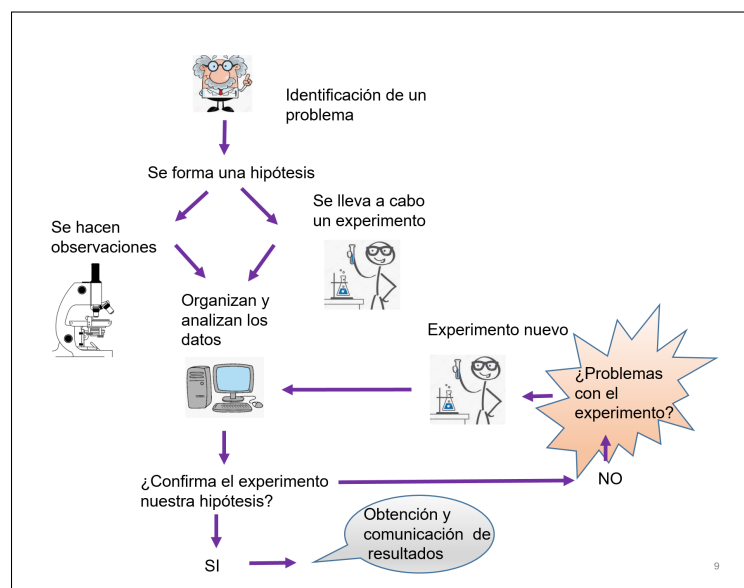


Figura 1.1: La estadística y el método científico<sup>1</sup>

<sup>1</sup> Fuente original:

Adaptado de <https://www.universetoday.com/74036/what-are-the-steps-of-the-scientific-method>

Figuras: <https://accionalaciencia.com/2017/05/06/conoces-a-un-cientifico/>;

<https://www.educima.com/dibujo-para-colorear-microscopio-i16110.html>;

[https://www.goconqr.com/es/p/1747965?dont\\_count=true&frame=true&fs=true](https://www.goconqr.com/es/p/1747965?dont_count=true&frame=true&fs=true)

## 1. Definiciones

### 1.1. Población y Muestra

Una **población** es el grupo de objetos o individuos sobre el que al investigador le gustaría que aplicaran los resultados de su experimento. En numerosas ocasiones es inviable tomar medidas de todos los elementos de la población, por eso se suele seleccionar un subconjunto de ésta denominado **muestra**. Si la muestra se selecciona aleatoriamente, entonces es razonable asumir que la muestra refleja las propiedades de la población y, consecuentemente, las inferencias estadísticas hechas a partir de los datos de la muestra van a poder aplicarse a la población. Esto no se puede garantizar si la muestra no se selecciona aleatoriamente.

---

Selecciona un artículo científico e identifica la población y la muestra que se describen en el mismo. Justifica tu respuesta con evidencias del artículo científico.

**Población**

**Muestra**

---

### 1.2. Tratamientos

En los experimentos, un **tratamiento** es el conjunto de circunstancias que el investigador quiere administrar a las unidades experimentales. El objetivo es evaluar la influencia de los tratamientos en una variable de interés a la que llamamos variable respuesta.

---

Identifica los tratamientos que se describen en el artículo científico seleccionado. Justifica tu respuesta con evidencias del dicho artículo.

**Tratamientos**

---

### 1.3. Unidades experimentales y observacionales

Una **unidad experimental** se define como la unidad más pequeña a la que se le ha aleatorizado un tratamiento de forma independiente. Una **unidad observacional** es la unidad más pequeña sobre la que se han tomado las mediciones. Las unidades experimentales y observacionales pueden o no coincidir.



Figura 1.2: Cuadrante en un campo de trigo

Considera el siguiente experimento. Se parcela un terreno y se asigna aleatoriamente distintas variedades de trigo a cada una de las parcelas. Se toman medidas de la cantidad de biomasa en cinco cuadrantes dentro de cada parcela.



"¿Cuáles son los tratamientos?" – las variedades son los tratamientos.

"¿Cuáles son las unidades experimentales?" – en este caso se dice que las variedades (tratamientos) se han aleatorizado a las parcelas del terreno. Por tanto, las parcelas de terreno son las unidades experimentales.

"¿Cuál es la variable respuesta?" – en este caso la cantidad de biomasa es lo que ha sido medido y es por tanto la variable respuesta.

"¿Cuáles son las unidades observacionales?" – se dice que la cantidad de biomasa se mide en cinco cuadrantes dentro de cada parcela. Es por ello que el cuadrante es la unidad física sobre la que tomamos las medidas y por tanto la unidad observacional. En este caso, la unidad observacional y experimental no coinciden ya que como se ha mencionado anteriormente, la unidad experimental es la parcela.

---

Identifica la unidad experimental y observacional que se describe en el artículo científico seleccionado. Justifica tu respuesta con evidencias del artículo.

**Unidades experimentales**

**Unidades observacionales**

---

## 1.4. Réplicas

Replicar hace referencia a reproducir de forma exacta e independiente un tratamiento. El número de réplicas de un tratamiento se contabiliza teniendo en cuenta el número de unidades experimentales a las que se les ha aplicado de forma independiente ese tratamiento. Replicar nos permite estimar la variabilidad de cada tratamiento. Incrementar el número de réplicas incrementa la información que se tiene de cada tratamiento, lo que usualmente resulta en menos variabilidad y mayor precisión en la estimación del tratamiento. La replicación también es una salvaguarda de la pérdida de información que se puede producir debido a imprevistos (por ejemplo: muestras dañadas por animales o enfermedades). Económicamente, más replicación implica un mayor coste, así que hay que mantener el balance entre aumentar el número de réplicas y la cantidad de dinero, tiempo y recursos que esto conlleva.

---

Identifica el número de réplicas de los tratamientos empleados en el artículo científico seleccionado. Justifica tu respuesta con evidencias del artículo.

**Número de réplicas**

---

### 1.4.1. Pseudoréplicas

La mayoría de los modelos requieren que los tratamientos se repliquen para llevar a cabo inferencia estadística sobre el efecto de los mismos. La replicación permite la estimación de la variabilidad de cada tratamiento. Sin estimar dicha variabilidad, es imposible usar la inferencia estadística. A veces el concepto de réplica se confunde con el de pseudoréplica. ¿Qué se entiende por pseudoréplica? El término pseudoréplica fue acuñado por Hurlbert para referirse a "el uso de la inferencia estadística para probar los efectos de los tratamientos cuando estos no se replican (aunque las unidades de muestreo puede que sí) o cuando las réplicas no son estadísticamente independientes." Generalmente, esto implica realizar múltiples medidas en las unidades experimentales y tratarlas como si fueran respuestas independientes a los tratamientos, lo cual resulta erróneo.

Imagina el siguiente experimento con solo dos parcelas donde en una parcela no se ha aplicado ningún tratamiento, y en la otra se ha aplicado un fertilizante. En este caso, sería erróneo probar el efecto del fertilizante usando la cantidad de grano de múltiples cuadrantes de cada una de las parcelas. Esto sería un ejemplo de pseudoréplica ya que no existe replicación de los tratamientos puesto que estos han sido aleatorizados una sola vez (cada tratamiento a una parcela).

Tratar a las pseudoréplicas como si fueran verdaderas réplicas de los tratamientos es uno de los errores más comunes en el diseño y análisis de los estudios biológicos.

## 1.5. Bloques

Agrupar las unidades experimentales más parecidas entre sí en grupos es una técnica muy común al diseñar un experimento. A estos grupos se les denomina bloques. Los bloques se deben elegir de forma que agrupen unidades experimentales que sean más parecidas dentro del mismo bloque que entre bloques distintos. Los bloques deben ser suficientemente grandes para que incluyan una réplica de cada uno de los tratamientos que se vayan a evaluar en el experimento. En el contexto de experimentos de mejora genética en plantas, ejemplos de bloques podrían ser:

- una serie de columnas adyacentes en el campo de manera que los bloques se alineen con un gradiente natural en el campo.
- macetas en un invernadero agrupadas de manera que se alineen con las condiciones de luz del invernadero.
- sesiones de mañana y de tarde en un laboratorio para llevar a cabo una prueba de calidad.

La agrupación de unidades experimentales en bloques es una herramienta muy importante que suele reducir la varianza de las comparaciones entre tratamientos. No utilizar bloques cuando el experimento así lo requiere resulta en estimaciones de los efectos de los tratamientos (por tanto, de las comparaciones) con variabilidad inflada, es decir, menor precisión y mayor variabilidad.

## 1.6. Aleatorización

### 1.6.1. Variabilidad y la media

Considera que se propone medir la altura de las plantas en un campo de colza. Medir la altura de cada una de las plantas es una tarea muy laboriosa que muchos de nosotros no estaríamos dispuestos a llevar a cabo. Lo más razonable sería medir la altura de una muestra de las plantas de dicho campo, asumiendo que la muestra es representativa de todas las plantas de ese campo.

Midiendo más de una planta se puede estimar no solo la altura de las plantas (calculando la media), sino que también podemos entender la variabilidad innata que existe en las alturas de las plantas de colza.

Considera dos campos de colza con dos variedades de colza diferentes. Se puede medir la altura de las plantas tomando una muestra de cada campo. La altura media de la variedad Ruby se ha estimado en 129.6 cm y en 164.7 cm para la variedad Warriot CL. La gráfica de las medias se muestra en la Figura 1.3:

Observando la Figura 1.3 podríamos concluir que las alturas difieren. Sin embargo, mirando la escala del eje de las  $y$  podríamos pensar que las diferencias han sido acentuadas ya que si el eje de las  $y$  comenzara en cero, las diferencias parecerían mucho menos pronunciadas. De esta manera, sin tener una idea de la variabilidad de la altura de las plantas de colza es difícil concluir si existen diferencias o no.

Considera las Figuras 1.4 & 1.5, ¿se puede asegurar que las alturas de las plantas difieren en alguna de estas dos gráficas?

En la Figura 1.4 vemos que las medias están bastante separadas y que la variabilidad es menor, los diagramas de cajas de ambas variedades no se superponen. Esto es una indicación de que la altura de las plantas podría ser diferente – ciertamente existe evidencia de una diferencia en las muestras que fueron seleccionadas. No

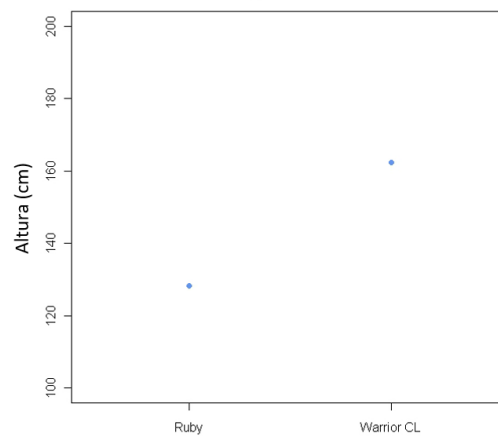


Figura 1.3: Altura media de la variedad Ruby y Warrior CL en plantas de colza

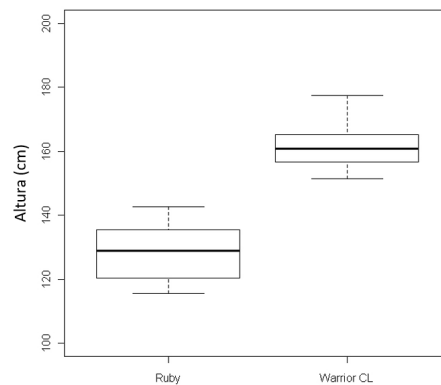


Figura 1.4: Altura media de Ruby y Warrior CL en plantas de colza.

obstante, en los diagramas de cajas de la Figura 1.5 se muestra el caso en el que las alturas son mucho más variables. En este caso no concluiríamos que las alturas son distintas. Las medias de las muestras no parecen ser distintas.

Solo cuando tenemos en cuenta la media y la variabilidad de los datos es cuando podemos determinar si existen diferencias entre los grupos. En investigación, es frecuente preguntarse si existen diferencias, y si existen, saber lo grandes que son. Este proceso de establecer juicios en relación a un parámetro de la población se llama inferencia estadística. McKillup (2006) lo describe como "*la probabilidad de obtener la diferencia observada, o incluso más extrema, entre muestras (o entre valores esperados y observados) si una hipótesis específica (normalmente la hipótesis nula de no diferencias) es verdadera.*" Es la asociación de la probabilidad a los resultados lo que es importante en la inferencia estadística. Una vez que se sabe cómo de probable es el resultado bajo la hipótesis original que se está comprobando, se puede establecer conclusiones sobre las diferencias o relaciones.

### 1.6.2. Aleatorización

Lo primero que hay que tener en cuenta es que aleatorio no es lo mismo que caótico o sin planear.

Una verdadera asignación aleatoria de tratamientos a unidades experimentales se define como aquella en la

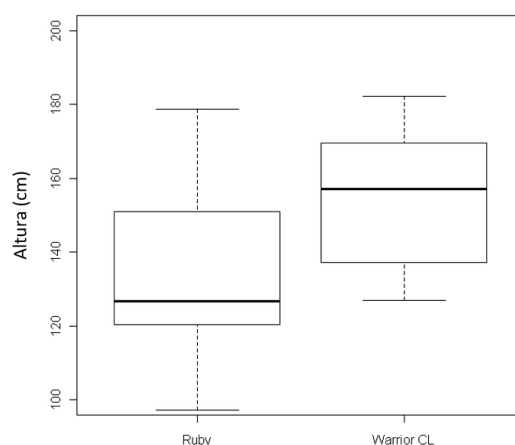


Figura 1.5: Altura media de la variedad Ruby y Warrior CL en las plantas de colza.

que todas las unidades experimentales tienen igual probabilidad de recibir cada uno de los tratamientos. Los detalles de cómo se lleva a cabo esta aleatorización depende del diseño del experimento. Aleatorio no significa que se asignan los tratamientos de forma subjetiva o caótica – la aleatorización debe basarse en la probabilidad para asegurar su validez.

### 1.7. Factores confundidos

Se dice que dos factores están confundidos cuando sus efectos no se pueden disociar unos de otros y en el diseño de experimentos esto se puede apreciar cuando las variables siempre varían juntas. La Figura 1.6 muestra una base de datos donde la columna **rep** y **block** tienen efectos confundidos. La Figura 1.7 muestra una base de datos con factores tratamiento confundidos.

Para evitar confundir efectos accidentalmente primero hay que identificar todas las variables del estudio que están sujetas a poder ser confundidas. Se recomienda llevar a cabo una lista de todas las variables que podrían influir en la variable respuesta. Especialmente, lo que se ha de evitar es que se confundan los efectos de los tratamientos con otra variable.

En algunos casos, utilizar efectos confundidos puede ser una herramienta útil. Por ejemplo, imagina un experimento con un número muy grande de macetas de manera que no sea posible plantar todas las plantas en las macetas en un solo día. Se pueden plantar las macetas por bloques, de manera que los días de plantación estén confundidos con los efectos de los bloques. El efecto del bloque y el día de plantación nunca podrán disociarse, pero si los objetivos del experimento no involucran conocer nada sobre el día de plantación, esta estrategia sería razonable. No obstante, los tratamientos de interés nunca han de confundirse con otra variable.

### 1.8. Heterogeneidad

Al diseñar un experimento es conveniente que las unidades experimentales que se usen para comparar los tratamientos sean relativamente uniformes (homogéneas). De esta forma, se reduce la variabilidad en la respuesta de estas unidades experimentales, lo que conlleva que aumente la precisión y posibilidad de detectar diferencias entre los tratamientos. Si las unidades experimentales son muy diversas entre ellas (heterogéneas), puede que produzcan respuestas muy distintas y no se pueda apreciar las diferencias entre los tratamientos.

### 1.9. Factores y niveles

Las variables explicativas pueden ser cuantitativas o cualitativas. Las variables explicativas cualitativas se denominan **factores** y los distintos grupos o categorías dentro de un factor se denominan **niveles**. Por ejemplo, si se evalúa la productividad de distintas variedades de trigo, la variedad sería un factor y los niveles de dicho factor podrían ser: Beacom, Emu Rock y Manning.

rep	block	trt	yield
R1	B1	p	45
R1	B1	k	55
R1	B1	d	53
R1	B1	npk	36
R1	B1	dnk	41
R1	B1	dnp	48
R1	B1	dpk	55
R1	B1	n	42
R1	B1	dp	50
R1	B1	nk	44
R1	B1	dk	43
R1	B1	pk	51
R1	B1	dnpk	44
R1	B1	-1	58
R1	B1	dn	41
R1	B1	np	50
R2	B2	npk	43
R2	B2	d	42
R2	B2	p	39
R2	B2	dnk	34
R2	B2	n	47
R2	B2	dnp	52
R2	B2	k	50
R2	B2	dpk	44
R2	B2	nk	43
R2	B2	dp	52

Figura 1.6: Un archivo de EXCEL mostrando variables confundidas

---

Identifica los factores y sus niveles involucrados en el experimento del artículo científico seleccionado. Justifica tu respuesta con evidencias del artículo.

**Factores**

**Niveles**

---

### 1.10. Efectos principales e interacciones

Los efectos principales y las interacciones son relevantes cuando en un experimento se quiere evaluar la influencia de dos o más factores en la variable respuesta. Los **efectos principales** para un factor representan los efectos de cada uno de sus niveles, cuando se promedia sobre los otros niveles de los otros factores. Dos factores se consideran que actúan independientemente si el efecto de aplicarlos juntos es equivalente al efecto de sumar sus efectos principales. En otro caso, se dice que los factores interactúan y existe, por tanto, un efecto **interacción** entre ellos. Las siguientes figuras muestran las situaciones que podemos encontrar en experimentos con dos factores: las medias de dos factores sin efecto interacción (Figure 1.8), con un efecto interacción grande

Block	Trt	Zadoks	Response
1	9	60	9.9
5	11	56	10.5
2	12	62	8.9
3	11	56	9.7
1	3	63	8.8
5	10	53	12.0
1	6	58	7.7
1	7	61	11.2
3	9	60	9.3
5	9	60	10.1
1	8	59	9.8
6	9	60	10.5
6	7	61	13.9
4	9	60	9.1
5	5	54	8.7
3	3	63	10.2
6	5	54	9.4
4	7	61	11.7
5	8	59	10.3
2	10	53	12.1
3	8	58	8.6
4	6	58	7.8
3	6	58	7.5
1	10	53	12.8

Figura 1.7: Un archivo de EXCEL mostrando variables de tratamientos confundidos

(Figure 1.9) y con un efecto interacción pequeño (Figure 1.10).

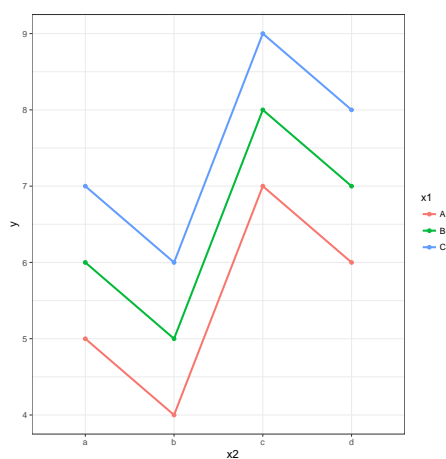


Figura 1.8: Patrón de respuesta solo con efectos principales

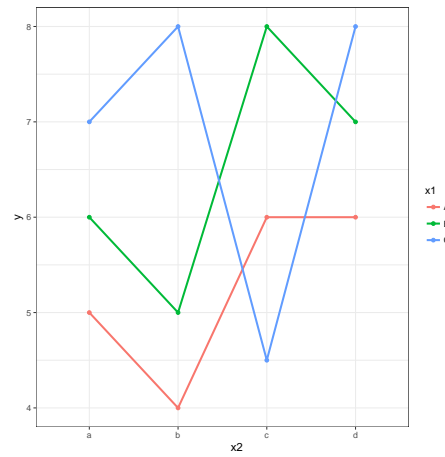


Figura 1.9: Patrón de respuesta con un efecto interacción grande

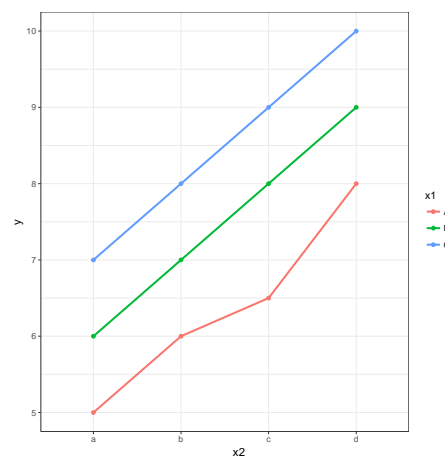


Figura 1.10: Patrón de respuesta con un efecto interacción pequeño

La siguiente gráfica muestra cuando es correcto afirmar causalidad y cuando las inferencias aplican a la población de interés. La mayor parte de los estudios agrónomos (desde nuestro punto de vista) se clasificarían dentro del cuadro de abajo a la izquierda.



10

Figura 1.11: Inferencias estadísticas permitidas por el diseño. Ramsey & Schafer. The Statistical Sleuth (pg 9)



## Capítulo 2

# Diseños

Agricolae (de Mendiburu, 2016) es el paquete de R (R Development Core Team, 2010) que utilizaremos para diseñar los experimentos.

La descripción del paquete dice:

Agricolae ofrece una amplia funcionalidad en diseños experimentales especialmente en estudios agrícolas y de mejora genética, que además pueden usarse con otros propósitos. Con este paquete se puede llevar a cabo diseños de retícula, alpha, cíclico, bloques completos aleatorizados, cuadrado latino, grecolatinos, bloques aumentados, factoriales, diseño de parcelas divididas y diseños de parcelas en banda. Incluye también varias herramientas para el análisis de datos, por ejemplo, comparaciones de tratamientos, varios test no paramétricos, índices de biodiversidad y grupos de consenso.

Para instalar este paquete en R

```
install.packages("agricolae")  
library(agricolae)
```

o en R Studio

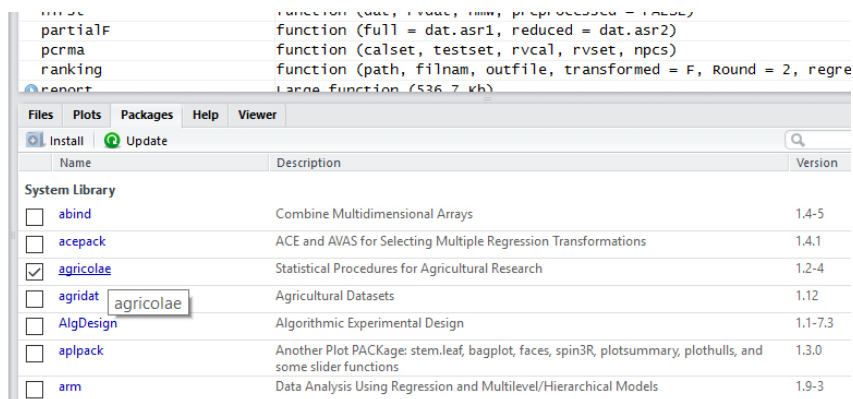


Figura 2.1: El paquete Agricolae en R Studio

### 1.11. Generando números aleatorios

En internet existen múltiples formas de generar números aleatorios, pero es más sencillo hacerlo con R. Generar un conjunto de números aleatorios en R puede llevarse a cabo con la función **sample**:

```
sample()  
sample(1:10) # Aleatoriza los números del 1 al 10  
sample(1:10,2) # Selecciona aleatoriamente dos números del 1 al 10
```

### 1.12. Diseños totalmente aleatorizados

#### 1.12.1. Características del diseño

Los diseños totalmente aleatorizados, en inglés completely randomised design (CRD), son los diseños estadísticos más simples. Los tratamientos se asignan aleatoriamente a las unidades experimentales sin ningún tipo de restricción. Cada tratamiento tiene la misma probabilidad de ser asignado a cada una de las unidades experimentales. La principal ventaja de este diseño es que es fácil de realizar y simple de analizar.

#### 1.12.2. Uso apropiado

El diseño totalmente aleatorizado es apropiado cuando las unidades experimentales son homogéneas y no tienen ningún tipo de estructura.

#### 1.12.3. Creando este tipo de diseños

##### Ejemplo 1

Un experimento tiene como objetivo investigar el efecto de distintas concentraciones de calcio en el suelo en el crecimiento de las raíces de las plantas. Se va a utilizar 20 macetas, cada una con una planta, expuestas a condiciones controladas y uniformes de temperatura, humedad, luz etc. Las macetas se dispondrán en un rectángulo de cuatro filas y cinco columnas de macetas. Los tratamientos consisten en cuatro concentraciones distintas de calcio: - 1, 5, 10 y 20. Cada tratamiento se va a aplicar de forma aleatoria a cinco macetas llevando a cabo un diseño totalmente aleatorizado.



Figura 2.2: Efecto del calcio en las raíces de las plantas

En resumen:

- **Objetivo:** evaluar el efecto de las concentraciones de calcio en el crecimiento de las raíces de las plantas
- **Observaciones:** 20 macetas, cada una contiene una planta
- **Disposición:** Dispuestas en un rectángulo de 4 filas y 5 columnas de macetas
- **Tratamientos:** concentración de calcio - 1, 5, 10 y 20.

- Réplicas: 5
- Diseño: diseño totalmente aleatorizado

Para crear el diseño en R se utiliza el siguiente código:

```
trt <- c(1, 5, 10, 20)
rep <- 5
outdesign <- design.crd(trt, r = rep)
```

#### 1.12.4. Comprobación del diseño

Siempre es necesario comprobar el diseño obtenido. La Figura 2.3 muestra un diseño totalmente aleatorizado que cumple con todos los requisitos expuestos en el Ejemplo 1.

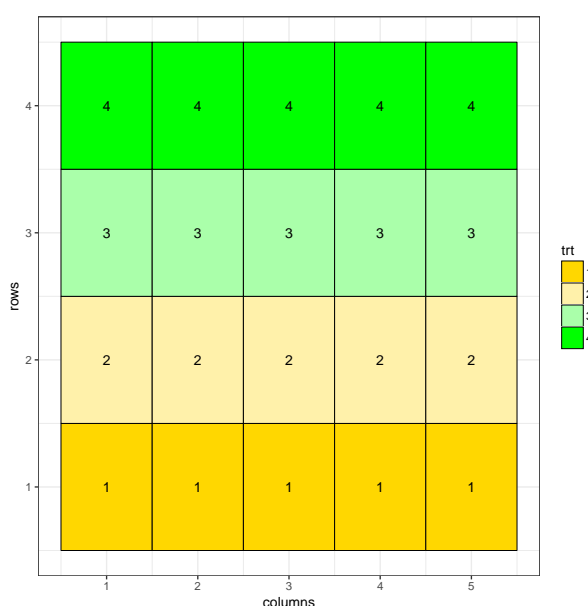


Figura 2.3: Diseño totalmente aleatorizado donde los tratamientos se confunden con las filas

No obstante, todos los tratamientos se confunden con las filas por lo que el objetivo del experimento, investigar los efectos de las distintas concentraciones de calcio en el crecimiento de las raíces de las plantas, no podría ser dilucidado, ya que no se puede dissociar los efectos de las filas de los efectos de los tratamientos. Posiblemente, nunca se obtendría un diseño tan peculiar como este, pero es posible que todas las réplicas de un tratamiento aparezcan en un extremo o en las esquinas haciendo que tal diseño fuera inapropiado. Con el ejemplo de la Figura 2.3 se ha querido recalcar la importancia de comprobar los diseños obtenidos.

Se puede utilizar el siguiente código para crear una gráfica como la Figura 2.4 y así poder visualizar el diseño obtenido y para comprobar el esqueleto de la tabla ANOVA para evaluar los grados de libertad del error.

```
des.out<-des.info(design.obj=outdesign, nrow = 4, ncol = 5)
```

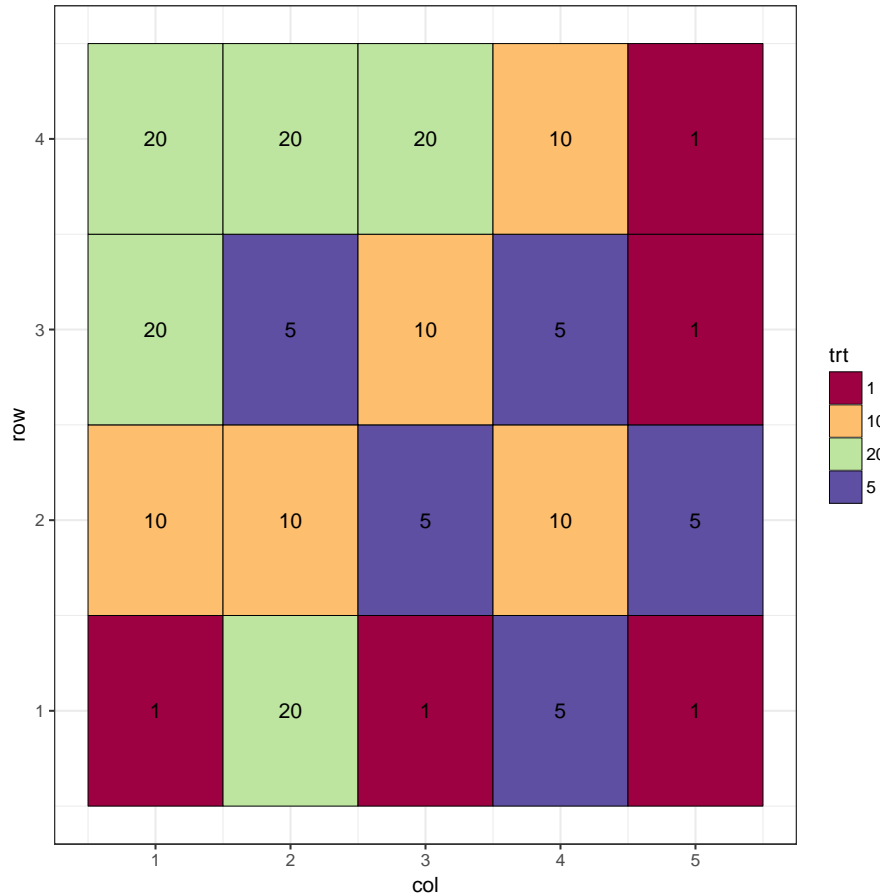


Figura 2.4: Diseño totalmente aleatorizado para el experimento del calcio

A continuación, se muestra la tabla producida:

Fuente de variación	gl
trt	3
Residuo	16
Total	19

El esqueleto de la tabla ANOVA es como la tabla que se obtendría al analizar los datos usando los métodos del análisis de varianza, pero sin datos, solo aparece información relacionada con los grados de libertad. Los grados de libertad se obtienen de la siguiente forma:

$$df_{trt} = \text{número de niveles de tratamientos} - 1 = 4 - 1 = 3$$

$$df_{Total} = \text{número de observaciones} - 1 = 20 - 1 = 19$$

$$df_{Residuos} = df_{Total} - df_{tratamientos}$$

En general, para asegurar una estimación razonable de la variabilidad de los residuos, el número de réplicas ha de ser suficientemente grande para asegurar al menos 15 grados de libertad asociados a los residuos. Hay poco beneficio en tener más de 20 grados de libertad para los residuos ya que a partir de esa cifra la ganancia en precisión aumenta muy lentamente conforme aumentan los grados de libertad.

### 1.12.5. archivo EXCEL

El diseño se escribe en un archivo EXCEL de manera que las medidas de las variables que se tomen durante el experimento se añadan a dicho archivo quedando una base de datos lista para ser analizada.

En R se utiliza el siguiente código:

```
setwd("la ruta donde se guardará el archivo que contiene el diseño")
write.csv2(des.out, "nombre del archivo.csv", row.names = FALSE)
```

o en R Studio

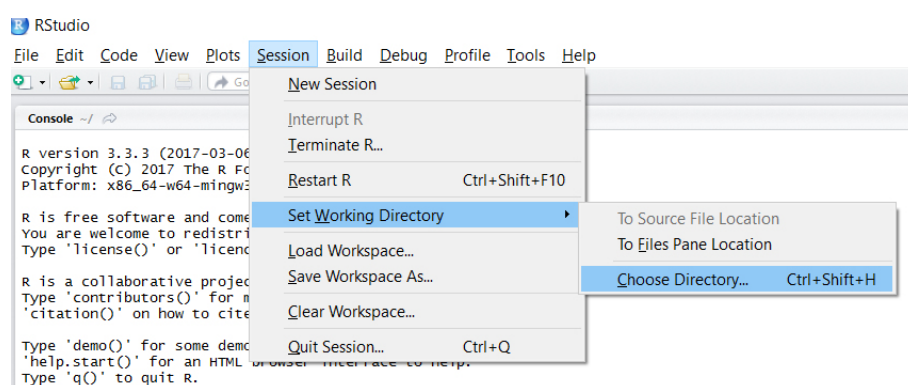


Figura 2.5: Establecer el directorio de trabajo en R Studio

```
write.csv2(des.out$design, "nombre del archivo.csv", row.names = FALSE)
```

**Ejercicio 1** Un técnico tiene que medir la acidez de cuatro suelos. Se le proporcionan tres muestras de cada suelo. El técnico propone procesar primero todas las muestras del suelo A, luego todas las muestras del suelo B y así sucesivamente. Justifica por qué la idea del técnico no es buena y **crea** un diseño para dicho experimento. Asume que todas las muestras pueden procesarse en el mismo día. Pista: crea un diseño totalmente aleatorizado con una columna.

### Ejercicio 2

Se requiere un diseño totalmente aleatorizado para evaluar la efectividad de siete tratamientos para paliar una enfermedad de la piel de las patatas. El experimento se lleva a cabo en el campo y las patatas se plantan en parcelas dispuestas en un terreno rectangular de 6 filas y 14 columnas de parcelas. Crea un diseño para este experimento.

### 1.13. Diseño en bloques completos aleatorizados

#### 1.13.1. Características del diseño

Diseño en bloques completos aleatorizados, en inglés randomised complete block design (RCBD), es el diseño más simple que incluye bloques y es probablemente el más usado. En este diseño el número de unidades experimentales en cada bloque debe ser igual al número de tratamientos. Dentro de cada bloque, los tratamientos son aleatorizados a las unidades experimentales que comprenden el bloque. Se lleva a cabo una aleatorización separada (y por tanto distinta) para cada bloque.

El diseño en bloques completos aleatorizados es útil porque permite tener en cuenta la heterogeneidad entre las unidades experimentales, siendo todavía un diseño bastante sencillo de llevar a cabo. El análisis es simple porque cada tratamiento ocurre una vez en cada bloque (los tratamientos son ortogonales a los bloques) y balanceado (todas las comparaciones entre los tratamientos tienen la misma precisión).

#### 1.13.2. Uso apropiado

Un diseño en bloques completos aleatorizados se usa solo cuando el número de tratamientos es igual al tamaño del bloque.

#### 1.13.3. Creando el diseño

##### Ejemplo 2

Un experimento tiene como objetivo evaluar las diferencias en la producción de 11 variedades distintas de trigo:

- |            |           |              |
|------------|-----------|--------------|
| ■ Eagle    | ■ Oxley   | ■ Songlen    |
| ■ Gamenya  | ■ WW-11B  | ■ Timson     |
| ■ Heron    | ■ RAC-88  | ■ 64-WO-2-29 |
| ■ Pinnacle | ■ RAC-255 |              |

Cada una de estas variedades fueron replicadas cuatro veces. Las unidades experimentales son parcelas dispuestas en un terreno rectangular de 11 filas y 4 columnas de parcelas. Con el objetivo de crear un diseño en bloques completamente aleatorizado, cada bloque debe de tener 11 filas por 1 columna.

En resumen:

- **Objetivo:** investigar las diferencias en la producción de 11 variedades de trigo
- **Observaciones:** 44 parcelas
- **Disposición:** terreno rectangular de 11 filas y 4 columnas de parcelas
- **Tratamientos:** 11 variedades
- **Réplicas:** 4
- **Diseño:** diseño en bloques completos aleatorizados

Para crear el diseño en R utilizamos el siguiente código:

```
trt <- c("Eagle", "Gamenya", "Heron", "Pinnacle", "Oxley", "WW-11B",
"RAC-88", "RAC-255", "Songlen", "Timson", "64-WO-2-29")
rep <- 4
outdesign <- design.rcbd(trt, r=rep)
```

### 1.13.4. Comprobación del diseño

Se puede utilizar el siguiente código para crear una gráfica como la Figura 2.4 y así poder visualizar el diseño obtenido así como para generar el código para el esqueleto de la tabla ANOVA.

```
des.out <- des.info(design.obj=outdesign, nrows=11, ncols=4, brows=11, bcols=1)
```

A continuación, se muestra la tabla producida:

Fuente de variación	gl
=====	
Estrato del bloque	3
-----	
trt	10
Residuo	30
=====	
Total	43

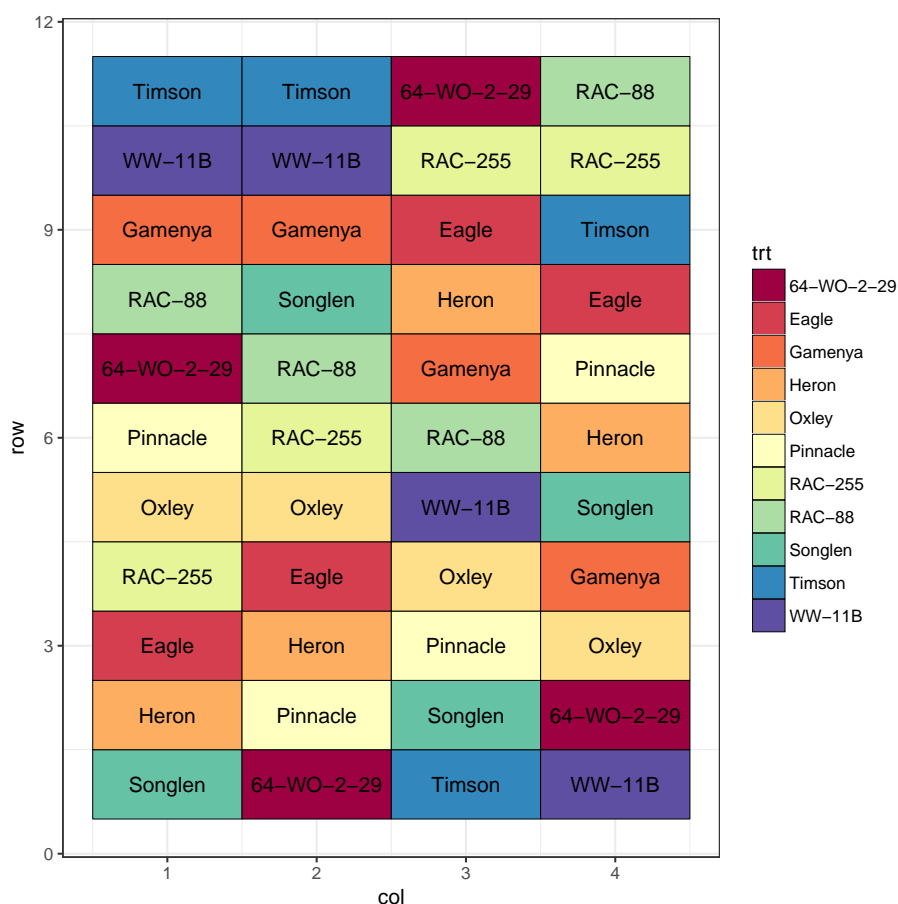


Figura 2.6: Diseño en bloques completos aleatorizados para el experimento de trigo

### 1.13.5. archivo EXCEL

El diseño se escribe en un archivo EXCEL de manera que las medidas de las variables que se tomen durante el experimento se añadan a dicho archivo quedando una base de datos lista para ser analizada.

```
write.csv2(des.out$design, "nombre del archivo.csv", row.names = FALSE)
```

**Ejercicio 3** Un experimento tiene como objetivo comparar los distintos efectos de cuatro sistemas de cultivo en la cantidad de biomasa producida por plantas de col. El ensayo se va a llevar a cabo en un terreno rectangular de 4 columnas y 5 filas de parcelas. Se sabe que existe una tendencia de fertilidad en las filas del terreno así que se propone realizar un diseño en bloques completos aleatorizados.

1. ¿Por qué es beneficioso usar un diseño en bloques completos aleatorizados?
2. Diseñar el experimento y comprobar el esqueleto de la tabla ANOVA.

**Ejercicio 4** Considera un experimento de campo para evaluar la respuesta de un cultivo a cinco ratios de fertilizante distintos (0, 50, 100, 150 y 200 kg/ha de N). Crea un diseño en bloques completos aleatorizados con 8 réplicas y guárdalo en un archivo EXCEL para enviárselo al investigador. El experimento se lleva a cabo en un terreno rectangular de 4 columnas por 10 filas de parcelas.

## 1.14. Cuadrado latino

### 1.14.1. Características del diseño

El cuadrado latino es un diseño donde el número de filas y de columnas es igual al número de tratamientos, y por tanto al número de réplicas. Es por ello que si se tienen cuatro tratamientos se necesitan cuatro filas y cuatro columnas para crear un diseño en cuadrado latino. El resultado es un diseño donde cada tratamiento aparece una vez en cada fila y en cada columna. Este diseño se originó para disposiciones con formas cuadradas, es por ello que los factores estructurales se suelen llamar factor **fila** y factor **columna**, pero no necesariamente son filas y columnas físicas ya que los factores estructurales podrían ser otro tipo de factores, por ejemplo, el **observador** y el **tiempo**.

Algunos libros de diseño de experimentos contienen diseños en cuadrados latinos estándares, sin embargo, es más fácil generar este tipo de diseños con R.

### 1.14.2. Uso apropiado

A veces queremos incluir dos factores de bloques en el diseño para reducir más la variabilidad del error en nuestra variable respuesta. El número de tratamientos debe ser igual al número de filas y columnas. La asignación de tratamientos se realiza de forma que cada tratamiento aparece una vez en cada fila y en cada columna.

La principal desventaja de un diseño en cuadrado latino es la restricción de que el número de filas, columnas y tratamientos han de ser igual.

### 1.14.3. Creando el diseño

#### Ejemplo 3

Un experimento tiene como objetivo investigar el efecto de los tipos de suelos en el crecimiento de los altramuces. El experimento se lleva a cabo en macetas dispuestas en una bancada del invernadero que tiene una tendencia sistemática a lo largo de la bancada (de izquierda a derecha) como resultado de un gradiente de temperatura y a lo ancho de la bancada (de arriba a abajo) debido a una diferencia en la iluminación. Las macetas se colocan unas al lado de otras formando un cuadrado donde las filas y las columnas se consideran bloques. Un diseño en cuadrado latino se considera apropiado para este experimento. Cuatro tratamientos (S1, S2, S3 y S4) fueron utilizados representando 4 tipos de suelo distintos.

En resumen:

- **Objetivo:** investigar los efectos de los distintos suelos en el crecimiento de los altramuces
- **Observaciones:** 16 macetas
- **Disposición:** 4 filas × 4 columnas





Figura 2.7: Altramuces

- **Tratamientos:** 4 tipos distintos de suelo
- **Réplicas:** 4
- **Diseño:** diseño en cuadrado latino
- **Disposición de los bloques:** 1 fila  $\times$  4 columnas y 4 filas  $\times$  1 columna

Para crear el diseño en R utilizamos el siguiente código:

```
trt <- c("S1", "S2", "S3", "S4")
outdesign <- design.lsd(trt)
```

#### 1.14.4. Comprobando el diseño

Se puede utilizar el siguiente código para crear una gráfica como la Figura 2.8 y así visualizar el diseño obtenido. Es importante asegurar que hemos incluido suficientes colores para todos los tratamientos. Este código también sirve para generar el esqueleto de la tabla ANOVA.

```
des.out<-des.info(design.obj=outdesign, nrows=4, ncols=4)
```

A continuación, se muestra la tabla producida:

Fuente de variación	gl
=====	
Fila	3
Columna	3
trt	3
Resido	6
=====	
Total	15

#### 1.14.5. Archivo EXCEL

El diseño se guarda en un archivo EXCEL de manera que las medidas de las variables que se tomen durante el experimento se añadan a dicho archivo quedando una base de datos lista para ser analizada.

```
write.csv2(des.out$design, "design file name.csv", row.names = FALSE)
```

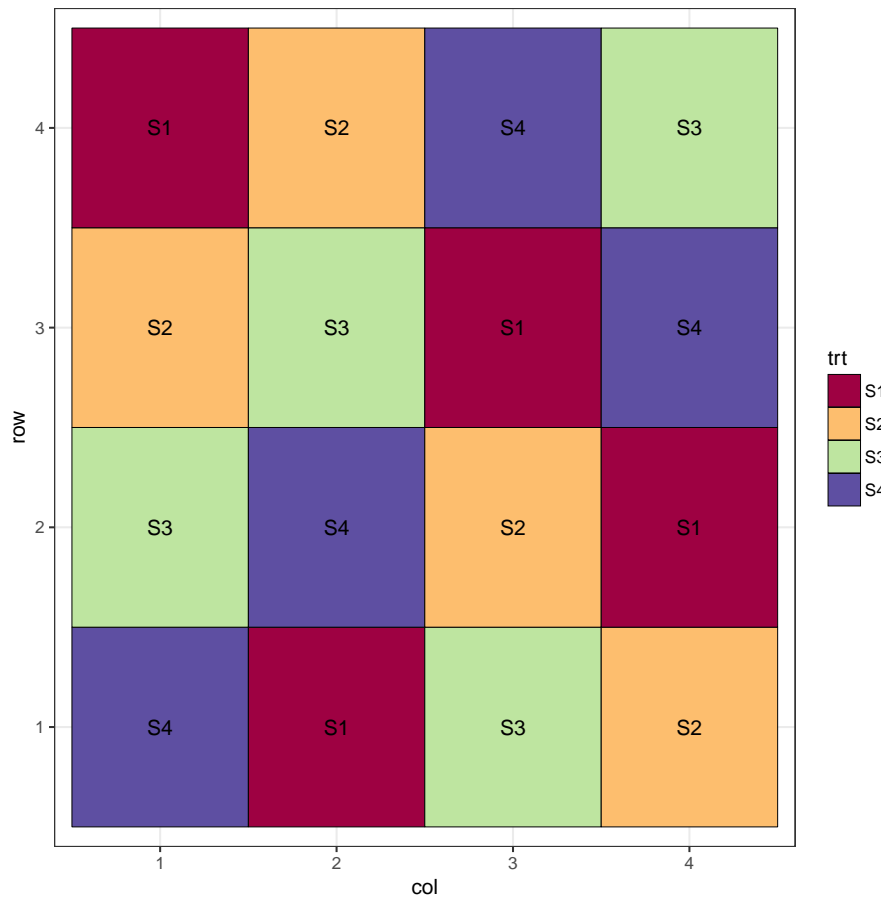


Figura 2.8: Diseño en cuadrado latino para el experimento de los altramuces

### Ejercicio 5

El gusano del alambre es una plaga muy dañina en el cultivo de la patata. Esta plaga solo puede controlarse mediante químicos pero dichos químicos son a largo plazo un peligro para el medio ambiente. Se quiere comprobar el efecto de unos tratamientos alternativos a los químicos. Se propone llevar a cabo un experimento con los siguientes cuatro tratamientos:

1. un control (no se aplica tratamiento);
2. el químico estándar;
3. un método alternativo para elevar el suelo alrededor de la planta (alternativa 1)
4. otro método alternativo que también consiste en elevar el suelo alrededor de la planta (alternativa 2)

para evaluar los efectos de estos tratamientos en la proporción de plantas seriamente afectadas por el gusano del alambre en la temporada de la cosecha de la patata. El experimento se lleva a cabo en un terreno rectangular de 16 parcelas, dispuestas en 4 filas  $\times$  4 columnas de parcelas. Cada parcela contiene 5 filas  $\times$  10 plantas, y la variable respuesta es la proporción de plantas seriamente afectadas por el gusano del alambre. Existe una pendiente de un lado al otro del terreno. Existe también un gradiente en la textura del suelo que es perpendicular a la pendiente del terreno. Se asume que tanto la pendiente como la textura del suelo podría influenciar la variable respuesta. Crea un diseño en cuadrado latino para este experimento.

## Ejercicio 6

Se lleva a cabo un experimento para investigar el efecto del color de los pétalos de rosa en la afluencia de insectos. Se utilizan cinco tonos distintos de colores de pétalos y se utiliza un diseño en cuadrado latino que va a tener en cuenta la dirección migratoria (desconocida) de los insectos. Crea un diseño en cuadrado latino para este experimento.

### 1.15. Estructura de los tratamientos

En los diseños que se han descrito hasta ahora, se han considerado estructuras de tratamiento muy simples. Los tratamientos no tienen necesariamente una estructura sencilla, a veces su estructura puede ser bastante compleja. Esta complejidad tiene lugar cuando dos o más factores dan lugar a los tratamientos. Por ejemplo, considera un experimento donde se tienen distintas variedades de canola, así como distintos niveles de irrigación. Es la combinación de los niveles de irrigación con las distintas variedades de canola lo que define los distintos tratamientos.

Experimentos con la misma estructura de tratamientos pueden llevarse a cabo utilizando distintos tipos de diseños experimentales, por ejemplo, completamente aleatorizados, en bloques completos aleatorizados, en cuadrado latino etc. Es la estructura de los tratamientos junto a la aleatorización de los mismos, fruto de un diseño experimental, lo que verdaderamente describe un experimento.

Los diseños con más de un factor pueden tener sus factores cruzados o anidados. Veamos que significan estos dos conceptos.

#### 1.15.1. Estructura de tratamiento con factores cruzados

En el experimento con distintas variedades de canola y distintos niveles de irrigación, las variedades de canola son las siguientes:

1. Victory
2. Hurricane
3. Bravo
4. Tarcoola

Existen dos niveles de irrigación, seco o regadío. El factor variedad tiene pues cuatro niveles y el factor irrigación dos. En este experimento cada variedad se cultiva bajo los dos niveles de irrigación. Cuando todos los niveles de un factor aparecen con todos los niveles del otro factor se dice que los factores están **cruzados** y que existe una estructura factorial de los tratamientos. En el experimento que se ha planteado, los factores variedad e irrigación están cruzados, ver la Tabla 2.1 :

Variedad	Irrigación	
	Seco	regadío
Victory	✓	✓
Hurricane	✓	✓
Bravo	✓	✓
Tarcoola	✓	✓

Cuadro 2.1: Factores cruzados

Una estructura de tratamiento con factores cruzados tiene todas las posibles combinaciones de los niveles de los factores presente en dicha estructura. Si una o más celdas de la tabla de arriba estuvieran vacías la estructura de los tratamientos no sería factorial.

### 1.15.2. Estructura de tratamiento con factores anidados

Otro tipo de estructura de tratamientos que se utiliza a menudo es aquella con factores anidados. Este tipo de estructura tiene lugar cuando los niveles de un factor se encuentran anidados con los niveles de otro factor o en otras palabras cada nivel de uno de los factores aparece asociado a un único nivel del otro factor. Considera un estudio de nutrición en el ganado. Se administran distintas dietas basadas en distintas variedades de trigo y avena. El tipo de grano es un factor con dos niveles: trigo y avena. El otro factor es el factor variedad cuyos niveles podrían ser:

- Janz
- Currawong
- Lang
- Blackbutt
- Eurabbie

Las variedades no pueden ser trigo y avena a la vez, es por ello que las variedades se encuentran anidadas con el tipo de grano, ver Tabla 2.2. En esencia, la estructura de los tratamientos con factores anidados puede tratarse como si fuera una estructura de tratamientos simples.

Variedad	Especies	
	Trigo	Avena
Janz	✓	
Currawong	✓	
Lang	✓	
Blackbutt		✓
Eurabbie		✓

Cuadro 2.2: Estructura de tratamientos con factores anidados

### 1.15.3. Creando el diseño

**Ejemplo 4** Un experimento tiene como objetivo investigar el efecto de la variedad y la irrigación en la cantidad de grano producido en plantas de canola. Las tres variedades utilizadas son:

1. Victory
2. Hurricane
3. Bravo

Existen dos tratamientos de riego, ausencia o presencia de riego. El factor variedad tiene tres niveles y el factor irrigación tiene dos. Esta estructura de tratamientos se denomina estructura factorial 3 x 2. En este experimento cada variedad se cultiva bajo los dos regímenes de riego y por tanto los factores variedad e irrigación están cruzados.

Los tratamientos serán replicados tres veces y se utilizará un terreno rectangular de seis filas por tres columnas de parcelas utilizando un diseño completamente aleatorizado con tres réplicas por tratamiento.

En resumen:

- **Objetivo:** investigar el efecto de la variedad y la irrigación en la producción de grano en plantas de canola

- **Observaciones:** 18 parcelas
- **Disposición:** 6 filas  $\times$  3 columnas
- **Tratamientos:** 2 factores - variedad (3) e irrigación (2)
- **Réplicas:** 3
- **Diseño:** estructura de tratamientos factorial, en un diseño completamente aleatorizado
- **Disposición de los bloques:** NA

Para crear el diseño en R utilizamos el siguiente código:

```
trt <- c(3, 2) # factorial 3 x 2
rep <- 3
outdesign <- design.ab(trt, r = rep, design = "crd")
```

#### 1.15.4. Comprobación del diseño

Se puede utilizar el siguiente código para crear una gráfica como la Figura 2.9 y así poder visualizar el diseño obtenido. Este código también genera el esqueleto de la tabla ANOVA.

```
des.out<-des.info(design.obj=outdesign, nrows=6, ncols=3)
```

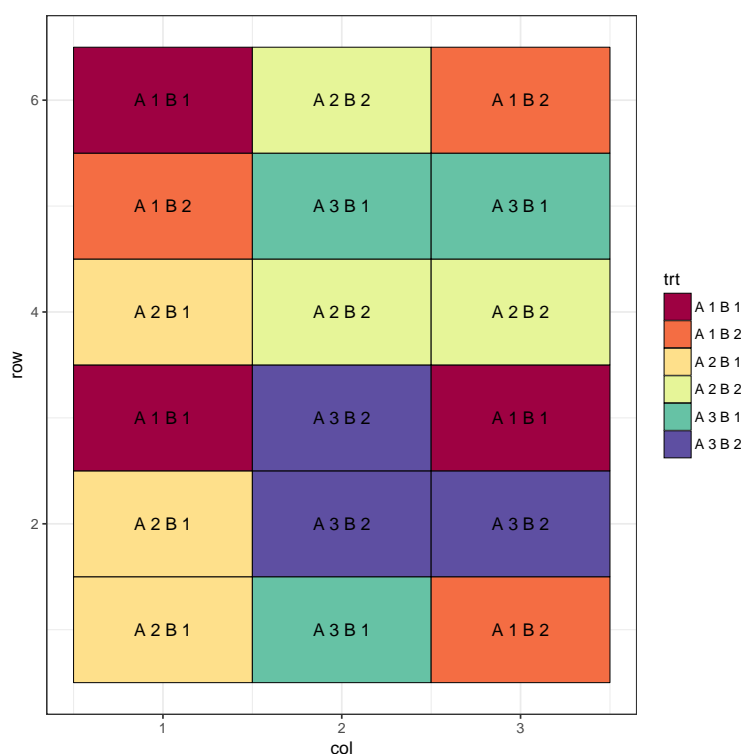


Figura 2.9: Diseño factorial completamente aleatorizado para el experimento de irrigación en canola

A continuación, se muestra la tabla producida:

Fuente de variación	gl
=====	=====
A	2
B	1
AB	2
Residuo	12
=====	=====
Total	17

#### 1.15.5. Archivo EXCEL

El diseño se guarda en un archivo EXCEL de manera que las medidas de las variables que se tomen durante el experimento se añadan a dicho archivo quedando una base de datos lista para ser analizada.

```
write.csv2(des.out$design, "design file name.csv", row.names = FALSE)
```

#### Ejemplo 5

*Orobanche aegyptiaca* es una especie de planta que parasita las raíces de otras plantas. Esta planta parasitaria no tiene clorofila y crece bajo la raíz de la planta hospedadora. Sus semillas permanecen latentes en el suelo hasta que ciertos componentes de las otras plantas estimulan su germinación. Se quiere estudiar dos genotipos de estas plantas parasitarias, *O. aegyptiaca* 73 y *O. aegyptiaca* 75. Sus semillas fueron embadurnadas con dos extractos preparados de una planta de judías y una planta de pepino. El diseño experimental es un diseño factorial 2x2, donde cada tratamiento será replicado 6 veces. El objetivo del experimento es investigar el efecto del genotipo y el extracto de la planta en la propagación de esta especie parasitaria.



Figura 2.10: *Orobanche aegyptiaca*

En el experimento se utilizaron 24 macetas, cada una de ellas conteniendo una planta. Las macetas fueron dispuestas en un rectángulo de cuatro filas por seis columnas de macetas. Las macetas se colocaron en una bancada del invernadero que tiene una tendencia sistemática de un lado al otro de la bancada como consecuencia de un gradiente de temperatura. Es por ello que se utilizó un diseño en bloques completos aleatorizados.

En resumen:

- **Objetivo:** investigar el efecto del genotipo y el extracto de planta en la propagación de la planta parasitaria.
- **Observaciones:** 24 macetas
- **Disposición:** 4 filas × 6 columnas
- **Tratamientos:** 2 factores – genotipo (2) y extracto de planta (2)
- **Réplicas:** 6
- **Diseño:** estructura de tratamiento factorial; diseño en bloques completos aleatorizados
- **Disposición de los bloques:** bloques igual a columnas

Para crear el diseño en R utilizamos el siguiente código:

```
trt <- c(2, 2) # factorial 2 x 2
rep <- 6
outdesign <- design.ab(trt, r = rep, design = "rcbd")
```

#### 1.15.6. Comprobación del diseño

Se puede utilizar el siguiente código para crear una gráfica como la Figura 2.11, y así poder visualizar el diseño obtenido. Este código también sirve para generar el esqueleto de la tabla ANOVA.

```
des.out<-des.info(design.obj=outdesign, nrows=4, ncols=6, brows=4, bcols=1)
```

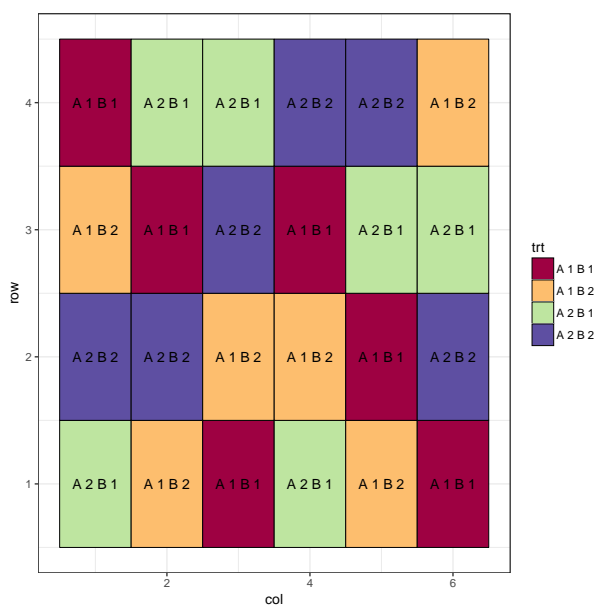


Figura 2.11: Diseño factorial en bloques completos aleatorizados para el experimento de la propagación de *Orobanche aegyptiaca*

A continuación, se muestra la tabla producida:

Fuente de variación	gl
=====	=====
Estrato del bloque	5
-----	-----
A	1
B	1
AB	1
Residuo	15
=====	=====
Total	23

#### 1.15.7. Archivo EXCEL

El diseño se guarda en un archivo EXCEL de manera que las medidas de las variables que se tomen durante el experimento se añadan a dicho archivo quedando una base de datos lista para ser analizada.

```
write.csv2(des.out$design, "design file name.csv", row.names = FALSE)
```

#### Ejemplo 6

Se llevó a cabo un experimento con el objetivo de evaluar el efecto patógeno de una serie de diferentes cepas de hongos sobre las plántulas de canola. Las cepas se tomaron de distintas especies de *Brassica* etiquetadas como A y B, con cuatro cepas distintas de la especie A y tres de la especie B. Cada tratamiento se replicó 7 veces utilizando un diseño en cuadrado latino, siendo el **tiempo** y la **posición** en la incubadora las dos estructuras de bloques para el diseño.

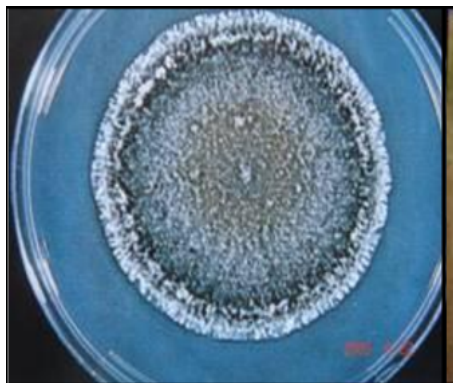


Figura 2.12: Cepa de hongo

En este ejemplo los factores especie de hongo y cepas se encuentran anidados. Las cepas no pueden pertenecer a ambas especies de hongos. Es por ello por lo que existen siete tratamientos y un diseño en cuadrado latino 7x7 es apropiado.

En resumen:

- **Objetivo:** evaluar el efecto patógeno de una serie de cepas de hongos sobre las plántulas de canola
- **Observaciones:** 49 placas de petri
- **Disposición:** 7 filas × 7 columnas
- **Tratamientos:** anidados – las cepas en las especies (7)



- Réplicas: 7
- Diseño: cuadrado latino
- Disposición de los bloques: 1 fila  $\times$  7 columnas y 7 filas  $\times$  1 columna

Para crear el diseño en R utilizamos el siguiente código:

```
trt <- c("A1", "A2", "A3", "A4", "B1", "B2", "B3")
outdesign <- design.lsd(trt)
```

#### 1.15.8. Comprobación del diseño

Para poder visualizar el diseño obtenido, se puede utilizar el siguiente código para crear una gráfica como la que aparece en la Figura 2.13. Este código también sirve para generar el esqueleto de la tabla ANOVA.

```
des.out<-des.info(design.obj=outdesign, nrow=7, ncol=7)
```

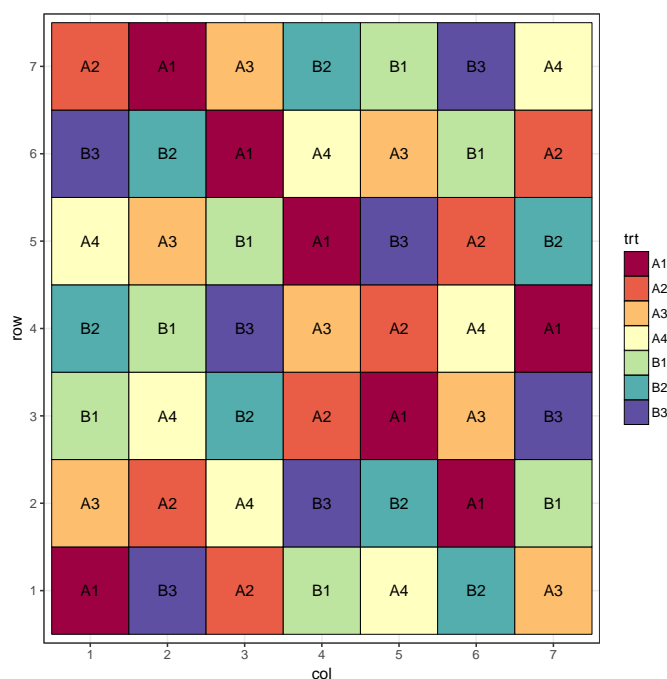


Figura 2.13: Diseño en cuadrado latino para el experimento de las cepas de hongos

A continuación, se muestra la tabla producida:

Fuente de variación	gl
=====	
Fila	6
Columna	6
trt	6
Residuo	30
=====	
Total	48

### 1.15.9. Archivo EXCEL

El diseño se escribe en un archivo EXCEL de manera que las medidas de las variables que se tomen durante el experimento se añadan a dicho archivo quedando una base de datos lista para ser analizada.

```
write.csv2(des.out$design, "design file name.csv", row.names = FALSE)
```

### Ejercicio 7

La mayoría de las variedades comerciales de sorgo forrajero tienen talle alto. El uso de talle bajo es limitado, principalmente por su baja producción de forraje. Recientemente, una nueva variedad de sorgo forrajero de talle bajo denominada Tal ha sido desarrollada en Israel. El estudio que se describe a continuación tiene como objetivo examinar los niveles de irrigación y la densidad de las plantas en la producción de forraje de la variedad Tal. El factor irrigación incluye los niveles de 20, 100 and 180 mm y el factor densidad incluye niveles de 200, 260 y 330 K plantas/ha. Crea un diseño completamente aleatorizado para este experimento asumiendo que los niveles de irrigación y la densidad de plantas pueden ser aleatorizados a parcelas individuales. Todos los niveles del factor irrigación aparecen con todos los niveles del factor densidad. Las parcelas se distribuyen en un terreno rectangular de 6 filas por 6 columnas. Crea el diseño y guárdalo en un archivo Excel.

### Ejercicio 8

Para el ejercicio 7, el técnico piensa que puede haber una tendencia en el terreno – a lo largo de las filas. Se requiere crear un nuevo diseño en bloques completos aleatorizados para este experimento. Nota que el nuevo diseño se llevaría a cabo en un terreno rectangular de 4 filas por 9 columnas de parcelas. Genera el nuevo diseño.

### Ejercicio 9

Las plantas del algodón son mayormente auto fértiles y auto polinizadoras y algunas de ellas necesitan un agente externo como los insectos para llevar a cabo su polinización. Cada flor del algodón puede ser polinizada solo el día que se abre la flor; la corola y la columna estaminal se caen al segundo día. Se propone llevar a cabo un experimento para evaluar el efecto de cuatro variedades distintas de algodón y dos niveles de irrigación en el nivel de actividad de las abejas. Se cree conveniente utilizar un diseño en cuadrado latino ya que la dirección de las abejas podría variar dependiendo de la dirección del viento. Crea un diseño para este experimento.

### Ejercicio 10

Se lleva a cabo un experimento con el objetivo de evaluar el impacto de distintas estrategias de control de la maleza junto con distintos cultivos para controlar el afloramiento de la Phalaris que es considerada una mala hierba. Se incluirán tres prácticas de control de la maleza que se utilizarán en dos variedades de trigo, dos de cebada y una de avena. El experimento incluye cuatro réplicas y se dispone en un diseño en bloques completos aleatorizados. Sugiere distintos tamaños de parcelas para dicho experimento y crea un diseño para una de las posibilidades sugeridas.

## 1.16. Diseño de parcelas divididas

### 1.16.1. Características del diseño

En algunos experimentos, los niveles de los factores pueden requerir unidades experimentales de distinto tamaño. Por ejemplo, cierto tipo de maquinaria o tratamientos de irrigación pueden necesitar áreas de gran tamaño mientras que las variedades pueden ser cultivadas en áreas más pequeñas. Una solución para este tipo de experimentos es usar un diseño de parcelas divididas lo que en inglés se conoce como split-plot. Este tipo de diseños es apropiado para experimentos factoriales donde los niveles de un factor requieren unidades experimentales más grandes mientras que los niveles del otro factor pueden ser aplicados a unidades experimentales más pequeñas.

Un diseño de parcelas divididas suele tener bloques, que contienen un conjunto completo de tratamientos. Cada bloque se divide en parcelas grandes (en inglés *whole plots* o *main plot*), con los niveles de un factor A aleatorizados a estas parcelas grandes con una aleatorización distinta dentro de cada bloque. Finalmente, cada parcela grande se divide en un número de parcelas pequeñas (en inglés *subplot*) y los niveles de un factor B se aleatorizan a estas parcelas pequeñas con una aleatorización distinta dentro de cada parcela grande.

### 1.16.2. Uso apropiado

Un diseño de parcelas divididas debería utilizarse solo cuando existen verdaderas restricciones en el tamaño de las unidades experimentales para aplicar ciertos tratamientos, ya que el diseño es menos eficiente que el correspondiente a un diseño en bloques completos aleatorizados basados en el mismo número de unidades experimentales. A veces, los diseños en parcelas divididas se pueden presentar para poder añadir un nuevo factor a un experimento ya existente.

### 1.16.3. Creando el diseño

#### Ejemplo 7

Se lleva a cabo un experimento para estudiar el efecto competitivo de tres especies de malezas junto con un control en un campo de trigo bajo dos regímenes de estrés hídrico. La variación en el estrés hídrico se define con la presencia o ausencia de irrigación. Los dos niveles de irrigación solo pueden aplicarse a grandes áreas de terreno mientras que las especies de maleza pueden aplicarse a parcelas más pequeñas. Un diseño de parcelas divididas con cuatro réplicas se considera apropiado para esta situación. Existen dos tratamientos de irrigación (factor **irrigación**, con dos niveles) aplicados a parcelas grandes (factor **WholePlots**, con dos niveles dentro de cada bloque). Cada parcela grande se divide en cuatro subparcelas (factor **Subplot**, con cuatro niveles). Las especies de maleza (factor **maleza**, con niveles C (control), Am, Ga and Sm) se aleatorizan a las subparcelas dentro de cada parcela grande. Las subparcelas se disponen en un terreno rectangular de 8 filas y 4 columnas de subparcelas.



Figura 2.14: Trigo

En resumen:

- **Objetivo:** estudiar el efecto competitivo de tres especies de maleza distinta en un campo de trigo bajo dos niveles de estrés hídrico.
- **Observaciones:** 32 parcelas
- **Disposición:** 8 filas  $\times$  4 columnas
- **Tratamientos:** factorial – especies de maleza (4) y estrés hídrico (2)
- **Réplicas:** 4
- **Diseño:** diseño de parcelas divididas
- **Disposición de los bloques:** 8 bloques en 8 filas  $\times$  1 columna con dos parcelas grandes dentro de cada bloque y 4 subparcelas dentro de cada parcela grande.

Para crear el diseño en R utilizamos el siguiente código:

```
irrigation <- c("R","I")
weed <- c("C","Am","Ga","Sm")
outdesign <- design.split(irrigation, weed, r = 4)
```

#### 1.16.4. Comprobación del diseño

Para poder visualizar el diseño obtenido, se puede utilizar el siguiente código para crear una gráfica como la de la Figura 2.15. Este código también sirve para generar el esqueleto de la tabla ANOVA.

```
des.out<-des.info(design.obj=outdesign, nrows=8, ncols=4, brows=8, bcols=1)
```

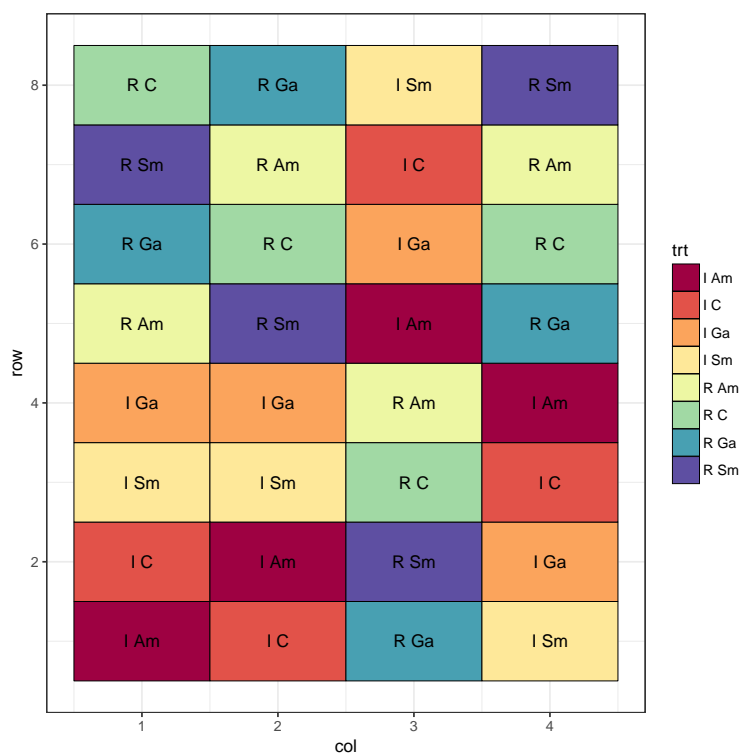


Figura 2.15: Diseño de parcelas divididas para el experimento de la maleza bajo distintos niveles de estrés hídrico

A continuación, se muestra la tabla producida:

Fuente de variación	gl
=====	
Estrato bloque	3
-----	
Estrato parcela grande	
irrigación	1
Residuo parcela grande	3
=====	
Estrato parcela pequeña	
maleza	3
irrigación:maleza	3
Residuo parcela pequeña	18
=====	
Total	31

### 1.16.5. Archivo EXCEL

El diseño se guarda en un archivo EXCEL de manera que las medidas de las variables que se tomen durante el experimento se añadan a dicho archivo quedando una base de datos lista para ser analizada.

```
write.csv2(des.out$design, "nombre_archivo.csv", row.names = FALSE)
```

### Ejercicio 11

La mejora de la calidad de la malta es importante si se quiere que la cebada de sureste de Australia siga siendo competitiva en los mercados de exportación. Se busca que el grano produzca altos niveles de extracto de malta y poder diastático y que tenga niveles moderados de proteína. Para examinar el efecto de los cultivos y el ambiente y, especialmente de la fertilización con nitrógeno, en los parámetros indicativos de la calidad de la malta y sus correlaciones, se lleva a cabo el siguiente experimento. Se consideran siete variedades distintas de cebada que se van a cultivar bajo cinco niveles de fertilización de nitrógeno. Los niveles de fertilizante se van a aleatorizar a parcelas grandes (whole plots) y las variedades de cebada a parcelas pequeñas (subplots). Crea el diseño para este experimento con tres réplicas. Las parcelas se disponen en un terreno rectangular de 7 columnas por 15 filas de parcelas.

### Ejercicio 12

Cuando la mayor parte del fósforo (P) se localiza cerca de la superficie del suelo, la deshidratación de la superficie del suelo que se produce en primavera en las áreas de clima mediterráneo puede reducir la disponibilidad del P para las plantas y causar reducciones en su productividad. Para aquellas especies que requieren un suministro extra de fósforo durante la primavera, la inyección de fertilizante de P en profundidad ha demostrado ser un método efectivo para mejorar la disponibilidad del P e incrementar la productividad de las plantas. Se lleva a cabo un experimento para evaluar el efecto de la disposición del P, así como la irrigación en la productividad de las plantas. Se utiliza tres disposiciones de P distintas y tres regímenes de irrigación distintos después de la floración, siendo cada tratamiento replicado tres veces. El experimento se va a llevar a cabo en un invernadero con un diseño de parcelas divididas donde los regímenes de agua se aleatorizan a unidades experimentales más grandes y la disposición del fertilizante a unidades experimentales más pequeñas. Las macetas se disponen en bancadas de 3 filas por 9 columnas de macetas. Crea un diseño para este experimento.

## 2. Grados de libertad del error

El diseño que utilicemos tiene un efecto en los grados de libertad del error, que puede influenciar si detectamos o no el efecto de un tratamiento. En el ejemplo 8 vamos a comparar el esqueleto de la tabla ANOVA para el mismo experimento utilizando diseños distintos.

### Ejemplo 8

El efecto del encharcamiento de agua en la producción de algodón fue evaluado en un experimento de campo llevado a cabo en suelo arcilloso. En el experimento se empleó dos tratamientos de encharcamiento de agua y tres variedades con tres réplicas. El  $O_2$  del suelo fue medido tras 24 horas del encharcamiento en cada parcela. Las parcelas en el campo se disponen en un terreno rectangular de seis filas por tres columnas de parcelas.

#### 2.1. Diseño totalmente aleatorizado

En resumen:

- **Objetivo:** determinar el efecto del encharcamiento de agua y las variedades en el  $O_2$  del suelo.
- **Observaciones:** 18 parcelas
- **Disposición:** 6 filas  $\times$  3 columnas
- **Tratamientos:** encharcamiento (2) y variedades (3) – estructura factorial
- **Réplicas:** 3
- **Diseño:** diseño totalmente aleatorizado
- **Disposición de los bloques:** NA

Para crear el diseño en R utilizamos el siguiente código:

```
trt <- c(2, 3) # factorial 2 x 3
rep <- 3
outdesign <- design.ab(trt, r = rep, design = "crd")
```

A continuación, se muestra el esqueleto de la tabla ANOVA para este diseño:

Fuente de variación	gl
A	1
B	2
AB	2
Residuo	12
Total	17

#### 2.2. Diseño en bloques completos aleatorizados

En resumen:

- **Objetivo:** Determinar el efecto del encharcamiento del agua y las variedades en el  $O_2$  del suelo.

- **Observaciones:** 18 parcelas
- **Disposición:** 6 filas  $\times$  3 columnas
- **Tratamientos:** encharcamiento de agua (2) y variedades (3) – estructura factorial
- **Réplicas:** 3
- **Diseño:** diseño en bloques completos aleatorizados
- **Disposición de los bloques:** 6 filas  $\times$  1 columna

Para crear el diseño en R utilizamos el siguiente código:

```
trt <- c(2, 3) # factorial 2 x 3
rep <- 3
outdesign <- design.ab(trt, r = rep, design = "rcbd")
```

A continuación, se muestra el esqueleto de la tabla ANOVA para este diseño:

Fuente de variación	gl
Estrato del bloque	2
A	1
B	2
AB	2
Residuo	10
Total	17

### 2.3. Parcelas divididas

En resumen:

- **Objetivo:** determinar el efecto del encharcamiento de agua y las variedades en el  $O_2$  del suelo.
- **Observaciones:** 18 parcelas
- **Disposición:** 6 filas  $\times$  3 columnas
- **Tratamientos:** encharcamiento de agua (2) y variedades (3) – estructura factorial
- **Réplicas:** 3
- **Diseño:** parcelas divididas
- **Disposición de los bloques:** 6 filas  $\times$  1 columna con dos parcelas grandes (whole plot) dentro de cada bloque y tres parcelas pequeñas (subplots) dentro de cada parcela grande.

Para crear el diseño en R utilizamos el siguiente código:

```
waterlog <- c("W1", "W2")
variety <- c("A", "B", "C")
rep <- 3
outdesign <- design.split(waterlog, variety, r = rep)
```

A continuación, se muestra el esqueleto de la tabla ANOVA para este diseño:

Fuente de variación	gl
=====	
Estrato del bloque	2
-----	
Estrato parcela grande	
encharcamiento	1
Residuo parcela grande	2
=====	
Estrato parcela pequeña	
variedad	2
encharcamiento:variedad	2
Residuo parcela pequeña	8
=====	
Total	17

Conforme incrementa la complejidad del diseño, los grados de libertad del error disminuyen. El tamaño del experimento influye en el tipo de diseño seleccionado, no obstante, la heterogeneidad de las unidades experimentales también se tiene que tener en cuenta.



### 3. Diseños más complejos

Los diseños descritos hasta el momento son los diseños más simples. Estos diseños suelen ser bastante comunes en el área de agricultura. En esta sección se introducen las características y el uso apropiado de otros diseños un poco más complejos. Se recomienda consultar con un estadístico si se quiere llevar a cabo este tipo de diseños para asegurar que los experimentos se diseñan de forma apropiada y se alcanzan los objetivos establecidos con el experimento.

#### 3.0.1. Diseño de parcelas en banda o de bloques divididos

Existen ocasiones donde se quiere conocer el efecto de dos factores dentro de un mismo experimento y ambos factores requieren unidades experimentales grandes. Por ejemplo, supón que existen cuatro niveles de espaciado entre semillas y tres niveles de arado de los que se quiere evaluar su efecto en la productividad de las plantas. En este experimento ambos factores requieren unidades experimentales grandes. Si la combinación factorial de los dos factores se aleatoriza en un diseño en bloques completos aleatorizados, las parcelas que se requerirían deberían ser muy grandes y por tanto heterogéneas entre sí.

Si se utiliza un diseño de parcelas divididas y el factor (espaciado entre semillas) se aleatoriza a parcelas grandes y el otro factor (método de arado) se aleatoriza a subparcelas dentro de la parcela grande, las subparcelas tendrían también que ser muy grandes. Es por ello que un diseño de parcelas divididas tampoco sería adecuado. En este tipo de situaciones se adopta lo que se llama un diseño de parcelas en banda también conocido como en bloques divididos y en inglés strip plot design. El área experimental se divide en tres tipos de parcelas, principalmente llamadas banda vertical, banda horizontal y parcela intersección de ambas bandas. Se aleatoriza los niveles del factor A y B, respectivamente, a las bandas verticales y horizontales, y esto permite que la parcela intersección acomode la interacción de ambos factores. Como en el diseño de parcelas divididas, las parcelas horizontales y verticales son perpendiculares unas de otras. No obstante, en el diseño de parcelas en banda la relación de tamaño de las bandas verticales y horizontales no es tan distinta a la de las parcelas grandes y subparcelas en los diseños de parcelas divididas. La aleatorización de los tratamientos a las parcelas pequeñas no ocurre de forma independiente como en el diseño de parcelas divididas sino que es fruto de la intersección de los niveles de los dos factores que han sido aleatorizados a las bandas. La unidad experimental de la interacción entre ambos factores es la más pequeña en tamaño.

En el diseño de parcelas en banda, cada bloque se divide en un número de bandas horizontales y verticales dependiendo de los niveles de los dos factores que intervienen en el experimento. Sea A el factor horizontal con  $a$  niveles, B el factor vertical con  $b$  niveles y  $r$  el número de réplicas. El área experimental se divide en  $r$  bloques. Cada bloque se divide en  $a$  bandas horizontales y los  $a$  niveles del factor A se aleatorizan a estas bandas de forma separada e independiente dentro de los  $r$  bloques. Seguidamente, cada bloque se divide en  $b$  bandas verticales y los  $b$  niveles del factor B se aleatorizan a estas bandas de forma separada e independiente dentro de cada uno de los  $r$  bloques. Una posible disposición de un diseño de parcelas en banda con  $a = 5$  ( $a_1, a_2, a_3, a_4$ , and  $a_5$ ),  $b = 3$  ( $b_1, b_2$ , and  $b_3$ ) y 4 réplicas se muestra en la siguiente página.

El diseño de parcelas en banda sacrifica la precisión de los efectos principales de ambos factores para proporcionar mayor precisión del efecto interacción que generalmente será determinado de forma más exacta que con los diseños en bloques completos aleatorizados o en parcelas divididas. Es por ello que no se recomienda este tipo de diseños a no ser que haya restricciones en la aplicación de los niveles de los factores o el objetivo principal del experimento sea estudiar la interacción entre ambos factores.

Se lleva a cabo un experimento con el objetivo de estudiar el efecto de tres niveles de espaciado entre semillas y cinco niveles de arado. El diseño seleccionado para dicho experimento es un diseño de parcelas en banda. Cada bloque tiene tres columnas y cinco filas. Dentro de cada bloque, los niveles de espaciado de semilla se aleatorizan a las columnas y los niveles de arado a las filas. El experimento se dispone en un terreno rectangular de doce columnas y cinco filas.

En resumen:

- **Objetivo:** determinar el efecto de tres niveles de espaciado entre semillas y cinco niveles de arado
- **Observaciones:** 60 parcelas

- **Disposición:** 12 filas  $\times$  5 columnas
- **Tratamientos:** espaciado entre semillas (3) y arado (5) – estructura factorial
- **Réplicas:** 4
- **Diseño:** bloques divididos/ parcelas en bandas
- **Disposición de los bloques:** 5 filas  $\times$  3 columnas. Dentro de cada bloque los niveles de espaciado de semilla se aleatorizan a las columnas y los niveles de arado a las filas.

A continuación, se muestra la gráfica del diseño remarcando los tratamientos:

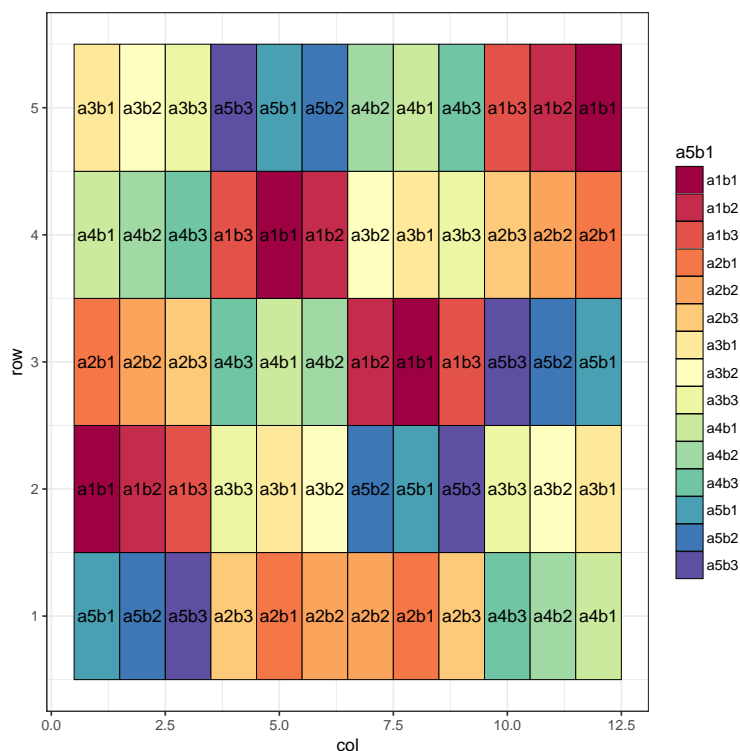


Figura 2.16: Diseño de parcelas en banda remarcando los tratamientos para el estudio de la interacción entre niveles de espaciado de semilla y niveles de arado

A continuación, se muestra la gráfica del diseño remarcando los niveles del factor arado:

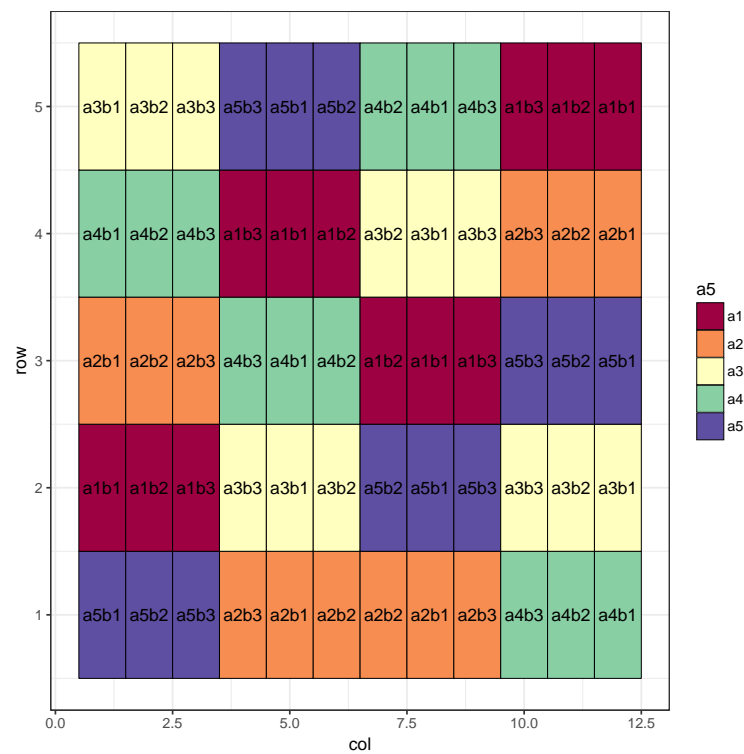


Figura 2.17: Diseño de parcelas en banda remarcando los niveles del factor arado

A continuación, se muestra la gráfica del diseño remarcando los niveles del factor espaciado de semillas:

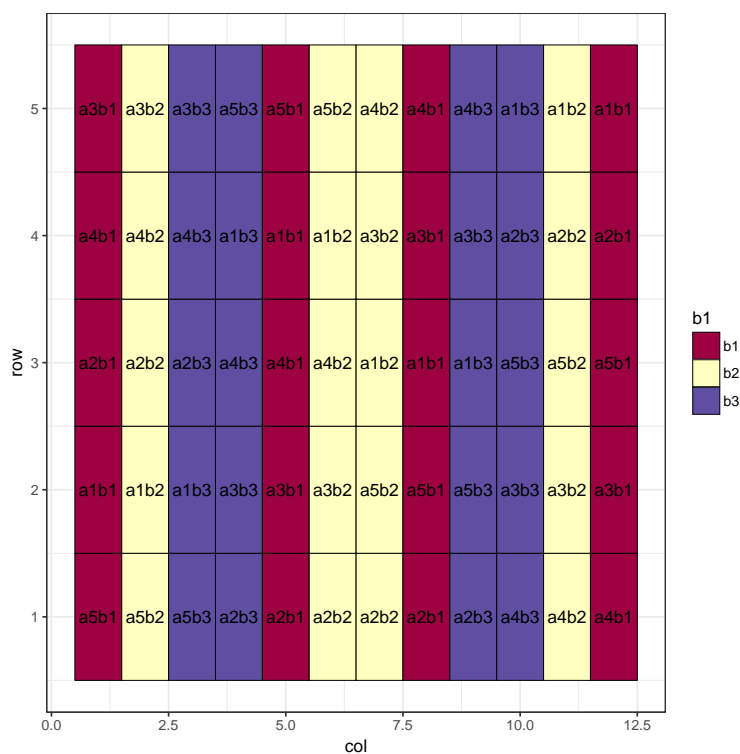


Figura 2.18: Diseño de parcelas en banda remarcando los niveles del factor espaciado de semillas

A continuación, se muestra una gráfica remarcando los niveles donde el factor asignado a las bandas horizontales no ha sido aleatorizado en cada uno de los bloques:

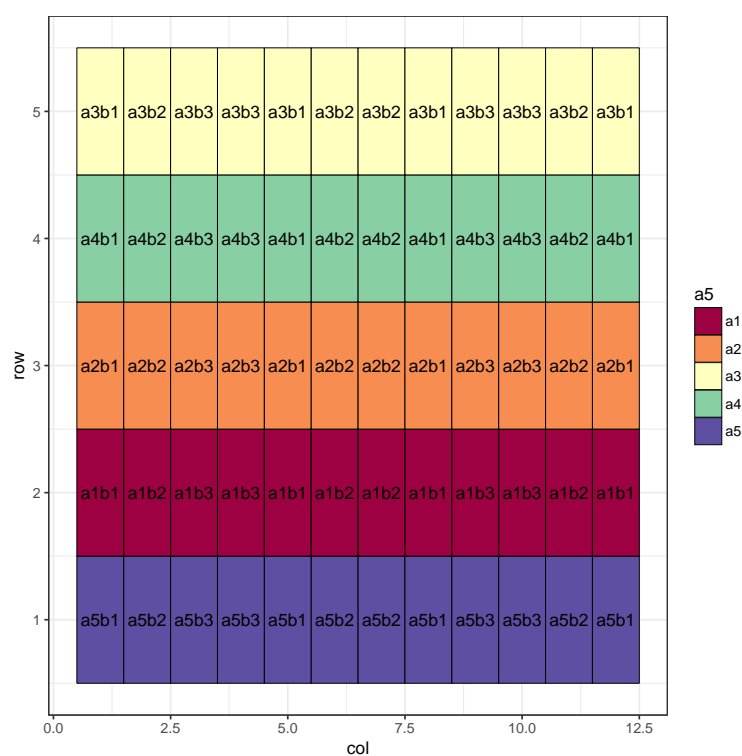


Figura 2.19: Diseño de parcelas en banda con una aleatorización incorrecta

### 3.0.2. $p$ -rep

Los ensayos de primera generación o, en otras palabras, aquellos que se producen al principio de un programa de mejora genética de variedades se llevan a cabo para explorar variables cuantitativas como puede ser la producción de grano en variedades de trigo nuevas. En este tipo de ensayos puede que no haya suficiente cantidad de semilla para replicar una nueva variedad y es por ello que se adoptan lo que se llaman diseños no replicados. En este tipo de ensayos era común llevar a cabo diseños en grid donde se interpone una red de parcelas que contiene variedades estándares que sí se replican y otras variedades nuevas que no se replican entre las parcelas restantes.

Por ejemplo, los datos del experimento de campo no replicado llevado a cabo en Tullibigeal, Nueva Gales del Sur, Australia en 1987-88 mostrado abajo, se aprecia que la variedad Kite fue replicada, otras seis variedades estándar fueron aleatoriamente intercaladas en las parcelas del experimento. Cada parcela tenía una dimensión de 15m x 1.8m, “orientada con el lado más largo con las filas”. (Los datos se obtuvieron del paquete **agridat**).

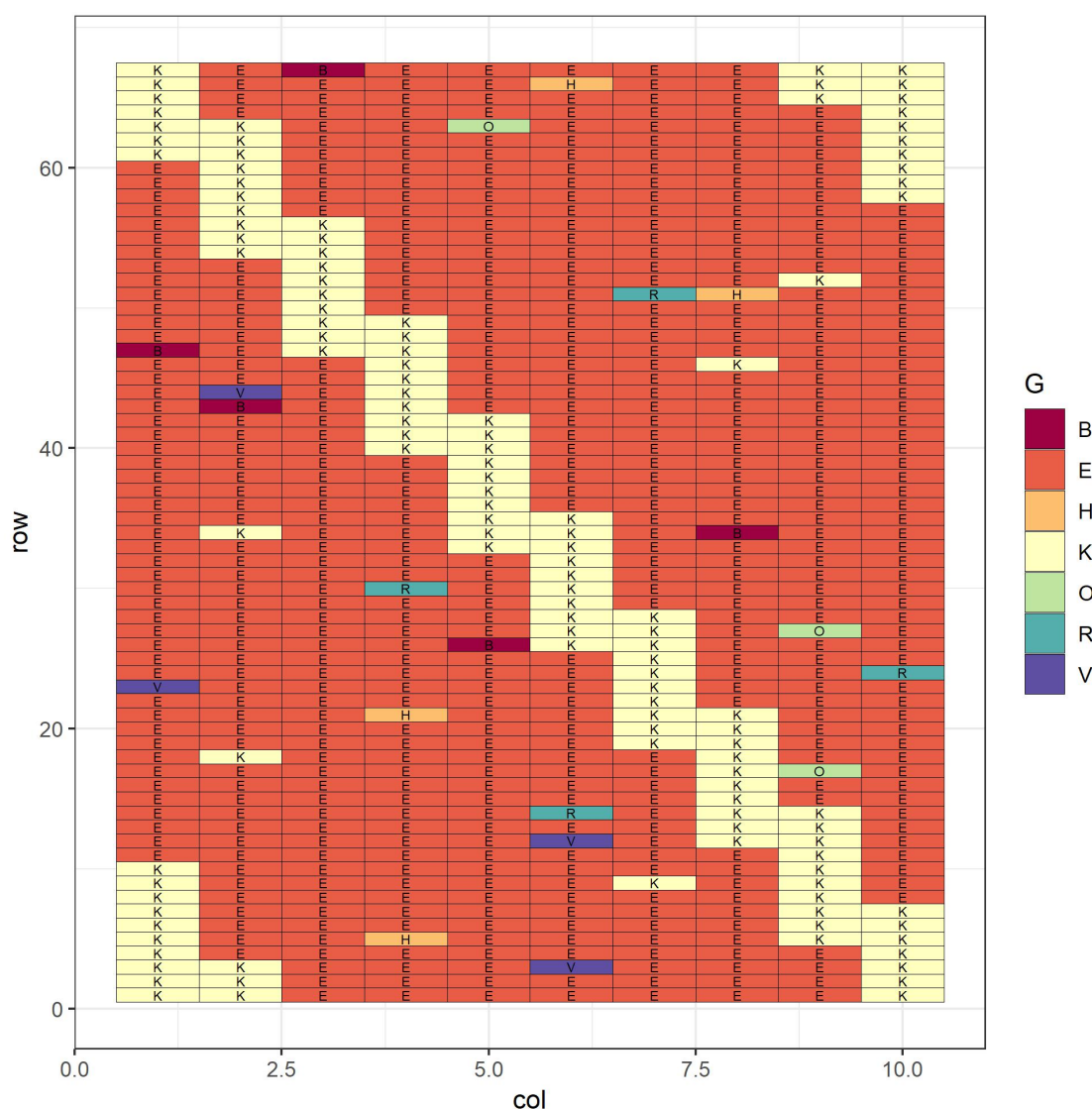


Figura 2.20: Experimento de campo no replicado

Cullis, Smith, y Coombes (2006) introdujo el concepto de diseños parcialmente replicados ( $p$ -rep) como una alternativa a esos diseños en grid. Los diseños  $p$ -rep se llevan a cabo replicando un porcentaje  $p$  de las variedades de interés. Normalmente en un diseño donde las variedades estándares forman un grid, la ratio de variedades estándares frente a las nuevas variedades de interés es de 1:4. Se sugiere que para los diseños parcialmente replicados el porcentaje de variedades replicadas sea al menos 25% y cada variedad replicada se replique dos

veces. Este diseño básico se puede modificar fácilmente incrementando el porcentaje de replicación, así como el número de réplicas de las variedades para acomodar los requerimientos específicos de cada experimento.

Considera un experimento de campo con 96 variedades de las que se quiere cuantificar su rendimiento en cuatro sitios distintos con cinco réplicas en total. En este escenario, se propone un diseño  $p$ -rep donde en cada sitio se repliquen 24 variedades y el resto de las 72 variedades no se repliquen ( $p = 25\%$ ) tal y como se muestra en la siguiente tabla.

	Entradas 1 - 24	Entradas 25 - 48	Entradas 49 - 72	Entradas 73 - 96
Sitio A	2	1	1	1
Sitio B	1	2	1	1
Sitio C	1	1	2	1
Sitio D	1	1	1	2

Las Figuras 2.21 y 2.22 muestran ejemplos de aleatorización para estos experimentos.

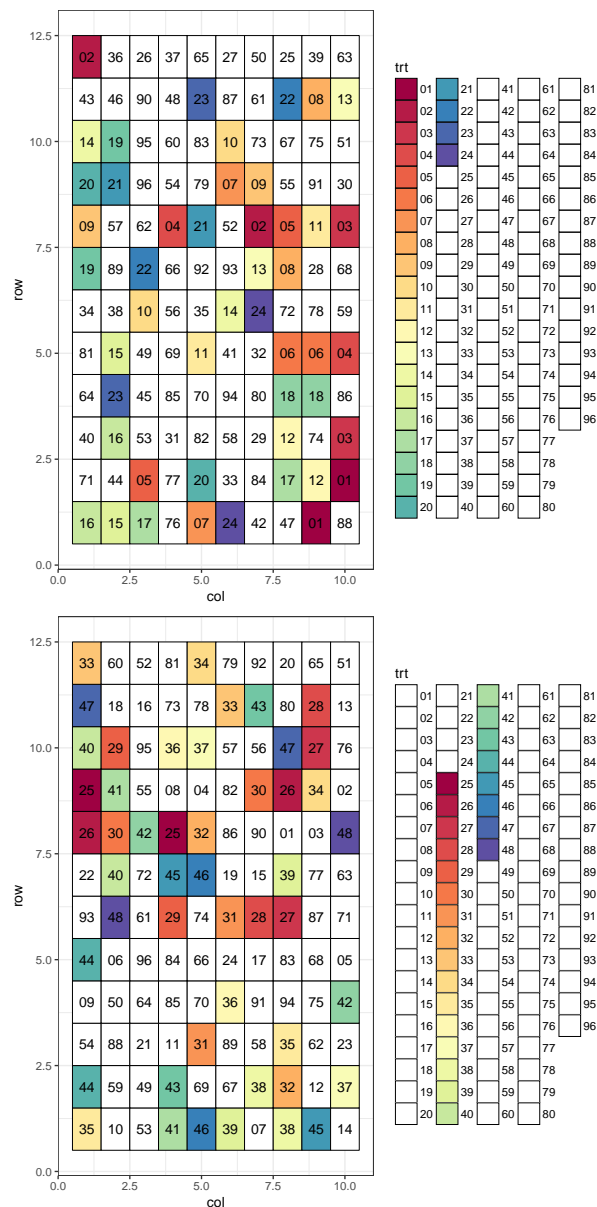
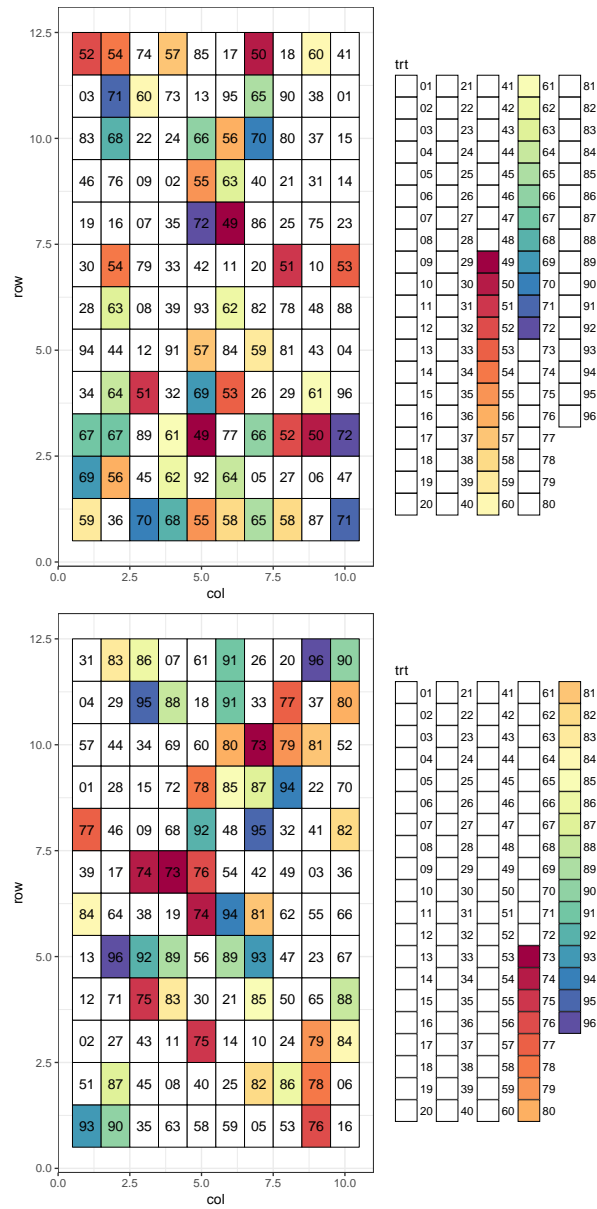


Figura 2.21:  $p$ -rep diseño obtenidos para el sitio A y B

Figura 2.22:  $p$ -rep diseño obtenidos para el sitio C y D



## Referencias

- Cullis, B. R., Smith, A. B., y Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(4), 381–393.
- de Mendiburu, F. (2016). agricolae: Statistical procedures for agricultural research [Manual de software informático]. Descargado de <https://CRAN.R-project.org/package=agricolae> (R package version 1.2-4)
- Hofseth, L. J. (2018). Getting rigorous with scientific rigor [Journal Article]. *Carcinogenesis*, 39(1), 21–25. Descargado de <http://dx.doi.org/10.1093/carcin/bgx085> doi: 10.1093/carcin/bgx085
- Lenth, R. V. (2007). Statistical power calculations. *Journal of animal science*, 85(13\_suppl), E24–E29.
- McKilup, S. (2006). *Statistics explained: an introductory guide for life scientists*. Cambridge, United Kingdom: Cambridge University Press.
- R Development Core Team. (2010). R: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria. Descargado de <http://www.R-project.org> (ISBN 3-900051-07-0)
- Welham, S. J., Gezan, S. A., Clark, S., y Mead, A. (2014). *Statistical methods in biology: Design and analysis of experiments and regression*. CRC Press.