



**Universidad de Valladolid**

Facultad de Ciencias

Trabajo Fin de Grado

Grado en Estadística

**Estudio de diseños experimentales  
split-plot y su aplicación en  
reforestación**

*Beatriz Fernández Miguel*

*Tutora:*

*Itziar Fernández Martínez*



## Agradecimientos

*A mí familia y amigos, por confiar en mí y apoyarme en todo momento.*

*A mi tutora Itziar Fernández por toda la ayuda y consejos proporcionados durante estos meses.*

*A Irene Ruano y Felipe Bravo por haberme cedido los datos para este trabajo.*



# Índice

Resumen .....	7
Abstract.....	8
Introducción.....	9
1. Diseño de experimentos .....	11
1.1 Elementos esenciales en el diseño de experimentos .....	11
1.2 Principios básicos del diseño de experimentos.....	11
1.3 Fases en el diseño experimental.....	12
2. Diseño split-plot (Parcelas divididas).....	14
2.1 Planteamiento del diseño.....	14
2.2 Modelo y Análisis.....	17
3. Herramientas.....	23
3.1 R.....	23
3.1.2 librería agricolae .....	23
3.1.2.1 Función desing.split .....	23
3.1.2.2 Función sp.plot .....	26
3.1.3 librería agricolaeplotr .....	27
3.1.3.1 plot_split_rcbd.....	27
3.1.4 Ajuste de los datos.....	29
3.1.4.1 Función glmer ().....	29
3.2 SAS .....	31
3.2.1 Creación de un diseño experimental split-plot.....	31
3.2.2 Análisis .....	34
4. Aplicación.....	36
4.1 Regeneración forestal.....	36

4.1.1	Introducción.....	36
4.1.2	Estudio de la regeneración natural de “ <i>Pinus Pinaster</i> ” en arenales del centro de España .....	37
4.1.2.1	Situación del experimento.....	37
4.1.2.2	Datos.....	37
4.1.2.3	Experimentación mediante Split-plot.....	42
4.1.2.3.1	Análisis descriptivo .....	45
4.1.2.3.2	Ajuste del modelo.....	47
5.	Discusión .....	51
6.	Conclusiones .....	52
7.	Bibliografía.....	53

## INDICE DE FIGURAS

Figura 2.1. Ejemplo de división de parcelas según el método de riego. ....	16
Figura 2.2. Ejemplo de división de parcelas según el abono y el método de riego.	16
Figura 3.1: sentencia y parámetros para realizar un diseño Split-plot.....	23
Figura 3.2: Salida de R del diseño generado mediante Split-plot. ....	25
Figura 3.3: Sentencia y parámetros para analizar un diseño Split-plot. ....	26
Figura 3.4: Sentencia y parámetros para analizar el diseño del apartado anterior	26
Figura 3.5: Sentencia y parámetros para graficar un diseño Split-plot.....	27
Figura 3.6: Planteamiento del diseño sobre el terreno. ....	28
Figura 3.7: Sentencia y parámetros para la orden glmer .....	29
Figura 3.8: Creación de un diseño split-plot con el programa SAS.....	31
Figura 3.9.1: Pasos para la creación del diseño.....	32
Figura 3.9.2: Pasos para la creación del diseño.....	33
Figura 3.10: Procedimiento de SAS para analizar el diseño.....	34
Figura 3.11: Descripción del conjunto de datos .....	34
Figura 3.12: Salida análisis split-plot .....	35
Figura 4.1: Mediciones del árbol a las trampas.....	40
Figura 4.2: Esquema de la distribución de parcelas según la intensidad de corta.	43
Figura 4.3: División sistemática de cada una de las parcelas.....	44
Figura 4.4: Numero de semillas recogida en cada parcela indicando el nivel de corta y box-plot. ....	45
Figura 4.5: Numero de semillas recogida en cada parcela en cada año, indicando el nivel de corta.....	46
Figura 4.6: Distintos gráficos de diagnóstico.....	50

## INDICE DE TABLAS

Tabla 2.1 ANOVA Diseño split-plot .....	19
Tabla 2.2 Estimadores y errores estándar.....	20
Tabla 2.3: Error estándar para la diferencia entre dos medias .....	20
Tabla 3.1: Tabla resumen Anova .....	26
Tabla 4.1: División de las variables en el proceso de maduración. ....	39
Tabla 4.2: Variables más relevantes.....	40
Tabla 4.3: Clasificación de las parcelas según el % de corta.....	43
Tabla 4.4: Resultado de los Efectos aleatorios.....	49
Tabla 4.5: Resultado de los Efectos fijos.....	49

## **Resumen**

En este proyecto se abordará la explicación del diseño de experimentos Split-Plot, así como el uso de las herramientas estadísticas R y SAS, que servirán para diseñar y analizar estos experimentos.

Finalmente se mostrará un ejemplo práctico y real de cómo se puede utilizar este tipo de diseño en el ámbito de la Ingeniería forestal para estudiar la regeneración natural en coníferas situadas en los arenales de Castilla y León.

### **Palabras clave**

Diseño de experimentos, Split-Plot, parcelas, subparcelas, modelo mixto.

## **Abstract**

This project is focused on the use of the Split-Plot designs, as well as its implementation in statistical software such as R and SAS, which will be used to design and analyze these experiments. Finally, a real-world example of the field of forest engineering will be used to illustrate how this type of experimental design can be support to study natural regeneration in conifers located in the sands of Castilla y León.

### **Palabras clave**

Design of experiment, Split-Plot, plot, subplot, mixed model.

# Introducción

En la actualidad la tecnología está muy presente en el día a día de cualquier proceso, y siendo la estadística un apoyo fundamental para el desarrollo de dicha tecnología, conlleva a que la estadística también esté muy presente en ese día a día.

La experimentación, que es el conjunto de pruebas que se realizan sobre una serie de variables de entrada con el objetivo de observar cambios en la variable(s) respuesta, forma parte intrínseca del desarrollo de cualquier experimento y mediante la estadística somos capaces de optimizar el modelo experimental mediante el diseño de experimentos. [1]

Se considera pionero en el diseño de experimentos a Ronald Fisher (S.XX), quien los usó en sus estudios relacionados con la agronomía, en concreto en la mejora de producción de patatas.

El diseño de Experimentos [2] es una técnica estadística, en la cual se busca definir la estrategia experimental óptima para poder obtener respuestas en relación a la investigación, de manera que éstas sean lo más eficientes, claras y concisas posible. Asimismo, se evalúan los resultados experimentales obtenidos para poder garantizar la máxima fiabilidad en la interpretación de los resultados.

Todas estas premisas permiten mejorar el rendimiento de los procesos y reducir la variabilidad.

Estos diseños son muy usados en muchos y muy diferentes ámbitos, como por ejemplo la ecología, medicina, industria, entre otros.

Un tipo de diseño de experimentos es el denominado Split-plot, el cual permite obtener información bastante precisa sobre un factor y la interacción con un segundo, pero renunciando a la precisión de este último y que está muy

recomendado para todos aquellos experimentos en los que se use maquinaria y procesos agrícolas sobre parcelas de cultivo.

Esta memoria está dividida en cuatro capítulos: En el primer capítulo se relata cuáles son los elementos principales del diseño de experimentos y los pasos a seguir en la creación de un diseño. En el segundo capítulo se explica que es un diseño split-plot, como se plantea este diseño y como es el modelo y el análisis de este. A continuación, se explican las herramientas, es decir, los programas estadísticos que se usan para poder tanto diseñar como analizar este tipo de experimentos. Y por último se muestra un caso real de aplicación del diseño split-plot en ingeniería forestal en el que se estudia la regeneración de las coníferas en los arenales de Castilla y León.

# 1. Diseño de experimentos

## 1.1 Elementos esenciales en el diseño de experimentos

Los elementos esenciales para implementar un diseño son:

- **Factores:** Son variables mediante la cuales se puede medir el efecto experimental. Estas son independientes pudiendo haber interacción entre ellas.
- **Niveles o tratamiento:** Estados que puede llegar a tener un factor. Si estos elementos los elige el experimentador se les identifica como efectos fijos, pero en cambio sí son elegidos al azar, serán denominados efectos aleatorios.
- **Bloque:** Grupo de unidades experimentales análogas frente a un factor concreto. [\[1\]](#)

## 1.2 Principios básicos del diseño de experimentos.

Cuando se realice un experimento, siempre habrá que tener en consideración los principios básicos del diseño de experimentos: aleatorización, formación de bloques y realización de réplicas.

- **Aleatorización:** Es una pieza fundamental de la experimentación, en ella se asignan de forma aleatoria la unidad experimental al tratamiento, con el objetivo de reducir la influencia negativa de condiciones externas no controlables. [\[3\]](#)
- **Formación de bloques:** Se usarán los bloques para dividir las observaciones, de tal manera que cada bloque tenga unas condiciones experimentales similares. Por lo tanto, los factores a bloquear serán aquellos cuya influencia sobre la variable respuesta no sea de interés, ya que dicha influencia será

mayoritariamente negativa en la variable respuesta, lo que aumentará la variabilidad.

- **Réplicas:** Son repeticiones con el mismo diseño, independientes entre sí y en las que se usan los mismos factores. [4]

### 1.3 Fases en el diseño experimental

En primer lugar, se tendrá que fijar cual es el objetivo del experimento. Esta parte es fundamental ya que es el elemento esencial para poder empezar a plantear el diseño experimental. Teniendo claro cuál es el objetivo se deberán de seguir los siguientes pasos:

1. **Reconocer el problema:** Será importante estudiar únicamente aquellas cuestiones que sean fundamentales, de lo contrario, el experimento puede llegar a complicarse en exceso. [1]
2. **Identificar la variable respuesta:** Inicialmente, se indicará cual es la(s) variables(s) respuesta(s). Esta parte es fundamental, ya que en ella se recogen los resultados de las pruebas.
3. **Seleccionar los factores y sus niveles:** El siguiente paso consiste en indicar cuales son los factores que debemos estudiar para ver que influencia tienen sobre la variable respuesta.  
  
Primeramente, se puede introducir una combinación de factores de “todos con todos”, aunque no siempre es necesario, para más tarde, poder determinar cuáles de ellos tiene un mayor efecto sobre la variable respuesta. También se tendrá que indicar cuantos niveles tiene cada factor.
4. **Elección del diseño experimental:** Una vez fijados los puntos anteriores deberán definirse las características del experimento, en ellas se fijarán el tamaño de la muestra y secuencia de ejecución de éste.

5. **Realización del experimento:** Una vez aclaradas las premisas anteriores, puede realizarse el trabajo experimental.

Si el tamaño del conjunto de datos es suficientemente grande se podrá, tanto realizar un experimento piloto para detectar posibles problemas que pudiera haber en los datos, como ayudar a la hora de seleccionar el modelo estadístico más adecuado.

6. **Análisis de los datos y conclusión:** La metodología empleada más frecuentemente en el diseño de experimentos es el Análisis de la Varianza (ANOVA) mediante el que se analiza de forma comparativa resultados en distintos grupos. Finalmente, tras tener los datos analizados se extraen las conclusiones.[\[1\]](#)

## 2. Diseño split-plot (Parcelas divididas)

### 2.1 Planteamiento del diseño.

Este diseño proviene del entorno agrícola y lo desarrolló Ronald Fisher, haciendo referencia a las parcelas de cultivo donde estas son de grandes extensiones y suelen estar divididas en subparcelas. [5]

La definición de parcela según la RAE es: [7]

“Porción pequeña de terreno, de ordinario sobrante de otra mayor que se ha comprado, expropiado o adjudicado”.

Cada subparcela se obtiene a través de una parcelación de la parcela, una parcelación según la RAE es: [7]

“Dividir una finca grande para venderla o arrendarla en porciones más pequeñas”.

El diseño Split-plot se emplea generalmente cuando se busca estudiar dos tipos de factores: uno de ellos será más fácil de variar que el otro[5].

Se trata de experimentos multifactoriales donde no es fácil que todas las ejecuciones sean completamente aleatorizadas [6]. Por ello, es un caso particular de los experimentos factoriales ya que, en este diseño, no hay combinación aleatoria entre los factores ya que, están anidados. A cada parcela se le asignan uno o dos factores principales y a cada subparcela se le asigna uno de los niveles del factor que corresponde a las subparcelas. Por lo tanto, cada parcela es un bloque para los niveles de las subparcelas.

Respecto a la aleatorización, se asignará a cada parcela uno de los niveles del factor de manera aleatoria y a continuación, se aleatorizan los niveles del factor de la subparcela, dividiendo cada parcela en tantas partes como niveles tenga el factor correspondiente a las subparcelas. [10]

Es extremadamente importante asignar los factores de manera correcta a las parcelas o a las subparcelas, en función del tamaño de la parcela, el grado de precisión de la medición de los efectos y el material experimental disponible: [10]

**Grado de precisión:** Se buscará tener más grado de precisión en un factor que en otro, por lo tanto, este factor será el que se asigne a las subparcelas.

**Tamaño relativo de los efectos principales:** Si se sabe que el efecto principal de un factor y su detección va a ser más sencilla que la de otro factor, este se asigna a la parcela y el de menor efecto principal a la subparcela.

**Material experimental:** Aquel factor que necesite mayor número de material experimental o más terreno se asigna la parcela.

Para ilustrar este diseño se utiliza el siguiente ejemplo, mediante las Figuras 2.1 y 2.2:

Se quiere saber los efectos que tienen dos métodos de riego (Factor 1) y dos abonos (Factor 2) en el rendimiento de una plantación de patatas:

Como se puede deducir, es difícil regar de forma distinta las diferentes partes de una parcela de forma estrictamente controlada, pero sí que es fácilmente posible aplicar distintos tipos de abono a éstas. A continuación, en las Figuras 2.1 y 2.2 se muestra el planteamiento sobre el terreno en un diseño formado por 2 bloques completos al azar:

Inicialmente, hay 4 parcelas donde se indica aleatoriamente que tipo de método de riego se usa, identificándose como R1 y R2

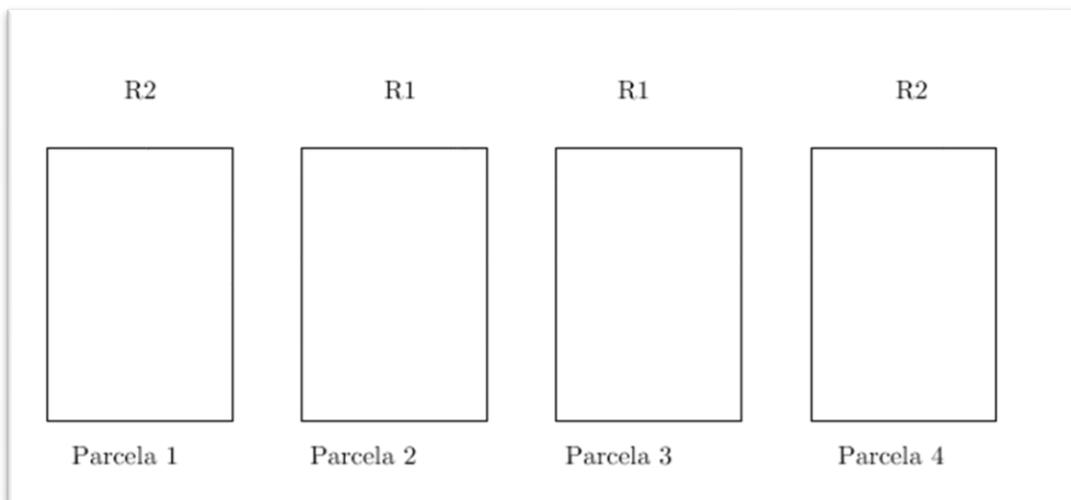


Figura 2.1. Ejemplo de división de parcelas según el método de riego.

Se divide en dos cada parcela para así poder asignar aleatoriamente a cada parte un tipo de abono teniendo, por tanto, 4 parcelas y en cada una de estas 2 subparcelas. A cada tipo de abono se le identifica como tratamiento de la subparcela (B1, B2).

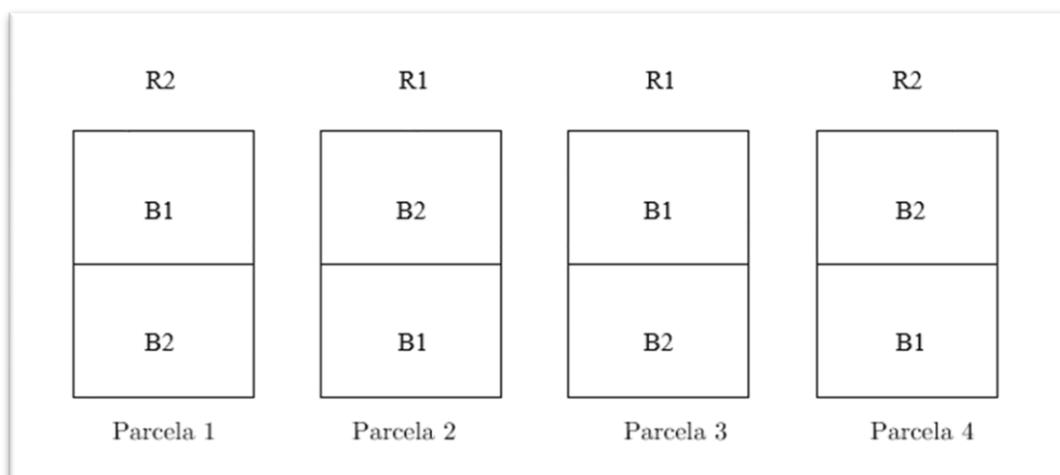


Figura 2.2. Ejemplo de división de parcelas según el abono y el método de riego.

Cuando el diseño está formado por 3 factores, se denomina diseño de parcelas subdivididas donde cada parcela se divide en subparcela y sub-subparcela. Este diseño no se suele utilizar con mucha frecuencia ya que al aumentar los factores los cálculos y las interpretaciones de los resultados se complican.

## 2.2 Modelo y Análisis

El modelo y el análisis se dividen en dos partes, la que hace referencia a las parcelas y la relacionada con las áreas subparceladas.

Entre las distintas parcelas puede haber correlación. En agricultura, por ejemplo, esto demuestra que las parcelas tienen propiedades agrícolas relacionadas.

También puede haber correlación entre los errores experimentales de las subparcelas. [2]

Suponiendo un diseño con  $r$  réplicas,  $a$  parcelas y  $b$  subparcelas, el modelo estadístico es el siguiente,: [8]

$$Y_{ijl} = \mu + \alpha_i + \beta_j + \delta_l + (\alpha\beta)_{ij} + (\beta\delta)_{jl} + \varepsilon_{ijl}$$

$$i = 1, \dots, r; l = 1, \dots, a; j = 1, \dots, b$$

Se describen los siguientes elementos pertenecientes al modelo:

- $Y_{ijl}$ : variable respuesta, donde se recogerán los resultados de la unidad experimental.
- $\mu$ : parámetro que corresponde a la media general del modelo.

Conforme a la parcela:

- $\alpha_i$  : parámetro que corresponde a las réplicas.
- $\beta_j$  : El efecto del factor de la parcela
- $(\alpha\beta)_{ij}$ : El error de la parcela.

En relación a las subparcelas tenemos los elementos:

- $\delta_l$  : el factor de la subparcela,
- $(\beta\delta)_{il}$ : la interacción,

- $\epsilon_{ijl}$  : los errores aleatorios.

Todos los errores son independientes y seguirán la distribución  $N(0, \delta^2)$ .

La principal diferencia entre el error de la parcela y las subparcelas son los grados de libertad.

- Siendo  $a$  los grados de libertad del factor de la parcela y  $r$  los grados de libertad del bloque:  $(r-1) \times (a-1)$  grados de libertad.
- Siendo  $b$  los grados de libertad del factor de la subparcela y  $r$  los grados de libertad del bloque:  $(r-1) \times (b-1)$  grados de libertad.

Se prueban hipótesis nulas de interés:

- $H_0: \beta_j = 0$
- $H_0: \delta_l = 0$
- $H_0: (\beta\delta)_{il} = 0$  , si se rechaza esta última hipótesis las otras dos restantes carecerán de interés.

Los efectos se probarán mediante el estadístico test para la hipótesis nula de que no existen diferencias en las medias de los tratamientos:

Los efectos de la parcela se calculan:

$$F = \frac{MS\beta_j}{MS(\alpha\beta)_{ij}}$$

Siendo  $MS\beta_j$  el cuadrado medio de la parcela y  $MS(\alpha\beta)_{ij}$  el cuadrado medio del error de la parcela.

Los efectos de la subparcela y la interacción de los efectos de la parcela y la subparcela se probarán mediante el cuadrado medio del error de la subparcela.

$$F = \frac{MS\delta_l}{MS(\beta\delta)_{il}}$$

Para calcular el valor de la  $F$  para el error de la subparcela, se calculará de la siguiente forma:

$$F = \frac{MS(\beta\delta)_{il}}{MS(\beta\delta)_{il}}$$

Si estos tres valores anteriores tienen un resultado próximo a uno, su p-valor será superior 0.05 y por lo tanto no se rechazará la hipótesis nula de no efecto entre los factores y la interacción entre ambos. En el caso contrario, es decir cuando el valor de las F sea mayor que 1, su p-valor será inferior a 0.05 y por lo tanto se rechazará la hipótesis diciendo que el efecto del factor o de la interacción es significativa.

A continuación, se muestra la tabla Anova en la Tabla 2.1 [9]:

Tabla 2.1 ANOVA Diseño split-plot

	Suma de cuadrados	G. L	cuadrado medio	F
<b>BLOQUE</b>	SSA <sup>1</sup>	r-1	$\sigma_{\varepsilon}^2 + ab\sigma_{\alpha}^2$	
<b>FACTOR PARCELA</b>	SSP	a-1	$\sigma_{\varepsilon}^2 + b\sigma_{\alpha\beta}^2 + \frac{pb \sum \beta^2}{a-1}$	MSP <sup>2</sup> /MSEr
<b>Error (p)</b>	SSEa	(r-1) (a-1)	$\sigma_{\varepsilon}^2 + b\sigma_{\alpha\beta}^2$	
<b>FACTOR SUBPARCELA</b>	SSB	b-1	$\sigma_{\varepsilon}^2 + \frac{pa \sum \delta_l^2}{ab-1}$	MSb/MSEab
<b>INTERACCIÓN</b>	SSpb	(a-1) (b-1)	$\sigma_{\varepsilon}^2 + \frac{p \sum \sum (\beta\delta)_{il}^2}{(a-1)(b-1)}$	MSab/MSEab
<b>ERROR (PB)</b>	SSEpb	a(r-1) (b-1)	$\sigma_{\varepsilon}^2$	
<b>TOTAL</b>	SST	rab-1		

Destacar también, que el cuadrado medio de las parcelas será mayor que el relacionado con las subparcelas. [11]

<sup>1</sup> SS<sub>x</sub> corresponde a la suma de cuadrados en inglés “sum of Squares”

<sup>2</sup> MS<sub>x</sub> hace referencia cuadrado medio en inglés “middle square”

A continuación, en la Tabla 2.2 se muestra el cálculo de los estimadores y los errores estándar: [10]

Tabla 2.2 Estimadores y errores estándar.

FACTORES	ESTIMADORES	ERROES ESTANDAR
FACTOR PARCELA	$\bar{y}_{i..} = \frac{\sum_{i=1}^p y_{i..}}{rb}$	$\sqrt{\frac{MSEa}{rb}}$
FACTOR SUBPARCELA	$\bar{y}_{.j.} = \frac{\sum_{i=1}^p y_{.j.}}{ra}$	$\sqrt{\frac{MSEab}{ra}}$
INTERACCIÓN	$\bar{y}_{ij.} = \frac{\sum_{i=1}^p \sum_{j=1}^b y_{ij.}}{r}$	$\sqrt{\frac{MSEab}{r}}$

A continuación, se muestra en la Tabla 2.3 el cálculo de los errores estándar de la diferencia entre dos medias: [10]

Tabla 2.3: Error estándar para la diferencia entre dos medias

Diferencia entre:	Cálculo
Dos medias del factor de la parcela	$\sqrt{\frac{2MSEa}{rb}}$
Dos medias del factor de la subparcela	$\sqrt{\frac{2MSEab}{ra}}$
Dos medias del factor de la subparcela para el mismo nivel del factor de la parcela	$\sqrt{\frac{2MSEab}{r}}$
Dos medias del factor de la parcela para el mismo o diferente nivel del factor de la subparcela	$\sqrt{\frac{2x(b-1)xMSEa + MSEa}{rb}}$

Es de vital importancia comprobar que el modelo es el adecuado para los datos, verificando que se cumplen las siguientes hipótesis:

- Hipótesis de Normalidad.
- Hipótesis de Independencia.
- Hipótesis de Homocedasticidad.

Para la hipótesis de normalidad esta se puede comprobar mediante dos métodos:

- Un gráfico QQ-plot donde si hay normalidad los residuos están distribuidos en torno a la recta  $y=x$ . Donde los pares de puntos que se representan son los siguientes, siendo  $e_{(r)}$ , el  $r$ -ésimo residuo ordenado y  $\Phi^{-1}$  es la función de distribución de un  $N(0,1)$ :

$$\left( e_{(r)}, \Phi^{-1} \frac{r - 0.5}{n} \right)$$

- Representando los datos mediante un histograma y superponiendo la curva de una distribución normal con la misma media y varianza que los datos.
- Mediante distintos test de Normalidad como pueden ser: Shapiro-Wilk, Kolmogórov-Smirnov, Cramer-von Mises, entre otros. Donde se contrasta la hipótesis nula de que los residuos proceden de una muestra normal.

En relación a la independencia:

- Se realiza un gráfico de los residuales dispuestos en el orden temporal de la recolección de estos datos, así se puede observar si existen correlaciones entre los residuos. Si los puntos graficados no presentan ningún patrón habrá independencia.

Por último, para comprobar si hay homocedasticidad:

$$H_0 = \sigma_1^2 = \dots = \sigma_n^2$$

$$H_1 = \sigma_i^2 \neq \sigma_j^2 \quad \text{siempre que } i \neq j$$

- Se comprueba, mediante el gráfico de residuos frente predichos donde se observa si los puntos de la gráfica no presentan ningún patrón y se distribuyen entorno al 0, lo que indica que las varianzas son iguales.
- Al igual que la normalidad, la homocedasticidad también se puede comprobar mediante contrastes de hipótesis como por ejemplo el test de Levene. [\[1\]](#)

## 3. Herramientas

### 3.1 R

En el software estadístico R [12], se encuentran distintas librerías donde se pueden tanto crear diseños de experimentos split-plot y posteriormente analizarlos.

#### 3.1.2 librería agricolae

Esta librería está implementada en la Universidad Nacional de Ingeniería de Lima (Perú). Está diseñada para procedimientos estadísticos enfocados a la investigación Agropecuaria. El paquete está formado por distintos tipos de funciones bien, para analizar datos que han sido diseñados mediante un diseño split-plot, bien para generar el análisis split-plot, entre otros.[13]

##### 3.1.2.1 Función `desing.split`

Esta función [14] permite crear un diseño split plot. Donde como se muestra a continuación se le dan distintos parámetros a la función dependiendo del diseño que se quiera crear. Se indican como parámetros principales los factores de las parcelas, las subparcelas y mediante que método estarán distribuidos estos efectos: distribución aleatoria (crd), bloques totalmente aleatorizados (rcbd) o cuadrados latinos (lsd). También se indican el número de repeticiones.

La función y sus parámetros para la generación de este se presenta en la Figura 3.1:

```
design.split(trt1, trt2,r=NULL, design=c("rcbd", "crd", "lsd"),serie = 2,  
seed = 0, kinds = "Super-Duper", first=TRUE, randomization=TRUE)
```

Figura 3.1: sentencia y parámetros para realizar un diseño Split-plot

- trt1: Tratamientos en Parcelas
- trt2: Tratamientos en subparcelas
- r: Replicas o bloques
- design: Diseño experimental
- serie: gráfica numérica: Etiqueta que indica a que bloque pertenece, 1: 11,12; 2: 101,102; 3: 1001,1002. Por ejemplo: 1:11,12, indica que se ha elegido el formato 1 de etiquetado, donde se indica que es el bloque 1, y las filas 1: 11 y dos: 12.
- seed: semilla
- kinds: método para aleatorizar, estos métodos pueden ser: "Wichmann-Hill", "Marsaglia-Multicarry", "Super-Duper", "Mersenne-Twister", "Knuth-TAOCP", "suministrado por el usuario", "Knuth-TAOCP-2002", " por defecto"
- first: VERDADERO o FALSO - aleatorizar repetición 1
- randomization: VERDADERO o FALSO – aleatorizar [\[14\]](#)

A continuación, se muestra un ejemplo práctico:

El diseño que se genera en este caso es un diseño de un experimento split-plot mediante bloques completos aleatorizados.

Este está formado por un factor parcela que tiene 5 niveles, (TP), un factor subparcelas que tiene 4 niveles (TS) y 3 bloques (block).

A continuación, se muestra el diseño obtenido en la Figura 3.2:

```

$parameters
$parameters$design
[1] "split"

$parameters[[2]]
[1] TRUE

$parameters$trt1
[1] "T1" "T2" "T3" "T4" "T5"

$parameters$applied
[1] "rcbd"

$parameters$r
[1] 3

$parameters$serie
[1] 2

$parameters$seed
[1] 45

```

	plots	splots	block	TP	TS
1	101	1	1	T5	4
2	101	2	1	T5	3
3	101	3	1	T5	2
4	101	4	1	T5	1
5	102	1	1	T3	1
6	102	2	1	T3	2
7	102	3	1	T3	3
8	102	4	1	T3	4
9	103	1	1	T2	3
10	103	2	1	T2	2
11	103	3	1	T2	1
12	103	4	1	T2	4
13	104	1	1	T4	1
14	104	2	1	T4	2
15	104	3	1	T4	4
16	104	4	1	T4	3
17	105	1	1	T1	2
18	105	2	1	T1	3
19	105	3	1	T1	1
20	105	4	1	T1	4
21	106	1	2	T2	2
22	106	2	2	T2	1
23	106	3	2	T2	4
24	106	4	2	T2	3
25	107	1	2	T3	1

Figura 3.2: Salida de R del diseño generado mediante Split-plot.

Inicialmente se muestra un resumen, donde se indica el diseño usado: “Split”, los parámetros por los que está formado el modelo, el método que se ha usado en este caso: “rcbd”, el número de réplicas o repeticiones del experimento, el formato que se usará para nombrar a las parcelas y subparcelas y la semilla.

### 3.1.2.2 Función `sp.plot`

Esta función lo que realiza es el análisis de la varianza (ANOVA) del diseño en cuestión, para ello hay que introducir los siguientes parámetros como en la Figura 3.3:

```
sp.plot (bloque , pplot, splot, Y )
```

Figura 3.3: Sentencia y parámetros para analizar un diseño Split-plot.

- `block`: las réplicas o bloques que tiene el modelo
- `pplot`: Tratamiento principal de la parcela
- `splot`: Tratamiento principal de la subparcela
- `Y`: La variable respuesta [15]

Se usa como ejemplo (Figura 3.4) el diseño generado en el apartado anterior, la variable respuesta (`var`) se ha generado con la orden de R `runif ()` de tal manera que:

```
sp.plot(block, TP, TS, var)
```

Figura 3.4: Sentencia y parámetros para analizar el diseño del apartado anterior

Donde el resultado obtenido se muestra en la Tabla 3.1:

Tabla 3.1: Tabla resumen Anova

	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>	<b>F value</b>	<b>Pr(&gt;F)</b>
<b>block</b>	2	1462.9	731.45		
<b>TP</b>	4	5908.9	1477.23	1.7454	0.2329
<b>Ea</b>	8	6770.9	846.37		
<b>TS</b>	3	2988.5	996.17	1.5234	0.2286
<b>TP:TS</b>	12	7406.9	617.24	0.9439	0.5191
<b>Eb</b>	30	19616.8	653.89		

### 3.1.3 librería agricolaeplotr

Esta librería, está implementada por la Universidad de Lima, y lo que permite es graficar el diseño split-plot. [16]

#### 3.1.3.1 plot\_split\_rcbd

Permite graficar aquellos diseños que hayan sido realizados mediante bloques completamente aleatorizados.

Los parámetros que se introducirán en dicha función, serán los que se muestran en la Figura 3.5:

```
plot_split_rcbd (design, nrows, ncols, factor_name_1 = "T1", factor_name_2  
= "T2", labels='plots', subplots = TRUE, width = 1, height = 1, space_width  
= 0.95, space_height = 0.85, reverse_y = FALSE, reverse_x = FALSE)
```

Figura 3.5: Sentencia y parámetros para graficar un diseño Split-plot

Donde:

- design: correspondiente al dataframe del diseño.
- nrows : Numero de filas que forman el diseño
- ncols: Número de columnas que forman el diseño
- factor\_name\_1: nombre del factor correspondiente a la parcela
- factor\_name\_2: nombre del factor correspondiente a la subparcela
- labels : etiquetas que aparecerán en el grafico
- subplots: TRUE, si se quiere que devuelva las subparcelas
- width : el ancho que mide cada parcela
- height : la altura que tiene cada parcela
- space\_width : Indica si hay espacio entre las parcelas. 0=solo espacio, 1=sin espacio entre parcelas en términos de ancho

- space\_height. 0=sólo espacio, 1=sin espacio entre parcelas en términos de altura
- reverse\_y: para trasponer el factor correspondiente al eje de las y
- reverse\_x: para trasponer el factor correspondiente al eje de las x

Se genera de nuevo un diseño experimental split-plot, similar al del apartado anterior, pero con una dimensión menor, para facilitar su comprensión. [17]

Se genera un diseño con 3 factores para las parcelas y dos a la subparcela. Este planteamiento sobre el terreno se puede ver en la Figura 3.6:

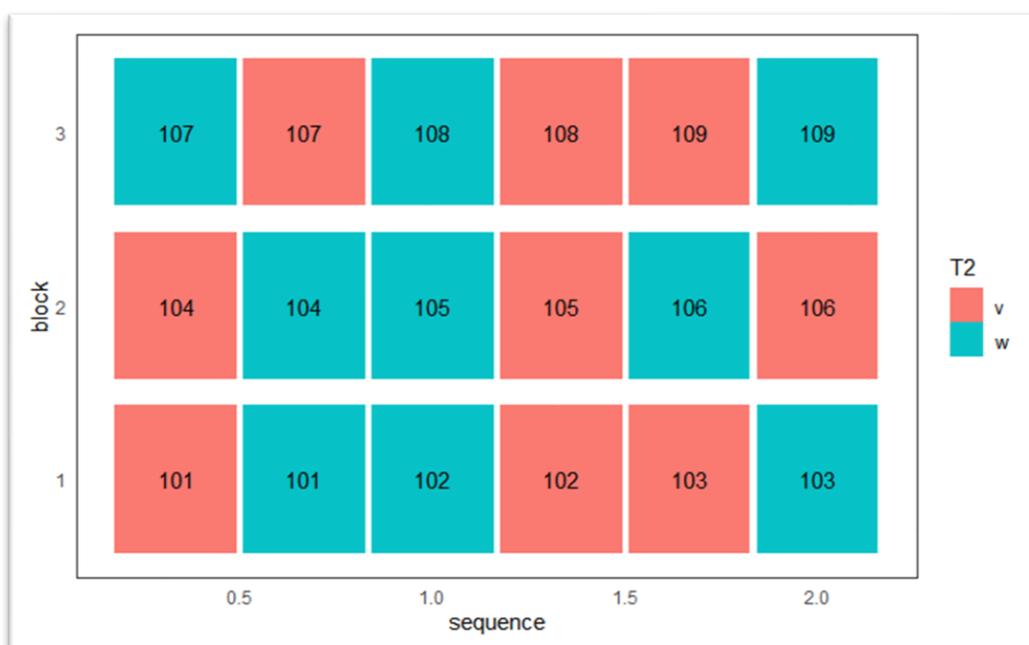


Figura 3.6: Planteamiento del diseño sobre el terreno.

Donde v y W son los niveles del factor de las subparcelas.

### 3.1.4 Ajuste de los datos

Los modelos se ajustan a partir de un conjunto de datos. La finalidad es encontrar el mejor modelo para los datos en cuestión.

Los diseños split-plot suelen ajustarse de manera correcta mediante modelos mixtos debido a la estructura de estos diseños, ya que el “factor parcela” es de efectos fijos mientras el “factor subparcela” de efectos aleatorios.

Para ajustar estos datos mediante R una de las mejores librerías será “lme4” ya que está bastante completa. [18]

Esta librería contiene las siguientes funciones para ajustar y analizar los modelos mixtos:

- Lineales
- Lineales generalizados
- No lineales

#### 3.1.4.1 Función glmer ()

A continuación, se explica la función que ajusta los modelos generalizados mixtos ya que se usará posteriormente.

La orden que se utilizará en R se muestra en la Figura 3.7.

```
glmer (formula, data = NULL, family = gaussian)
```

Figura 3.7: Sentencia y parámetros para la orden glmer

Siendo los parámetros principales los siguientes:

- Formula: Es la ecuación del modelo estadístico, donde se indican tanto los efectos fijos como los aleatorios. La variable respuesta se separa del resto de variables mediante el símbolo “~”, y el resto se separa mediante el símbolo “+”, los efectos aleatorios se indicarán mediante “|”, que separa aquellas

expresiones que se usan en las matrices de diseño para los factores de agrupación.

- Data: el conjunto de datos donde están las variables que se utilizan en el parámetro fórmula.
- Family: indica a que tipo de familia de distribución pertenecen los datos, estos pueden pertenecer a la familia: binomial, gaussiana, gamma, gaussiana inversa, poisson, quasi, quasibinomial, quasipoisson.

Posteriormente en el apartado de Aplicaciones se mostrará un ajuste de un conjunto de datos obtenidos mediante un diseño split-plot y analizados con esta función. [19]

## 3.2 SAS

Mediante el uso del programa SAS [20] con licencia aportada por la Universidad de Valladolid. Se explica:

- Como crear un diseño experimental split-plot
- Como analizar el diseño split plot.

### 3.2.1 Creación de un diseño experimental split-plot

Gracias a esta herramienta, es posible realizar un diseño experimental de distintas formas:

- A través de código en el Editor del programa.
- Usando las Herramienta de creación de diseño de experimentos que posee el programa: Adx Interface for Design of Experiments. [20]

Para realizar el diseño a través del editor, se hace de la siguiente manera:

A continuación, se describe el diseño usado como ejemplo:

- 3 tratamientos en cada parcela, a los cuales corresponde la variable TP en código SAS.
- 3 tratamientos en relación a cada subparcela, indicados por la letra TS.
- el número de bloques serán 3, es decir las repeticiones del experimento, a los cuales corresponde la letra B.

```
data diseno_split_plot;
  do TS = 1 to 3;
    do B = 1 to 3;
      do TP = 1 to 3;
        input y @@;
        output; end; input; end; end;
      cards;
17.332361 90.605582 15.778959
50.772665 65.850389 45.092746
42.300356 35.991771 81.297283
83.118172 19.447388 78.907000
18.339572 83.899496 62.995186
67.062971 61.382015 25.545740
69.937083 27.814339 74.806531
52.637881 68.039313 11.882574
26.372541 89.126446 91.472531
;
```

Figura 3.8: Creación de un diseño split-plot con el programa SAS

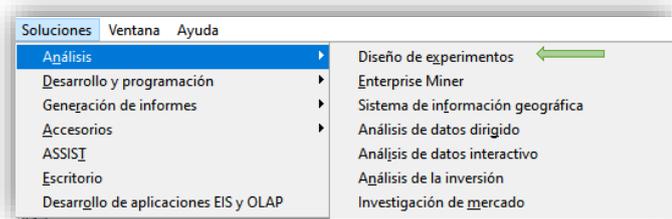
La segunda forma para realizar un diseño de experimentos split-plot es la que se muestra desde la Figura 3.9 [17]

PASO 1

PASO 2

Pinchar sobre las siguientes pestañas:  
Soluciones- Análisis- Diseño de experimentos

Seleccionando sobre el siguiente icono se puede empezar a desarrollar el diseño.



PASO 3

PASO 4

Se indican los niveles del factor de la parcela

Se indican los niveles del factor de la subparcela

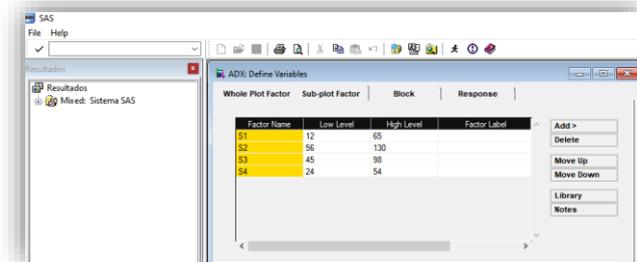
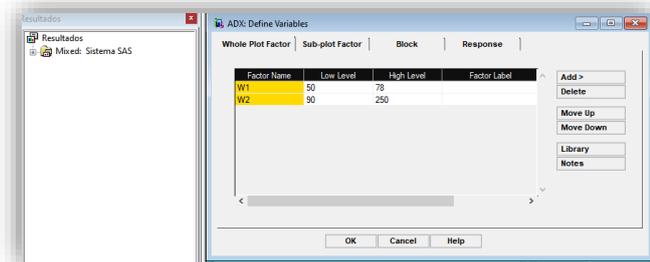


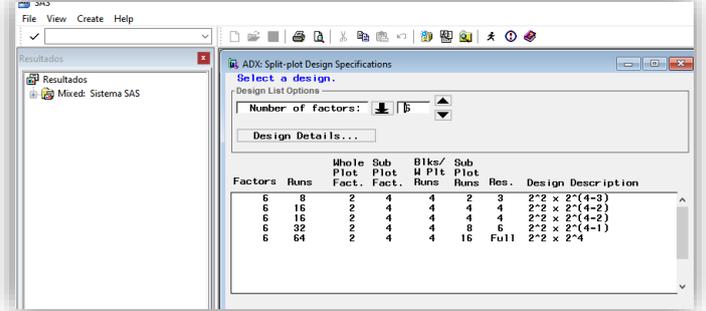
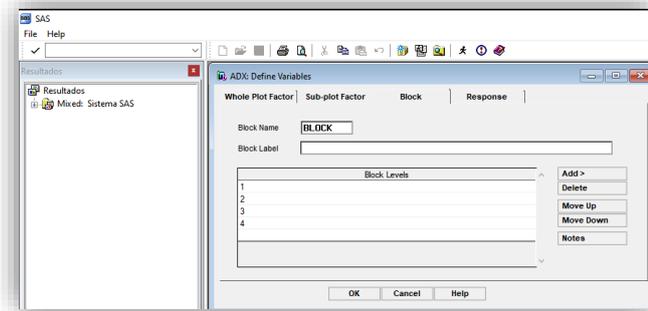
Figura 3.9.1: Pasos para la creación del diseño

PASO 5

PASO 6

Se indican en número de bloques

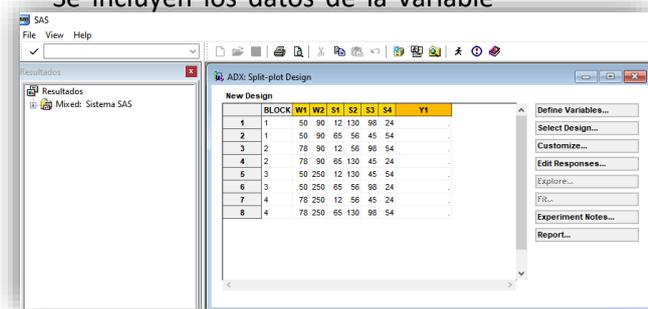
Se indica el número de réplicas.



PASO 7

PASO 8

Se incluyen los datos de la variable



Tras obtener el diseño estos datos se analizar de manera predefinida mediante un modelo mixto.

Figura 3.9.2: Pasos para la creación del diseño

### 3.2.2 Análisis

Al igual que con el software R, con SAS también se podrá realizar el ajuste y el análisis para los datos recogidos a través de un diseño split-plot.

Como ejemplo se usará el generado en el apartado 3.2.1

Para ajustar los datos se hará de la siguiente manera (Figura 3.10) en función del modelo utilizado, y se usará procedimiento **proc mixed**:

```
proc mixed;  
  class TP TS B;  
  model Y = TP TS TP*TS;  
  random B TP*B;  
run;
```

Figura 3.10: Procedimiento de SAS para analizar el diseño.

La salida obtenida es la que se muestra entre las Figura 3.11 y 3.12

Información del nivel de clase		
Clase	Niveles	Valores
TP	3	1 2 3
TS	3	1 2 3
B	3	1 2 3

Dimensiones	
Parámetros de covarianza	3
Columnas en X	16
Columnas en Z	12
Asuntos	1
Obs máx por Tema	27

Número de observaciones	
Número de observaciones leídas	27
Número de observaciones usadas	27
Número de observaciones no usadas	0

Figura 3.11: Descripción del conjunto de datos

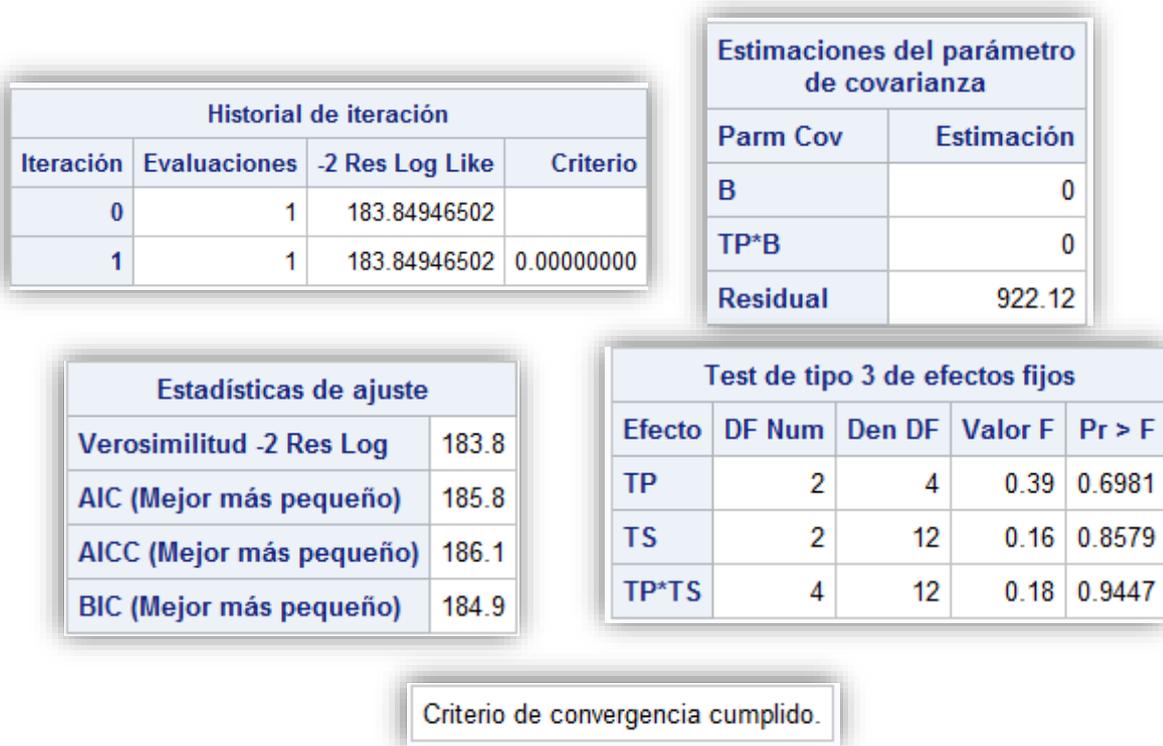


Figura 3.12: Salida análisis split-plot

La tabla historial de iteraciones muestra los pasos del proceso de optimización, en concreto el método que usa es el algoritmo de Newton-Raphson, donde este usa las derivadas primera y segunda de la función objetivo para así poder encontrar mediante iteraciones el mínimo. La siguiente tabla muestra la estimación de los parámetros de la covarianza. A continuación, se muestra la tabla con los estadísticos de ajuste, donde están los diferentes criterios de comparación de modelos. Por último, la tabla donde se muestran los efectos fijos probados mediante funciones estimables de Tipo 3.

## **4. Aplicación**

En esta sección se detallan los métodos y resultados obtenidos de una aplicación real sobre la regeneración forestal del “Pinus Pinaster” en Castilla y León, que ilustra el uso de los diseños experimentales Split-plot. [21],

### **4.1 Regeneración forestal**

#### **4.1.1 Introducción**

La regeneración forestal se define como la capacidad que tienen los árboles para producir semillas, y en consecuencia a la aparición de nuevos árboles, permitiendo que la especie se perpetúe. La regeneración puede ser tanto natural, cuando las semillas se esparcen sin la intervención experimental, como artificial, cuando las semillas se esparcen de manera manual.

Es posible estudiar la regeneración forestal de un área de muchas maneras posibles, las más común es a través de la lluvia de semillas. Se define como lluvia de semillas a aquellas semillas que caen desde el árbol al suelo.

En la aplicación que se presenta en este trabajo, la regeneración fue natural. Puesto que se trata de un proceso natural, se deben de tener en cuenta diferentes condiciones naturales que afectan a los resultados, tales como la climatología o los depredadores entre otros.

## 4.1.2 Estudio de la regeneración natural de “*Pinus Pinaster*” en arenales del centro de España

### 4.1.2.1 Situación del experimento

En el año 2004 se instaló el sitio experimental de Cuéllar (Segovia, España) para el proyecto AGL2001-1780, que pertenece al Instituto Universitario Forestal (IuFor) de la Universidad de Valladolid en Palencia. El lugar concreto donde se realizó el experimento es en el Monte N<sup>o</sup> 32 “COMUN DE LA TORRE Y JARAMIELA” (Cuéllar).

El experimento está situado en la Meseta Castellana donde el clima es bastante seco debido a las pocas precipitaciones y un periodo de heladas que dura al menos 3 meses, y donde el terreno está formado por arenas silíceas. En esta zona abundan 3 tipos distintos de pinos “*Pinus Pinaster*”, “*Pinus pinea*” y “*Quercus ilex*”. o “*Juniperus communis*”.

### 4.1.2.2 Datos

En este estudio se midieron y recogieron diferentes variables climáticas que afectaban a la dispersión de las semillas. Para realizar una primera selección de las variables, se calcularon los coeficientes de correlación de Pearson entre cada una de las variables climáticas y la variable respuesta: Número de semillas recogidas en cada una de las trampas. De esta manera se pudieron obtener 17 variables, fueron recogidas durante 10 años, 2004-2013, ambos inclusive. [22]

Estas variables como se puede ver en la tabla 4.1 se dividieron en 5 etapas que formaron parte del proceso fisiológico clave para la maduración de los conos (piñas) los cuales fueron:

- **Crecimiento primario:** Se dio 2 años antes de la dispersión. En este periodo, se desarrolló el estado embrionario de la piña, siendo clave para su desarrollo.
- **Crecimiento secundario:** Se dio 1 año antes de la dispersión. Es el periodo de engorde de la piña. En el que se formó el piñón propiamente dicho.
- **Maduración final de cono:** Ocurrió un año antes de la dispersión. Este fue un momento clave para el desarrollo de la semilla. Cualquier variación pudo alterar la viabilidad del piñón.
- **Floración:** Esta ocurrió 2 años después de la dispersión de la semilla. Momento en el que los conos floríferos masculinos y femeninos emergieron y se produjo la fecundación de los óvulos.
- **Liberación de semillas:** Se dio en el momento de la apertura de la piña donde los piñones cayeron al suelo para, en el momento de humedad y temperatura adecuados, poder germinar.

Tabla 4.1: División de las variables en el proceso de maduración.

Etapa	Variables climáticas <sup>3</sup>
Crecimiento primario	1. Temperatura mínima en octubre. 2. Índice de sequía de Walter´s en verano <sup>4</sup> . 3. Número de días en los que heló en octubre
Crecimiento Secundario	4. Precipitación total en el mes de abril. 5. Número de días en los que heló en el mes de marzo. 6. Número de días en los que heló en el mes de marzo de forma severa <sup>5</sup> .
Maduración final del cono	7. Precipitación total en el mes de octubre y noviembre.
Floración	8. Precipitación total en abril y mayo 9. Número de días que hubo precipitaciones en abril y mayo 10. Velocidad máxima del viento en el mes de mayo 11. Temperatura media en el mes de abril 12. Temperatura máxima en el mes de abril 13. Número de días en los que heló en el mes de mayo <sup>6</sup> .
Liberación de semillas	14. Número de días en los que hubo precipitaciones en mayo 15. Velocidad máxima del viento en verano 16. Temperatura media en agosto 17. Número de días en los que heló en abril

<sup>3</sup> Fuente: Estación meteorológica cercana al sitio, datos aportados por el Instituto Tecnológico agrario de Castilla y León.

<sup>4</sup>  $(\text{temperaturas medias diarias de junio a agosto}) - (0,333 \times (\text{media lluvia diaria de junio a agosto}))$ . Los valores más altos indican una sequía más intensa.

<sup>5</sup> los días de heladas severas se definen como aquellos días con una temperatura mínima inferior a  $-5\text{ }^{\circ}\text{C}$

<sup>6</sup> los días de helada se definen como aquellos días con una temperatura mínima inferior a  $0\text{ }^{\circ}\text{C}$ .

Finalmente, de la totalidad de las variables climáticas las más relevantes fueron las presentadas en la Tabla 4.2:

Tabla 4.2: Variables más relevantes.

Etapa	Nombre de la variable	Acrónimo
Crecimiento primario	Temperatura mínima en octubre.	$T_{minOCT}$
Crecimiento secundario	Precipitación total en el mes de abril.	$p_{APR}$
Maduración final del cono	Precipitación total en el mes de octubre y noviembre.	$p_{OCTNOV}$

También se analizó como dice Ruano, I. (2016). “el efecto de la densidad de la masa se estimó mediante un índice espacial calculado como la suma del inverso de las distancias entre la trampa y cada árbol potencialmente fuente de semilla, con un límite de 20 metros” (p.71) como se puede observar en la Figura 4.1. Para realizar esta medición solo se usaron los árboles en los que se contaban las semillas. A esta variable se la identificó como “fuente de semilla”.

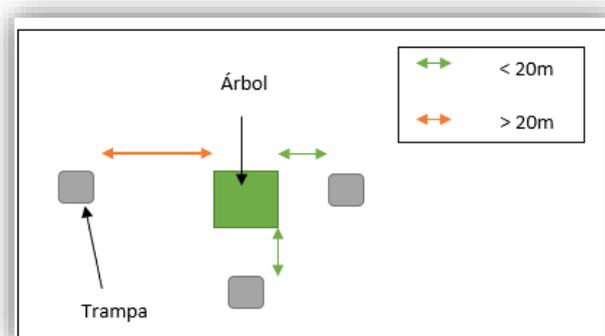


Figura 4.1: Mediciones del árbol a las trampas

Tal y como dice Ruano, I. (2016):

La variable respuesta, la cual era el número de semillas recogidas en cada año en las trampas sigue una distribución de Poisson ya que esta variable representa un conteo, en este caso de semillas. No se asumió la independencia de los datos ya que los mismos fueron recogidos en un diseño de parcelas, las cuales siempre han sido las mismas y por lo tanto se supo que las observaciones que pertenecían a la misma parcela, misma trampa y mismo año, no lo iban a ser. (p.59)

Donde se denotaron las parcelas, las trampas y los años con los índices  $i, j, k$ .

El número de parcelas que se tuvo fueron 10 por lo tanto, se denoto  $i$  como  $i=1\dots 10$ .

Respecto al número de trampas estas fueron 9 por lo tanto la variable  $j$  figura entre los índices  $j=1\dots 9$  En relación al número de años que duró el experimento estos fueron 10 por lo tanto la variable  $k$  fue entre los índices  $k= 1\dots 10$ . [22]

$$Y_{ijk} = Poisson(\lambda_{ijk})$$

Donde  $\lambda_{ijk}$ :

$$\lambda_{ijk} = exp^{X_{ijk}*\beta}$$

Se tuvo  $X_{ijk}$  como un vector fila de covariables que podrían afectar a la dispersión de las semillas y  $\beta$  es el vector columna del parámetro.

### 4.1.2.3 Experimentación mediante Split-plot

El terreno estaba dividido de manera artificial en 10 parcelas, con una superficie total de 4.900m<sup>2</sup>, correspondiendo cada parcela a un área cuadrada de 70 m de lado.

Se utilizó un diseño experimental Split-plot para asignar uno de 4 tratamientos posibles a cada parcela. El tratamiento que se asignó estaba relacionado con la intensidad de corta, considerando los niveles:

- 100%: Se cortó toda el área basimétrica de la parcela.
- 50%: Se cortó el 50% del área basimétrica.
- 25%: Se cortó el 25% del área basimétrica.
- 0%: No se cortó nada.

Siendo el área basimétrica o G, el parámetro expresado en m<sup>2</sup>/ha de la sección de los árboles medidos a 1,30 m de altura en una zona forestal. La cual se calculará de la siguiente manera:

$$G = \frac{\pi}{4} \sum_{i=1}^n (d_i^2)$$

Donde i es el número de árboles en el área, y d es el diámetro de cada uno de los árboles que la forman.

La aplicación de la corta se hizo según los siguientes criterios:

Se extrajeron de la masa los árboles que tenían las siguientes características:

- Aquellos que hacían competencia a árboles con copa bien conformada,
- Los que estaban enfermos,
- Aquellos que no estaban rectos,
- También árboles con heridas o resquebrajamientos, caídos o tumbados.

Una vez extraídos estos, si hizo falta se eliminaron aquellos que:

- más competían entre ellos, favoreciendo a los que se quedan para poder conformar una buena copa para producir semilla.

Todas las cortas se repitieron 3 veces. Excepto la corta al 0% que únicamente se aplicó sobre la parcela número 10, ya que ésta era la parcela control.

La información completa de la asignación de las parcelas puede consultarse en la Tabla 4.3

Tabla 4.3: Clasificación de las parcelas según el % de corta.

Intensidad de corta	Parcelas afectadas
100%	1,3,9
50%	6,7,8
25%	2,4,5
0% (Control)	10

La distribución de las parcelas se esquematiza en la Figura 4.2.

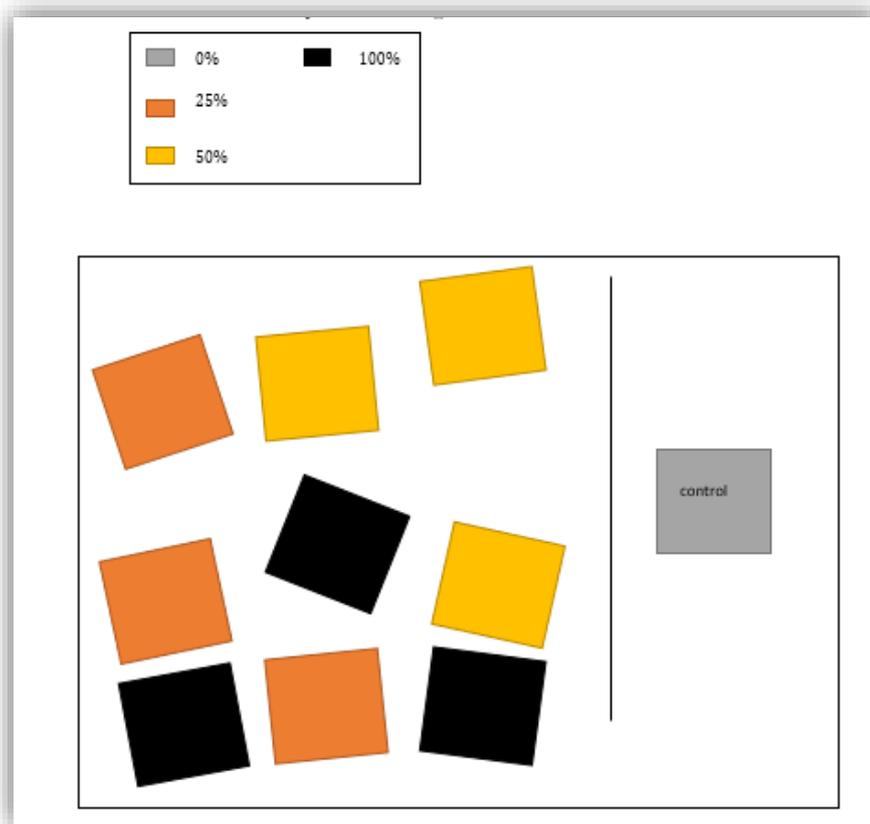


Figura 4.2: Esquema de la distribución de parcelas según la intensidad de corta.

En cada una de las parcelas se instalaron 9 trampas para la recolección posterior de las semillas usando un replanteo sistemático.

Siendo el replanteo sistemático: el trazado o marcado en el terreno de la disposición de las trampas de semillas que se utilizaron en el experimento. Es decir, se denomina como manera en la que se distribuyeron las trampas en el terreno. Más adelante se representará un esquema con la situación de éstas en cada parcela.

El replanteo fue el siguiente: Se dejó la distancia de seguridad de 10 m del margen de la parcela, para que se asegurase que la recogida de la semilla fuese de la misma parcela y no del resto del monte y así se evitó la recogida de semillas que no pertenecían a la parcela en cuestión.

Como se puede observar en la Figura 4.3 dentro de un área cuadrado de 50m de lado se hará la distribución de las trampas (replanteo sistemático) donde se colocará una trampa cada 25m.

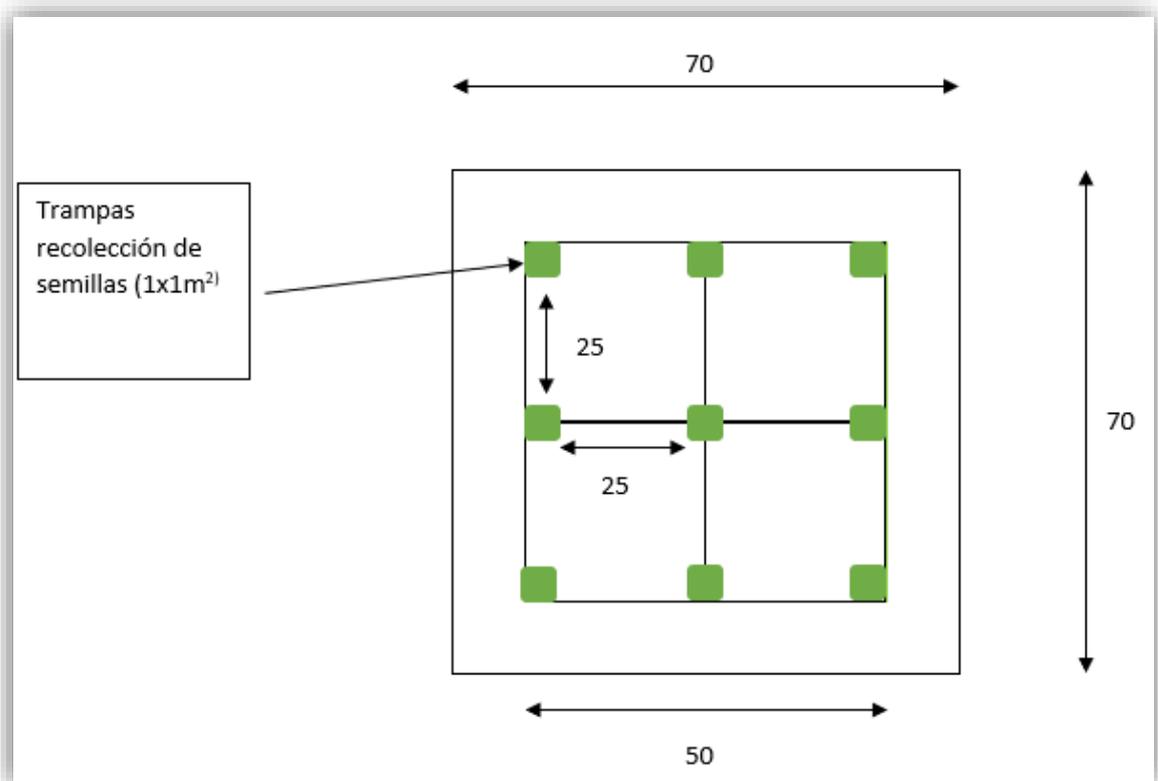


Figura 4.3: División sistemática de cada una de las parcelas.

### 4.1.2.3.1 Análisis descriptivo

A continuación, se realizará un análisis descriptivo de los datos, para realizar las gráficas se usará la librería ggplot2. [23]

En la Figura 4.4 se observa la densidad de semillas recogidas en cada parcela en función de los 4 niveles de corta. Se puede observar que como cabía esperar, es la parcela 10 en la que no se realiza corta, en la que se recoge más semillas, seguida de las parcelas 6 y 7, las cuales comparten un 50% de nivel de corta.

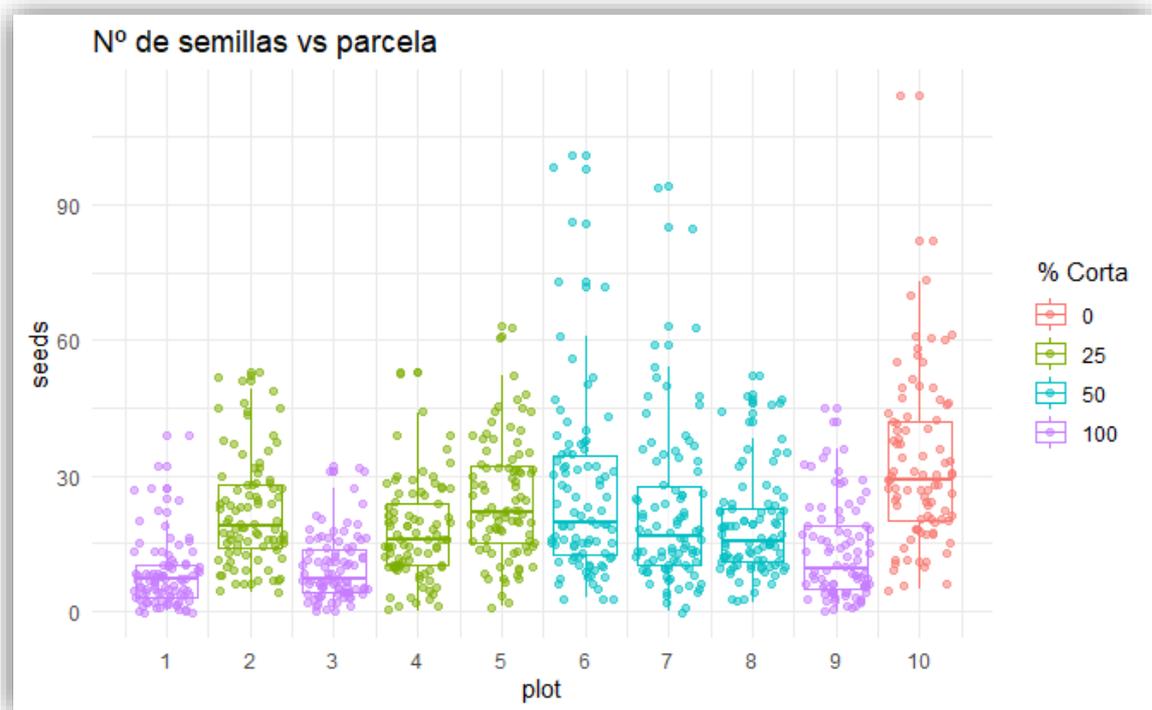


Figura 4.4: Numero de semillas recogida en cada parcela indicando el nivel de corta y box-plot.

En la Figura 4.5 se representa en número de semillas recogidas en cada parcela, por cada año en el que realizó la recogida de muestra. Se puede observar que en los primeros años se recogió menos semillas que los años correspondientes al final del estudio.

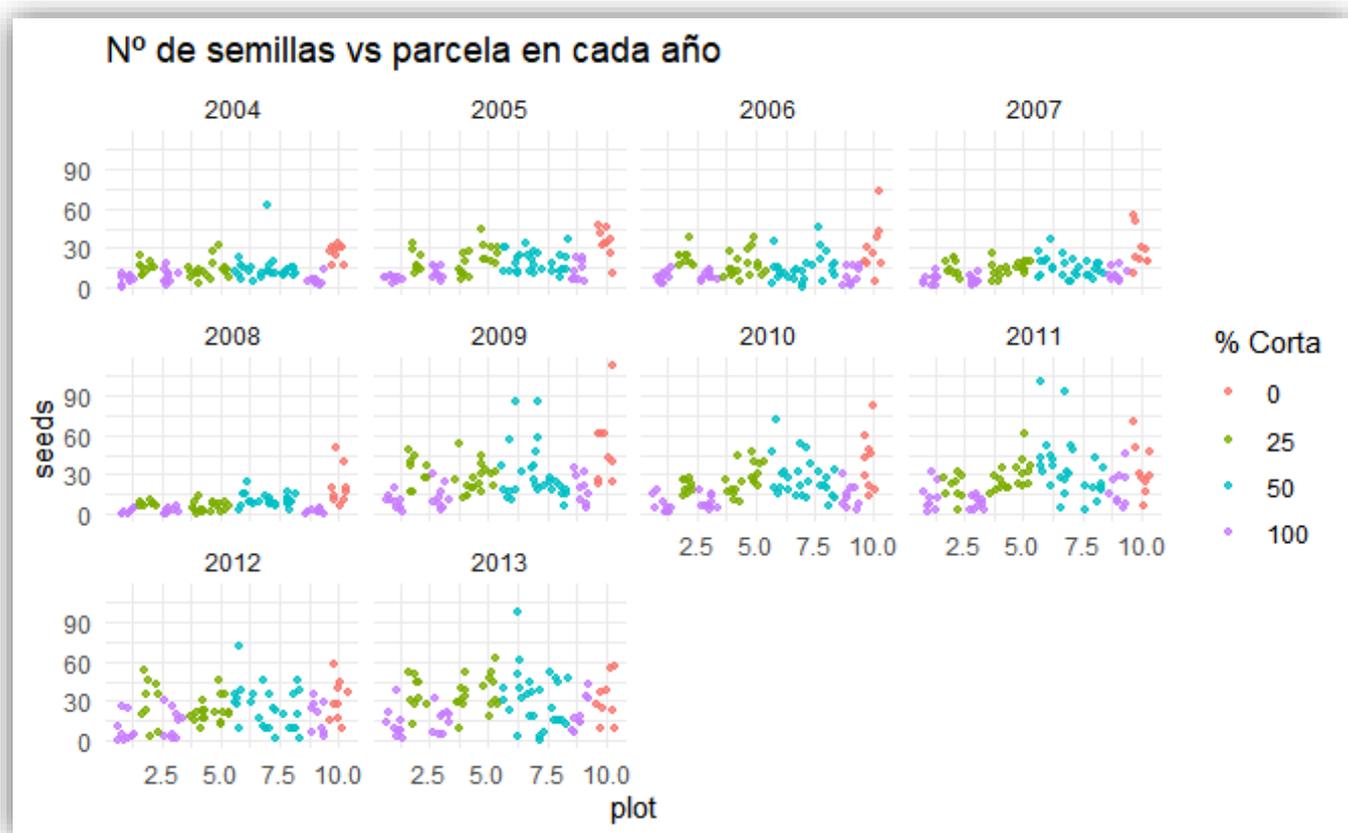


Figura 4.5: Numero de semillas recogida en cada parcela en cada año, indicando el nivel de corta.

### 4.1.2.3.2 Ajuste del modelo

Tras la obtención de las posibles variables que podrían formar parte del ajuste del modelo, se buscó el mejor modelo para realizar el ajuste de estos datos para así poder ver como afectaban las variables climáticas a las fases de la maduración de las piñas, ya que ahí es donde están situadas las semillas.

A la hora de buscar que variables eran aquellas que ajustaban mejor, este proceso se dividió en 3 pasos:

- Primer paso:

Inicialmente se ajustaron y compararon distintos modelos preliminares para cada una de las variables explicativas y así se pudo ver como afectaban estas a la variable respuesta y así poder ver cuáles de ellas pasarían a formar parte del ajuste.

A cada una de las variables climáticas candidatas se les aplicaron las siguientes transformaciones:

$$x^2, x+x^2, \log x, x+\log x, \sqrt{x}, x+\sqrt{x}$$

También se reescalaron de estas variables para evitar tener problemas numéricos.

Ya que las variables climáticas generalmente están correlacionadas, estas no se probaron todas al mismo tiempo.

Para poder observar que variables climáticas o en el caso de que necesitasen transformaciones en ellas se evaluaron mediante gráficos de residuos estudentizados.

- Segundo paso:

Las distintas variables climáticas y sus transformaciones se fueron incluyeron secuencialmente. A continuación, se incluyeron como efectos aleatorios las parcelas, las trampas.

Para la elección del mejor modelo se utilizó el criterio de Akaike y todos los posibles modelos se evaluaron con una significación estadística de  $\alpha=0.05$

El modelo genérico resultante tras incluir los efectos aleatorios donde  $U_i \sim N(0, \sigma^2_{parcela})$  hizo referencia a la parcela y  $U_{ij} \sim N(0, \sigma^2_{trampa})$  a las trampas:

$$\lambda_{ijk} = \exp^{X_{ijk} * \beta + U_i + U_{ij}}$$

Los efectos fijos seleccionados fueron precipitación total de abril, precipitación de diciembre y noviembre y la precipitación mínima de octubre.

- Tercer paso:

Una vez se esclarecieron efectos fijos y aleatorios que formarían parte del modelo. Se realizaron dos ajustes con funciones distintas del programa R usando la librería “lme4”<sup>[23]</sup> esta sirve para ajustar modelos lineales de efectos mixtos, el primero fue mediante la función “glm” (modelo lineal generalizado) en este ajuste no estuvieron presentes los efectos aleatorios y la segunda fue mediante la función “glmer” (modelo lineal mixtos generalizado), mientras que en este ajuste estuvieron presentes los efectos aleatorios.

Siendo el de menor AIC el segundo, el cual era de 7128.3. El AIC del primer ajuste fue 9278.7.

El modelo que se presento fue el mostrado a continuación ya que es preciso y su sesgo y error cuadrático medio son pequeños.

El modelo ajustado tiene la expresión<sup>7</sup>:

$$E[Y_{ijk}] = \exp^{\widehat{\beta}_0 + \widehat{\beta}_1 * SSI + \widehat{\beta}_2 * \log(SSI + 1) + \widehat{\beta}_3 * \frac{T_{minOCT}}{10} + \widehat{\beta}_4 * \frac{p_{APR}}{100} + \widehat{\beta}_5 * \frac{p_{OCTNOV}}{100} + \widehat{\beta}_6 * \log\left(\frac{p_{OCTNOV}}{100}\right) + u_i + u_{ij}}$$

Tal y como se muestra en Tabla 4.4 y Tabla 4.5 el modelo ajustado.

---

<sup>7</sup>SSI: índice de fuente de semilla; T\_inOCT: Temperatura mínima en octubre; p\_APR: Precipitación total en el mes de abril; p\_OCTNOV: Precipitación total en el mes de octubre y noviembre.

Tabla 4.4: Resultado de los Efectos aleatorios.

	<b>Random effects:</b>	
<b>Groups Name</b>	<b>Variance</b>	<b>Std.Dev</b>
Trampa (Intercept)	0.12484	0.3533
Parcela (Intercept)	0.01389	0.1179

Tabla 4.5: Resultado de los Efectos fijos

		<b>Fixed effects:</b>		
	<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>Pr(&gt; z )</b>
(Intercept)	2.59140	0.13790	18.792	< 2e-16 ***
SSI	-1.61155	0.38415	-4.195	2.73e-05 ***
Log (SSI +1)	4.30480	0.73044	5.893	3.78e-09 ***
P_OCTNOV	-0.83312	0.03648	-22.841	< 2e-16 ***
Log(P_OCTNOV)	0.40528	0.02422	16.733	< 2e-16 ***
T_minOCT	-0.42195	0.07061	-5.976	2.29e-09 ***
P_APR	0.69201	0.04609	15.013	< 2e-16 ***

Tras calcular los residuos del modelo, se pudieron ver los siguientes gráficos de diagnóstico en la Figura 4.6.

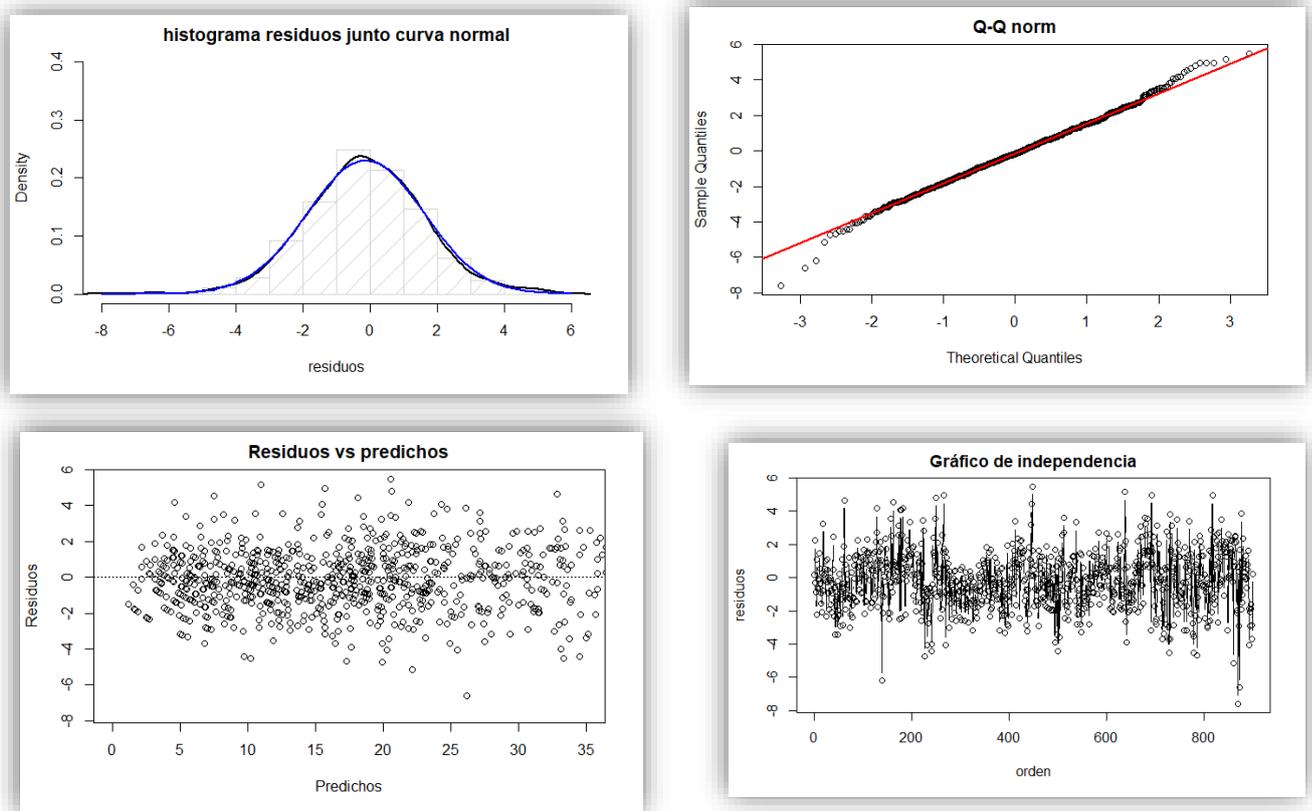


Figura 4.6: Distintos gráficos de diagnóstico.

Donde se puede observar el histograma de los residuos junto con la curva Normal, se puede observar que hay algunos puntos que no se ajustan a la recta, por ello se realiza el test de “Cramer-von Mises”, con el cual se obtuvo un p-valor de 0.3369, por lo tanto, no hay evidencias significativas para rechazar la normalidad. Respecto a la hipótesis de homocedasticidad, se observa que los residuos se distribuyen de forma aleatoria, además en test de Levene se obtuvo un p-valor de 0.20, por lo tanto, no hay evidencias significativas para rechazar la hipótesis nula. En la última gráfica no se observa ningún patrón, por lo tanto, por lo tanto, no hay evidencias significativas para rechazar la independencia.

## 5. Discusión

En consideración con el diseño de experimentos visto de manera global, se pueden destacar distintas ventajas desarrolladas a continuación: El análisis de estos diseños podría ser realizado por etapas y esto aportaría datos útiles para combinaciones posteriores. Así mismo, el experimento podría optimizarse reduciendo el número de pruebas a realizar, lo cual repercutiría en menor coste económico.

En relación a las desventajas, la investigación previa a realizar, supone una alta inversión tanto en tiempo, como en financiación económica, resultando por ello un tipo de experimentos muy costosos. Otro gran problema que observo es la incorrecta manipulación de variables, lo que puede llevar a situaciones imprecisas donde el modelo tenga errores tanto en el diseño, como en la estimación.

En relación al uso de los diseños experimentales split-plot, en este caso enfocado para ámbito forestal.

El diseño de experimentos Split-plot es recomendable cuando, nos encontramos ante al menos un factor cuya aleatorización no es posible completamente. Además, son aconsejables, cuando cada factor se aplica a terrenos de diferente superficie, como es en el caso práctico expuesto anteriormente, donde las parcelas a las que se aplica el “factor corta” tienen una dimensión de un cuadrado de 70m de lado, mientras que las trampas donde se recogen las semillas tienen una dimensión de un cuadrado de 1m de lado.

Con este tipo de diseño se llega a obtener una muy buena precisión en relación a la estimación de los factores, lo que se traduce en una gran eficiencia estadística y precisión. Respecto a las desventajas en estos experimentos, recalco la pérdida de precisión por parte del “factor parcela” y que el análisis estadístico puede llegar a ser complicado debido a que las diferentes comparaciones tendrán diferentes errores de variación.

## 6. Conclusiones

El diseño de experimentos Split-plot es un diseño mayoritariamente usado en el entorno forestal y agrícola ya que permite adaptar fácilmente los distintos factores al terreno, donde en la mayoría de los casos en este ámbito habrá un factor de difícil variación el cual estará situado en las parcelas y otros factores que se permitirán que se puedan varían de manera más sencillas entre las parcelas, pasando a formar parte estos factores de las subparcelas.

Se ha visto que este tipo de diseños se puede tanto realizar como analizar sus resultados mediante distintos softwares estadísticos: R y SAS Con el programa SAS se podrán usar dos tipos de procedimientos PROC GLM y PROC MIXED para analizar estos diseños, también se podrán realizar los diseños. El programa R nos permitirá usar distintas librerías para poder diseñar este tipo de experimentos y también poder analizarlos posteriormente con las distintas funciones que aporta la librería agricolae.

## 7. Bibliografía

- [1] Montgomery, Douglas C. (2005). *Diseño y análisis de experimentos*. 2a ed. México: Limusa - Wiley.
- [2] Cochran, William G., and Gertrude M. Cox. (1987). *Diseños experimentales* [trad. Alfonso Carrillo Liz...(et al.)]. 1ª ed., 10ª reimpr. México [etc.: Trillas, Print.
- [3] *¿Qué es aleatorización?* - Minitab. (2021). (C) Minitab, LLC.
- [4] *Réplicas y repeticiones en experimentos diseñados* - Minitab. (2021). (C) Minitab, LLC.
- [5] Z. (2020,). *What is a Split-Plot Design? (Explanation & Example)*. Statology.
- [6] *14.3 - The Split-Plot Designs / STAT 503*. (s. f.). PennState: Statistics Online Courses. Universidad de Pensilvania.
- [7] REAL ACADEMIA ESPAÑOLA: Diccionario de la lengua española, 23.<sup>a</sup> ed., [versión 23.5 en línea].
- [8] Lawal, Bayo. (2014). *Applied Statistical Methods in Agriculture, Health and Life Sciences by Bayo Lawal*. 1st ed. Cham: Springer International Publishing, 2014. Web.
- [9] Rajender Parsad\*, V K Gupta, P K Batra, R Srivastava, Rajinder Kaur, Ajit Kaur Bhatia, Praveen Arya, ICAR-Indian Agricultural Statistics Research Institute (2003). A diagnostic study of design and analysis of field experiments
- [10] Escobar, R. F., Casas, A. T., Giménez, J. D., Pesca, A. D. C. A. Y., & Andalucía. Consejería de Agricultura y Pesca. (2010). *Experimentación en agricultura*. Consejería de Agricultura y Pesca, Servicio de Publicaciones y Divulgación.

- [11] Diseño de Parcela Divididas. (s.f) Facultad de Ciencias Agrarias y Forestales, Universidad nacional de la Plata.
- [12] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- [13] Mendiburu. F (2021). agricolae: Statistical Procedures for Agricultural Research. R package version 1.3-5.
- [14] Mendiburu, F. (2014). Split Plot Design. R package version 1.3-5.
- [15] Mendiburu, F. (2021). Split-Plot analysis. R package version 1.3-5.
- [16] Harbers, J (2021), agricolaeplotr: a package for visualization of design of experiments. R package version 0.10
- [17] Harbers, J (2021) (2014). *Plot Split Plot Designs with rcbd*.
- [18] Bates. D, Maechler. M, Bolker. B, Walker.S (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- [19] *glmer function - RDocumentation*. (s. f.). R-Documentation.
- [20] SAS Institute & SAS Institute. (2004). Getting Started with the SAS 9.1 ADX Interface for Design of Experiments. SAS Institute.
- [21] Ruano, I., Manso, R., Fortin, M., & Bravo, F. (2015). Extreme climate conditions limit seed availability to successfully attain natural regeneration of *Pinus pinaster* in sandy areas of central Spain. *Canadian Journal of Forest Research*, 45(12), 1795-1802.

- [22] Benito, I. R. (2016, 9 noviembre). Factores claves para la gestión de la regeneración natural en pinares mediterráneos.
- [23] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag NewYork, 2016.
- [24] Íñiguez Lobeto, Carlos, and Manuel José Gayoso Rodríguez. (2017) *Guía teórico-práctica para elaborar tus publicaciones científicas: (artículos, revisiones, tesis doctorales, paneles, comunicaciones, etc.)* / Carlos M. Íñiguez Lobeto, Manuel J. Gayoso Rodríguez. Valladolid: Ediciones Universidad de Valladolid, Print.