



---

**Universidad de Valladolid**

Facultad de Ciencias

**TRABAJO FIN DE GRADO**

Grado en Estadística

**Estudio de la influencia de la densidad de  
siembra sobre el rendimiento de forraje  
en cultivos asociados de cebada y  
guisante**

**Autor:**

Elvira Delgado de Paz





---

**Universidad de Valladolid**

Facultad de Ciencias

**TRABAJO FIN DE GRADO**

Grado en Estadística

**Estudio de la influencia de la densidad de  
siembra sobre el rendimiento de forraje  
en cultivos asociados de cebada y  
guisante**

**Autor:**

Elvira Delgado de Paz

**Tutores:**

Abel Barrios Casado

Agustín Mayo Íscar



# Agradecimientos

Este trabajo se ha realizado gracias al proyecto “Nuevas estrategias en la diversificación de cultivos: cultivos intercalados para la producción sostenible de forrajes de calidad” (FEADER, PDR CyL 2014-2020, submedida 16.2).

Quiero agradecer a todos los profesores que me han formado, ya sea en la Universidad como en el Colegio e Instituto, por ayudarme a desarrollar habilidades que ni yo sabía que poseía.

Gracias Abel por descubrirme que el mundo de los “bichos” y de las “plantitas” podía ser tan interesante y por dejarme conocer la gran persona que eres. Gracias al ITACyL, por acogerme durante dos grandes meses y por los datos proporcionados para este trabajo.

Gracias a todas las personas que he conocido a lo largo de estos años en la universidad y fuera de ella y que me han enseñado y ayudado a ser quien soy ahora mismo.

Gracias a mi familia, de los que aprendo cada día y de los que seguiré aprendiendo. Por último, gracias a todas las mujeres de mi vida, luchadoras incansables



## Resumen

El cultivo de enormes extensiones de terreno con únicamente un cultivo -agricultura de monocultivo- es un invento relativamente reciente de la agricultura moderna. Los cultivos mixtos o cultivos asociados son la práctica de cultivar dos o más cultivos en el mismo terreno, además de ser la forma más antigua de producción agrícola sistematizada. No obstante, los cultivos asociados se han ido reemplazando por los monocultivos, especialmente en los países más desarrollados. La competencia entre especies que proporcionan los cultivos mixtos ha sido muy estudiada dentro del mundo de la ecología. Asimismo, existe un gran interés dentro del ámbito de la investigación de las posibles ventajas o inconvenientes que pueden ofrecer los cultivos intercalados. Esto es debido a que requieren menos aportes de nutrientes, control de plagas, poda... que los monocultivos y, como resultado, suelen ser más rentables.

Por todo lo anterior, el objetivo principal de este proyecto se basa en la evaluación del rendimiento de la materia seca, utilizada como forraje para alimentar al ganado, de cebada - guisante como cultivo asociado, según diferentes ambientes de cultivo. Por último, se pretende encontrar una combinación de densidades de cultivo que optimice el rendimiento y la comparación del cultivo intercalado con ambos monocultivos.





---

## **Abstract**

The cultivation of huge fields with only one crop - monoculture farming - is a recent invention of modern agriculture.

Mixed crops or intercropping are the practice of growing two or more crops on the same land, as well as being the oldest form of systematized agricultural production. However, associated crops have been replaced by monocultures, especially in more developed countries.

The competition between species provided by mixed crops has been extensively studied within the world of ecology. Also, there is great interest within researchers, about the possible advantages or disadvantages that intercropping can offer. This is because they require less input of nutrients, pest control, pruning... than monocultures and, consequently, they are usually more profitable.

For all of the above, the main objective of this project is based on the evaluation of the yield of dry matter, used as fodder to feed cattle, of barley - pea as an associated crop, according to the different cultivation environments. Finally, it aims to find a combination of crop densities that optimizes yield and comparison of intercropping with both monocultures.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Contexto . . . . .	1
1.2. Objetivos . . . . .	3
1.3. Herramientas utilizadas . . . . .	3
<b>2. Materiales y métodos</b>	<b>5</b>
2.1. Datos . . . . .	5
2.1.1. Descripción . . . . .	5
2.1.2. Depuración de los datos . . . . .	8
2.2. Prueba de Rosner . . . . .	8
2.3. Diseño de superficies de respuesta . . . . .	9
2.3.1. Diseños y modelos para Superficie de Respuesta . . . . .	9
2.3.2. Diseños de primer orden . . . . .	9
2.3.3. Diseño Factorial $2^k$ . . . . .	10
2.3.4. Diseños de segundo orden . . . . .	10
2.3.5. Optimización de la superficie respuesta . . . . .	12
2.4. Relación de equivalencia de la tierra (LER) . . . . .	16
2.5. Diseño de series de reemplazo . . . . .	16
<b>3. Análisis y resultados</b>	<b>19</b>
3.1. Análisis descriptivo . . . . .	19
3.1.1. Detección de outliers . . . . .	20
3.1.2. Relación entre variables numéricas . . . . .	22
3.2. Análisis del rendimiento según el tratamiento y la localidad . . . . .	26
3.2.1. Test no paramétrico para el análisis del rendimiento . . . . .	28
3.2.2. Análisis del rendimiento según localidades . . . . .	30
3.3. Evaluación usando diseño de series con reemplazo . . . . .	33
3.3.1. Evaluación del rendimiento de cultivos mixtos de forma conjunta . . . . .	33
3.3.2. Evaluación del rendimiento de cultivos mixtos para diferentes ambientes . . . . .	35
3.4. Evaluación mediante análisis de superficie respuesta . . . . .	37
3.4.1. Modelo completo sin interacción . . . . .	40
3.4.2. Modelo completo sin interacción y sin un término cuadrático . . . . .	42
3.4.3. Elección del mejor modelo . . . . .	44

<b>4. Conclusiones</b>	<b>45</b>
4.1. Conclusiones . . . . .	45
4.2. Trabajo Futuro . . . . .	46
<b>Apéndices</b>	<b>49</b>
<b>A. Manual de utilización</b>	<b>49</b>
<b>B. Script del análisis</b>	<b>51</b>

# Índice de figuras

2.1. Croquis del diseño de bloques aleatorizado para cada localidad. . . . .	6
2.2. Cultivo mixto guisante - cebada en Zamadueñas (Valladolid) . . . . .	7
2.3. Diseño experimental utilizado en las tres fincas. . . . .	7
2.4. Diseño de tres factores para el diseño compuesto central, con los 3 tipos de puntos [18].	11
2.5. Diseño de tres factores para el diseño compuesto central, con los 3 tipos de puntos [20]. Los puntos rosas son los puntos de estrella y los puntos azules los puntos factoriales. . .	12
2.6. Ejemplo de los tipos de superficie respuesta . . . . .	14
2.7. Ejemplo de representación gráfica del análisis de cordillera . . . . .	15
2.8. Diseño de series con reemplazo para los distintos tipos de tratamientos . . . . .	17
3.1. Distribución de la materia seca según la localidad en la que se ha realizado el ensayo. . .	19
3.2. Distribución de la materia seca según el tratamiento utilizado. . . . .	20
3.3. Diagrama de cajas de la variable respuesta materia seca con outliers. . . . .	21
3.4. Matriz de correlación con p-valor, diagramas de dispersión y diagramas de densidades con histogramas. . . . .	23
3.5. Gráfico de dispersión de la <i>materia seca VS LER según la localidad</i> . . . . .	24
3.6. Distribución del <i>LER</i> según la <i>localidad</i> en la que se ha realizado el ensayo. . . . .	24
3.7. Gráficos de dispersión de materia seca según localidad o tratamiento. . . . .	25
3.8. Gráficos de dispersión según la localidad o el tratamiento. . . . .	25
3.9. Grafico de interacción de las variables tratamiento y localidad. . . . .	26
3.11. Grafico de dispersión de la cebada con la recta de regresión ajustada. . . . .	33
3.12. Grafico de dispersión del guisante con la recta de regresión ajustada. . . . .	35
3.13. Diseño CCC de las densidades relativas de ambos cultivos. . . . .	38
3.14. Gráfico 3D del modelo 2 de superficie respuesta de las densidades relativas de la cebada y el guisante. . . . .	41
3.15. Gráfico de contorno del modelo 2 de las densidades relativas de la cebada y el guisante.	41
3.16. Gráfico 3D del modelo 3 de superficie respuesta de las densidades relativas de la cebada y el guisante. . . . .	43
3.17. Gráfico de contorno del modelo 3 de las densidades relativas de la cebada y el guisante.	43



# Índice de tablas

2.1. Total de tratamientos realizados . . . . .	6
3.1. Prueba de Rosner con $k = 2$ . . . . .	21
3.2. ANOVA bidireccional con interacción. . . . .	26
3.3. Prueba de Kruskal-Wallis para la materia seca en función de la localidad. . . . .	28
3.4. Prueba de Kruskal-Wallis para la materia seca en función del tratamiento. . . . .	28
3.5. Comparaciones post-hoc para la variable Localidad. . . . .	28
3.6. Media de materia seca según cada localidad. . . . .	29
3.7. Comparaciones post-hoc entre los diferentes tratamientos. . . . .	29
3.8. Comparaciones post-hoc en Zamadueñas entre los diferentes tratamientos. . . . .	30
3.9. Comparaciones post-hoc en Viñalta entre los diferentes tratamientos. . . . .	31
3.10. Comparaciones post-hoc en Castrillo entre los diferentes tratamientos. . . . .	31
3.11. Resultados del ajuste de regresión para el cultivo cebada. . . . .	34
3.12. Resultados del ajuste de regresión para el cultivo guisante. . . . .	34
3.13. Resumen de cada regresión estimada para cada localidad . . . . .	36
3.14. Número de observaciones por encima y por debajo del umbral ( $LER$ ) para cada localidad. . . . .	36
3.15. Modelo completo de superficie de respuesta ajustado. . . . .	38
3.16. Comparación de los 3 modelos de superficie respuesta . . . . .	39
3.17. P-valor de cada modelo ajustado. . . . .	39
3.18. Modelo completo sin interacción de superficie de respuesta ajustado. . . . .	40
3.19. Modelo completo sin interacción y sin un término cuadrático de superficie respuesta ajustado. . . . .	42





# Capítulo 1

## Introducción

### 1.1. Contexto

Los cultivos mixtos o asociados son aquellos que se desarrollan en la misma parcela durante el mismo tiempo, evolucionando simultáneamente de forma más o menos sincronizada. Pueden estar formados por especies herbáceas, leñosas o una combinación de ambas. Algunas asociaciones tradicionales de especies son: maíz - judías, olivo - garbanzo, respectivamente.

Se pueden distinguir tres tipos de cultivos mixtos según la disposición de las parcelas y el tiempo de crecimiento: cultivo mixto o asociado propiamente dicho, cultivo intercalado en bandas o surcos, y cultivos intercalado de relevo o escalonados [1]. En el cultivo mixto o asociado propiamente dicho las especies están completamente mezcladas dentro de cada surco o han sido sembradas aleatoriamente. Los cultivos intercalados en bandas o franjas, se desarrollan de forma simultánea. Por último, en los cultivos intercalados de relevo o escalonados, se siembra un segundo cultivo antes de la maduración del primero.

El cultivo mixto es una práctica tradicional, que fue relegada por grandes áreas de monocultivo, típicos de una agricultura moderna, que consumen recursos y son de alto rendimiento [2]. Sin embargo, su uso sigue siendo común en países donde aún se practica agricultura de subsistencia con poca mecanización, y en algunas regiones del mundo los cultivos mixtos permanecen como la forma dominante de agricultura [3]. Normalmente, este tipo de cultivo es realizado por agricultores que practican la agricultura de bajos insumos, (aquellas que explotan recursos sin aplicar medidas que los repongan) y bajo rendimiento en pequeñas parcelas [4]. Aunque parecen estar vinculados a la agricultura de subsistencia, los cultivos mixtos pueden aportar mucho en la actual agricultura moderna, donde los recursos son cada vez más escasos y caros. Por lo tanto, estos laboreos apuntan hacia un gran potencial debido a su gran eficiencia y sostenibilidad, algo muy demandado por la sociedad hoy en día.

El porqué de los cultivos mixtos se debe a un mejor aprovechamiento de los recursos disponibles y/o por complementariedad de especies. Así, en el cultivo de veza – avena, ya estudiado, el cereal proporciona un tutor a la leguminosa, interceptando más luz, mejorando la calidad del forraje y facilitando la siega mecánica [5]. Existen numerosos trabajos acerca de los grandes beneficios que producen y que

justifican la aplicación de esta técnica, especialmente ante situaciones donde se practica agricultura de subsistencia, con escasos recursos y bajos insumos.

Los cultivos asociados proporcionan numerosas ventajas, estudiadas anteriormente y que justifican la aplicación de este método. Una de las principales ventajas sería la mayor facilidad de captura de luz por parte de los cultivos para realizar la fotosíntesis. Bedoussac y Justes [6], en Auzeville (Francia) con trigo duro - guisante de invierno, observaron un aumento hasta del 10 % respecto del trigo duro como cultivo único.

En lugares donde existe poca cantidad de terreno cultivable, los cultivos mixtos permiten maximizar el uso de este terreno [7]. La "maximización" del terreno se mide con el *Land Equivalent Ratio* (LER). Para obtener más información acerca de esta medida, ir al capítulo 2. Además, los cultivos mixtos influyen en la disponibilidad y eficiencia de uso de los nutrientes. La simbiosis y la fijación de nitrógeno atmosférico por parte de los rizobios de las leguminosas aumentan cuando se intercalan con otras plantas en comparación con el cultivo único [8].

Además, los cultivos mixtos con leguminosas son capaces de modificar la cantidad de nitrógeno atmosférico fijada, como se ha observado en guisante - cebada en Roskilde (Dinamarca), donde hubo aumentos del 40 – 80 % en la fijación de nitrógeno atmosférico respecto del monocultivo de guisante [9].

Asimismo, estos cultivos son una práctica sostenible que puede conservar y mejorar la calidad, salud y fertilidad del suelo [10], reduciendo las pérdidas de nitrógeno y aumentando el contenido de materia orgánica, ya que aumenta la biomasa de raíz que queda en el terreno después de la cosecha [11].

Por último, estos cultivos asociados reducen el efecto de plagas y enfermedades. Esto es debido a efecto de dilución (se reduce el número de huéspedes), y también, a un efecto barrera de las plantas resistentes, además de de la compensación de los bajos rendimientos de la especie afectada [12].

No obstante, a pesar de todos estos potenciales beneficios, los cultivos mixtos tienen que lidiar contra una enorme competencia de la agricultura intensiva de monocultivo a gran escala. Además, no todos los sistemas de cultivos asociados proporcionan todas las ventajas que se han enumerado anteriormente. Por ejemplo, en el sur de Australia, se llevaron a cabo ensayos con trigo - garbanzo, y se obtuvieron valores de LER (Land Equivalent Ratio) entre 1.01 y 1.02, concluyendo que no existían ventajas del cultivo mixto respecto al monocultivo [13].

Por todo lo anterior, existe un gran interés dentro del ámbito de la investigación de para poder comprender los procesos, mecanismos y posibles ventajas que ofrecen los cultivos intercalados. Este proyecto se va a centrar en la evaluación del rendimiento de la materia seca, utilizada como forraje, del cultivo mixto de cebada - guisante.

### 1.2. Objetivos

Los principales objetivos de este Trabajo de Fin de Grado son:

1. **Analizar el rendimiento** obtenido en función de las localidades y los tratamientos utilizados.
2. **Estudiar el comportamiento del cultivo** según la variación de las densidades del guisante y la cebada, mediante la evaluación del rendimiento.
3. **Obtener una combinación de densidades de cultivo** que optimice el rendimiento.
4. **Estimar los coeficientes de hacinamiento** de los cultivos mixtos frente al monocultivo.

Para el desarrollo de este TFG se ha contado con el Instituto Tecnológico Agrario de Castilla y León (ITACyL) que han proporcionado los datos que se han trabajado en este proyecto a partir de un ensayo experimental en el año 2021 en tres localidades, Castrillo de la Guareña (Zamora), Finca Viñalta (Palencia) y Finca Zamadueñas (Valladolid). Además, se ha contado con la ayuda de un investigador del ITACyL, Abel barrios, que ha ejercido como tutor del alumno.

### 1.3. Herramientas utilizadas

En esta sección se van a mencionar y explicar brevemente las herramientas tecnológicas utilizadas durante el desarrollo de este proyecto:

- **R:** Lenguaje de programación utilizado especialmente para el análisis de dato. Este software está formado por varias herramientas estadísticas y permite también la creación de gráficos. Es compatible con cualquier sistema operativo.
- **R-studio:** Interfaz de desarrollo que se encuentra integrado para el lenguaje R. Cuenta con consola y un editor para escribir el código y depurarlo.
- **Overleaf:** Editor de LaTeX online. Se ha decidido usar LaTeX para la creación de la memoria por la capacidad de este de generar documentos de alta calidad.
- **Excel:** Hojas de cálculos para crear, editar, ver o guardar los datos del proyecto. También permite crear hojas de datos online entre varios usuarios



## Capítulo 2

# Materiales y métodos

En esta sección del proyecto, se va explicar la estructura y el tipo de datos utilizados, así como la metodología utilizada durante el desarrollo del Trabajo de Fin de Grado.

### 2.1. Datos

#### 2.1.1. Descripción

Los datos empleados durante el desarrollo de este Trabajo de Fin de Grado han sido obtenidos de forma experimental durante el año 2021 en tres localidades Viñalta (Palencia), Castrillo de la Guareña (Zamora) y Zamadueñas (Valladolid). Se planteó un diseño de bloques aleatorizado en columnas con cuatro repeticiones (Figura 2.1) y se recogieron los resultados escritos a mano. Las dimensiones de cada parcela son de 1,5m x 8m u. e. (unidad experimental).

REP		I	I	I	I	I	I	II	II	II	II	II	II	III	III	III	III	III	III	IV	IV	IV	IV	IV	IV	REP	
COL	0	1	1	2	2	3	3	1	1	2	2	3	3	1	1	2	2	3	3	1	1	2	2	3	3	0	COL
	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	
1	borde	C33-G33	C33-G33	C33-G0	C33-G0	C100-G100	C100-G100	C100-G0	C100-G0	066-G33	066-G33	C100-G33	C100-G33	00-G66	00-G66	00-G100	00-G100	C33-G100	C33-G100	066-G100	066-G100	C100-G66	C100-G66	00-G33	00-G33	borde	1
	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	
2	borde	066-G66	066-G66	00-G66	00-G66	C33-G66	C33-G66	066-G100	066-G100	C100-G66	C100-G66	00-G100	00-G100	C100-G0	C100-G0	00-G33	00-G33	C33-G0	C33-G0	C33-G33	C33-G33	066-G0	066-G0	C100-G33	C100-G33	borde	2
	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	
3	borde	C100-G33	C100-G33	066-G100	066-G100	066-G33	066-G33	00-G33	00-G33	C33-G100	C33-G100	C33-G0	C33-G0	066-G0	066-G0	C33-G33	C33-G33	C100-G100	C100-G100	C100-G0	C100-G0	066-G66	066-G66	C33-G66	C33-G66	borde	3
	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	
4	borde	C33-G100	C33-G100	00-G33	00-G33	00-G100	00-G100	C33-G33	C33-G33	C33-G66	C33-G66	066-G0	066-G0	066-G66	066-G66	066-G100	066-G100	C100-G66	C100-G66	066-G33	066-G33	C100-G100	C100-G100	00-G66	00-G66	borde	4
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
5	borde	C100-G66	C100-G66	C100-G0	C100-G0	066-G0	066-G0	C100-G100	C100-G100	00-G66	00-G66	066-G66	066-G66	066-G33	066-G33	C33-G66	C33-G66	C100-G33	C100-G33	C33-G0	C33-G0	00-G100	00-G100	C33-G100	C33-G100	borde	5
MAQ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MAQ
COL	0	1	1	2	2	3	3	1	1	2	2	3	3	1	1	2	2	3	3	1	1	2	2	3	3	0	COL

Figura 2.1: Croquis del diseño de bloques aleatorizado para cada localidad.

Las especies utilizadas fueron **guisante** proteaginoso Viriato y una variedad local de **cebada**, en distintas proporciones de densidad de siembra. Ambas variedades, se encuentran adaptadas a las condiciones de suelo y clima de la región. Las siembras se realizaron en otoño en condiciones de cultivo ecológico. Se tomaron datos en distintos estados fenológicos de la planta. Sin embargo, este trabajo se encuentra enfocado al estudio del cultivo durante su estado lechoso, época durante la cual se recoge para su aprovechamiento como forraje que va ser utilizado para alimentar a la ganadería.

Por lo tanto, este conjunto de datos corresponde a mediciones del cultivo en estado lechoso y está formado por 390 observaciones recogidas de las tres localidades (130 observaciones de cada localidad) en el cual se han medido inicialmente 23 variables.

La variable *tratamiento* hace referencia a la densidad utilizada para una parcela según ambos cultivos. Hay un total de 15 tratamientos con una aplicación de 24 parcelas cada tratamiento excepto el tratamiento borde (50%:50%) que se ha aplicado en 30 parcelas. Los tratamientos utilizados encuentran en la siguiente Tabla 2.1. Los números indican las diferentes densidades de siembra sobre 100%, correspondiente al monocultivo. La letra delante de ellos el cultivo utilizado, C (cebada) ó G (guisante).

<i>C50-G50</i> (borde)	<i>C0-G100</i>	<i>C0-G33</i>	<i>C0-G66</i>
<i>C100-G0</i>	<i>C100-G100</i>	<i>C100-G33</i>	<i>C100-G66</i>
<i>C33-G0</i>	<i>C33-G100</i>	<i>C33-G33</i>	<i>C33-G66</i>
<i>C66-G0</i>	<i>C66-G100</i>	<i>C66-G33</i>	<i>C66-G66</i>

Tabla 2.1: Total de tratamientos realizados

## Materiales y métodos

Por último, se recogieron dos muestras por unidad experimental cuando la cebada estaba en estado de grano lechoso. Se midió la biomasa de ambos cultivos por separado y de las malas hierbas. Gracias a ello, se ha calculado el LER de cada observación. En la Figura 2.2 se puede observar una imagen del ensayo que se llevó a cabo en la finca Zamadueñas (Valladolid).



Figura 2.2: Cultivo mixto guisante - cebada en Zamadueñas (Valladolid)

Además se ha construido un gráfico que muestra los tratamientos utilizados en este proyecto (Figura 2.3). Cada punto representa un tratamiento diferente, es decir, combinaciones de densidades de siembra entre especies, guisante - cebada. La densidad es el número de individuos existentes por unidad de área.

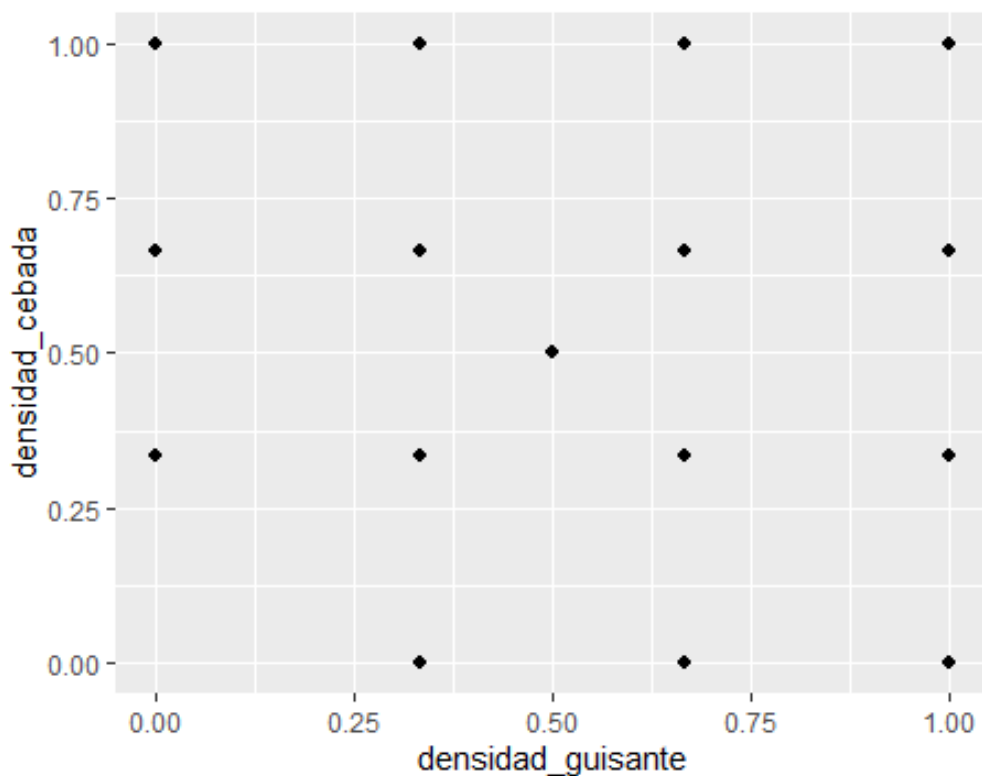


Figura 2.3: Diseño experimental utilizado en las tres fincas.

Otras variables de interés medidas para ambos cultivos individualmente han sido:

- **Nº de plantas:** Número de plantas de cebada o guisante obtenidas en cada parcela.
- **Peso materia seca** (gramos) : Peso total de la materia seca obtenida en cada parcela. Esta variable será la variable respuesta, ya que la materia seca obtenida será lo que proporcionará el forraje a la ganadería.
- **Materia seca/nº plantas** (gramos/ planta): Materia seca entre en nº de plantas en cada parcela para ambos cultivos.

### 2.1.2. Depuración de los datos

El proceso de corrección de errores dentro de las bases de datos, así como la identificación de datos incorrectos, incompletos o que no son de interés, se ha llevado a cabo mediante Excel en primera instancia y más tarde utilizando R.

Una vez eliminadas las columnas que no son de interés estadístico, se calcula el LER individual para cada unidad experimental, así como la densidad relativa del primer cultivo. Para la densidad relativa del segundo cultivo se ha usado R, observando los dos últimos caracteres de la variable que nos indica el tratamiento que hemos utilizado y estableciendo un número de 0 a 1 como densidad tal y como se puede ver en el código del Apéndice B:

Además, realizando un análisis general de la base de datos, se han encontrado inconsistencias que han sido corregidas. Estas correcciones se han llevado a cabo junto al tutor de las prácticas observando los cuadernos de campo con las anotaciones de cada variable medida y, viendo si se trataban de fallos humanos de escritura.

## 2.2. Prueba de Rosner

El test de Rosner es una prueba de detección de outliers que, puede detectar en una muestra más de un outlier al mismo tiempo. Esta prueba no aproxima de una forma adecuada cuando el número de muestras es pequeño ( $n < 25$ ) [14]. Además, no tiene problemas de enmascaramiento de observaciones que se encuentren muy próximas.



## 2.3. Diseño de superficies de respuesta

La **Metodología de Superficie de Respuesta** (MSR), introducida por Box y Wilcox (1951), es un conjunto de técnicas que permiten al investigador la modelización y el análisis de problemas en los que una variable de interés (que puede ser mostrada como una superficie) depende de otras variables que son factores cuantitativos. Es decir, el objetivo sería mostrar los valores óptimos para la variable respuesta mediante la maximización o minimización de las variables explicativas [15].

Normalmente, no se conoce la relación existente entre las variables independientes y la variable respuesta. Por lo tanto, se desea encontrar un modelo que aproxime la distribución entre la variable dependiente  $Y$  y las variables independientes [16].

### 2.3.1. Diseños y modelos para Superficie de Respuesta

La principal estrategia en el modelo MSR se basa en que la variable respuesta  $Y$  es una función  $f(x_1, x_2, \dots, x_k)$  del conjunto de variables de diseño  $x_1, x_2, \dots, x_k$  y esta función puede ser aproximada en algún punto de las variables independientes mediante un polinomio de primer o segundo orden [17]. Otra aproximación para modelos de orden  $n$  podría hacerse a partir de una expansión de las series de Taylor. Sin embargo, en la práctica y para regiones pequeñas los polinomios de primer y segundo grado son capaces de representar adecuadamente la función verdadera.

Los polinomios de primer grado se podrían representar de la siguiente forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.1)$$

Y los de segundo orden:

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i=1}^k \sum_{j=i+1}^k \beta_{ij} x_i x_j + \varepsilon \quad (2.2)$$

donde  $\varepsilon$  es el error aleatorio, y los coeficientes  $\beta_i$  se deben estimar a partir de una regresión lineal.

### 2.3.2. Diseños de primer orden

Estos diseños se utilizan durante las primeras etapas de análisis. El principal objetivo es la eliminación de factores cuando estos se encuentran en exceso y se quiere eliminar aquellos que no influyen significativamente en la variable respuesta [17].

El modelo de primer orden que se quiere ajustar tiene  $k+1$  factores para  $k$  parámetros a estimar. Por lo tanto, es necesaria la presencia de  $k+1$  observaciones.

Este modelo se sustituye por otro en el momento que no describe adecuadamente el comportamiento de la variable respuesta.

### 2.3.3. Diseño Factorial $2^k$

En este diseño se escogen 2 niveles del primer factor, 2 del segundo... y así sucesivamente. El total de combinaciones son  $2^k$ . Cada una de las combinaciones se convierte en una condición para el experimento [15].

### 2.3.4. Diseños de segundo orden

En muchas ocasiones, el modelo de primer orden no es suficiente para explicar la distribución que siguen las variables, ya que existe una curvatura.

Estos diseños son de utilidad en la última etapa del análisis, ya que permiten estudiar efectos lineales, de interacción y efectos cuadráticos, es decir un ajuste más preciso. Se quiere encontrar el punto estacionario que esta dentro de la región de optimización [17].

Hay diferentes tipos de modelos de segundo orden, y a continuación se van a explicar los más utilizados en la práctica.

#### Diseño Factorial $3^k$

Estos diseños están formados por  $k$  factores con 3 niveles cada uno, con un número total de observaciones  $3^k$ . En consecuencia, se requerirá un mayor número de experimentos que en el diseño  $2^k$  [15].

No obstante, debido a que es un modelo de segundo orden, permite modelar relaciones de forma cuadrática cuando el número de factores es pequeño.

#### Diseño de Composición Central

Este diseño es uno de los más utilizados en el MSR a día de hoy, ya que son diseños más económicos y son útiles en experimentos secuenciales [17].

Este diseño se basa en un diseño factorial fraccionado donde cada variable tiene dos puntos centrales que amplía con réplicas en el centro, en vez de aumentar el número de niveles de los factores. Por consiguiente, estos diseños constan de tres tipos de puntos (Figura 2.4):

- Consta de un diseño factorial  $2^k$  completo o fraccionado.
- $n_0$  puntos centrales.
- Dos puntos axiales en cada variable a una distancia  $\alpha$  del centro del diseño (puntos de estrella).

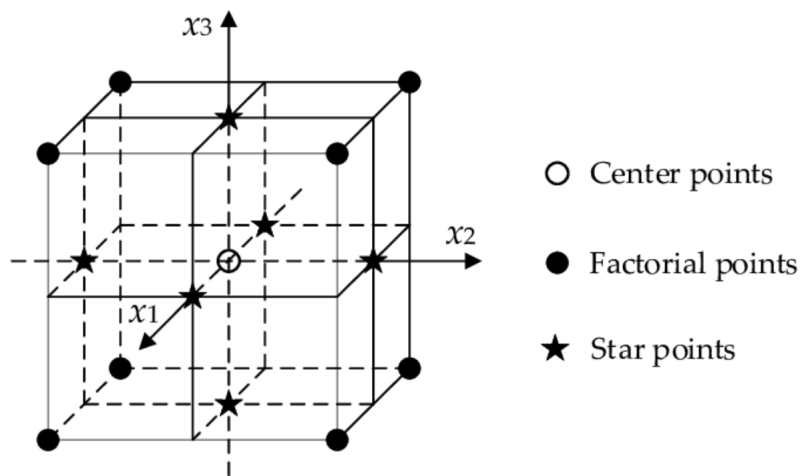


Figura 2.4: Diseño de tres factores para el diseño compuesto central, con los 3 tipos de puntos [18].

Finalmente, el número total de puntos de este diseño sería:  $N = 2^k + 2k + n_0$

Este diseño central compuesto (CCD) dispone de tres variantes: **circunscritos** (CCC), **inscritos** (CCI) y **centrados en la cara** (CCF) [19].

CCC involucra puntos factoriales, puntos centrales y puntos de estrella. Estos últimos representan los valores extremos de las variables. La distancia entre el punto central y el punto factorial es  $\pm 1$  y, entre el punto central y el punto de estrella es  $\alpha$ . Los puntos de diseño describen un círculo circunscrito al cuadrado factorial. Para tres factores, los puntos de diseño CCC describen una esfera alrededor del cubo factorial.

Para CCC,  $\alpha$  es mayor que 1, mientras que para CCI, los puntos de estrella toman los valores límite especificados de las variables.

En conclusión, el diseño CCC explora el espacio más grande y el CCI explora el espacio más pequeño. Tanto el CCC como el CCI son diseños rotables, mientras que CCF explora un espacio  $[-1, +1]$  pero sin giratorio.

En la Figura 2.5 se muestra la representación los tipos de diseños centrales compuestos.

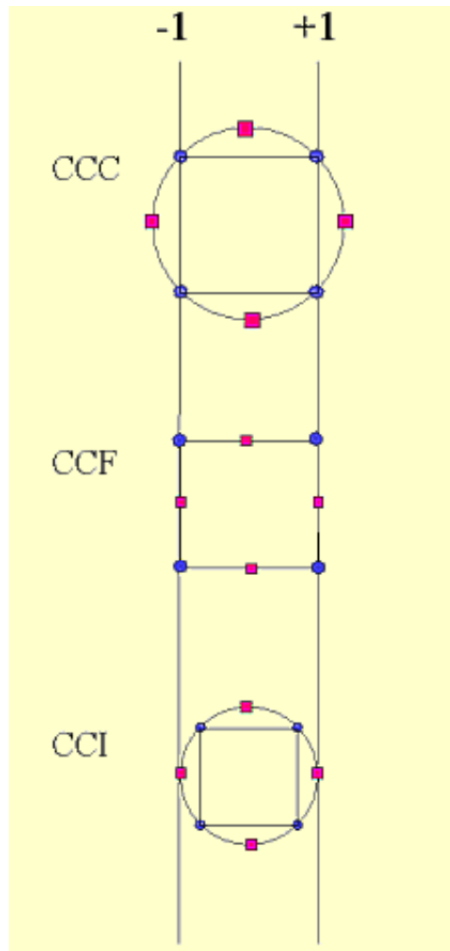


Figura 2.5: Diseño de tres factores para el diseño compuesto central, con los 3 tipos de puntos [20]. Los puntos rosas son los puntos de estrella y los puntos azules los puntos factoriales.

### 2.3.5. Optimización de la superficie respuesta

Una vez elegido y ajustado el modelo, se quiere encontrar la combinación de niveles de los factores que producen una respuesta óptima. Para ello se utilizan diferentes técnicas de optimización dependiendo del diseño elegido y la situación que exista en los datos.

#### Escalamiento Ascendente (o Descendente)

Este proceso de optimización se utiliza cuando el modelo seleccionado es de primer orden. Se basa en encontrar la dirección de máxima respuesta (o mínima) de la variable respuesta sobre el plano experimental [17].

**Análisis canónico**

Si el punto máximo en un modelo de segundo orden existe, entonces las derivadas parciales del modelo evaluadas en el conjunto de variables del diseño serán [21]:

$$\frac{\partial \hat{y}}{\partial x_1} = \frac{\partial \hat{y}}{\partial x_2} = \dots = \frac{\partial \hat{y}}{\partial x_k} = 0 \quad (2.3)$$

Este punto, en el que las derivadas parciales son 0 y por lo tanto, el plano tangente a la superficie tiene pendiente cero se le conoce como **punto estacionario**. Este punto puede representar un *máximo*, un *mínimo* o un *punto de inflexión*. Al análisis del modelo de segundo orden, donde en parte se encuentra este punto estacionario, se la llama **análisis canónico**. Este análisis transforma el modelo a un nuevo sistema de coordenadas con el origen en el punto estacionario.

Por lo tanto, el principal objetivo de este análisis es encontrar el punto estacionario del modelo, **dentro** de la superficie respuesta. Además, también se puede analizar cómo se relacionan las variables canónicas y las variables codificadas del modelo.

Por lo tanto, considerando un modelo de segundo orden como el dado en la Ecuación 2.2, en forma matricial se representaría [17]:

$$\hat{Y} = \hat{\beta}_0 + X^T \beta + X^T B X \quad (2.4)$$

Donde:

$$X^T = (x_1, x_2, \dots, x_k) \quad (2.5)$$

$$\beta = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k) \quad (2.6)$$

$$B = \begin{bmatrix} \hat{\beta}_{11} & \frac{\hat{\beta}_{12}}{2} & \dots & \frac{\hat{\beta}_{1k}}{2} \\ \frac{\hat{\beta}_{21}}{2} & \hat{\beta}_{22} & \dots & \frac{\hat{\beta}_{2k}}{2} \\ \dots & \dots & \dots & \dots \\ \frac{\hat{\beta}_{k1}}{2} & \frac{\hat{\beta}_{k2}}{2} & \dots & \hat{\beta}_{kk} \end{bmatrix} \quad (2.7)$$

Como la Ecuación 2.4 es diferenciable en el espacio  $R^k \rightarrow R$  el punto estacionario pertenece al conjunto de puntos  $x \in R$  que cumplen [17]:

$$grad \hat{Y}(x) = 0 \quad (2.8)$$

A partir de la forma bilineal simétrica también se puede obtener que:

1. Si todos los autovalores son **positivos**, el punto estacionario es un *mínimo local*.
2. Si todos los autovalores son **negativos**, el punto estacionario es un *máximo local*.
3. Si un autovalor es **positivo y otro negativo**, el punto estacionario es un *punto silla*.

En la siguiente Figura 2.6 se pueden observar los distintos tipos de superficies de respuesta:

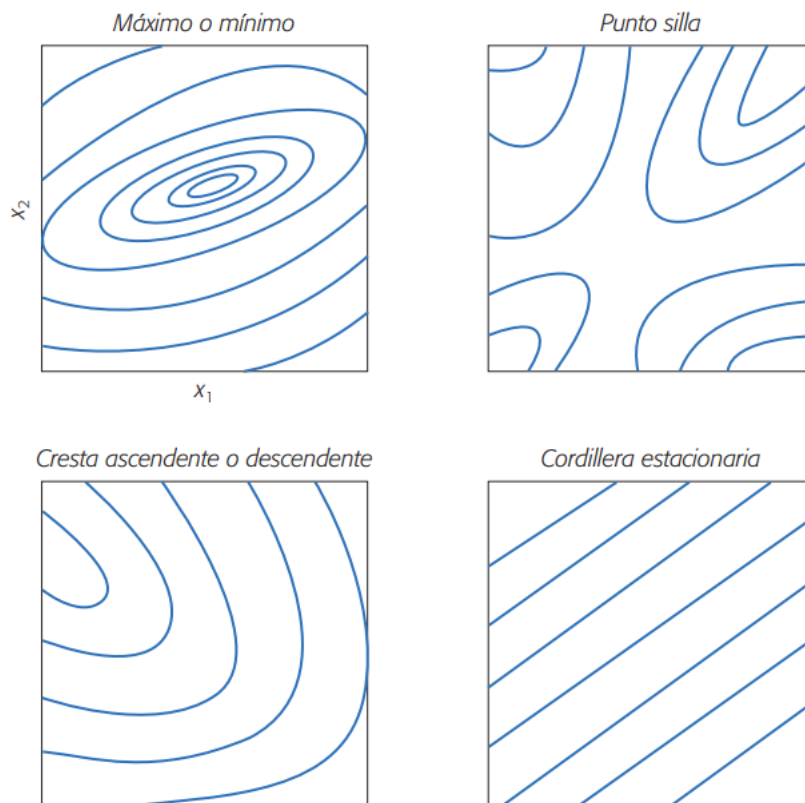


Figura 2.6: Ejemplo de los tipos de superficie respuesta .

### Análisis de cordillera

En ciertas ocasiones, se desea encontrar el punto estacionario aunque este no esté dentro de la superficie respuesta o cuando el punto estacionario no es el óptimo de la región experimental, por lo que se utiliza un **análisis de cordillera**. Este método también se utiliza para la optimización simultánea cuando se tiene varias variables repuestas.

Este método consiste en el cálculo de un máximo o un mínimo sobre esferas concéntricas al centro del modelo, cubriendo esferas de todo tipo de radios hasta llegar a una esfera que cuente con todos los puntos experimentales. Cuando solo hay dos factores, no son esferas, sino circunferencias [16].

## Materiales y métodos

---

Partiendo del modelo de segundo orden en forma matricial (Ecuación 2.4), se plantea la ecuación para encontrar el punto en la esfera, donde la respuesta es máxima (o mínima):

$$F = \hat{\beta}_0 + X^T \beta + X^T B X + \lambda(X^T X - R_i) \quad (2.9)$$

Donde  $\lambda$  es el *multiplicador de Lagrange* y  $R_i$  es el radio de la esfera que cumple la restricción:

$$R_i = X^T X = \sum_{i=1}^k x_i^2 \quad (2.10)$$

En la Figura 2.7 se puede observar un ejemplo de un análisis de cordillera.

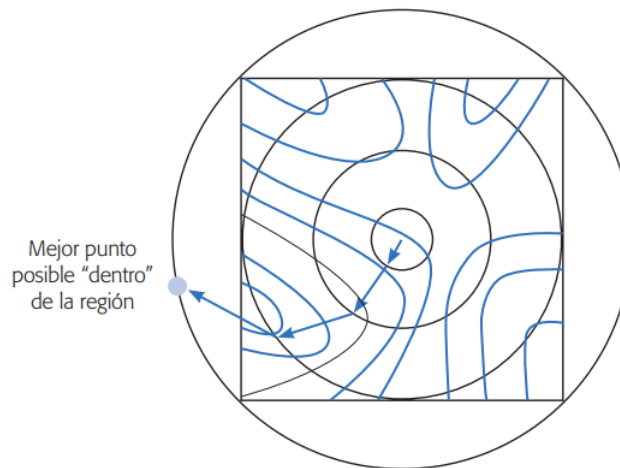


Figura 2.7: Ejemplo de representación gráfica del análisis de cordillera

## 2.4. Relación de equivalencia de la tierra (LER)

Un parámetro comúnmente utilizado para medir la superioridad o inferioridad de los cultivos mixtos sobre los monocultivos es el **LER** (*Land Equivalent Ratio*) [22].

A esta medida también se la puede llamar **Rendimiento Relativo Total** (*RYT*), siendo igual a la suma de los rendimientos máximos de cada intercultivo tal y como se puede ver en la Ecuación 2.11:

$$RYT_{max} = \frac{Y_{gc}^*}{Y_{gg}} + \frac{Y_{cg}^*}{Y_{cc}} \quad (2.11)$$

Donde  $Y_{gc}^*$  y  $Y_{cg}^*$  son los rendimientos máximos estimados en cultivos asociados y,  $Y_{gg}$  y  $Y_{cc}$  los rendimientos de los monocultivos de guisante y cebada, respectivamente.

Esta Ecuación 2.11 se puede sustituir por el LER de ambos intercultivos (Ecuación 2.12):

$$RYT_{max} = LER_{guisante} + LER_{cebada} \quad (2.12)$$

Un valor de LER superior a 1, indicaría que la producción es mayor en un sistema con cultivos mixtos que en un sistema de monocultivo. Un LER igual a 1 señala que no existe diferencia entre cultivar en monocultivos o en forma mixta.

En el caso de este trabajo, los cultivos utilizados son guisante y cebada, pero podría ser cualquier tipo de cultivo con las mismas ecuaciones.

## 2.5. Diseño de series de reemplazo

El **diseño de series de reemplazo** es muy utilizado en investigación de cultivos asociados. Cada parcela cuenta con proporciones de siembra de cada especie distintas, pero con una densidad constante de plantas en todas las parcelas. Este diseño se utiliza para medir y detectar la competencia entre especies [23].

Se puede elegir cualquier tipo de diseño, con tal de que la suma de ambas densidades sea 100%. En este trabajo, el diseño elegido consiste en cinco diferentes mezclas con diferentes densidades relativas de ambos cultivos, guisante y cebada. Las densidades utilizadas han sido (C-G): 100%:0%, 66%:33%, 50%:50%, 33%:66%, 0%:100%. En la Figura 2.8 se puede observar una imagen del diseño utilizado.

Para el análisis, se han utilizado regresiones del rendimiento entre especies del guisante (Ecuación 2.13) y de la cebada (Ecuación 2.14) [24]:

$$Y_{gc} = \frac{k_{gc}r_{gc}}{(1 - r_{gc}) + k_{gc}r_{gc}} Y_{gg} \quad (2.13)$$



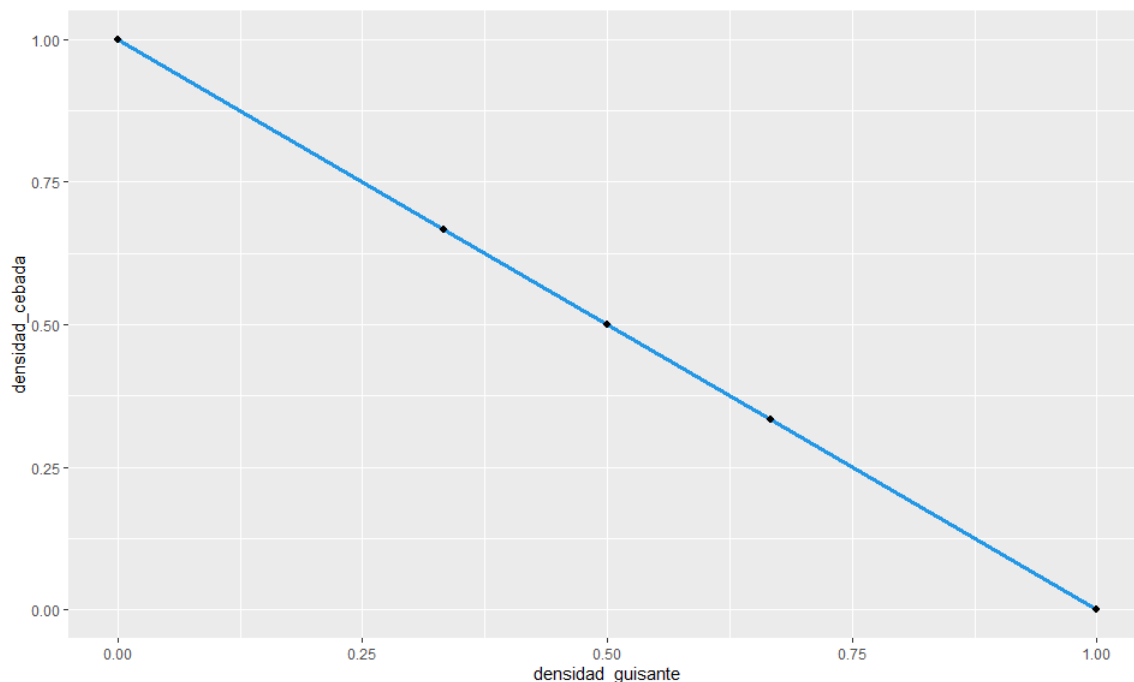


Figura 2.8: Diseño de series con reemplazo para los distintos tipos de tratamientos

$$Y_{cg} = \frac{k_{cg}r_{cg}}{(1 - r_{cg}) + k_{cg}r_{cg}} Y_{cc} \quad (2.14)$$

Donde  $Y_{gg}$  y  $Y_{cc}$  representan el rendimiento de monocultivo de guisante y cebada, respectivamente.

$r_{gc}$  y  $r_{cg}$  hacen referencia a las densidades relativas del cultivo mixto de guisante y cebada, respectivamente.

Por último,  $k_{cg}$  y  $k_{gc}$  son constantes, que describen la habilidad de competencia entre los cultivos intercalados de guisante y cebada, llamadas **coeficientes de hacinamiento** (*crowding coefficients*).

Estos coeficientes de hacinamiento son interesantes desde el punto de vista tanto práctico como teórico para saber cómo funcionan los cultivos entre especies. Son una poderosa herramienta para evaluar las interacciones de especies competitivas, para estimar las ventajas económicas en los cultivos mixtos y medir que especie es la dominante[25]. Para estimarlos, se han convertido las ecuaciones de rendimiento en dos regresiones lineales simples.

Hall [26] propuso el término coeficiente de hacinamiento relativo ( $k$ ). Este coeficiente proporciona una medida de si esa especie ha producido más o menos rendimiento de lo esperado. Si una especie tiene un coeficiente menor que, igual o mayor que uno, significa que ha producido menos rendimiento, el mismo rendimiento o más rendimiento que el rendimiento esperado, respectivamente. Para determinar si hay una ventaja de rendimiento de los cultivos mixtos, el producto del coeficiente es formado ( $K = k_a * k_b$ ). Si  $K$  es mayor, igual o menor que uno, hay una ventaja de rendimiento, no hay diferencia o hay desventaja de rendimientos, respectivamente [27].

En primer lugar, se ha sustituido el LER de ambos cultivos (Ecuación 2.11) en las Ecuaciones 2.13

y 2.14 y se han obtenido las siguientes fórmulas:

$$LER_{guisante} = \frac{k_{gc}r_{gc}}{(1 - r_{gc}) + k_{gc}r_{gc}} \quad (2.15)$$

$$LER_{cebada} = \frac{k_{cg}r_{cg}}{(1 - r_{cg}) + k_{cg}r_{cg}} \quad (2.16)$$

Se calcula la inversa de las dos ecuaciones anteriores, se simplifica y se obtienen las transformaciones de ambos cultivos:

$$Y_g = \frac{1}{LER_{guisante}} - 1 \quad (2.17)$$

$$X_g = \frac{1 - r_{gc}}{r_{gc}} \quad (2.18)$$

$$\beta_g = \frac{1}{k_{gc}} \quad (2.19)$$

Donde  $Y_g$  y  $X_g$  son la variable respuesta y la variable independiente, respectivamente; y  $\beta_g$  es el parámetro de interés, ya que es la inversa del coeficiente de hacinamiento.

Las Ecuaciones 2.17, 2.18 y 2.19 se refieren al cultivo del guisante.

Las regresiones interesantes a estimar para cada cultivo serían:

$$Y_i = \beta_i X_i \quad (2.20)$$

Para  $i = c, g$

## Capítulo 3

# Análisis y resultados

En este Capítulo se va a realizar un análisis estadístico de los datos y se van a mostrar los resultados obtenidos.

### 3.1. Análisis descriptivo

En primer lugar, se va a realizar un análisis descriptivo procediendo a la descripción del conjunto de datos.

En la siguiente Figura 3.1 se puede observar el diagrama de cajas de la variable respuesta materia seca total (MS) según la localidad en la que se encuentre. También, en la Figura 3.2 se muestra la distribución de la materia seca total (MS) en función del tipo de tratamiento utilizado (TRAT).

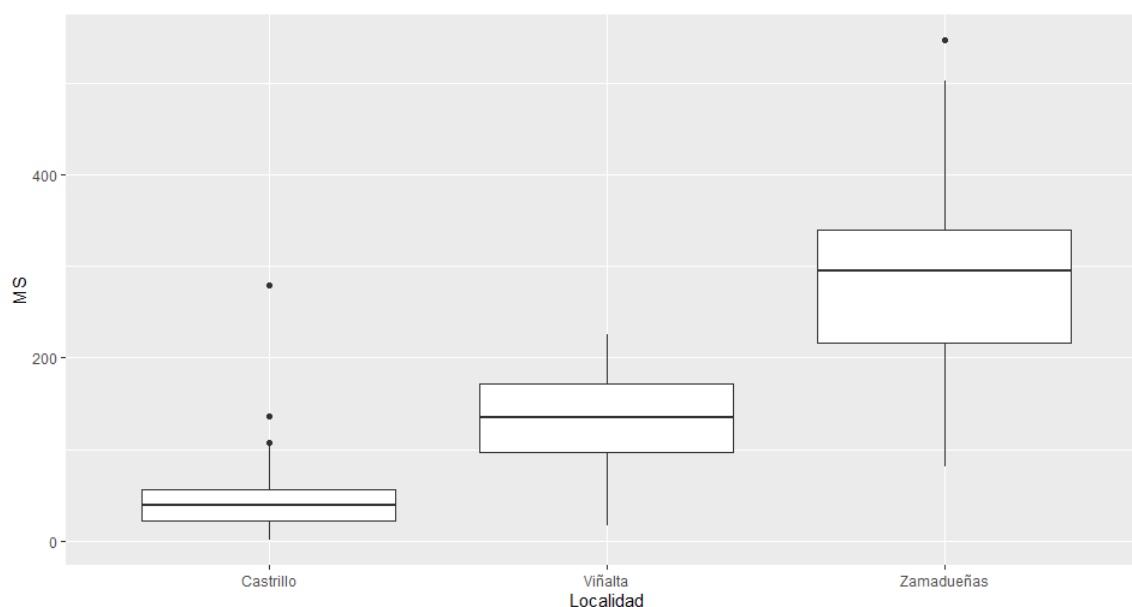


Figura 3.1: Distribución de la materia seca según la localidad en la que se ha realizado el ensayo.

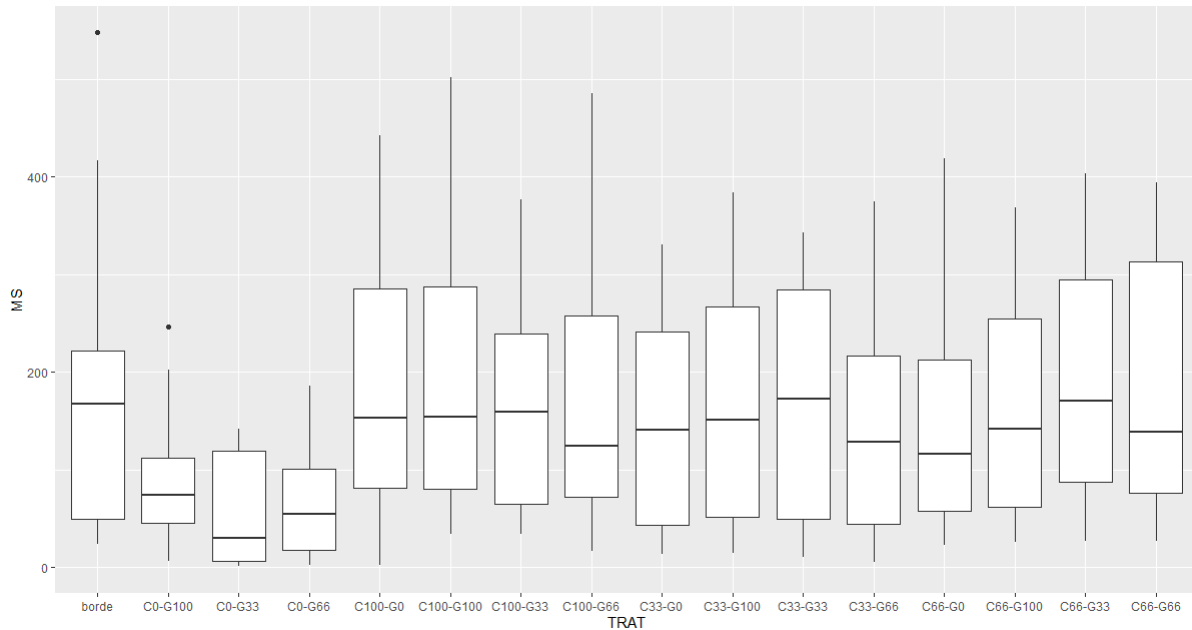


Figura 3.2: Distribución de la materia seca según el tratamiento utilizado.

En la Figura 3.1 se puede observar que existen muchas diferencias entre las desviaciones típicas de cada localidad. Se intuye que habrá que utilizar una transformación logarítmica para estabilizar las varianzas. Por último, mencionar que existen diferencias significativas ( $p$ -valor  $< 0.05$ ) entre localidades, siendo Zamadueñas la que mayor rendimiento produce, seguida de Viñalta y con Castrillo de la Guareña como última localización en cuanto a la producción de materia seca. Esto es debido a que son ambientes con diferente potencial productivo.

En cuanto a los tratamientos utilizados, examinando la Figura 3.2, el tratamiento que mejor resultados de materia seca obtiene es el C66 - G33, mientras que C0 - G33, proporciona la menor cantidad de materia seca.

### 3.1.1. Detección de outliers

En este apartado se va a realizar un análisis de detección de outliers mediante diversas herramientas gráficas y analíticas. Tal y como se puede examinar en la Figura 3.3, al dibujar la distribución de la variable respuesta materia seca en *gramos/muestra* siendo  $muestra = 0.25m^2$ , se hallan dos valores atípicos.

Estos valores atípicos corresponden a las observaciones con 547.33 y 501.94 de valor de materia seca, valores superiores a la media de materia seca existente (148.84).

Otra forma de cálculo de outliers es mediante el uso de test estadísticos basados en inferencia que compara valores críticos. En este caso, se va a usar la *prueba de Rosner*, ya que es capaz de detectar varios outliers. Se han observado 2 valores sospechosos por lo que se establece la  $k$  del test igual a 2.

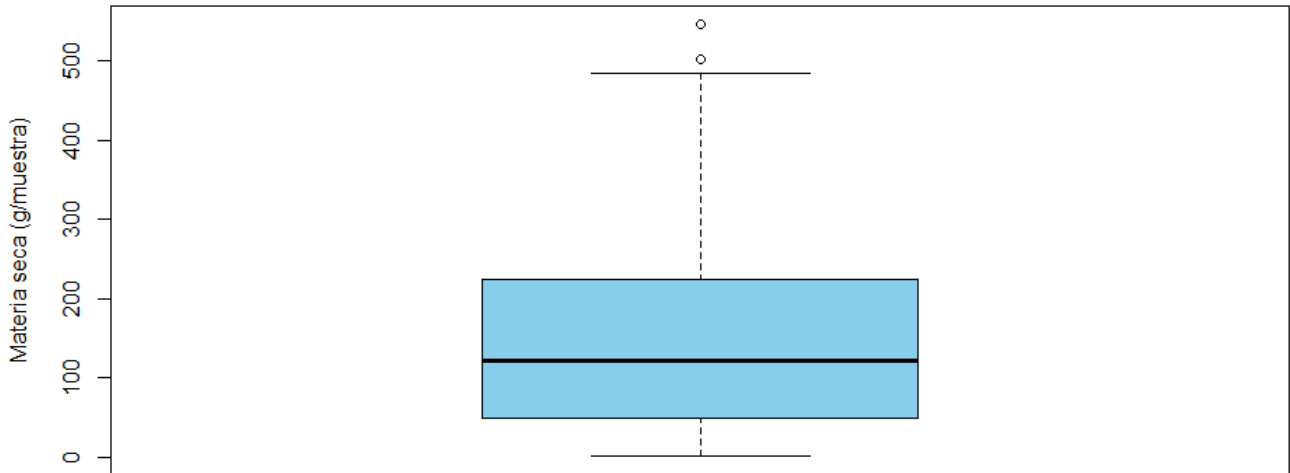


Figura 3.3: Diagrama de cajas de la variable respuesta materia seca con outliers.

El resultado del test se puede ver en la Tabla 3.1. Se puede concluir que los valores atípicos no afectan negativamente a los datos, por lo tanto, no han sido eliminados.

<b>i</b>	<b>Media.i</b>	<b>Desviación típica.i</b>	<b>Valor</b>	<b>Número observación</b>	<b>R.i+1</b>	<b>Lambda.i+1</b>	<b>OUTLIER</b>
0	148.843	117.257	547.33	262	3.399	3.789	FALSE
1	147.789	115.600	501.94	370	3.064	3.788	FALSE

$R_i$  : estadísticos que representan las desviaciones estudentizadas extremas calculadas a partir de muestras sucesivamente reducidas de tamaño  $n, n-1, \text{ etc. } \lambda_i$  : valor crítico para cada estadístico  $R_i$ .

Tabla 3.1: Prueba de Rosner con  $k = 2$ .

### 3.1.2. Relación entre variables numéricas

En este apartado, se quiere estudiar si existe algún tipo de relación o tendencia que sea de interés entre las variables numéricas del conjunto de datos.

Se ha realizado un *Test de Shapiro–Wilk* a las variables numéricas del modelo para comprobar si la población se encuentra normalmente distribuida. Como todas las variables numéricas del modelo no siguen una distribución normal ( $p$ -valor  $<0.05$ ), se ha calculado la **correlación de Spearman** para las variables del modelo.

La Figura 3.4 cuenta con los siguientes elementos:

- En la diagonal de la matriz, se encuentran los nombres de las variables.
- También en la diagonal, se muestran las gráficas de densidad con sus respectivos histogramas, para cada variable.
- Los gráficos de la esquina inferior izquierda muestran los gráficos de dispersión de cada par de variables.
- Los cuadros de la esquina superior derecha muestran el coeficiente de correlación entre cada par de variables.

Observando el gráfico cabe destacar que, las correlaciones más elevadas se encuentran entre la variable respuesta *materia seca* y, las variables numéricas, siendo significativa esta correlación. Existe una relación muy fuerte lineal positiva entre la materia seca total y el *nº de plantas de cebada* ( $0.53^{***}$ ), la *materia seca/planta cebada* ( $0.79^{***}$ ), y el *peso de materia seca de cebada* ( $0.88^{***}$ ). Se puede deducir que cuanto mayor sea el número de plantas de cebada y mayor tamaño tengan esas plantas, mayor materia seca total se obtendrá. No ocurre así lo mismo con el guisante, que a pesar de tener correlaciones positivas y significativas con la materia seca total obtenida, sus valores son muy bajos. Esto hace pensar que es el cultivo de cebada el que más contribuye a la obtención de materia seca total en el conjunto de la parcela, es decir, es la especie que más forraje aporta.

Otra correlación interesante sería la *materia seca total* con el *peso en materia seca de malas hierbas* ( $-0.87^{***}$ ). Cuanto mayor cantidad de malas hierbas se obtengan menor será la materia seca extraída de los cultivos. Es una prueba de la competencia que ejerce la flora adventicia sobre el rendimiento de los cultivos. De hecho, todas las correlaciones donde están implicadas las malas hierbas, son negativas.

Dentro de las correlaciones significativas, destacan dos por presentar una función de regresión muy ajustada a la dispersión de puntos. La recta de regresión resultado del ajuste del peso de la materia seca total con el peso de la materia seca, ajusta muy bien los datos, mediante una relación lineal. Esto supone que dentro de los valores estudiados, a más cebada, más materia seca obtengo. La otra relación que se quiere enfatizar es entre el peso de la materia seca total con el peso que aportan las malas hierbas: cuanto mayor peso aportan las hierbas, menor contenido en materia seca total se obtiene, pero en este caso la pendiente negativa no es constante, siendo al principio muy fuerte para ir convirtiéndose en asintótica. Esto significa que pequeñas infestaciones de malas hierbas, producen

## Análisis y resultados

disminuciones importantes de producción cuando el nivel de población de hierbas es bajo, mientras que con altos niveles de población de hierbas, pequeñas infestaciones apenas producen disminuciones del rendimiento en forraje.

El resto de correlaciones no son de interés ya que hay variables que se han calculado en función de otras o su correlación no es significativa.

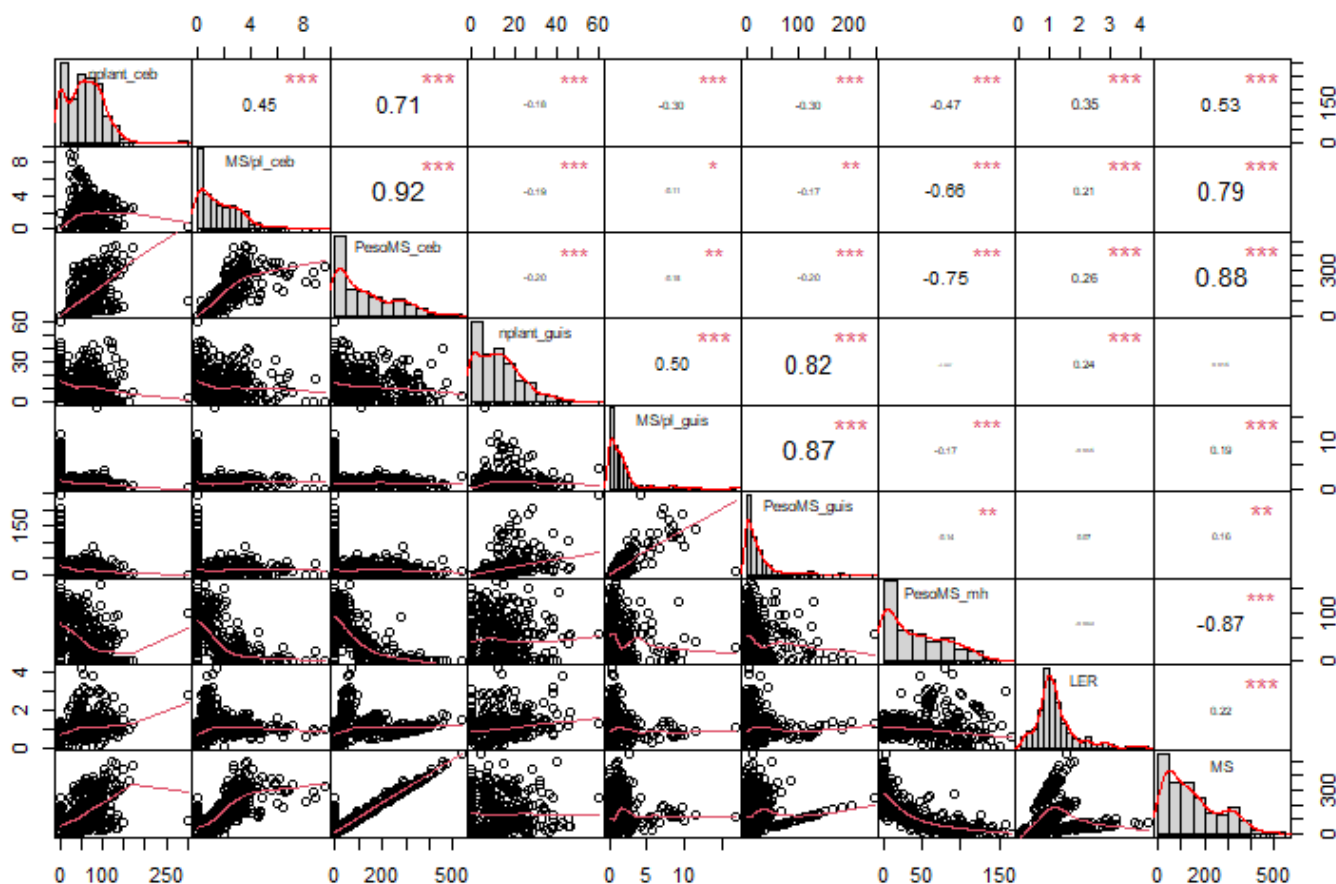


Figura 3.4: Matriz de correlación con p-valor, diagramas de dispersión y diagramas de densidades con histogramas.

Se van a analizar algunos gráficos de dispersión para intentar explicar las distribuciones subyacentes según alguna variable del conjunto de datos.

En la Figura 3.5 se observa el gráfico de dispersión de la *materia seca vs LER*, en función de las localidades. La materia seca obtenida en cada localidad es distinta, siendo Zamadueñas la localidad donde se obtiene un mayor rendimiento, y Castrillo la de menor, como se había deducido anteriormente. Además, en Viñalta y Zamadueñas se han obtenido valores de LER inferiores a 2. Sin embargo, en Castrillo existen valores de LER alrededor de 4.

Desde el punto de vista de la materia seca, en Zamadueñas se ha encontrado una alta variabilidad en los datos, mientras que en Viñalta y Castrillo una dispersión media y baja, respectivamente, de

los datos (Figura 3.1). No obstante, en el caso del LER la situación es opuesta. Castrillo consta con una dispersión muy amplia de las observaciones; mientras Zamadueñas y Viñalta con datos menos dispersos (Figura 3.6).

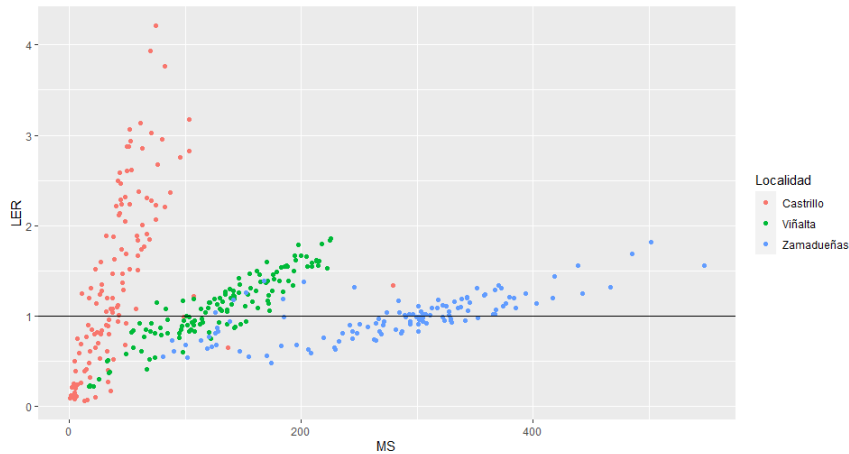


Figura 3.5: Gráfico de dispersión de la *materia seca* VS *LER* según la *localidad*.

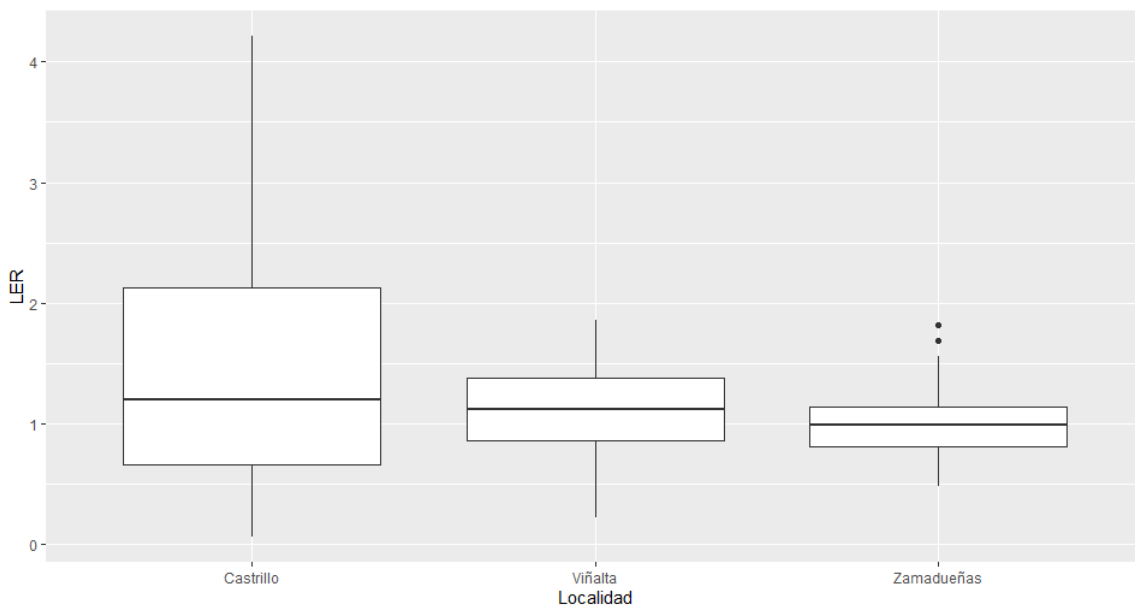


Figura 3.6: Distribución del *LER* según la *localidad* en la que se ha realizado el ensayo.

La Figura 3.7a muestra la dispersión entre la *materia seca* vs *peso de materia seca* en el cultivo de cebada según la *localidad*. En Zamadueñas es en donde se obtiene mayor cantidad de *materia seca* de cultivo de cebada debido a su mayor potencial de rendimiento como se ha indicado anteriormente.

En la Figura 3.7b se observa esta misma dispersión según el tratamiento utilizado. Dicha dispersión es muy semejante en las tres localidades, las cuales están muy bien definidas. Salvo para el monocultivo de guisante, no se observa que los tratamientos se encuentren ordenados como ocurre en la Figura 3.7a, en función de las localidades.



## Análisis y resultados

En ambas gráficas se guarda una relación lineal bastante definida.

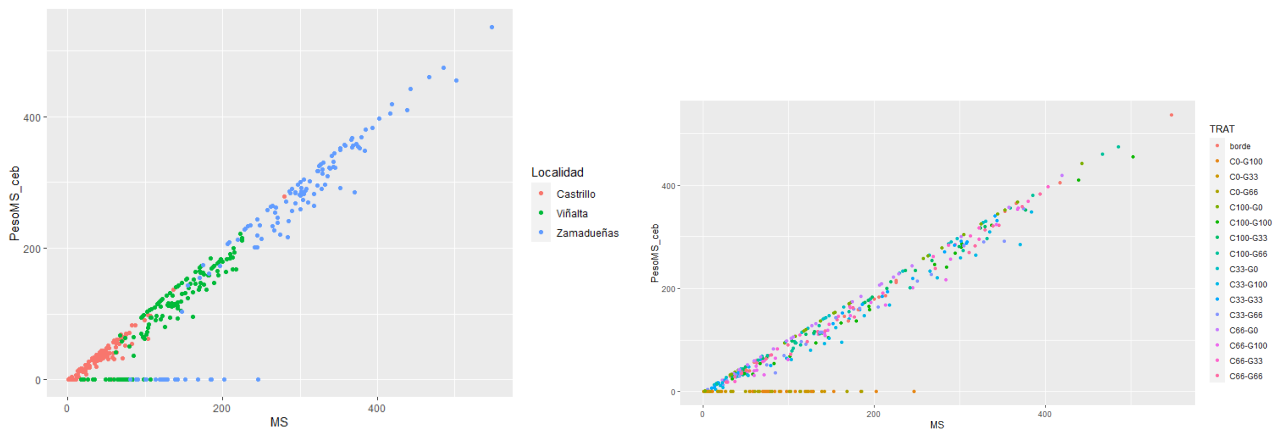


Figura 3.7: Gráficos de dispersión de materia seca según localidad o tratamiento.

En el caso del gráfico de dispersión de la *materia seca VS peso de materia seca del guisante* según *localidad o tratamiento* se obtiene la Figura 3.8. En este caso, se observa una mayor dispersión entre las observaciones. Además, se contempla otra prueba más de que la mayor cantidad de materia seca total se obtiene en Zamadueñas y que la mayor parte de materia seca total corresponde a la cebada.

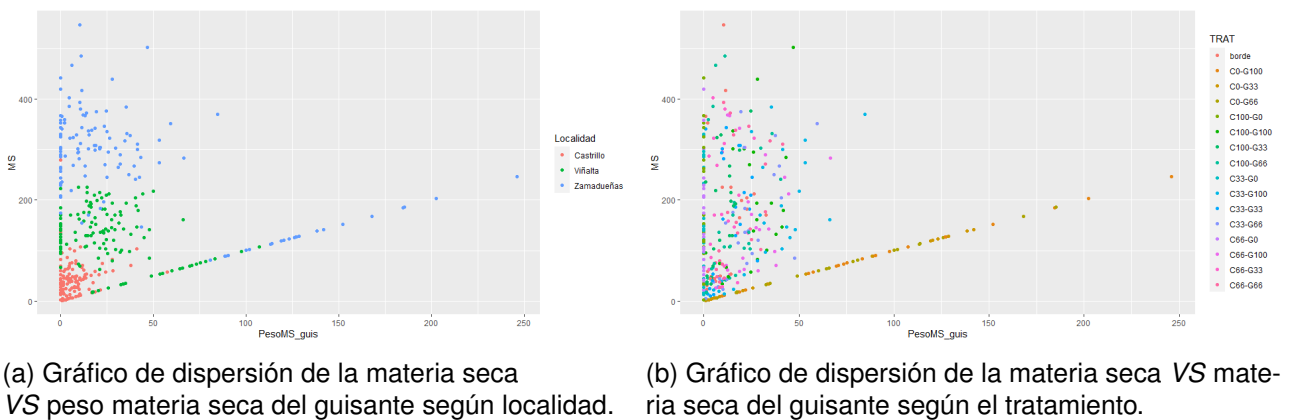


Figura 3.8: Gráficos de dispersión según la localidad o el tratamiento.

### 3.2. Análisis del rendimiento según el tratamiento y la localidad

En esta sección se va a realizar un análisis de la varianza (ANOVA) para tratar de explicar la relación entre dos variables independientes cualitativas (*tratamiento* y *localidad*), con la variable respuesta cuantitativa (*materia seca total*).

Para el análisis de la varianza también se va a medir la interacción entre los dos factores para ver si esta interacción influye significativamente en el modelo.

La Figura 3.9 muestra el gráfico de interacción de ambas variables y se puede observar que las líneas no son completamente paralelas, sin embargo, no se cruzan entre ellas. Se observa una ligera interacción entre ambos factores, lo que indicaría que la posible relación entre el tratamiento utilizado y la cantidad de materia seca depende de la localidad en la que se ha desarrollado el ensayo.

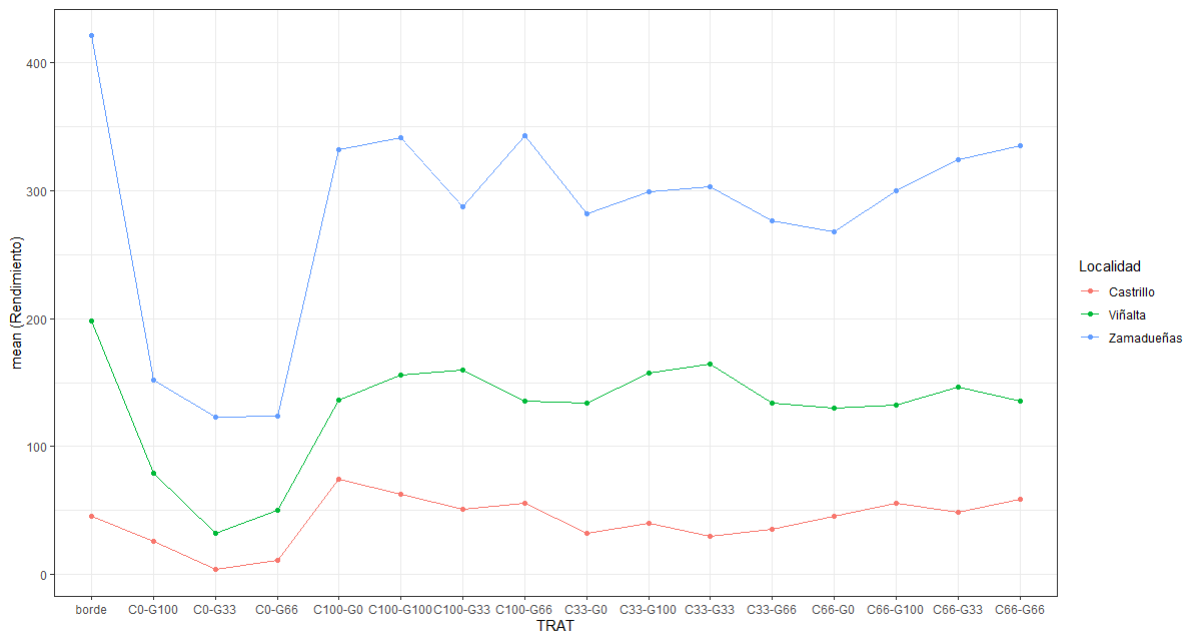


Figura 3.9: Grafico de interacción de las variables tratamiento y localidad.

Para comprobar el efecto de la interacción se realiza un ANOVA de dos vías que se muestra en la Tabla 3.2:

Columna	Df	Sum sq	Mean sq	F value	Pr(>F)
TRAT	15	665961	44397	21.936	<2e-16***
Localidad	2	3565571	1782785	880.846	<2e-16***
TRAT:Localidad	30	295767	9859	4.871	1.01e-13***
Residuals	331	669927	2024		

Df: grados de libertad. Sum Sq: valor de la suma de cuadrados. Mean Sq: media aritmética de los cuadrados. F value: valor del estadístico F. Pr(>F): p-valor de cada variable. TRAT: tratamiento.

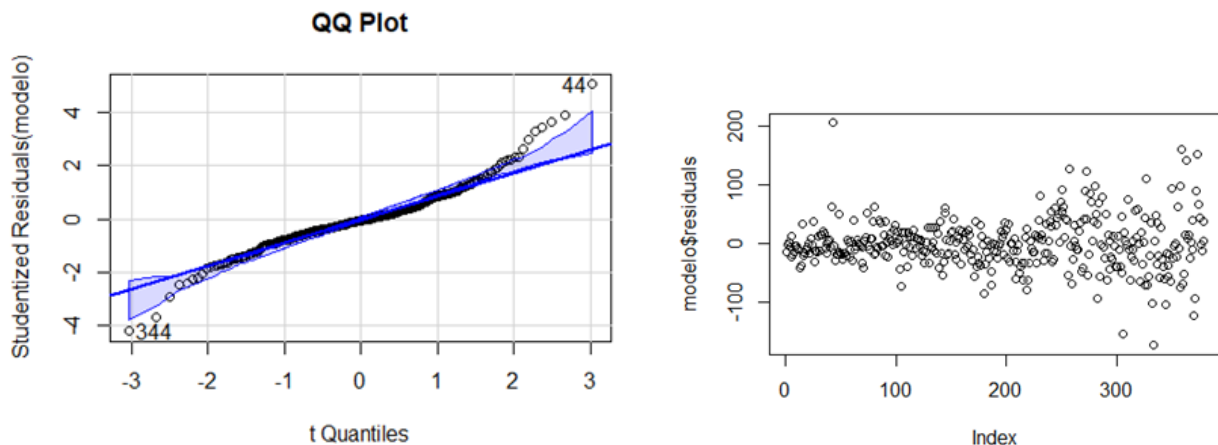
Tabla 3.2: ANOVA bidireccional con interacción.

## Análisis y resultados

Se puede determinar que tanto el tratamiento, como la localidad y la interacción entre ambos han resultado significativos en un análisis de la varianza a un nivel de confianza del 95%. Por lo tanto, tanto el cambio en el tipo de tratamiento como en la localidad influyen en la cantidad de materia seca obtenida. Además, tal y como se observó en el gráfico de interacción de la Figura 3.9, la relación entre ambos factores independientes influye en la variable respuesta. Para comprobar si el análisis realizado es correcto, se comprueban los supuestos del modelo: normalidad y homocedasticidad.

Tal y como se contempla en la Figura 3.10a, los residuos estandarizados no se ajustan a la distribución normal. Además, se ha realizado el test de Lilliefors (Kolmogorov-Smirnov) para probar la hipótesis nula de que los residuos provienen de una distribución normal. El valor del estadístico ha sido  $0.088$  y p-valor de la prueba de  $2.799e-07$ . Por lo tanto, se concluye que los residuos no siguen una distribución normal.

La Figura 3.10b muestra la distribución de los residuos del modelo. En esta imagen cabe destacar que en las primeras observaciones los residuos se distribuyen de una forma más o menos compacta y en los últimos esta misma se dispersa. Esta gran dispersión y distribución acampanada indica la presencia de heterocedasticidad en el modelo calculado.



(a) QQ-plot de los residuos estandarizados

(b) Distribución de los residuos del modelo

Como el modelo no es homocedástico ni normal, el ANOVA realizado no proporciona la información correcta, para ello habría que realizar una prueba robusta al no cumplimiento de estos supuestos.

Para intentar solucionar la falta de normalidad de los datos, se ha recurrido a una transformación de la variable respuesta. En la Figura 3.3 se observa una distribución asimétrica positiva de las observaciones, por lo que se ha empleado una **transformación logarítmica** de los datos. Sin embargo, esta transformación no mejora la normalidad ni la homocedasticidad del conjunto de datos, ya que invierte las observaciones y lo convierte en una asimetría negativa.

Por consiguiente, se ha optado por utilizar una prueba robusta a la falta de normalidad, para encontrar diferencias significativas: **Test de Kruskal-Wallis**.

### 3.2.1. Test no paramétrico para el análisis del rendimiento

El test no paramétrico que se va a realizar es una prueba de *Kruskal- Wallis*. Si el estadístico es significativo implica que al menos dos grupos entre los comparados difieren significativamente entre ellos. Sin embargo, esta prueba no indica entre que grupos existen diferencias significativas, para saberlo habría que realizar un test post-hoc.

---

**Kruskal-Wallis rank sum test**

---

Chi-Squared = 269.47

Df = 2

p-valor &lt;2.2e-16\*\*\*

**Chi- Squared:** valor del estadístico del test. **Df:** grados de libertad

Tabla 3.3: Prueba de Kruskal-Wallis para la materia seca en función de la localidad.

---

**Kruskal-Wallis rank sum test**

---

Chi-Squared = 56.143

Df = 15

p-valor = 1.146e-06\*\*\*

**Chi- Squared:** valor del estadístico del test. **Df:** grados de libertad

Tabla 3.4: Prueba de Kruskal-Wallis para la materia seca en función del tratamiento.

Examinando las Tablas 3.3 y 3.4 se puede inferir que existen diferencias significativas entre los diferentes tipos de tratamiento y de localidad para la variable respuesta *MS* (materia seca). Por lo tanto, se va a realizar un test post-hoc que compara los diferentes grupos para ver entre cuales hay diferencias.

La prueba de comparación más utilizada para el test de Kruskal-Wallis es el **test de Mann-Whitney** entre cada par de grupos con corrección de significancia. El método de corrección utilizado ha sido **Bonferroni**, ya que controla el aumento de los errores de Tipo I.

---

	<b>Viñalta</b>	<b>Castrillo</b>
<b>Viñalta</b>	<2e-16	-
<b>Castrillo</b>	<2e-16	<2e-16

---

Tabla 3.5: Comparaciones post-hoc para la variable Localidad.

Si se analiza la Tabla 3.5, se encuentran diferencias significativas entre todas las localidades siendo Zamadueñas la que obtiene mejores resultados y Castrillo como última localidad en cuanto a media de materia seca, tal y como se observa en la Tabla 3.6.

<b>Localidad</b>	<b>Materia seca</b>
Zamadueñas	277.35
Viñalta	130.05
Castrillo	42.32

Tabla 3.6: Media de materia seca según cada localidad.

La Tabla 3.7 separa por grupos de una forma más visual. Entre los grupos con la misma letra no existen diferencias significativas, mientras que grupos con distintas letras difieren significativamente entre ellos.

<b>Tratamiento</b>	<b>Materia seca</b>	<b>Grupos</b>
C66-G33	190,74	a
C100-G100	186,58	a
C100-G0	180,87	a
C100-G66	178,13	a
C66-G66	176,47	a
borde	169,28	a
C100-G33	166,14	a
C33-G33	165,72	a
C33-G100	165,57	a
C66-G100	162,46	a
C33-G0	149,44	a
C33-G66	148,57	a
C66-G0	147,77	a
C0-G100	85,70	b
C0-G66	61,81	b
C0-G33	53,21	b

Tabla 3.7: Comparaciones post-hoc entre los diferentes tratamientos.

Solo se han encontrado diferencias significativas entre el monocultivo de guisante respecto del resto de tratamientos, ya sean cultivos mixtos o monocultivo de cebada. Entre el resto de cultivos no existen diferencias. Es una prueba más de quien decide sobre la materia seca del cultivo, es la cebada. En ambos grupos destaca que no hay diferencias en cuanto a la densidad de siembra, es decir, desde el punto de vista productivo, me da igual sembrar al 100 % de la dosis, que al 33 %. Se podría deducir que la dosis de siembra no influye en el rendimiento y que los cultivos se comportan de forma plástica. Si siembro al 100 %, habrá más competencia entre individuos de la misma especie que si siembro al 33 %, donde tendrán los mismos recursos para menos plantas. Esto se traduciría en que habría pocas plantas pero más productivas en oposición al 100 %, muchas plantas pero menos productivas.

Para cerrar esta sección, se podría concluir que se obtienen mejores resultados para cultivos mixtos y monocultivo de cebada que para monocultivo de guisante.

#### 3.2.2. Análisis del rendimiento según localidades

Se quiere observar si existen diferencias significativas entre tratamientos en distintos tipos de ambientes. Para ello, se va a separar el conjunto de datos en tres distintos, según la localización de cada finca.

Se ha realizado un ANOVA para cada conjunto de datos, en los cuales se cumplen los supuestos de normalidad y homocedasticidad de los datos a un nivel 0.01. Además, en las tres localidades la variable tratamiento ha resultado significativa.

En las siguientes tablas, se puede observar el test de Tukey para cada localidad, según los tratamientos utilizados. La materia seca está medida en *gramos/muestra* siendo  $muestra = 0.25m^2$ .

En la primera Tabla 3.8, referente a la finca Zamadueñas, se puede deducir que salvo en el cultivo de cebada al 100% de su dosis de siembra, los mayores rendimientos los han obtenido los cultivos mixtos (grupo a), independientemente de su densidad de siembra. El monocultivo de guisante ha sido significativamente menor que el resto, independientemente también de su densidad de siembra (grupo c). Al igual que en las comparaciones totales, no hay diferencias entre cultivos mixtos y monocultivo de cebada.

Tratamiento	Materia seca	Grupos
borde	420.585	a
C100-G66	342.820	ab
C100-G100	341.285	ab
C66-G66	334.913	ab
C100-G0	331.978	ab
C66-G33	323.708	ab
C33-G33	302.609	ab
C66-G100	299.450	ab
C33-G100	299.209	ab
C100-G33	287.394	ab
C33-G0	282.074	b
C33-G66	276.640	b
C66-G0	267.390	b
C0-G100	152.156	c
C0-G66	123.643	c
C0-G33	123.413	c

Tabla 3.8: Comparaciones post-hoc en Zamadueñas entre los diferentes tratamientos.

En cuanto a la Tabla 3.9, correspondiente a la finca Viñalta, el monocultivo de guisante de densidad media y baja (66% y 33%, respectivamente) sí se ha diferenciado de forma significativa del resto de mezclas de cultivo (siendo los de menor rendimiento), en cambio, a diferencia de Zamadueñas, el monocultivo de guisante al 100% no se ha diferenciado de cualquier monocultivo de cebada (grupo c). El tratamiento del borde (50%:50%) ha sido el de mayor rendimiento, al igual que en Zamadueñas.

<b>Tratamiento</b>	<b>Materia seca</b>	<b>Grupos</b>
borde	198.011	a
C33-G33	164.433	ab
C100-G33	159.883	ab
C33-G100	157.233	ab
C100-G100	155.723	ab
C66-G33	146.714	ab
C100-G0	136.205	abc
C100-G66	135.675	bc
C66-G66	135.184	bc
C33-G0	133.849	bc
C33-G66	133.655	bc
C66-G100	132.486	bc
C66-G0	129.943	bc
C0-G100	79.111	cd
C0-G66	50.385	d
C0-G33	32.249	d

Tabla 3.9: Comparaciones post-hoc en Viñalta entre los diferentes tratamientos.

Por último, en la Tabla 3.10, referente a la localidad Castrillo de la Guareña, existe la menor separación entre los tratamientos considerados. Si en Zamadueñas el monocultivo de guisante se agrupa de forma bien definida, en Castrillo no sucede lo mismo. No hay diferencias significativas del monocultivo de guisante a baja densidad (33 %) con la mayoría de tratamientos, incluyendo algunas mezclas con cebada al 100 % de densidad (grupo c).

<b>Tratamiento</b>	<b>Materia seca</b>	<b>Grupos</b>
C100-G0	74.423	a
C100-G100	62.737	ab
C66-G66	59.303	ab
C100-G66	55.889	abc
C66-G100	55.450	abc
C100-G33	51.156	abc
C66-G33	48.444	abc
C66-G0	45.968	abc
borde	45.765	abc
C33-G100	40.241	abc
C33-G66	35.388	abc
C33-G0	32.383	abc
C33-G33	30.134	abc
C0-G100	25.831	abc
C0-G66	11.409	bc
C0-G33	3.953	c

Tabla 3.10: Comparaciones post-hoc en Castrillo entre los diferentes tratamientos.

### **3.2. Análisis del rendimiento según el tratamiento y la localidad**

---

Independientemente de significatividades, se han obtenido menores valores de rendimiento de materia seca en el monocultivo de guisante en todas las localidades. Por otro lado, si hay que elegir cuál es el mejor tratamiento, la combinación cebada – guisante al 33 % siempre se ha situado en los grupos más productivos en todas las localidades, y a falta de más información, es la que requiere menos semilla la hora de sembrar.

Por último, en Zamadueñas, hay una separación muy clara entre aquellos tratamientos que tienen cebada sobre aquellos que no (monocultivo de guisante). En el resto de localidades esta separación no está tan clara. Esto parece indicar que en Zamadueñas, el cultivo de cebada es muy dominante, más que en las otras dos localidades.



### 3.3. Evaluación usando diseño de series con reemplazo

En el Capítulo 2 se ha explicado el diseño de series con reemplazo que se va a realizar junto a los tratamientos (densidades relativas de ambos cultivos) utilizados. En este apartado, se va a calcular el coeficiente de hacinamiento relativo y total, para intentar explicar la dominancia entre especies y si existe una ventaja del cultivo mixto sobre el monocultivo.

En primer lugar, al calcular las variables de las regresiones ( $Y_c, Y_g, X_c, X_g, \beta_c, \beta_g$ ) con R, se han obtenido varios datos cuyo resultado es infinito (al dividir entre 0 cualquier número). Estos datos se han sustituido por NA (dato perdido).

Se van a realizar dos estimaciones. La primera para todo el conjunto de datos con dos regresiones, una para cada cultivo (cebada, guisante). En segundo lugar, seis regresiones, una para cada cultivo (cebada, guisante) y localidad (Finca Zamadueñas, Finca Viñalta, Castrillo de la Guareña) para explicar si existen diferencias de comportamiento entre diferentes ambientes.

#### 3.3.1. Evaluación del rendimiento de cultivos mixtos de forma conjunta

La primera regresión se ha realizado para el cultivo de cebada y se ha obtenido el gráfico de la Figura 3.11. En la gráfica se muestra la dispersión de las x's de cebada, resultado de la transformación de las densidades relativas; en función de las y's, variable transformada en función del LER, junto a la recta de regresión que ajusta los datos. Existe una alta variabilidad de los datos alrededor de la recta de regresión.

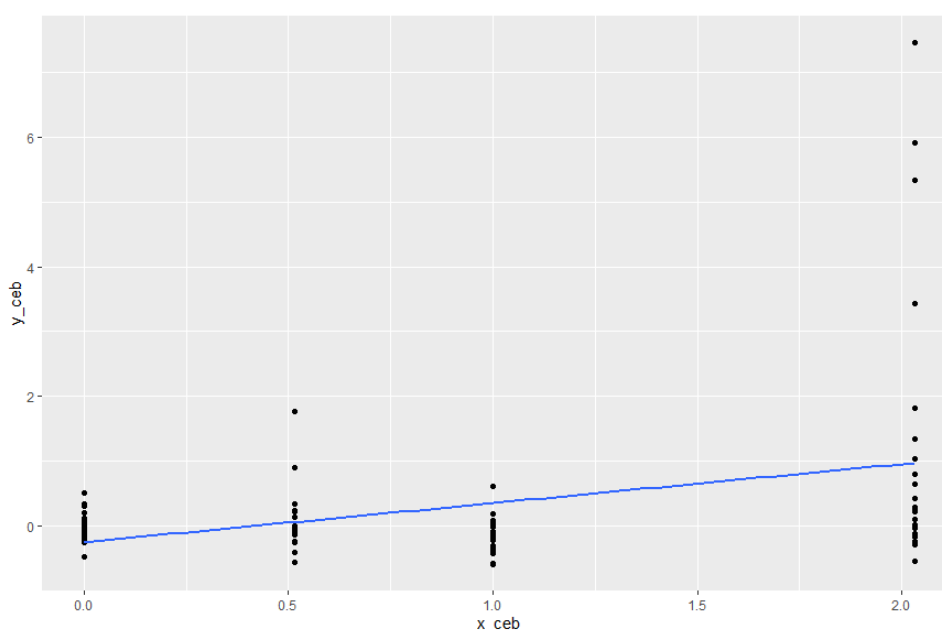


Figura 3.11: Grafico de dispersión de la cebada con la recta de regresión ajustada.

### 3.3. Evaluación usando diseño de series con reemplazo

	Estimate	Sd.	Error	t value	Pr(> t )
<i>Intercept</i>	-0.246		0.189	-1.298	0.198
<i>x_ceb</i>	0.601		0.160	3.742	0.0003 ***

Tabla 3.11: Resultados del ajuste de regresión para el cultivo cebada.

Una vez realizado el ajuste mediante regresión lineal, se obtiene un R-cuadrado de 0.136, un valor bajo que indica un mal ajuste del modelo a los datos, debido a la alta variabilidad de las observaciones. Sin embargo, el análisis de regresión ha resultado significativo ( $p$ -valor = 0.0003), al igual que el único regresor existente, tal y como se puede ver en la Tabla 3.11. Este regresor influye significativamente en la variable respuesta. Sustituyendo en la fórmula dada anteriormente 2.19:

$$k_{cg} = 1/\beta_c = 1/0.602 = 1.661$$

Donde  $k_{cg}$  es el coeficiente de hacinamiento de la cebada y  $\beta_c$  es el parámetro de la estimación de la regresión anteriormente ajustada.

Como este valor es mayor que 1, se puede concluir que la cebada produce más rendimiento del estimado.

La segunda regresión se ha realizado para el cultivo del guisante. El R-cuadrado del modelo obtenido es bajo ( $p$ -valor = 0.153), la recta no ajusta bien los datos, ya que existe una alta variabilidad entre los puntos del modelo.

Tal y como se aprecia en la Tabla 3.12, el regresor influye significativamente en la variable dependiente del modelo, cada vez que aumenta la  $x$  aumenta la  $y$  en 0.884 unidades.

	Estimate	Sd.	Error	t value	Pr(> t )
<i>Intercept</i>	0.261		0.873	0.299	0.766
<i>x_guis</i>	5.109		0.785	6.507	4.87e-09 ***

Tabla 3.12: Resultados del ajuste de regresión para el cultivo guisante.

Por último, si se analiza la Figura 3.12, se puede decir que hay una alta variabilidad de las observaciones con una tendencia lineal creciente de los datos.

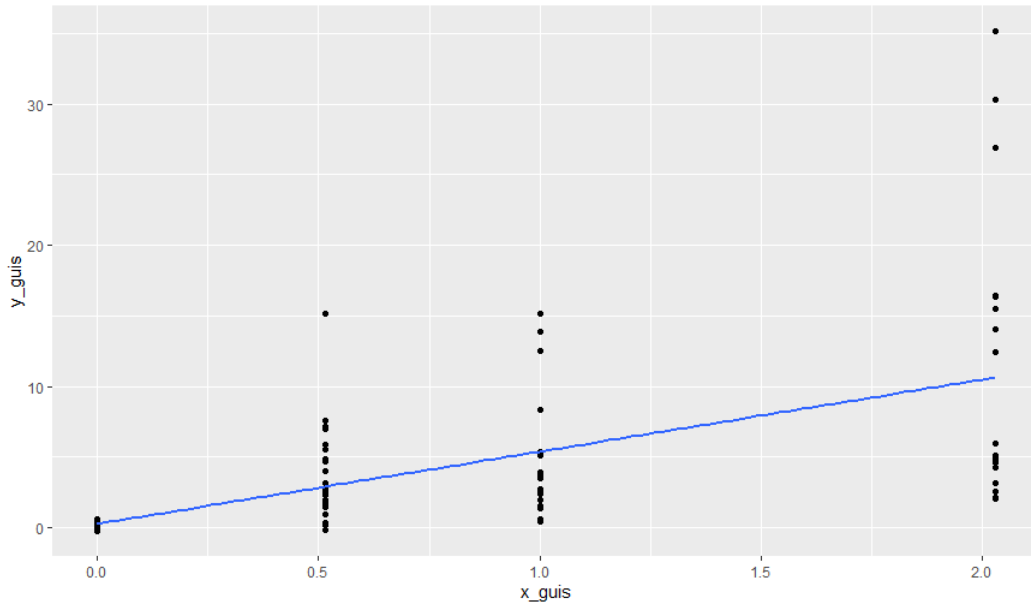


Figura 3.12: Grafico de dispersión del guisante con la recta de regresión ajustada.

Sustituyendo en la fórmula dada anteriormente 2.19:

$$k_{gc} = 1/\beta_g = 1/5.108 = 0.195$$

En el caso del guisante, el valor del coeficiente de hacinamiento es menor que 1, expresando que el guisante produce menos rendimiento del estimado.

Como ambos coeficientes de hacinamiento relativo son positivos, exhiben una competitividad entre especies débil. Además,  $k_{cg} > k_{gc}$ , indica que la cebada es el cultivo dominante y el guisante el cultivo dominado.

El coeficiente de hacinamiento global sería:  $K = k_{cg} * k_{gc} = 1.661 * 0.195 = 0.325$ , indicando que no existe una ventaja global de los cultivos mixtos respecto al monocultivo.

### 3.3.2. Evaluación del rendimiento de cultivos mixtos para diferentes ambientes

Se han calculado los coeficientes de hacinamiento para cada cultivo y localidad correspondiente. En la Tabla 3.13 se pueden los coeficientes de hacinamientos relativos a cada cultivo, el coeficiente de hacinamiento global, las pendientes estimadas de cada recta, y el R-cuadrado de cada regresión, según cada ambiente.

### 3.3. Evaluación usando diseño de series con reemplazo

Ambiente	Coeficientes de hacinamiento			Pendiente estimada		R- cuadrado	
	$k_{cg}$	$k_{gc}$	$K = k_{cg} * k_{gc}$	$\beta_c$	$\beta_g$	Cebada	Guisante
<b>Zamadueñas</b>	6.369	0.094	0.601	0.157	10.578	0.242	0.601
<b>Viñalta</b>	3.876	0.478	1.853	0.258	2.092	0.085	0.4716
<b>Castrillo</b>	0.751	0.665	0.5	1.332	1.504	0.255	0.217

$k_{cg}$  : coeficiente de hacinamiento relativo de la cebada.  $k_{gc}$  : coeficiente de hacinamiento relativo del guisante.  $K$  : coeficiente de hacinamiento total.  $\beta_c$  : pendiente estimada resultado de la regresión del cultivo de cebada.  $\beta_g$  : pendiente estimada resultado de la regresión del cultivo de guisante.

Tabla 3.13: Resumen de cada regresión estimada para cada localidad

Los valores de  $k_{cg}$  son mayores que los valores de  $k_{gc}$  en todos los ambientes. Este resultado indica que la cebada es el cultivo dominante y el guisante el cultivo dominado. Además, cuanto peores condiciones edafoclimáticas existen menor es la diferencia entre coeficientes de hacinamiento ( $k_{cg} - k_{gc}$ ) indicando menos diferencia entre rendimiento de cultivos.

En los casos de cultivos mixtos,  $K$  indica una tendencia similar al LER. Por lo tanto, existe ventaja de cultivos asociados únicamente en Viñalta, ya que  $K > 1$ .

Resumiendo, tal y como se ha visto anteriormente, la cebada es el cultivo dominante en todas las localidades. Sin embargo, el grado de dominancia depende de la localidad, existiendo en Castrillo muy poca diferencia de dominancia entre cultivos.

Se ha calculado el número de observaciones por encima del umbral del LER  $y = 1$  (Tabla 3.14) para cada localidad).

	<1	>1
<i>Castrillo</i>	49	78
<i>Viñalta</i>	51	77
<i>Zamadueñas</i>	67	58

Tabla 3.14: Número de observaciones por encima y por debajo del umbral ( $LER$ ) para cada localidad.

En función de los valores de  $K$ , únicamente es rentable utilizar cultivos mixtos en Viñalta, ya que como se ha observado en la Figura 3.5 y en la Tabla 3.14 hay más valores por encima del umbral de  $LER y=1$ , que por debajo y existe muy poca variabilidad entre los datos. En Castrillo, también existen más puntos por encima del umbral que por debajo (Tabla 3.14). Sin embargo, la gran desviación típica existente entre los datos resulta en que no existan diferencias entre cultivos mixtos y monocultivos (Figura 3.6).

### 3.4. Evaluación mediante análisis de superficie respuesta

Por último, para investigar el efecto que tiene el variar las densidades relativas dentro de los valores de la variable respuesta, se va a realizar un **análisis de superficie respuesta**.

En primer lugar, se quiere encontrar una combinación de densidades de guisante y cebada que optimice la materia seca obtenida. Para ello se han creado dos nuevas variables que indican las densidades relativas de guisante y cebada de cada observación. El rango sobre el cual va a trabajar cada factor será:

- $x_c$ : [-1, 1] (cultivo 1, cebada)
- $x_g$ : [-1, 1] (cultivo 2, guisante)

Seguidamente, se calcularán modelos de primer orden y de segundo orden comparando cuál se ajusta mejor a los datos. Por consiguiente, hay que realizar una transformación de las variables para que se encuentren en la escala [-1, 1]. La variable respuesta que se quiere analizar es la materia seca.

La transformación utilizada para codificar los factores se muestra a continuación (Ecuación 3.1):

$$x_i = \frac{\xi - [\max(\xi_i) + \min(\xi_i)]/2}{[\max(\xi_i) - \min(\xi_i)]/2} \quad (3.1)$$

Donde  $\xi_i$  son **variables naturales** en forma de factores de carácter cuantitativo.

Las variables codificadas resultantes de la transformación de las variables naturales serían:

$$x_1 = \frac{\xi_1 - 0.5}{0.5} \quad (3.2)$$

$$x_2 = \frac{\xi_2 - 0.5}{0.5} \quad (3.3)$$

Se pretende generar un **diseño central compuesto circunscrito** (CCC) para evaluar el ajuste de un polinomio completo de segundo grado. La función que se va a maximizar se ha detallado a continuación (Ecuación 3.4):

$$\hat{Y}_{ms} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1r} x_1^2 + \beta_{2r} x_2^2 + \beta_{12} x_1 x_2 \quad (3.4)$$

Donde  $Y_{ms}$  es la cantidad de materia seca que maximiza el modelo,  $x_i$  son las variables codificadas resultado de la transformación y cada  $\beta_i$  es el parámetro de estimación de cada término del modelo que maximiza la cantidad de materia seca total.

En el siguiente gráfico (Figura 3.13) se puede observar el diseño llevado cabo según las densidades

### 3.4. Evaluación mediante análisis de superficie respuesta

relativas de cada cultivo, en el rango [-1 , 1].

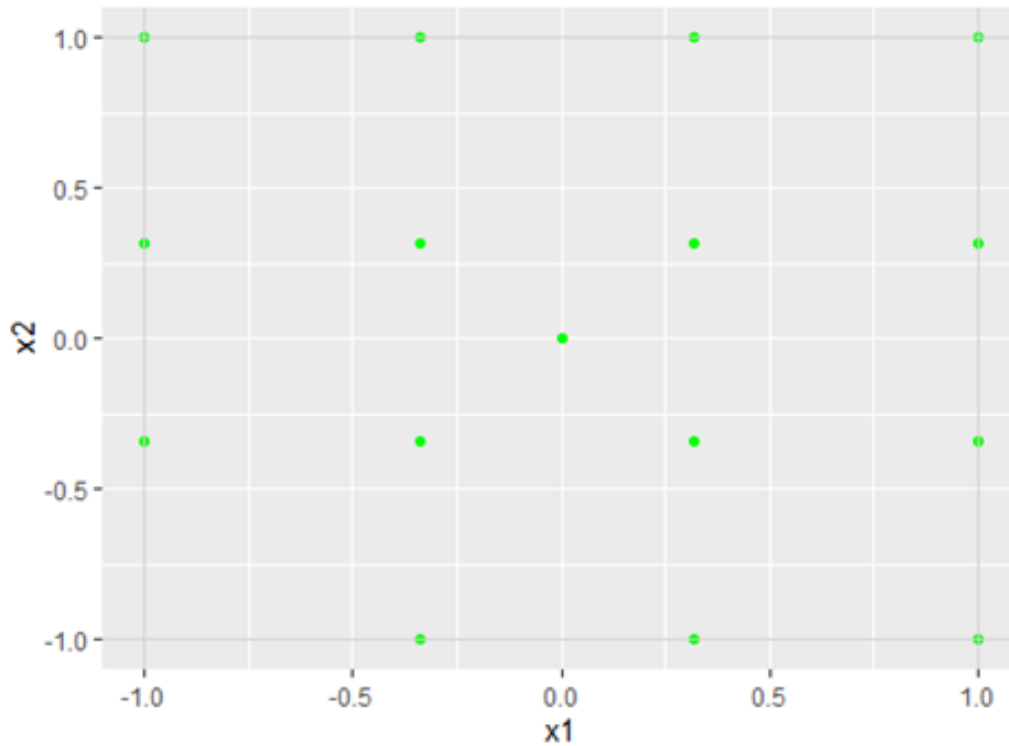


Figura 3.13: Diseño CCC de las densidades relativas de ambos cultivos.  
 $x_1$ : variable codificada del cultivo del guisante.  $x_2$ : variable codificada del cultivo de cebada

Se va a utilizar el paquete de R *rsm* para realizar el análisis y al ajustar el modelo cuadrático con interacción.

	Estimate	Std. Error	t value	Pr(> t )	
<i>Intercept</i>	169.200	10.800	15.6677	<2.2e-16	***
$x_1$	52.853	8.616	6.134	2.182e-14	***
$x_2$	7.034	8.610	0.817	0.414	
$x_1 : x_2$	-4.547	12.541	-0.363	0.717	
$x_1^2$	-47.592	13.148	-3.620	0.0003	***
$x_2^2$	-0.660	13.145	-0.050	0.960	

**Estimate:** estimación del parámetro del modelo. **Std. Error:** error estándar, que se calcula dividiendo la desviación estándar por la raíz cuadrada del número de observaciones. **t value:** valor del estadístico t. **Pr(>|t|):** p-valor de la variable. **Intercept:** término independiente.  $x_1$ : variable codificada del guisante.  $x_2$ : variable codificada de la cebada.  $x_1x_2$ : interacción de las variables codificadas del guisante y la cebada, respectivamente.  $x_1^2$ : variable codificada del guisante al cuadrado.  $x_2^2$ : variable codificada de la cebada al cuadrado

Tabla 3.15: Modelo completo de superficie de respuesta ajustado.

Una vez ajustado el modelo completo (Tabla 3.15), se ha observado que la interacción entre factores ( $\beta_{12}$ ) no ha resultado significativa a un nivel 0.05 ( $p$ -valor = 0.717). Por consiguiente, se van a comparar tres modelos: un modelo de primer orden lineal sin interacción (Modelo 1), un modelo de segundo orden eliminando la interacción ya que no ha resultado significativa en el modelo completo (Modelo 2) y por último, un modelo cuadrático sin interacción y solamente con el término cuadrático de la variable  $x_1$ , ya

## Análisis y resultados

que el coeficiente  $\beta_{2r}$  (coeficiente del término cuadrático  $x_2$ ) no es significativo en el modelo completo, con un  $p$ -valor = 0.960 (Modelo 3).

Las características de los tres modelos se han resumido en la siguiente Tabla 3.16:

	R - cuadrado	R-cuadrado ajustado	Autovalores	Fórmula
<b>Modelo 1</b>	0.084	0.079	-	$\hat{Y}_{ms} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
<b>Modelo 2</b>	0.115	0.105	$x_1 = 0.327,$ $x_2 = -46.606$	$\hat{Y}_{ms} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1r} x_1^2 + \beta_{2r} x_2^2$
<b>Modelo 3</b>	0.115	0.108	$x_1 = 0,$ $x_2 = -46.613$	$\hat{Y}_{ms} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1r} x_1^2$

Tabla 3.16: Comparación de los 3 modelos de superficie respuesta

Además, los p-valores de cada modelo se muestran en la siguiente Tabla 3.17:

Modelo	P-valor
Modelo 1	7.546e-08 ***
Modelo 2	2.860e-09 ***
Modelo 3	6.413e-10 ***

Tabla 3.17: P-valor de cada modelo ajustado.

A pesar de que los tres modelos han resultado significativos (Tabla 3.17), el modelo 1 (modelo de primer orden) queda descartado debido a que ajusta peor los datos, ya que el R-cuadrado y el R-cuadrado ajustado son más bajos que para los otros dos modelos.

Entre el modelo 2 y el modelo 3 no se encuentra mucha diferencia entre coeficientes de determinación, ya que los datos se ajustan de forma muy parecida en ambos. Por consiguiente, se van a desarrollar ambos modelos para cerciorarse cuál realiza un mejor ajuste de los datos.

3.4.1. Modelo completo sin interacción

El modelo 2 es un modelo significativo ( $p\text{-valor} = 2.51e-15$ , Tabla 3.17) pero con un coeficiente de determinación bastante pobre,  $R^2 = 0.180$  (Tabla 3.16), indicando una baja calidad del ajuste. La ecuación estimada se muestra en la siguiente Fórmula 3.5 (Tabla 3.18):

$$\hat{Y}_{ms} = 168.598 + 52.021x_1 + 6.210x_2 - 46.606x_1^2 + 0.327x_2^2 \tag{3.5}$$

Los resultados obtenidos del análisis ANOVA del *Diseño Central Compuesto (DCC)* (Tabla 3.18 han mostrado que la variable con una alta significancia es la densidad relativa de la cebada ( $p\text{-valor} = 9,868e-10$ ), así como el cuadrado de esta ( $p\text{-valor} = 0,0003$ ). Por otro lado, ni la densidad relativa del guisante ( $p\text{-valor} = 0,454$ ) ni su cuadrado ( $p\text{-valor} = 0,980$ ) han resultado significativas en el modelo, dando por sentado que altas o bajas concentraciones de guisante no afectan de forma considerable al rendimiento ( $p\text{-valor} < 0,05$ ).

	Estimate	Std. Error	t value	Pr(> t )	
<i>Intercept</i>	168.598	10.659	15.818	<2.2e-16	***
$x_1$	52.021	8.295	6.272	9.868e-10	***
$x_2$	6.210	8.290	0.750	0.454	
$x_1^2$	-46.606	12.849	-3.627	0.0003	***
$x_2^2$	0.327	12.848	0.025	0.980	

**Estimate:** estimación del parámetro del modelo. **Std. Error:** error estándar, que se calcula dividiendo la desviación estándar por la raíz cuadrada del número de observaciones. **t value:** valor del estadístico t. **Pr(>|t|):** p-valor de la variable. *Intercept:* término independiente.  $x_1$ : variable codificada del guisante.  $x_2$ : variable codificada de la cebada.  $x_{1^2}$ : variable codificada del guisante al cuadrado.  $x_{2^2}$ : variable codificada de la cebada al cuadrado

Tabla 3.18: Modelo completo sin interacción de superficie de respuesta ajustado.

R proporciona un análisis canónico (eficaz para visualizar la superficie respuesta), ya que se ha ajustado un modelo de grado dos. Las coordenadas del punto estacionario u óptimo, en variables naturales, serían:  $x_c = 0.779$ ,  $x_g = -4.244$ . Los autovalores del análisis canónico  $\lambda_i$  para la superficie de materia seca obtenidos son  $\lambda_1 = 0.327$  y  $\lambda_2 = -46.606$ . Como un coeficiente es positivo,  $\lambda_1 > 0$  y otro negativo  $\lambda_2 < 0$ , entonces el resultado de cualquier movimiento que se aleja del origen a lo largo del eje  $Z_1$  (variable  $x_1$  en el eje rotado) aumenta la respuesta, mientras que a lo largo del  $Z_2$  (variable  $x_2$  en el eje rotado) la disminuye. El punto estacionario es un **punto de respuesta de silla**.

En la la Figura 3.14 se puede observar el comportamiento individual de las variables del modelo, donde el aumento de la densidad de la cebada tiene un efecto muy positivo y cuadrático en la materia seca, ya que a una densidad cercana al 0.8, esta vuelve a decrecer. El incremento de la concentración de cultivo de guisante tiene un efecto lineal creciente en la variable respuesta, pero muy suave, que no resulta especialmente significativo en los resultados.

Además de lo mencionado anteriormente, el punto de ensilladura ( $x_c = 0.779$ ,  $x_g = -4.244$ ) encontrado en el modelo donde la pendiente es 0, establece que hay un punto máximo en una dirección y mínimo perpendicularmente. Sin embargo, este punto de ensilladura no es interesante, ya que no aparece en la Figura 3.14, porque el valor de la coordenada del guisante (-4.244) está por debajo del rango



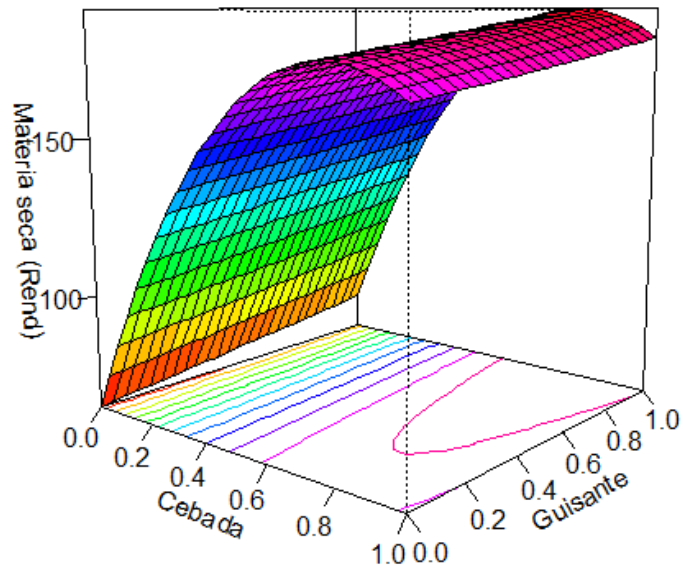


Figura 3.14: Gráfico 3D del modelo 2 de superficie respuesta de las densidades relativas de la cebada y el guisante.

establecido.

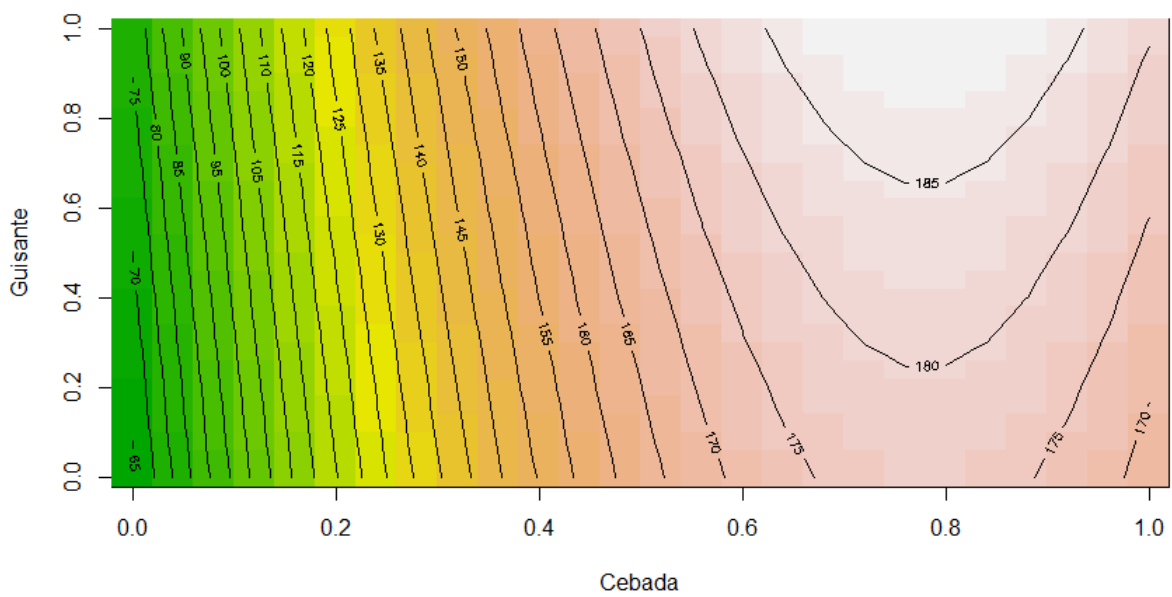


Figura 3.15: Gráfico de contorno del modelo 2 de las densidades relativas de la cebada y el guisante.

### 3.4. Evaluación mediante análisis de superficie respuesta

Por último, examinando el gráfico de contorno de la Figura 3.15 se deduce que las mayores cantidades de materia seca se obtienen con la presencia de cebada, aumentando el rendimiento, la presencia de guisante, únicamente alrededor de un 5%.

#### 3.4.2. Modelo completo sin interacción y sin un término cuadrático

El modelo 3 también es significativo ( $p$ -valor =  $6.652e-16$ , Tabla 3.17). La Ecuación 3.6 del modelo se muestra a continuación (Tabla 3.19):

$$\hat{Y}_{ms} = 168.761 + 52.046x_1 + 6.234x_2 - 46.613x_1^2 \quad (3.6)$$

	Estimate	Std. Error	t value	Pr(> t )	
<i>Intercept</i>	168.761	8.500	19.857	<2.2e-16	***
$x_1$	52.046	8.224	6.329	7.062e-10	***
$x_2$	6.234	8.221	0.759	0.449	
$x_1^2$	-46.613	12.828	-3.634	0.0003	***

**Estimate:** estimación del parámetro del modelo. **Std. Error:** error estándar, que se calcula dividiendo la desviación estándar por la raíz cuadrada del número de observaciones. **t value:** valor del estadístico t. **Pr(>|t|):** p-valor de la variable. *Intercept:* término independiente.  $x_1$ : variable codificada del guisante.  $x_2$ : variable codificada de la cebada.  $x_1^2$ : variable codificada del guisante al cuadrado.

Tabla 3.19: Modelo completo sin interacción y sin un término cuadrático de superficie respuesta ajustado.

Tal y como ocurría en el modelo sin interacción y como se puede observar en la Tabla 3.19, la variable más significativa es la densidad relativa de la cebada ( $p$ -valor =  $7.062e-10$ ), así como su cuadrado,  $p$ -valor =  $0.0003$ . La densidad relativa del guisante, no ha resultado significativa a un nivel de confianza del 95% ( $p$ -valor =  $0.449$ ), pero también se desea estudiar el efecto del guisante y de la mezcla de ambos cultivos, por lo que eliminarlo del modelo sería un error.

El análisis canónico proporcionado por R establece las coordenadas del punto estacionario u óptimo de las variables naturales en  $x_c = 0.779$  y  $x_g = 0.5$  y los autovalores en  $\lambda_1 = 0$  y  $\lambda_2 = -46.613$ . El punto estacionario tiene un mínimo por un lado y una línea horizontal por otro. El punto es candidato a ser **máximo relativo**.

La distribución de la materia seca según ambos cultivos se muestra en la Figura 3.16. En este caso, ya no existe un punto de silla en el gráfico, sino un punto en el que se podría optimizar la cantidad de materia seca obtenida. Este punto se encuentra en el  $x_c = 0.779$  y  $x_g = 0.5$ .

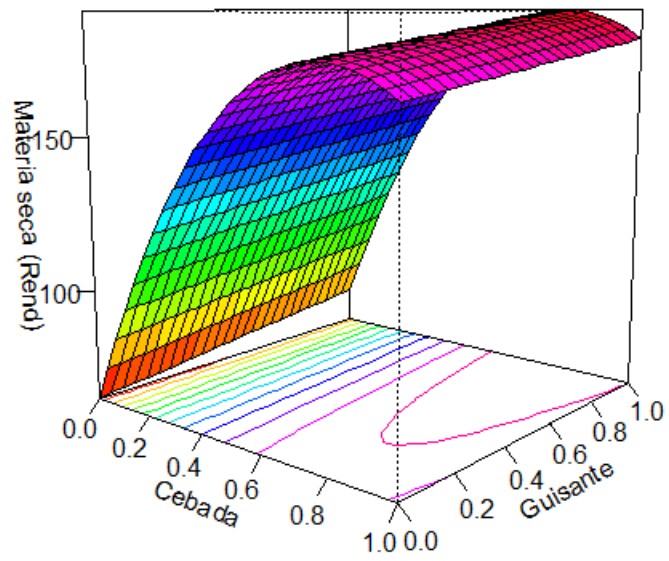


Figura 3.16: Gráfico 3D del modelo 3 de superficie respuesta de las densidades relativas de la cebada y el guisante.

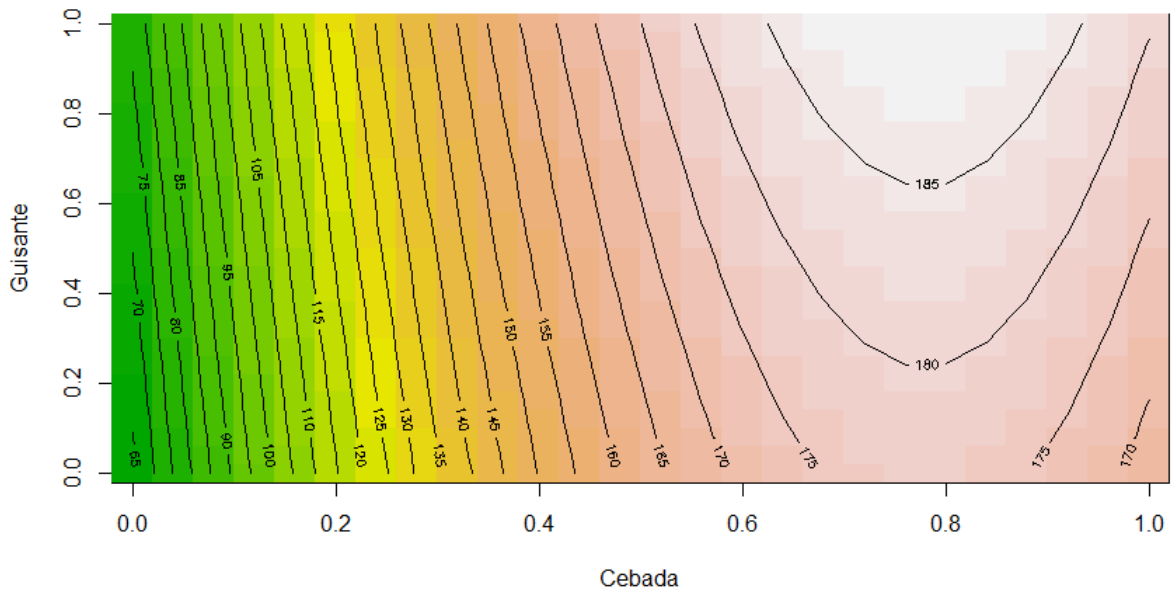


Figura 3.17: Gráfico de contorno del modelo 3 de las densidades relativas de la cebada y el guisante.

Por último, observando el gráfico de contorno (Figura 3.17) y el diagrama 3D (Figura 3.16), el cultivo de cebada se distribuye de una forma cuadrática con una pendiente lineal de 52.046 y una pendiente cuadrática de  $-46.613$ , mientras que el guisante se distribuye de una forma lineal creciente llegando con una pendiente lineal de 6.234.

#### 3.4.3. Elección del mejor modelo

En este apartado del proyecto, se va a elegir el modelo que mejor ajusta y explica los datos.

Tanto el modelo 2 como el modelo 3, tienen coeficientes de determinación excesivamente bajos (Tabla 3.16), lo que hace que se encuentren muy poco ajustados debido a la dispersión de los datos y su heterogeneidad. Sin embargo, los regresores de los modelos siguen siendo significativos y por lo tanto, también su interpretación. Este R-cuadrado tan bajo también se debe a que el modelo se ha ajustado con todas las localidades, existiendo múltiples diferencias entre ellas.

El R-cuadrado de ambos modelos es el mismo. Sin embargo, el modelo 3 cuenta con un regresor menos. Además, tanto los gráficos de contorno como los gráficos 3D son muy similares indicando un punto candidato a óptimo alrededor de la coordenada 0.8 de la cebada y 0.5 del guisante. Por lo tanto, el **modelo elegido sería el 3**, debido a que ajusta de una forma similar los datos, pero de una forma más sencilla.

El cultivo de cebada se distribuye de una forma cuadrática aumentando en 52.046 kg/ha la materia seca cada porcentaje que se aumenta la densidad de siembra de la cebada. Por otro lado, el guisante se distribuye de una forma lineal creciente. Por cada porcentaje que se aumenta la densidad de siembra del guisante, la materia seca total aumenta en 6.234 kg/ha.

Finalmente, cabe destacar que el cultivo de guisante no aportaría mucha cantidad de materia seca, ya que se obtiene casi la misma cantidad de materia seca con nulas concentraciones de guisante que con altas (diferencia en un 6% más de materia seca, de 175 a 185; Figura 3.17). De forma global, en las tres localidades, sería más interesante cultivar únicamente cebada.

## Capítulo 4

# Conclusiones

En este Capítulo, se van a exponer las diferentes ideas llevadas a cabo durante el desarrollo del proyecto, así como las posibles mejoras que se podrían llegar a realizar en un futuro.

### 4.1. Conclusiones

Este trabajo concluye que, la materia seca total utilizada como forraje, depende en gran medida de la materia seca del cultivo dominante, que en este caso ha sido la cebada. El grado de dominancia de la cebada depende de la localidad tal que, cuanto mayor potencial productivo existe en el ambiente mayor dominancia expresa la cebada sobre el resto de cultivos.

Por otro lado, el guisante ha sido el cultivo dominado en este estudio, ya que aportando poca cantidad de materia seca obtiene casi el mismo rendimiento con altas concentraciones de guisante que con bajas. Además, el monocultivo de guisante supondría una mala elección si el ambiente tiene unas condiciones edafoclimáticas aceptables o buenas. En el caso de una localidad con malas condiciones de suelo y clima (en este caso Castrillo de la Guareña), el guisante es capaz de fijar nitrógeno de mejor forma que la cebada, obteniendo cantidades de materia seca parecidas a las que obtiene en un buen ambiente.

El éxito de los cultivos mixtos depende en gran parte del ambiente. Hay lugares donde compensan estos cultivos asociados (Finca Viñalta), mientras que en otros no (Finca Zamadueñas y Castrillo de la Guareña). En un trabajo futuro sería interesante incluir factores climáticos para saber de que condiciones de suelo o clima dependen estos cultivos.

La cebada produce mayor rendimiento del esperado siguiendo una distribución cuadrática con una pendiente de 52.046, mientras que el guisante produce menor forraje del estimado, siguiendo una distribución lineal creciente con una pendiente suave de 6.234.

Por otro lado, la dosis de siembra utilizada no influye en el rendimiento de la materia seca total. Por lo tanto, los cultivos se comportan de forma plástica, obteniendo mejores resultados para los monocultivos

de cebada y los cultivos mixtos, pero sin diferencias entre ellos. Por consiguiente, sería interesante cultivar bajas densidades como C33 - G33, C33 - G0 o C66 - G33, ya que se produce un ahorro en semilla utilizada.

Por último, cuanto más contenido exista en materia seca de malas hierbas, menos materia seca total, ya que se halla una competencia de la flora adventicia. Asimismo, la población de malas hierbas encontrada presenta una distribución exponencial. Esta distribución indica que, cuando la población de malas hierbas es alta, pequeñas infestaciones de estas sobre el cultivo apenas producen disminución del rendimiento del forraje.

## 4.2. Trabajo Futuro

Después de la finalización del proyecto y de analizarlo de forma general, así como las ideas aportadas por los tutores, se han planteado algunas de las posibles mejoras para dar mayor calidad al proyecto:

- **Análisis de mayor cantidad de datos:** Realización del mismo análisis para los diferentes estados del cultivo.
- **Introducción de varias variables respuesta:** como los contenidos en proteínas, grasas, azúcares... de la materia seca obtenida. Todo ello para realizar un análisis más en profundidad.
- **Utilización de covariables climáticas** para saber de que condiciones de suelo o clima dependen cada cultivo.
- **Añadir un diseño de bloques aleatorizado** con mayor cantidad de tratamientos.

# Apéndices





## Apéndice A

# Manual de utilización

Los datos empleados durante el desarrollo del proyecto y el script de R donde se ha programado el trabajo se encuentran en el repositorio: <https://github.com/elvdelg/Estudio-del-rendimiento-de-cultivos-mixtos.git>



## Apéndice B

# Script del análisis

Se han utilizado diferentes librerías de R para la creación de este script. Algunas de ellas serían *ggplot2* [28] y *PerformanceAnalytics* [29] para crear las representaciones gráficas; *nortest* [30] y *lmtest* [31] para la realización de test de normalidad y homocedasticidad; *rsm* [32] para la realización de análisis de superficie respuesta; etc.

```
library(readxl)
library(dplyr)
library(ggplot2)
library(PerformanceAnalytics)
library(highcharter)
library(EnvStats)
library(robustbase)
library(MASS)
library(nortest)
library(lmtest)
library(rsm)

data_lechoso <- read_excel("forrajes_lechosos.xlsx")
str(data_lechoso)

#Convertir factores las variables categoricas
data_lechoso<- mutate_at(data_lechoso, vars(1:10), as.factor)

#Se elimina la fila que tiene una C como muestra, ya que es un error
data_lechoso <-data_lechoso[data_lechoso$MUE != 'C', ]

#Gráfico de la materia seca según localidad
ggplot(data_lechoso,
       aes(Localidad, MS)) + geom_boxplot()
```

---

```

aggregate(MS~ Localidad,data_lechoso, var)
aggregate(MS~ Localidad,data_lechoso, sum)

#Gráfico de la materia seca según tratamiento
ggplot(data_lechoso,aes(TRAT, MS)) + geom_boxplot()
aggregate(MS~ TRAT,data_lechoso, var)

#Gráfico materia seca VS Localidad y Tratamiento
ggplot(data_lechoso,
       aes(MS, TRAT)) + geom_boxplot(aes(colour = Localidad))

#Correlacion de las variables numéricas
chart.Correlation(data_lechoso[, c(11:19)], histogram = TRUE, method = "spearman")

#Materia seca VS LER según localidad. Se dibuja la #linea y = 1
ggplot(data_lechoso, aes(MS, LER)) + geom_point(aes(colour = Localidad)) + geom_hline(yintercept = 1)

#Materia seca VS numero de malas hierbas según #localidad
ggplot(data_lechoso, aes(MS, PesoMS_mh)) + geom_point(aes(colour = Localidad))

#Materia seca VS peso materia seca de cebada según localidad
ggplot(data_lechoso, aes(MS, PesoMS_ceb)) + geom_point(aes(colour = Localidad))

#Materia seca VS peso materia seca de cebada según tratamiento
ggplot(data_lechoso, aes(MS, PesoMS_ceb)) + geom_point(aes(colour = TRAT))

#Materia seca VS peso materia seca de guisante segun localidad
ggplot(data_lechoso, aes(PesoMS_guis, MS)) + geom_point(aes(colour = Localidad))

#Materia seca VS peso materia seca de guisante segun tratamiento
ggplot(data_lechoso, aes(PesoMS_guis, MS )) + geom_point(aes(colour = TRAT))

#Función que calcula outliers de una variable en un
#conjunto de datos
fun_outliers <- function(data, var, k = 2) {
  df <- data_to_boxplot(data, var, add_outliers=TRUE)
  highchart() %>%
    hc_add_series_list(df)

  print(boxplot.stats(var))
}

```

## Script del análisis

---

```
#Prueba de Rosner. Deteccion de outliers.
test <- rosnerTest(var, k, warn = TRUE)

print(test)
}

fun_outliers(data_lechoso, data_lechoso$MS)

#Gráfico de interacción entre localidad y tratamiento
ggplot(data = data_lechoso, aes(x = TRAT, y = MS, colour = Localidad,
                                group = Localidad)) +
  stat_summary(fun = mean, geom = "point") +
  stat_summary(fun = mean, geom = "line") +
  labs(y = 'mean (Rendimiento)') +
  theme_bw()

#Comprobacion de outliers de la variable respuesta
caja <- boxplot(data_lechoso$MS, col = "skyblue", ylab = "Materia seca (g/muestra)")
caja$out

#Regresión robusta
a1<-lmrob.control()
a1$k.max<-210
a1$refine.tol<-1e-07
a1$setting= "KS2014"

reg2rob<-lmrob(MS ~ TRAT + Localidad + TRAT*Localidad,
               data = data_lechoso, na.action=na.exclude,control=a1)
summary(reg2rob)

#Error del modelo
reg2rob$s #26.4569

#ANOVA
modelo <- aov(reg2rob, data = data_lechoso)
tabla<- as.matrix(summary(modelo))

plot(modelo)

#Comprobación de la normalidad y homocedasticidad
qqPlot(modelo, main="QQ Plot")
```

---

```

sresid <- studres(modelo)
hist(sresid, freq=FALSE,
     main="Distribution of Studentized Residuals")
xfit<-seq(min(sresid), max(sresid),length = 40)
yfit<-dnorm(xfit)
lines(xfit, yfit)

lillie.test( modelo$residuals)

bptest(modelo)

#Transformación logaritmica para solucionar la falta de normalidad y homocedasticidad
#data_lechoso$MS<- log(data_lechoso$MS)

#Test no paramétrico: Kruskal Wallis
kruskal.test(MS ~ Localidad, data = data_lechoso)
kruskal.test(MS ~ TRAT, data = data_lechoso)

#Test post-hoc
pairwise.wilcox.test(x = data_lechoso$MS, g = data_lechoso$Localidad, p.adjust.method = "bonferroni")
aggregate(MS ~ Localidad, data = data_lechoso, mean)
pairwise.wilcox.test(x = data_lechoso$MS, g = data_lechoso$TRAT, p.adjust.method = "bonferroni" )
a<-aggregate(MS ~ TRAT, data = data_lechoso, mean)
a_ordenado <- a[order(a$MS, decreasing = TRUE), ]

#Comparacion post-hoc con letras
library(agricolae)
HSD.test(modelo, "TRAT", group =TRUE, console = TRUE)

#Comparaciones post-hoc según localidades

#Zamadueñas
data_zam<- filter(data_lechoso, Localidad == "Zamadueñas")

modelo_zam <- aov(MS ~ TRAT ,
                 data = data_zam)
summary(modelo_zam)

lillie.test(modelo_zam$residuals)
bptest(modelo_zam)
HSD.test(modelo_zam, "TRAT", group =TRUE, console = TRUE)

```

## Script del análisis

---

```
#Castrillo
data_cas<- filter(data_lechoso, Localidad == "Castrillo")

modelo_cas <- aov(MS ~ TRAT ,
                 data = data_cas)
summary(modelo_cas)

lillie.test(modelo_zam$residuals)
bptest(modelo_cas)
HSD.test(modelo_cas, "TRAT", group =TRUE, console = TRUE)

#Viñalta
data_vin<- filter(data_lechoso, Localidad == "Viñalta")

modelo_vin <- aov(MS ~ TRAT ,
                 data = data_vin)
summary(modelo_vin)

lillie.test(modelo_vin$residuals)
bptest(modelo_vin)
HSD.test(modelo_vin, "TRAT", group =TRUE, console = TRUE)

#####Evaluación mediante diseño de series con reemplazo#####
#Estimación de coeficientes de hacinamiento

#Calculo de la segunda densidad
data_lechoso[data_lechoso$TRAT == "borde", "den_rel_cul2"] = 0.5
data_lechoso[substr(data_lechoso$TRAT,
                   nchar(as.character(data_lechoso$TRAT))-1,
                   stop = nchar(as.character(data_lechoso$TRAT)))== "00", "den_rel_cul2"] = 1

data_lechoso[substr(data_lechoso$TRAT,
                   nchar(as.character(data_lechoso$TRAT))-1,
                   stop = nchar(as.character(data_lechoso$TRAT)))== "33", "den_rel_cul2"] = 0.33

data_lechoso[substr(data_lechoso$TRAT,
                   nchar(as.character(data_lechoso$TRAT))-1,
                   stop = nchar(as.character(data_lechoso$TRAT)))== "66", "den_rel_cul2"] = 0.66
```

---

```

data_lechoso[substr(data_lechoso$TRAT,
  nchar(as.character(data_lechoso$TRAT))-1,
  stop = nchar(as.character(data_lechoso$TRAT)))=="G0", "den_rel_cul2"] = 0

#Para todas las localidades de forma conjunta

#Calculo de las variables transformadas
data_lechoso<-mutate(data_lechoso, y_ceb = (1/LER_cultivo1)-1,
  x_ceb = ((1-den_rel_cul1)/den_rel_cul1),
  y_guis = (1/LER_cultivo2)-1,
  x_guis = ((1-den_rel_cul2)/den_rel_cul2))

#Cultivos que cumplan el diseño del 100%
data_series<- filter(data_lechoso,( TRAT=="C66-G33" | TRAT=="C33-G66"
  | TRAT=="borde" | TRAT=="C100-G0" | TRAT=="CO-G100") )

plot(data_series$y_guis)
plot(data_series$y_ceb)

#Se sustituye el infinito por valor perdido (NA)
data_series[data_series == Inf] <- NA
data_series[data_series == -Inf] <- NA

#Cultivo de cebada

#Comprobacion de outliers
fun_outliers(data_series, data_series$y_ceb, 10)
#Eliminacion de outliers y valores que afectan al ajuste
data_series_ceb <- data_series[-c(38),]

reg1<- lm(y_ceb ~ x_ceb,data = data_series_ceb)
summary(reg1)

ggplot(data_series_ceb, aes(x_ceb, y_ceb)) + geom_point()+
  geom_smooth(method = "lm", se = FALSE)

#Cultivo de guisante

fun_outliers(data_series, data_series$y_guis, 10)
data_series_guis <- data_series[-c(85, 94, 118),]

reg2<- lm(y_guis ~ x_guis,data = data_series_guis)

```



## Script del análisis

---

```
summary(reg2)

ggplot(data_series_guis, aes(x_guis, y_guis)) + geom_point()+
  geom_smooth(method = "lm", se = FALSE)

###Separación por localidades###

#Zamadueñas
data_zam<- filter(data_series,( TRAT=="C66-G33" | TRAT=="C33-G66"
  | TRAT=="borde" | TRAT=="C100-G0" | TRAT=="CO-G100") & Localidad == "Zamadueñas")

#Cultivo cebada, Zamadueñas
fun_outliers(data_zam, data_zam$y_ceb, 10)

reg_zam_ceb<- lm(y_ceb ~ x_ceb,data = data_zam_ceb)
summary(reg_zam_ceb)

ggplot(data_zam_ceb, aes(x_ceb, y_ceb)) + geom_point()+
  geom_smooth(method = "lm", se = FALSE)

#Cultivo de guisante, Zamadueñas
#Outliers
fun_outliers(data_zam, data_zam$y_guis, 10)

data_zam_guis <- data_zam[c(-1,-10),]
ggplot(data_zam_guis, aes(x_guis, y_guis)) + geom_point()+geom_smooth(method = "lm", se = FALSE)

reg2<- lm(y_guis ~ x_guis,data = data_zam_guis, na.action = na.omit)
summary(reg2)

#Viñalta

data_vin<- filter(data_series,( TRAT=="C66-G33" | TRAT=="C33-G66"
  | TRAT=="borde" | TRAT=="C100-G0" | TRAT=="CO-G100") &

#Cultivo de cebada, Viñalta
fun_outliers(data_vin, data_vin$y_ceb, 10)
data_vin_ceb <- data_vin[c(-30, -28),]

reg3<- lm(y_ceb ~ x_ceb,data = data_vin_ceb)
summary(reg3)
```

---

```

ggplot(data_vin_ceb, aes(x_ceb, y_ceb)) + geom_point()+
  geom_smooth(method = "lm", se = FALSE)

#Cultivo de guisante, Viñalta
fun_outliers(data_vin, data_vin$y_guis, 10)
data_vin_guis <- data_vin[c(-14, -13),]

reg4<- lm(y_guis ~ x_guis,data = data_vin_guis)
summary(reg4)

ggplot(data_vin_guis, aes(x_guis, y_guis)) + geom_point()+
  geom_smooth(method = "lm", se = FALSE)

#Castrillo

data_cas<- filter(data_series,( TRAT=="C66-G33" | TRAT=="C33-G66"
                                | TRAT=="borde" | TRAT=="C100-G0" | TRAT=="CO-G100") & Loc

#Cultivo de cebada, Castrillo
fun_outliers(data_cas, data_cas$y_ceb, 10)
data_cas_ceb <- data_cas[c(-38),]

reg5<- lm(y_ceb ~ x_ceb,data = data_cas_ceb)
summary(reg5)

ggplot(data_cas, aes(x_ceb, y_ceb)) + geom_point()+
  geom_smooth(method = "lm", se = FALSE)

#Cultivo de guisante, Castrillo
fun_outliers(data_cas, data_cas$y_guis, 10)
data_cas_guis <- data_cas[c(-34),]

reg6<- lm(y_guis ~ x_guis,data = data_cas_guis)
summary(reg6)

ggplot(data_cas, aes(x_guis, y_guis)) + geom_point()+geom_smooth(method = "lm", se = FALSE)

#####Evaluación mediante diseño de superficie respuesta#####

```

## Script del análisis

---

```
names(data_lechoso)[22] = "Cebada"
names(data_lechoso)[23] = "Guisante"

#Codificacion de los factores
data_lechoso<- coded.data(data = data_lechoso,
                          x1 ~ (Cebada - 0.5)/0.5, x2 ~ (Guisante - 0.5)/0.5)

ggplot(data = data_lechoso, aes(x1, x2)) +
  geom_point(col = "green") +
  geom_vline(xintercept = c(-1,1), col = "lightgrey") +
  geom_hline(yintercept = c(-1,1) , col = "lightgrey")

#Diseño de primer orden
modelo_respuesta <- rsm(MS ~ FO(x1, x2), data = data_lechoso)
summary(modelo_respuesta)

#Diseño de segundo orden completo
modelo_respuesta <- rsm(MS ~ SO(x1, x2), data = data_lechoso)
summary(modelo_respuesta)

#Modelo 2 : diseño completo sin interaccion
modelo_respuesta <- rsm(MS ~ FO(x1, x2) +
                       PQ(x1, x2), data = data_lechoso)
summary(modelo_respuesta)

#Modelo 3: Diseño completo sin interaccion y sin el #termino cuadratico de la segunda variable
modelo_respuesta <- rsm(MS ~ FO(x1, x2) +
                       PQ(x1), data = data_lechoso)
summary(modelo_respuesta)

#Grafico de perspectiva
persp(modelo_respuesta, x2 ~ x1,
      zlab = "Materia seca (Rend)",
      contours = list(z = "bottom", col = "colors"), # posicion y color
      at = c(summary(modelo_respuesta$canonical$xs)),
      theta = 40, # coordenadas graficas
      phi = 10,
      nlevels = 20)

# Grafico de contornos
contour(modelo_respuesta, ~ x1 + x2 , image = TRUE, nlevels = 20)
```

---

```

#Diagnostico gráfico de los datos del modelo
par(mfrow = c(2, 2))
plot(modelo_respuesta)

####Gráficos de lineas y puntos####

#Gráfico 1
densidad_guisante <- c(0, 1/3, 0.5, 2/3, 1)
densidad_cebada <- c(1, 2/3, 0.5, 1/3, 0)
df <- data.frame(x = densidad_guisante, y = densidad_cebada)

ggplot(df, aes(x = densidad_guisante, y = densidad_cebada)) +
  geom_line(color = 4,      # Color de la línea
            lwd = 1.2,     # Ancho de la línea
            linetype = 1)+
  geom_point(lwd = 1.7)

#Gráfico 2
densidad_guisante <- c(0, 1/3, 0.5, 2/3, 1, 0, 0, 0,
                      0,1/3, 1,
                      1/3, 1/3, 2/3, 2/3,2/3, 1, 1)
densidad_cebada <- c(1, 2/3, 0.5, 1/3,
                    0,1/3,2/3, 1,
                    1/3, 1/3, 1/3,
                    0,1, 0, 2/3,1, 2/3, 1 )
df <- data.frame(x = densidad_guisante, y = densidad_cebada)

ggplot(df, aes(x = densidad_guisante, y = densidad_cebada)) +
  geom_point(lwd = 1.7)

```





# Bibliografía

- [1] L. Li, L. Zhang, and F. Zhang, "Crop mixtures and the mechanisms of overyielding," in *Encyclopedia of Biodiversity*. Elsevier, 2013, pp. 382–395.
- [2] T. P. Tomich, S. Brodt, H. Ferris, R. Galt, W. R. Horwath, E. Kebreab, J. H. Leveau, D. Liptzin, M. Lubell, P. Merel, R. Michelmore, T. Rosenstock, K. Scow, J. Six, N. Williams, and L. Yang, "Agroecology: A review from a global-change perspective," *Annual Review of Environment and Resources*, vol. 36, no. 1, pp. 193–222, 2011.
- [3] R. W. Brooker, A. E. Bennett, W.-F. Cong, T. J. Daniell, T. S. George, P. D. Hallett, C. Hawes, P. P. M. Iannetta, H. G. Jones, A. J. Karley, L. Li, B. M. McKenzie, R. J. Pakeman, E. Paterson, C. Schöb, J. Shen, G. Squire, C. A. Watson, C. Zhang, F. Zhang, J. Zhang, and P. J. White, "Improving intercropping: a synthesis of research in agronomy, plant physiology and ecology," *New Phytologist*, vol. 206, no. 1, pp. 107–117, 2014.
- [4] A. R. Ngwira, J. B. Aune, and S. Mkwinda, "On-farm evaluation of yield and economic benefit of short term maize legume intercropping systems under conservation agriculture in malawi," *Field Crops Research*, vol. 132, pp. 149–157, 2012.
- [5] C. I. G.-V. Eva María Córdoba Jiménez, *Guías de Cultivo Serie Leguminosas: la Veza Común*, 1st ed. JUNTA DE ANDALUCÍA. Instituto de Investigación y Formación Agraria y Pesquera, 2016.
- [6] L. Bedoussac and E. Justes, "Dynamic analysis of competition and complementarity for light and n use to understand the yield and the protein content of a durum wheat–winter pea intercrop," *Plant and Soil*, vol. 330, no. 1-2, pp. 37–54, 2010.
- [7] H. Knörzer, S. Graeff-Hönninger, B. Guo, P. Wang, and W. Claupein, "The rediscovery of intercropping in china: A traditional cropping system for future chinese agriculture – a review," in *Climate Change, Intercropping, Pest Control and Beneficial Microorganisms*. Springer Netherlands, 2009, pp. 13–44.
- [8] A. Neumann, K. Schmidtke, and R. Rauber, "Effects of crop density and tillage system on grain yield and n uptake from soil and atmosphere of sole and intercropped pea and oat," *Field Crops Research*, vol. 100, no. 2-3, pp. 285–293, 2007.
- [9] J. Ehrmann and K. Ritz, "Plant: soil interactions in temperate multi-cropping production systems," *Plant and Soil*, vol. 376, no. 1-2, pp. 1–29, 2013.

- [10] T. Chapagain and A. Riseman, "Barley–pea intercropping: Effects on land productivity, carbon and nitrogen transformations," *Field Crops Research*, vol. 166, pp. 18–25, 2014.
- [11] W.-F. Cong, E. Hoffland, L. Li, J. Six, J.-H. Sun, X.-G. Bao, F.-S. Zhang, and W. V. D. Werf, "Intercropping enhances soil carbon and nitrogen," *Global Change Biology*, vol. 21, no. 4, pp. 1715–1726, 2014.
- [12] J. Borg, L. Kiær, C. Lecarpentier, I. Goldringer, A. Gauffreteau, S. Saint-Jean, S. Barot, and J. Enjalbert, "Unfolding the potential of wheat cultivar mixtures: A meta-analysis perspective and identification of knowledge gaps," *Field Crops Research*, vol. 221, pp. 298–313, 2018.
- [13] Anil, Park, Phipps, and Miller, "Temperate intercropping of cereals for forage: a review of the potential for growth and utilization with particular reference to the UK," *Grass and Forage Science*, vol. 53, no. 4, pp. 301–317, 1998.
- [14] B. Rosner, "Percentage Points for a Generalized ESD ManyOutlier Procedure," *Technometrics*, vol. 25, no. 2, 2003.
- [15] G. F. Preciado, "Optimización de una superficie de respuesta utilizando JMP IN," *Mosaicos Matemáticos*, vol. 11, 2003.
- [16] H. G. P. y Román de la Vara Salazar, *Análisis y diseño de experimentos*, 2nd ed. McGRAW-HILL/INTERAMERICANA, 2008.
- [17] J. V. J. R. René Castro Montoya, Mario Castro Flores, "Metodología Superficie de Respuesta: Tres aplicaciones a conjuntos de datos reales," *Academia Journals*, 2019.
- [18] G. T. Wensheng Wang, Yongchun Cheng, "Design optimization of SBS-modified asphalt mixture reinforced with Eco-friendly basalt fiber based on response surface methodology," *Materials (Basel)*, vol. 11, no. 8, 2018.
- [19] S. Bhattacharya, *Response Surface Methodology in Engineering Science*, 1st ed. IntechOpen, 2020.
- [20] S. S. Ranade and P. Thiagarajan, "Selection of a design for response surface," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 263, 2017.
- [21] D. G. S. Julian David Muñoz, Diego Alejandro Camargo, "Aplicación de la metodología de superficie de respuesta en un proceso de absorción del  $CO_2$  de un biogás en una solución alcalina," *DYNA*, vol. 159, 2009.
- [22] P. V. Sánchez Vallduví, Tamagno L. Nora, "Siembra en mezcla de híbridos de girasol. Una alternativa para un manejo agroecológico de los recursos," *Resumos do V CBA*, vol. 2, 2007.
- [23] N. R. S. Hamilton, "Replacement and Additive Designs for plant competition studies," *Journal of Applied Ecology*, vol. 31, pp. 599–603, 1994.
- [24] R. R. Angelika Neumann, Jochen Werner, "Evaluation of yield–density relationships and optimization of intercrop compositions of field-grown pea–oat intercrops using the replacement series and the response surface design," *Elsevier B.V.*, vol. 114, pp. 286–294, 2009.



- [25] C. T. D. Wit, *On competition*, 2nd ed. Centrum voor Landbouwpublikaties, 1964.
- [26] H. R. L., "Analysis of the nature of interference between plants of different species 1. concepts and extension of the de wit analysis to examine effects," *Australian J. Agric. Res.*, vol. 25, 1974.
- [27] E. Taha and A. El-Mahdy, "Land equivalent ratio as a reference for relative crowding coefficient and aggressivity of intercropped plant species," *Middle East Journal of Agriculture Research*, no. 3, pp. 576–585, 2014.
- [28] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [Online]. Available: <https://ggplot2.tidyverse.org>
- [29] B. G. Peterson and P. Carl, *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis*, 2020, r package version 2.0.4. [Online]. Available: <https://CRAN.R-project.org/package=PerformanceAnalytics>
- [30] J. Gross and U. Ligges, *nortest: Tests for Normality*, 2015, r package version 1.0-4. [Online]. Available: <https://CRAN.R-project.org/package=nortest>
- [31] A. Zeileis and T. Hothorn, "Diagnostic checking in regression relationships," *R News*, vol. 2, no. 3, pp. 7–10, 2002. [Online]. Available: <https://CRAN.R-project.org/doc/Rnews/>
- [32] R. V. Lenth, "Response-surface methods in R, using rsm," *Journal of Statistical Software*, vol. 32, no. 7, pp. 1–17, 2009.



