

**ANÁLISIS ESTADÍSTICO DE DISEÑOS
EXPERIMENTALES CON APOYO
COMPUTACIONAL**

Presentado por:
ANA MARÍA PALENCIA PRETEL

Director:
GUILLERMO MARTÍNEZ FLÓREZ
Universidad de Córdoba

Codirector:
CARLOS BARRERA CAUSIL
Instituto Tecnológico Metropolitano - ITM

Universidad de Córdoba
Facultad de Ciencias Básicas
Departamento de Matemáticas y Estadística
Programa Estadística
Montería-2023

Dedicado a mi madre Carmen y padre Alfonso, agradezco la lucha personal que llevaron a cabo para verme alzar mis sueños, a mis hermanos Hans y Jesús por impulsarme a crecer, a mi pareja Alejandro por acompañarme y darme ánimos para no rendirme, por último, gracias a mis amigos por acompañarme en este camino.

Contenido

1. RESUMEN	8
2. ABSTRACT	8
3. INTRODUCCIÓN	8
4. OBJETIVOS	9
4.1. Objetivo general	9
4.2. Objetivos específicos	9
5. Entorno R.	9
5.1. Lectura de datos de un archivo	10
5.1.1. La Función <code>read.table()</code>	10
5.2. Operaciones básicas en R	11
5.3. Estadística con R	12
5.3.1. Ilustración para las Estadísticas con R	13
6. Concepto y análisis estadísticos en R	16
6.1. Diseño experimental	16
6.1.1. Conceptos básicos del diseño experimental	16
6.1.2. Estructuras del diseño experimental	17
6.2. Diseño completamente al azar (DCA)	19
6.2.1. Modelo estadístico y supuestos	20
6.2.2. Modelos de efectos fijos y aleatorios	21
6.2.3. Estimación de parámetros	22
6.2.4. Hipótesis de interés	23
6.2.5. Análisis de varianza (ANAVA)	24
6.2.6. Pruebas de hipótesis y errores estándar	25
6.2.7. Ejemplo en R para un DCA.	26
6.3. Pruebas de validación de los supuestos del modelo	28
6.3.1. Prueba de normalidad de los errores de Shapiro - Wilk	29
6.3.2. Ejemplo en R de la prueba de Shapiro - Wilk	30
6.3.3. Pruebas de homogeneidad de varianzas	30
6.3.4. Prueba de Hartley	30
6.3.5. Ejemplo en R de la prueba de Hartley	31
6.3.6. Prueba de Bartlett	32
6.3.7. Ejemplo en R de la prueba de Bartlett	32

6.3.8.	Prueba de Levene	32
6.3.9.	Ejemplo en R de la prueba de Levene	33
6.4.	Comparaciones múltiples	34
6.4.1.	Método de las diferencias mínimas significativas (DMS)	34
6.4.2.	Ejemplo en R para DMS	35
6.4.3.	La prueba de intervalos múltiples de Duncan	36
6.4.4.	Ejemplo en R para la prueba de intervalos múltiples de Duncan	37
6.4.5.	Prueba honesta de Tukey	39
6.4.6.	Ejemplo en R para la prueba honesta de Tukey	39
6.4.7.	Prueba de Scheffé	40
6.4.8.	Ejemplo en R para la prueba de Scheffé	41
6.4.9.	Prueba de Bonferroni	42
6.4.10.	Ejemplo en R para la prueba de Bonferroni	42
6.4.11.	Comparación de medias de tratamientos con un control (prueba de Dunnett)	43
6.4.12.	Ejemplo en R para la prueba de Dunnett	44
6.4.13.	Función (<code>ea1</code>) en R	45
6.4.14.	Ejemplo en R para un diseño completamente aleatorio usando la función (<code>ea1</code>).	46
6.5.	Contrastes	48
6.5.1.	Función (<code>fit.contrast</code>) en R	51
6.5.2.	Ejemplo en R para contrastes con un DCA usando la función (<code>fit.contrast</code>).	52
6.6.	Polinomios ortogonales	54
6.6.1.	Función (<code>crd</code>) en R	56
6.6.2.	Ejemplo en R para un diseño completamente aleatori- zado de un factor usando la función (<code>crd</code>)	58
6.7.	Transformaciones	62
6.7.1.	Transformaciones más frecuentes para homogenizar va- rianzas	63
6.7.2.	Ejemplo en R de transformaciones	64
6.8.	Diseño en bloques completamente aleatorizados (DBCA).	70
6.8.1.	Aleatorización en el diseño en bloques al azar	71
6.8.2.	Modelo estadístico	71
6.8.3.	Pruebas de hipótesis y análisis de varianza	72
6.8.4.	Ejemplo en R para un diseño en bloque completamente aleatorio usando la función (<code>ea1</code>).	73

6.8.5.	Función (<code>rbd</code>) en R	76
6.8.6.	Ejemplo en R para un diseño en bloque completamente aleatorio usando la función (<code>rbd</code>).	77
6.8.7.	Ejemplo con un factor cuantitativo en R para un diseño en bloque completamente aleatorio usando la función <code>ea1</code>	79
6.8.8.	Ejemplo con un factor cuantitativo en R para un diseño en bloque completamente aleatorio usando la función <code>rbd</code>	82
6.9.	Muestreo en unidades experimentales	84
6.9.1.	Muestreo de unidades experimentales en un DCA	84
6.9.2.	Descomposición de la suma de cuadrados total DCA	85
6.9.3.	Función (<code>Sub.DCA</code>)	87
6.9.4.	Ejemplo en R de muestreo de unidades experimentales en un DCA	88
6.9.5.	Muestreo de unidades experimentales en un DBCA	90
6.9.6.	Función (<code>Sub.DBCA</code>)	92
6.9.7.	Ejemplo en R de muestreo de unidades experimentales en un DBCA	93
6.10.	Diseño Cuadrado Latino	95
6.10.1.	Aleatorización en un Diseño Cuadrado Latino	95
6.10.2.	Modelo Estadístico e Hipótesis	97
6.10.3.	Análisis de Varianza	98
6.10.4.	Función (<code>latsd</code>) en R	99
6.10.5.	Ejemplo en R para un diseño cuadrado latino usando la función (<code>latsd</code>)	101
6.10.6.	Juego de Cuadrados Latinos	103
6.10.7.	Ejemplo en R para un diseño cuadrado cuadrado latino replicado usando la función (<code>ea2</code>).	103
6.11.	Análisis de covarianza	110
6.11.1.	Análisis de covarianza en un DCA	110
6.11.2.	Análisis de covarianza en un DBCA	113
6.11.3.	Ejemplo en R de análisis con covarianza con un DCA.	114
6.11.4.	Ejemplo en R de análisis con covarianza con un DBCA.	117
6.12.	Experimentos factoriales	121
6.12.1.	Conceptos básicos	121
6.12.2.	Efectos de un factor	122
6.12.3.	Experimento factorial bajo un DCA	122

6.12.4.	Análisis de varianza	123
6.12.5.	Errores estándar para medias de efectos principales e interacción	125
6.12.6.	Función (<code>fat2.crd</code>) en R	128
6.12.7.	Ejemplo en R para análisis para factores cualitativo \times cuantitativo.	129
6.12.8.	Función (<code>ea2</code>) en R	140
6.12.9.	Ejemplo en R para un experimento factorial con un DCA.	142
6.12.10.	Experimento factorial bajo un DBCA	154
6.12.11.	Ejemplo en R para un experimento factorial con un DBCA	156
6.13.	Experimentos con diferentes tamaños de parcelas	164
6.13.1.	Aleatorización	165
6.13.2.	Parcelas divididas en el espacio	166
6.13.3.	Análisis de varianza	167
6.13.4.	Errores estándar para la comparación de medias de tratamientos	170
6.13.5.	Función (<code>split2.crd</code>) en R	174
6.13.6.	Ejemplo en R de parcelas divididas para un DCA . . .	175
6.13.7.	Función (<code>split2.rbd</code>) en R	180
6.13.8.	Ejemplo en R de parcelas divididas para un DBCA . .	182
6.14.	Medidas repetidas en el tiempo	200
6.14.1.	Supuestos del modelo	200
6.14.2.	Errores estándar para comparaciones múltiples. . . .	202
6.14.3.	Función (<code>MRT.DCA</code>)	203
6.14.4.	Ejemplo en R de medidas repetidas bajo un DCA . . .	204
6.14.5.	Medidas repetidas bajo un DBCA	214
6.14.6.	Función (<code>MRT.DBCA</code>)	216
6.14.7.	Ejemplo en R de medidas repetidas bajo un DBCA . .	217
6.15.	Introducción a la metodología de superficie de respuesta . . .	224
6.15.1.	Mejor tratamiento y punto óptimo	225
6.15.2.	Elementos de la metodología de superficie de respuesta	226
6.15.3.	Diseños de superficie de respuesta	227
6.15.4.	Diseños para modelos de primer orden	227
6.15.5.	Diseños para modelos de segundo orden	228
6.15.6.	Técnicas de optimización	228
6.15.7.	Ejemplo en R de superficie de respuesta	229

7. CÓDIGOS	237
8. BIBLIOGRAFÍA.	254

1. RESUMEN

Este trabajo busca implementar el análisis estadístico de diseños experimentales con apoyo computacional a través del software Estadístico R. Para el respectivo análisis de estos diseños usaremos las librerías `ExpDes`, `easynova`, entre otras, las cuales contienen una gran cantidad de paquetes que nos ayudarán con el principal objetivo de este trabajo. Con esta implementación computacional se realizarán varios ejemplos, los cuales serán plasmados en un blog de acceso libre, para que toda la comunidad que necesite esta información pueda acceder a ella y puedan utilizarla para el respectivo análisis de algún diseño experimental que se le presente en una situación real.

Palabras claves : Diseños experimentales, entorno R, blog.

2. ABSTRACT

This work seeks to implement the statistical analysis of experimental designs with computational support through the R Statistical software. For the respective analysis of these designs we will use the libraries `ExpDes`, `easynova`, among others, which contain a large number of packages that will help us with the main objective of this work. With this computational implementation, several examples will be made, which will be reflected in a free access blog, so that all the community that needs this information can access it and can use it for the respective analysis of an experimental design that is presented to them in a real situation.

keywords : Experimental designs, R environment, blog.

3. INTRODUCCIÓN

El diseño de experimentos en R es una poderosa metodología que permite planificar, ejecutar y analizar experimentos de manera eficiente y reproducible. R proporciona una variedad de herramientas estadísticas y gráficas para realizar análisis experimentales de alta calidad. Los investigadores pueden aplicar diferentes diseños experimentales, realizar pruebas de hipótesis, análisis de varianza, regresiones y visualizar los resultados. Esta combinación entre R y el diseño de experimentos facilita la toma de decisiones basada en evidencia y contribuye al avance del conocimiento científico.

La metodología del diseño de experimentos se enfoca en modificar las condiciones de un proceso empírico con el fin de incrementar las posibilidades de identificar cambios significativos en la respuesta. Los métodos utilizados en el diseño de experimentos son vitales para el desarrollo y optimización de procesos, buscando mejorar su rendimiento y lograr una mayor consistencia y robustez. El objetivo final es alcanzar un proceso que sea más eficiente y confiable (Montgomery, 2012).

4. OBJETIVOS

4.1. Objetivo general

Elaborar funciones e incluir otras ya existentes en las librerías de RStudio en R-Project utilizando paquetes y funciones especiales contenidos en este software para el respectivo análisis estadístico de diseños experimentales.

4.2. Objetivos específicos

- Definir la teoría estadística que soporte el desarrollo de los métodos utilizados en el diseño de experimentos.
- Construir algoritmos para el análisis estadístico de un diseño de experimento utilizando funciones y paquetes especiales contenidos en este software.
- Elaborar un blog que contenga todas las funciones creados y las ya existentes aportadas por R, para que toda la comunidad académica e investigadores tengan acceso a esta herramienta.

5. Entorno R.

Este programa proporciona funciones para la manipulación de datos, cálculo y generación de gráficos. Los principales aspectos que ofrece son:

- Facilidad para el manejo y el almacenamiento de datos.
- Un conjunto de operadores para cálculo con arrays y matrices.

- Una colección extensa e integrada de herramientas intermedias para el análisis de datos.
- Multitud de facilidades gráficas.
- Un lenguaje de programación simple y efectivo que incluye las estructuras de control clásicas, funciones recursivas y facilidades para el entrada y salida de datos y resultados.

R es un entorno altamente dinámico, y a menudo se concibe como un vehículo para desarrollar (nuevos) métodos interactivos de análisis de datos.

Ventajas:

- R dispone de una comunidad de desarrolladores/usuarios que se dedican constantemente a la mejora y a la ampliación de las funcionalidades y capacidades del programa. Cualquier usurario con conocimiento del lenguaje podría ser desarrollador de rutinas en R.
- Incorporación constante de nuevos métodos.
- Es software libre.

5.1. Lectura de datos de un archivo

Los datos suelen leerse desde archivos externos sin tener que digitarlos de modo interactivo. Las capacidades de lectura de archivos de R son sencillas y sus requisitos son bastante estrictos cuando no inflexibles. Se presupone que el usuario es capaz de modificar los archivos de datos con otras herramientas, por ejemplo, con editores de texto, para ajustarlos a las necesidades de R. Generalmente esta tarea es muy sencilla.

5.1.1. La Función `read.table()`

Para poder leer una hoja de datos directamente, el archivo externo debe reunir las condiciones adecuadas. La forma más sencilla es:

- La primera línea del archivo debe contener el nombre de cada variable de la hoja de datos.

- En cada una de las siguientes líneas, el primer elemento es la etiqueta de la fila, y a continuación deben aparecer los valores de cada variable.
- En R, el carácter comúnmente utilizado para separar las columnas en conjuntos de datos es la coma (","). Esta convención se basa en el formato de archivo CSV (**Comma-Separated Values**). Sin embargo, se pueden utilizar otros caracteres como separadores, como tabulaciones o punto y coma (";"). Es importante especificar el carácter de separación correcto al leer o escribir archivos en RStudio, utilizando los argumentos adecuados en las funciones correspondientes, como `read.csv()` y `write.csv()`. Esto garantiza una correcta interpretación y manipulación de los datos en el entorno de trabajo.

Es importante considerar el formato en el que se presentan los datos al acceder a ellos en R. Para facilitar esta tarea, R proporciona diferentes funciones para leer y cargar datos en diferentes formatos. A continuación, se detallan las funciones necesarias para leer archivos en formatos específicos, así como las librerías requeridas para utilizar estas funciones:

Si los datos se encuentran en formato de archivo de texto (`.txt`), se puede utilizar la función `read.table()` para leerlos. No se requiere ninguna librería adicional, ya que esta función está incluida en la librería base de R.

Si los datos están en formato de archivo CSV (`.csv`), se utiliza la función `read.csv()`. Esta función también se encuentra en la librería base de R, por lo que no es necesario cargar ninguna librería adicional.

Para leer archivos en formato de hoja de cálculo de Microsoft Excel (`.xls`), se necesita la librería `readxl`. Puedes instalarla utilizando el siguiente comando:

5.2. Operaciones básicas en R

R además de ser un lenguaje de programación y ofrecer herramientas para el análisis de datos, también nos permite calcular operaciones entre valores o vectores algunas operaciones que se pueden realizar en R son las siguientes.

`a+b`: Retorna la suma entre a y b.

`a-b`: Retorna la diferencia entre a y b.

`a*b`: Retorna el producto entre a y b.
`a/b`: Retorna el cociente entre a y b.
`abs()`: Retorna el valor absoluto.
`a ^ n`: Retorna la n-ésima potencia del valor o vector a.
`log()`: Retorna el logaritmo.
`sin()`: Retorna el seno.
`cos()`: Retorna el coseno.
`tan()`: Retorna la tangente.
`asin()`: Retorna el arcseno.
`acos()`: Retorna el arccoseno.
`atan()`: Retorna la arctangente.
`sum()`: Retorna la suma de los elementos de un vector.

5.3. Estadística con R

El software R nos ofrece una amplia gama de funciones que nos facilitan el análisis de datos, dentro de estas funciones se tienen las siguientes.

`mean()`: Retorna la media aritmética de un vector de datos.
`median()`: Retorna la mediana de un vector de datos.
`var()`: Retorna la varianza de un vector de datos.
`sd()`: Retorna la desviación estándar de un vector de datos.
`quantile()`: Retorna los cuantiles de un vector de datos.
`summary()`: Muestra un resumen de estadísticos descriptivos: mínimo, máximo, media, mediana, primer y tercer cuartil.
`skewness()`: Retorna el coeficiente de asimetría de la distribución de una variable.
`kurtosis()`: Retorna la curtosis de la distribución de una variable.
`barplot()`: Retorna un diagrama de barras de un vector de datos.
`hist()`: Retorna un histograma de un vector de datos.
`pie()`: Retorna un diagrama circular de un vector de datos.
`stem()`: Retorna un diagrama de tallos y hojas de un vector de datos.
`boxplot()`: Retorna un diagrama de caja de un vector de datos.
`cov()`: Retorna la covarianza entre dos variables.
`cor()`: Retorna el coeficiente de correlación entre dos variables.
`lm()`: Ajusta un modelo lineal.
`table()`: Tabla de frecuencias de un vector.
`prop.table(table())`: Tabla de frecuencias relativas de un vector.

`table(x,y)`: Tabla de frecuencias cruzadas de `x` por `y`.
`prop.table(table(x),margin=i)`: Tabla de frecuencias relativas. Las frecuencias relativas se calculan por fila si `margin=1`, por columna si `margin=2` o globales si no se especifica `margin`.

5.3.1. Ilustración para las Estadísticas con R

Doce inspectores midieron el diámetro de un cojín usando dos calibradores diferentes. Los resultados fueron los siguientes. (Montgomery, 2012, p.63).

Tabla 1: Resultados obtenidos de cada inspector

Inspector	Calibrador 1	Calibrador 2
1	0,265	0,264
2	0,265	0,265
3	0,266	0,264
4	0,267	0,266
5	0,267	0,267
6	0,265	0,268
7	0,267	0,264
8	0,267	0,265
9	0,265	0,265
10	0,268	0,267
11	0,268	0,268
12	0,265	0,269

El objetivo de estos datos es ilustrar algunas funciones nombradas en el ítem 5.3.

- Antes de realizar cualquier procedimiento se debe cargar la base de datos, entonces, primero se le asigna el nombre `Base_0` y se carga utilizando la función `read_excel`, se encuentra formato de `.xlsx`, está compuesta por tres variables llamadas `Inspector`, `Calibrador 1` y `Calibrador 2`.

```
1 > Base_0 <- read_excel("C:/Users/Hp E2/Desktop/MONOGRAFIA  
2 -TODO/Base_0.xlsx")
```

- Podemos comenzar los análisis de este conjunto de datos haciendo un resumen con algunas estadísticas básicas, esto se consigue en R o usando la función `summary()` como se ilustra a continuación.

```

1 > summary(Base_0)
2   Inspector      Calibrador 1      Calibrador 2
3 Min.   : 1.00    Min.   :0.2650    Min.   :0.2640
4 1st Qu.: 3.75    1st Qu.:0.2650    1st Qu.:0.2647
5 Median : 6.50    Median :0.2665    Median :0.2655
6 Mean   : 6.50    Mean   :0.2662    Mean   :0.2660
7 3rd Qu.: 9.25    3rd Qu.:0.2670    3rd Qu.:0.2672
8 Max.   :12.00    Max.   :0.2680    Max.   :0.2690

```

Los siguientes ejemplos se llevarán a cabo utilizando las variables Calibrador 1 y 2 que aporta la base de datos. Se hallará la media, mediana, varianza, desviación estándar y se graficará el histograma. Véase a continuación.

- Función `mean()`:

```

1 > mean(Base_0$'Calibrador 1')
2 [1] 0.26625
3 > mean(Base_0$'Calibrador 2')
4 [1] 0.266
5 >

```

[language=R]

- Función `median()`:

```

1 > median(Base_0$'Calibrador 1')
2 [1] 0.2665
3 > median(Base_0$'Calibrador 2')
4 [1] 0.2655

```

- Función `var()`:

```

1 > var(Base_0$'Calibrador 1')
2 [1] 1.477273e-06
3 > var(Base_0$'Calibrador 2')
4 [1] 3.090909e-06

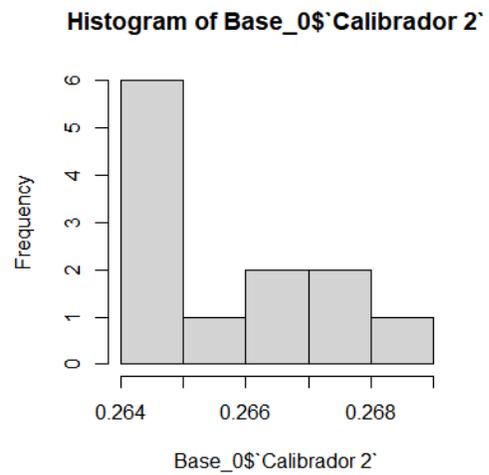
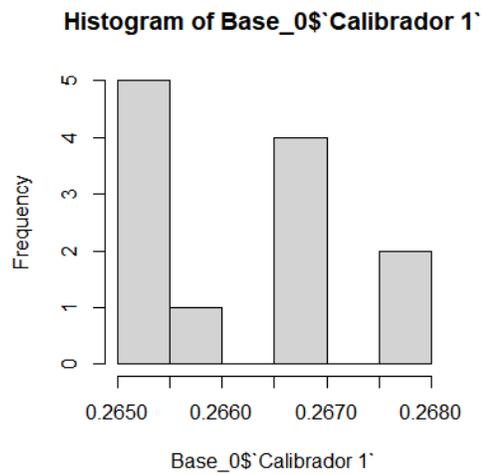
```

- Función `sd()`:

```
1 > sd(Base_0$`Calibrador 1`)  
2 [1] 0.001215431  
3 > sd(Base_0$`Calibrador 1`)  
4 [1] 0.001215431
```

■ Función hist():

```
1 > hist(x = Base_0$`Calibrador 1`)  
2 > hist(x = Base_0$`Calibrador 2`)
```



6. Concepto y análisis estadísticos en R

6.1. Diseño experimental

En la búsqueda planeada para obtener nuevos conocimientos o para confirmar o no resultados de experimentos, los investigadores se apoyan en las herramientas que brinda el diseño experimental para seguir realizando sus experimentos. Los objetivos se establecen sobre la base de preguntas que han de responderse, hipótesis que han de probarse, y efectos que han de estimarse.

6.1.1. Conceptos básicos del diseño experimental

Los conceptos básicos que son importantes en la construcción y ejecución de un diseño experimental.

- **Unidad experimental**

Una unidad experimental es la división más pequeña del material experimental, la unidad experimental constituye una repetición del tratamiento. (Ciencia UANL, 2012, p.2).

- **Tratamientos**

Los tratamientos son los procedimientos que se aplican a las unidades experimentales, dentro de este concepto es importante destacar el concepto de factor, este se conoce como un grupo del tratamiento, existen dos tipos de factores cualitativos y cuantitativos. (Ciencia UANL, 2012, p.2).

- **Error experimental (EE)**

Es una medida de la variación existente entre las respuestas de las unidades experimentales tratadas en forma similar. (Mendoza H.,2016).

De acuerdo con Fisher (1935), para obtener un buen diseño de experimento si se tienen en cuenta los siguientes principios básicos:

- **Repeticiones y sus Funciones**

Cuando un tratamiento aparece más de una vez en un experimento se dice que esta repetido.

Las funciones de la repetición son:

1. Permitir estimar el error experimental

2. Mejorar la precisión de un experimento mediante la reducción de la desviación estándar de una media de tratamiento.
3. Aumentar el alcance de la inferencia del experimento a través de la selección y uso apropiado de unidades experimentales más viables.
4. Ejercer control sobre la varianza del error

■ **Control del error experimental**

El control del error experimental tiene como objetivo la escogencia de un diseño experimental que posea buenas características de homogeneidad, por medio de diseños como el de bloques completos aleatorizados (DBCA) o cuadrados latinos (DCL), el uso de observaciones concomitantes (Covariable) y una apropiada elección del tamaño y la forma de las unidades experimentales. (Mendoza H.,2016)

■ **Aleatorización**

La aleatorización se refiere a la distribución aleatoria del material experimental y al orden en que se llevan a cabo los ensayos del experimento. En otras palabras, implica asignar de forma aleatoria los tratamientos a las unidades experimentales. Este proceso garantiza la dispersión equitativa de los tratamientos en el área experimental. (Ciencia UANL, 2012, p.4)

6.1.2. Estructuras del diseño experimental

Teniendo en cuenta el aporte de Hinkelmann & Kempthorne (1994), cuando se realiza un diseño es esencial tener en cuenta dos estructuras básicas, estas son:

■ **Estructura de tratamiento**

Según la naturaleza del experimento, en esta fase se debe determinar, el número de tratamientos, combinaciones de tratamientos que se van a estudiar y a comparar. Los tratamientos están determinados por el número de factores que se van a comparar para medir sus efectos. Si es un factor únicamente, la estructura de tratamientos está formada por un conjunto de tratamientos, en este caso se habla de una vía de clasificación. La estructura del tratamiento experimental con un solo factor se refiere a cómo se diseñan y organizan los diferentes tratamientos en un experimento en el que se analiza el efecto de un único factor.

A continuación, se presenta una descripción general de la estructura del tratamiento experimental con un solo factor:

■ **Estructura de diseño**

La elección del diseño adecuado depende de los material experimental, la naturaleza de las variables y las restricciones prácticas. Cada diseño tiene sus ventajas y limitaciones, y es importante seleccionar el más apropiado para responder a las preguntas de investigación y obtener resultados válidos y confiables.

La estructura de diseño tiene un impacto directo en la validez interna y externa de un estudio y en la capacidad para obtener conclusiones confiables. Algunos aspectos importantes de la estructura de diseño incluyen:

1. Asignación de tratamientos: Se refiere a cómo se asignan los diferentes tratamientos a los sujetos o unidades experimentales. Puede ser aleatoria, secuencial, por bloques, entre otros métodos.
2. Replicación: Implica la repetición de las unidades experimentales o la realización de múltiples ensayos con el mismo tratamiento para reducir el error experimental y aumentar la precisión de las estimaciones.
3. Control de variables: Consiste en controlar y mantener constantes ciertas variables que podrían influir en los resultados del estudio, excepto aquellas que se están manipulando o evaluando como tratamientos.
4. Aleatorización: Implica la asignación aleatoria de los sujetos o unidades experimentales a los diferentes tratamientos para minimizar los sesgos y aumentar la validez de los resultados.
5. Balance: Se refiere a la distribución equitativa de los tratamientos en todo el diseño experimental. Un diseño balanceado asegura que cada tratamiento se aplique en el mismo número de ocasiones y

en las mismas condiciones.

6. Tamaño de la muestra: Determina el número de unidades experimentales necesarias para obtener resultados estadísticamente significativos y confiables.

La elección adecuada de la estructura de diseño depende del objetivo del estudio, las variables involucradas y las limitaciones prácticas. Un diseño adecuado y bien planificado es esencial para obtener conclusiones válidas y confiables en el análisis estadístico.

6.2. Diseño completamente al azar (DCA)

El diseño completamente al azar se caracteriza por realizar todas las corridas experimentales en un orden completamente aleatorio (Gutiérrez y De La Vara, 2012).

En este tipo de diseño, se busca eliminar cualquier posible sesgo o influencia sistemática en la asignación de las condiciones experimentales. Cada corrida experimental tiene la misma probabilidad de ser asignada a cualquier grupo o condición, lo que ayuda a garantizar que los resultados obtenidos sean más representativos y generalizables.

El diseño completamente al azar se utiliza cuando se desea minimizar el impacto de factores desconocidos o variables no controladas en el estudio. Al asignar aleatoriamente las corridas experimentales, se asegura que cualquier variación o diferencia observada entre los grupos sea atribuible principalmente a las condiciones experimentales en lugar de otros factores.

Si se tienen t tratamientos y n unidades experimentales, con $n > t$, se puede aplicar el primer tratamiento a r_1 unidades experimentales escogidas al azar del total n , se asigna el segundo tratamiento a r_2 unidades experimentales a partir de las $n - r_1$ restantes, así hasta llegar a asignar el tratamiento t a la r_t últimas unidades experimentales. (Martínez y Robles, 2013, p.64).

6.2.1. Modelo estadístico y supuestos

Sea Y_{ij} la respuesta al tratamiento j -ésimo en la unidad experimental i -ésimo. Se puede considerar $Y_{11}, Y_{21}, Y_{31}, \dots, Y_{r_{j1}}$ como una muestra aleatoria de tamaño r_1 tomada de una población con media μ_1 y varianza σ_1^2 , también se puede suponer una segunda muestra aleatoria donde se tiene un de tamaño r_2 tomada de una población con media μ_2 y varianza σ_2^2 , el proceso puede seguir en este caso hasta t . (Martínez y Robles, 2013, p.64).

En un caso simple en el que todas las varianzas, denotadas como $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 = \sigma^2$, son iguales a σ^2 , se presenta una situación en la que la aplicación del tratamiento j -ésimo en la unidad experimental i -ésima afecta la media de la respuesta, pero no su varianza (Martínez y Robles, 2013, p.64).

El modelo estadístico que describe las respuestas de las n observaciones es:

$$\begin{aligned} Y_{ij} &= \mu + \tau_j + \epsilon_{ij} \\ i &= 1, 2, \dots, r_j \quad \text{y} \quad j = 1, 2, \dots, t, \end{aligned} \quad (1)$$

donde:

- Y_{ij} : Es la observación obtenida en la i -ésima unidad experimental del j -ésimo. tratamiento.
- μ : Es un parámetro común a todos los tratamientos y corresponde a la media general.
- τ_j : Es el efecto del tratamiento j -ésimo.
- ϵ_{ij} : Es el efecto de error en la i -ésima. unidad experimental del j -ésimo. tratamiento.
- r_j : Es el número de repeticiones del tratamiento j - *ésimo*.
- t : Es el número de tratamientos.

La ecuación (1) se puede escribir de la forma $Y_{ij} = \mu_j + \epsilon_{ij}$, donde, $\mu_j = \mu + \tau_j$, esta representa la media del tratamiento j -ésimo, entonces $\epsilon_{ij} = Y_{ij} - \mu_j$, por lo cual $E(\epsilon_{ij}) = E(Y_{ij}) - \mu_j = \mu_j - \mu_j = 0$ y $V(\epsilon_{ij}) = \sigma_{\epsilon_{ij}}^2 = \sigma_j^2$. Como todos

los tratamientos están sometidos a las mismas condiciones, se supone que $\sigma_{\epsilon_{ij}}^2 = \sigma_j^2 = \sigma^2$, para $j = 1, 2, \dots, t$, esto significa que los errores presentan una varianza constante igual a σ^2 . (Martínez y Robles, 2013, p.64).

Para llevar a cabo las pruebas de hipótesis sobre los parámetros del modelo se necesita asignarle una distribución a los errores, la más común es utilizar la distribución normal, esto permite que los errores sean independientes e idénticamente distribuidos, tienen media cero, varianza constante. (Martínez y Robles, 2013, p.64). Todas estas características de los errores se resume en la siguiente notación:

$$\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2).$$

6.2.2. Modelos de efectos fijos y aleatorios

El modelo de efectos fijos y/o aleatorios es escogido por el investigador dependiendo de los objetivos de la investigación.

- **Modelo de efectos fijos**

El objetivo primordial es saber si existen diferencias entre los tratamientos aplicados, lo cual equivale a comparar los promedios de los tratamientos para determinar si existen diferencias entre ellos. Como τ_j es la variación de la media del tratamiento j con respecto a la media general, $\mu_j = \mu + \tau_j$, entonces bajo este modelo se tiene el supuesto $\sum_{j=1}^t \tau_j = 0$. (Martínez y Robles, 2013, p.64). Las hipótesis de interés son:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_t = \mu$$

contra

$$H_1 : \text{Al menos un par de medias de tratamiento difieren.}$$

- **Modelo de efectos aleatorios**

El objetivo no se centra en las medias de los tratamientos, sino en la variabilidad de éstos, porque al no existir tal variabilidad, el comportamiento de los tratamientos sería similar. Ahora el objetivo central es estimar la variación entre las medias de los tratamientos. (Martínez y

Robles, 2013, p.64). La hipótesis de interés es: $H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 = \sigma^2$ contra H_1 : Al menos una de las varianzas $(\sigma_1^2, \sigma_2^2, \dots, \sigma_t^2)$ no es igual a σ^2 .

6.2.3. Estimación de parámetros

En el análisis de datos para comparar tratamientos, es importante tener en cuenta las siguientes cantidades.

$Y_{.j}$ = Suma de las observaciones del tratamiento j .

$\bar{Y}_{.j}$ = Media de las observaciones del tratamiento j .

$Y_{..}$ = Suma total de las n mediciones.

$\bar{Y}_{..}$ = Media global o promedio de todas las observaciones.

Para cada tratamiento $j = 1, 2, \dots, t$, la muestra de tamaño $r_j > 1$, proporciona un estimativo de la media y la varianza de ese tratamiento mediante las expresiones usuales:

$$\hat{\mu}_j = \bar{Y}_{.j} = \frac{\sum_{i=1}^{r_j} Y_{ij}}{r_j} = \frac{Y_{.j}}{r_j} \quad \text{y} \quad \hat{\sigma}_j^2 = S_j^2 = \frac{\sum_{i=1}^{r_j} (Y_{ij} - \bar{Y}_{.j})^2}{r_j - 1} \quad \text{siendo}$$

$$Y_{.j} = \sum_{i=1}^{r_j} Y_{ij}$$

Dado el supuesto que la variabilidad de los tratamientos es la misma, e igual a σ^2 , entonces un promedio ponderado de los t estimadores S_j^2 da el mejor estimador insesgado de σ^2 . Así el mejor estimador insesgado de σ^2 viene dado por:

$$\hat{\sigma}^2 = CM_{\epsilon\epsilon} = \frac{\sum_{j=1}^t (r_j - 1) S_j^2}{\sum_{j=1}^t (r_j - 1)} = \frac{\sum_{j=1}^t (r_j - 1) S_j^2}{n - t}$$

Ahora:

$$(r_j - 1) S_j^2 = \sum_{i=1}^{r_j} (Y_{ij} - \bar{Y}_{.j})^2 = \sum_{i=1}^{r_j} Y_{ij}^2 - r_j \bar{Y}_{.j}^2 = \sum_{i=1}^{r_j} Y_{ij}^2 - \frac{Y_{.j}^2}{r_j} = SC_{EE_j}, \text{ siendo}$$

SC_{EE_j} la suma de cuadrados de los errores para el tratamiento j -ésimo.

Entonces,

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^t SC_{EE_j}}{n - t}.$$

Para el estimador de μ_j , $\hat{\mu}_j = \bar{Y}_{.j}$ se tiene que:

$$\hat{\mu}_j \sim N\left(\mu_j, \frac{\sigma^2}{r_j}\right) \quad \text{para } j = 1, 2, \dots, t.$$

Entonces la variable aleatoria

$$t = \frac{\hat{\mu}_j - \mu_j}{\frac{\hat{\sigma}_j}{\sqrt{r_j}}} \sim t_{n-t}.$$

6.2.4. Hipótesis de interés

En el modelo de efectos fijos se desea probar si los efectos de los diferentes tratamientos son distintos de cero o también si las medias de los tratamientos presentan diferencias estadísticas. El experimentador puede probar los siguientes tipos de hipótesis.

1. $H_0 : \sum_{j=1}^t c_j \mu_j = a$, para un conjunto de constantes c_1, c_2, \dots, c_t y a .
2. $H_0 : \mu_1 = \mu_2 = \dots = \mu_t$ y
3. $H_0 : \mu_j = \mu_{j'}$ para algún $j \neq j'$

Para el primer caso se puede probar que el error estándar de $\sum_{j=1}^t c_j \hat{\mu}_j$ es:

$$ee\left(\sum_{j=1}^t c_j \hat{\mu}_j\right) = \hat{\sigma} \sqrt{\sum_{j=1}^t \frac{c_j^2}{r_j}}$$

Por lo tanto la estadística

$$t = \frac{\sum_{i=1}^t c_j \hat{\mu}_j - \sum_{i=1}^t c_j \mu_j}{\hat{\sigma} \sqrt{\sum_{i=1}^t \frac{c_j^2}{r_j}}} \sim t_{n-t}$$

es decir, para probar la hipótesis nula $H_0 : \sum_{j=1}^t c_j \mu_j = a$ se debe calcular

$$t_c = \frac{\sum_{j=1}^t c_j \hat{\mu}_j - a}{ee \left(\sum_{j=1}^t c_j \hat{\mu}_j \right)} = \frac{\sum_{j=1}^t c_j \hat{\mu}_j - a}{\hat{\sigma} \sqrt{\sum_{j=1}^t \frac{C_j^2}{r_j}}},$$

rechazándose H_0 sí $|t_c| > t_{(1-\alpha/2, n-t)}$. Las pruebas de hipótesis de los casos 2 y 3 se obtienen como casos particulares de la hipótesis del caso 1.

6.2.5. Análisis de varianza (ANAVA)

Si se tienen r_j observaciones para cada tratamiento, los datos del experimento se pueden representar mediante la matriz dada en la Tabla 2, donde r_j es el número de repeticiones por tratamiento y t es el número de tratamientos, $\sum_{i=1}^{r_j} Y_{ij} = Y_{.j}$ es el total del tratamiento j -ésimo para $j = 1, 2, \dots, t$.

Tabla 2: Datos y totales para un DCA

Repetición	t_1	t_2	t_3	...	t_j	...	t_t
1	Y_{11}	Y_{12}	Y_{13}	...	Y_{1j}	...	Y_{1t}
2	Y_{21}	Y_{22}	Y_{23}	...	Y_{2j}	...	Y_{2t}
3	Y_{31}	Y_{32}	Y_{33}	...	Y_{3j}	...	Y_{3t}
\vdots	\vdots	\vdots	\vdots	...	\vdots	...	\vdots
r_j	$Y_{r_1 1}$	$Y_{r_2 2}$	$Y_{r_3 3}$...	$Y_{r_j j}$...	$Y_{r_t t}$
Total	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$...	$Y_{.j}$...	$Y_{.t}$
Media	$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	$\bar{Y}_{.3}$...	$\bar{Y}_{.j}$...	$\bar{Y}_{.t}$

También se tiene que $Y_{..} = \sum_{i=1}^{r_j} \sum_{j=1}^t Y_{ij} = \sum_{j=1}^t Y_{.j}$, llamado total general, $\bar{Y}_{.j} =$

$\frac{Y_{.j}}{r_j}$ y $\bar{Y}_{..} = \frac{Y_{..}}{n}$, que representa la media general siendo $n = \sum_{j=1}^t r_j$.

■ **Sumas de cuadrados (SC) y grados de libertad (gl)**

En este diseño la variación total se divide en dos componentes

$$SC_{Total} = SC_{Tratamiento} + SC_{Error Experimental},$$

que se puede simplificar un poco escribiendo

$$SC_{Total} = SC_{TTO} + SC_{EE}$$

$$\text{donde } SC_{Total} = \sum_{i=1}^{r_j} \sum_{j=1}^t (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{r_j} \sum_{j=1}^t Y_{ij}^2 - \frac{Y_{..}^2}{n} = \sum_{i=1}^{r_j} \sum_{j=1}^t Y_{ij}^2 - FC$$

Donde:

$$FC = \frac{Y_{..}^2}{n}, \text{ llamado factor de corrección.}$$

$$SC_{TTO} = \sum_{i=1}^{r_j} \sum_{j=1}^t (\bar{Y}_{.j} - \bar{Y}_{..})^2 = \sum_{j=1}^t r_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 = \sum_{j=1}^t \frac{Y_{.j}^2}{r_j} - \frac{Y_{..}^2}{n} = \sum_{j=1}^t \frac{Y_{.j}^2}{r_j} - FC$$

$$\text{y } SC_{EE} = SC_{Total} - SC_{TTO} = \sum_{i=1}^{r_j} \sum_{j=1}^t Y_{ij}^2 - \sum_{j=1}^t \frac{Y_{.j}^2}{r_j}.$$

También $gl_{Total} = gl_{TTO} + gl_{EE}$, donde $gl_{Total} = \sum_{j=1}^t r_j - 1$ y $gl_{TTO} = t - 1$.

6.2.6. Pruebas de hipótesis y errores estándar

Como se mencionó anteriormente el interés está centrado en comparar las medias de los tratamientos, para verificar la hipótesis nula que entre las medias tratamientos no existen diferencias significativas. Esta hipótesis se puede enunciar como:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_t = \mu \text{ vs } H_1 : \mu_j \neq \mu_{j'} \text{ para algún } j \neq j',$$

la estadística utilizada para esta prueba es:

$$F_c = \frac{CM_{TTO}}{CM_{EE}} \sim F_{(t-1, n-t)}$$

y se rechaza H_0 , al nivel de significancia α , sí $F_c > F_{\alpha, (t-1, n-t)}$.

Todos estos procedimientos se pueden resumir en la Tabla 3 de análisis de varianza, dada a continuación.

Tabla 3: ANAVA para un DCA

F.V.	gl	SC	CM	F_c	$F_{5\%}$	$F_{1\%}$
TTO	$t - 1$	SC_{TTO}	CM_{TTO}	$\frac{CM_{TTO}}{CM_{EE}}$	$F_{5\%, (t-1, n-t)}$	$F_{1\%, (t-1, n-t)}$
EE	$n - t$	SC_{EE}	CM_{EE}			
Total	$n - 1$	SC_{Total}				

Sí no se rechaza H_0 al nivel de significancia del α se dice que no existen diferencias estadísticas significativas entre los tratamientos, lo cual se nota con NS , pero aún así se pueden hablar de diferencias al 6 %, 7 %, 8 %, 9 %, y hasta del 10 % que también es muy usual.

Así mismo, cuando el investigador desea llevar a cabo cualquier prueba para un tratamiento en particular, esta se puede realizar utilizando en este caso el error estándar de estimación para una media de tratamiento, el cual viene dado por:

$$S_{\bar{Y}.j} = \sqrt{\frac{CM_{EE}}{r_j}}$$

y cualquier inferencia para el $j - \text{ésimo}$ tratamiento se debe realizar con este error.

6.2.7. Ejemplo en R para un DCA.

Para el caso del ingeniero que desea maximizar la resistencia a la tensión de una nueva fibra sintética, esta toma seis concentraciones distintas o porcentajes de algodón para la fabricación de la nueva fibra de 15 %, 20 %, 25 %, 30 % y 35 %. La información se encuentra en la Tabla 4. (Montgomery, 2012)

Tabla 4: Porcentajes de algodón

% algodón	Resistencia				
	1	2	3	4	5
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

Se llevará a cabo este procedimiento en R teniendo en cuenta los siguientes pasos:

1. Al archivo que contiene los datos se le asigna el nombre `Base_1` y se carga utilizando la función `read.csv`, se encuentra formato de texto, está compuesto por dos variables llamadas porcentaje y resistencia.

```
1 > Base_1 <- read.csv("C:/Users/Hp E2/Desktop/MONOGRAFIA -
  TODO
2 /Base_1.txt", sep=" ")
```

2. Ahora usando la función `head`, se observa los primeros 6 elementos de la nuestra base de datos.

```
1 > head(Base_1)
2   Porcentaje Resistencia
3 1         15      7
4 2         15      7
5 3         15     15
6 4         15     11
7 5         15      9
8 6         20     12
```

3. Luego se convierte la variable porcentaje en un factor usando la función `factor`, para que R evalúe el número de factores que hay en vez de tratarlos de modo independiente.

```
1 > Porcentaje <- factor(Base_1$Porcentaje)
```

4. Se utiliza la función `aov` esta función ajusta el modelo de análisis de varianza, utilizando la función `summary` se obtiene el análisis de varianza dado en la Tabla 3.

```

1 > ANOVA=aov(Resistencia~Porcentaje,data = Base_1)
2 > summary(ANOVA)
3           Df Sum Sq Mean Sq F value    Pr(>F)
4 Porcentaje   4  475.8   118.94   14.76 9.13e-06 ***
5 Residuals  20   161.2     8.06
6 ---
7 Signif. codes:  0   '***'   0.001   '**'   0.01   '*'
                  0.05   '.'   0.1     1

```

El $p_{valor} = 9,13e-06$ indica que hay una diferencia significativa entre los grupos en términos de la variable Resistencia. Ahora en la parte inferior de los resultados, se presentan los códigos de significancia. En este caso, se muestra "***", lo que significa que la diferencia entre los grupos es estadísticamente significativa a un nivel de confianza muy alto.

En resumen, los resultados indican que la variable Porcentaje tiene un efecto significativo en la variable Resistencia. La diferencia entre los grupos de porcentajes evaluados es estadísticamente significativa, lo que implica que el porcentaje tiene un impacto en la resistencia medida.

6.3. Pruebas de validación de los supuestos del modelo

Existen ciertos supuestos sobre los errores del modelo en los que se basa el ANAVA, los cuales son:

1. Los errores tienen distribución normal con media cero, $E(\epsilon_{ij}) = 0$ y varianza constante, $V(\epsilon_{ij}) = \sigma^2$.
2. Son variables aleatorias independientes, es decir, para $i \neq i'$, $Cov(\epsilon_i, \epsilon_{i'}) = 0$.

En la práctica los errores se estiman con los datos experimentales que se obtienen en campo; para un DCA la estimación se lleva a cabo mediante la expresión

$$e_{ij} = \hat{\epsilon}_{ij} = Y_{ij} - \bar{Y}_{.j}, \quad i = 1, 2, \dots, r_j, \quad j = 1, 2, 3, \dots, t, \quad (2)$$

es decir, los errores estimados se calculan por tratamiento, restándole a los datos obtenidos de cada tratamiento la media de ese tratamiento.

Los dos supuestos del modelo más importantes a cumplirse, se refieren a la normalidad y la homogeneidad de varianza de los errores, puesto que de hecho los tratamientos son asignados aleatoriamente a las unidades experimentales al inicio de todo experimento y esto garantiza, en cierta forma, la independencia de los errores. A continuación se presentan los algoritmos para llevar a cabo las pruebas de homogeneidad de varianza y normalidad. (Martínez y Robles, 2013, p.72).

6.3.1. Prueba de normalidad de los errores de Shapiro - Wilk

La hipótesis a probar en este caso es:

H_0 : Los errores provienen de una población con distribución normal

H_1 : Los errores no provienen de una población con distribución normal.

Para llevar a cabo esta prueba inicialmente se calculan los errores del modelo bajo un diseño completamente al azar, los cuales vienen dados por la expresión.

$$e_{ij} = Y_{ij} - \bar{Y}_{.j} \quad \text{para cada } j = 1, 2, \dots, t,$$

posteriormente se ordenan los errores de menor a mayor, $e_{(1)}, e_{(2)}, \dots, e_{(n)}$, eliminando el término que queda en la mitad de los datos, si el número de observaciones fuese impar. El paso siguiente es realizar las diferencias $d_{(1)} = e_{(n)} - e_{(1)}, d_{(2)} = e_{(n-1)} - e_{(2)}, \dots$, es decir, el más grande menos el más pequeño, el segundo más grande menos el segundo más pequeño, etc; hasta llegar a los dos datos del centro; posteriormente estas diferencias se ponderan, en ese orden, por los términos $a_{k,n}$, los cuales se encuentran en la Tabla dada por Shapiro - Wilk, es decir se realizan los productos $a_{k,n} * d_{(k)}$, posteriormente se calcula la estadística

$$SW = \frac{1}{SC_{EE}} \left[\sum_k (a_{k,n} * d_{(k)}) \right]^2 \quad (3)$$

La hipótesis nula, H_0 , es rechazada sí $SW < W_{\alpha, n}$ siendo $W_{\alpha, n}$ el valor crítico de Shapiro - Wilk al nivel de significancia α . (Montgomery, 2012)

6.3.2. Ejemplo en R de la prueba de Shapiro - Wilk

El ejemplo 6.2.7. se tendrá en cuenta para realizar la prueba de Shapiro - Wilk en los residuales del modelo.

1. Las hipótesis estadísticas son las siguientes:

H_0 : Los errores provienen de una población con distribución normal.

H_1 : Los errores no provienen de una población con distribución normal.

2. Utilizando el comando `modelo$residuals` está seleccionando los residuales del modelo, así la función `shapiro.test` hace los cálculos para verificar la normalidad.

```
1 > shapiro.test(resid(ANOVA))
2
3 Shapiro-Wilk normality test
4
5 data: resid(ANOVA)
6 W = 0.94387, p-value = 0.1818
```

Como el $p_{valor} = 0.1818$ es mayor que el nivel de significancia comúnmente utilizado $\alpha = 0,05$, no se dispone de evidencia suficiente para rechazar la hipótesis nula. Por lo tanto, no hay suficiente evidencia para afirmar que los residuos del modelo ANOVA no siguen una distribución normal. En otras palabras, los residuos se ajustan razonablemente bien a una distribución normal.

En resumen, según el test de normalidad de Shapiro-Wilk, no se encontraron evidencias suficientes para rechazar la hipótesis de normalidad de los residuos del modelo ANOVA.

La información sobre la función `shapiro.test` se obtiene de la página web (Rpubs, 2019).

6.3.3. Pruebas de homogeneidad de varianzas

6.3.4. Prueba de Hartley

Para llevar a cabo la prueba de hipótesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 = \sigma^2$$

H_1 : las varianzas no son homogéneas,

donde σ_j^2 representa la varianza del j –ésimo tratamiento y además donde todos los tratamientos tienen el mismo número de repeticiones, es decir $r_1 = r_2 = \dots = r_t = r$. Hartley propone la siguiente estadística para llevar a cabo la prueba de hipótesis.

$$F_{Max} = \frac{Max \{S_j^2\}}{Min \{S_j^2\}},$$

conocida como la estadística F máxima de Hartley. En esta prueba H_0 se rechaza sí $F_{Max} > F_{Max(\alpha, t, v)}$ donde $F_{Max(\alpha, t, v)}$ es el valor en la Tabla F_{Max} al nivel de significancia α , para t tratamientos y $v = r - 1$ (Montgomery, 2012).

6.3.5. Ejemplo en R de la prueba de Hartley

El ejemplo 6.2.7., se tendrá en cuenta para realizar la prueba de Hartley. Para acceder a la función `hartleyTest`, se debe cargar la librería `PMCMRplus`.

- Utilizando la función `hartleyTest`, la cual realiza la prueba de la razón F máxima de Hartley con la hipótesis nula de que las varianzas en cada uno de los tratamientos son iguales.

```
1 > hartleyTest(resid(ANOVA) ~ Porcentaje)
2
3 Hartley's maximum F-ratio test of homogeneity of
  variances
4
5 data: resid(ANOVA) by Porcentaje
6 F Max = 2.6047, df = 4, k = 5, p-value = 0.8964
```

Como el p_{valor} es 0.8964, y éste es un valor mayor al compararlo con un $\alpha = 0,05$, este indica que no encuentra diferencias significativas entre las varianzas de los tratamientos.

La información sobre la función `hartleyTest` se obtiene de la página web (*Hartley's maximum F-ratio test of homogeneity of variances*, s.f.).

6.3.6. Prueba de Bartlett

Es una prueba más liberal que la anterior. Para llevar a cabo la prueba de hipótesis de igualdad de medias o de varianza, los tratamientos no necesariamente deben tener el mismo número de repeticiones. La estadística dada por Bartlett está definida por:

$$\chi^2 = \frac{1}{C} \left[v \ln(CM_{EE}) - \sum_{j=1}^t v_j \ln(S_j^2) \right]; j = 1, 2, \dots, t$$

donde: $v_j = r_j - 1$; $v = \sum_{j=1}^t v_j$ y $C = 1 + \frac{1}{3(t-1)} \left(\sum_{j=1}^t \frac{1}{v_j} - \frac{1}{v} \right)$ y la hipótesis nula H_0 se rechaza si $\chi^2 > \chi_{(\alpha, t-1)}^2$. (*Bartlett's test for homogeneity*, s.f.)

6.3.7. Ejemplo en R de la prueba de Bartlett

Siguiendo con el ejemplo 6.2.7., la librería que se va a utilizar es `stats`.

- Utilizando la función `bartlett.test`.

```
1 > bartlett.test(resid(ANOVA) ~ Porcentaje)
2
3 Bartlett test of homogeneity of variances
4
5 data: resid(ANOVA) by Porcentaje
6 Bartlett's K-squared = 0.93309, df = 4, p-value = 0.9198
```

Como el p_{valor} es 0.9198, y éste es un valor mayor al compararlo con un $\alpha = 0,05$, este indica que no encuentra diferencias significativas entre las varianzas de los tratamientos.

La información sobre la función `bartlett.test` se obtiene de la página web (*Bartlett's test for homogeneity*, s.f.)

6.3.8. Prueba de Levene

Levene propuso efectuar un análisis de varianza simple sobre las variables Z_{ij} . Si el estadístico F_c resulta significativo, se rechaza la hipótesis de homogeneidad de varianzas. Este test consiste en transformar la variable de

respuesta y posteriormente se calcula una tabla ANOVA para las variables transformadas.

El estadístico de prueba se define como:

$$F_c = \frac{(N - t) \sum_{j=1}^t r_j (\bar{Z}_{.j} - \bar{Z}_{..})^2}{(t - 1) \sum_{j=1}^t \sum_{i=1}^{r_j} (Z_{ij} - \bar{Z}_{.j})^2}$$

donde Z_{ij} puede tener una de las siguientes definiciones:

- $Z_{ij} = |Y_{ij} - \bar{Y}_{.j}|$ donde $\bar{Y}_{.j}$ es la media del j-ésimo tratamiento.
- $Z_{ij} = |Y_{ij} - \tilde{Y}_{.j}|$ donde $\tilde{Y}_{.j}$ es la mediana del j-ésimo tratamiento.

$\bar{Z}_{..}$ es la media general del Z_{ij} y $\bar{Z}_{.j}$ es la media del j-ésimo tratamiento de los Z_{ij} , N es el número de unidades experimentales, t el número de tratamientos y r_j es el número de repeticiones en el j-ésimo tratamiento.

En la prueba de Levene se rechaza la hipótesis de que las varianzas son iguales con un nivel de significancia α si $F_c > F_{(\alpha, t-1, N-t)}$ donde $F_{(\alpha, t-1, N-t)}$ es el valor crítico superior de la distribución F con $t - 1$ grados de libertad en el numerador y $N - t$ grados de libertad en el denominador a un nivel de significancia α . (Ciencia UANL, 2012)

6.3.9. Ejemplo en R de la prueba de Levene

Siguiendo con el ejemplo anterior, se puede aplicar con la función `leveneTest`, esta se obtiene instalando y cargando la librería `car`.

- Utilizando la función `leveneTest`.

```

1 > leveneTest(resid(ANOVA) ~ Porcentaje)
2 > Levene's Test for Homogeneity of Variance (center =
  median)
3      Df F value Pr(>F)
4 group  4  0.3179 0.8626
5      20

```

Como el p_{valor} es 0.8626, y éste es un valor mayor al compararlo con un $\alpha = 0,05$, este indica que no encuentra diferencias significativas entre las

varianzas de los tratamientos.

La información sobre la función `leveneTest` se obtiene de la página web (*Test de Levene*, s.f.)

6.4. Comparaciones múltiples

Supóngase que al efectuar un análisis de varianza para un modelo de efectos fijos la hipótesis nula es rechazada, esto implica que existen diferencias significativas entre las medias de los tratamientos, aunque no se especifique exactamente cuáles de ellas son diferentes. Cuando se presenta esta situación puede ser útil realizar comparaciones adicionales entre grupos de medias de los tratamientos. Se debe tener en cuenta que la media del j -ésimo tratamiento se define mediante $\mu_j = \mu + \tau_j$ y su estimación es $\bar{Y}_{\cdot j}$. Las comparaciones entre medias de tratamientos se realizan en términos de los totales de tratamientos $Y_{\cdot j}$ o de los promedios de tratamientos $\bar{Y}_{\cdot j}$ (Martínez y Robles, 2013, p.83).

6.4.1. Método de las diferencias mínimas significativas (DMS)

Supone que después de haber rechazado la hipótesis nula, con la prueba F del análisis de varianza, se desea probar que:

$$H_0 : \mu_j = \mu_{j'}, \text{ para } j \neq j'.$$

Esto puede hacerse empleando la estadística

$$t_0 = \frac{\bar{Y}_{\cdot j} - \bar{Y}_{\cdot j'}}{\sqrt{CM_{EE} \left(\frac{1}{r_j} + \frac{1}{r_{j'}} \right)}}.$$

Entonces, suponiendo una hipótesis alterna bilateral $\mu_j \neq \mu_{j'}$, la pareja de medias μ_j y $\mu_{j'}$ se consideran diferentes si

$$|\bar{Y}_{\cdot j} - \bar{Y}_{\cdot j'}| > DMS$$

donde la cantidad

$$DMS = t_{(1-\frac{\alpha}{2}, n-t)} \sqrt{CM_{EE} \left(\frac{1}{r_j} + \frac{1}{r_{j'}} \right)}$$

se denomina mínima diferencia significativa (*DMS* o *LSD*). Sí el diseño es balanceado, entonces $r_1 = \dots = r_t = r$ y

$$DMS = t_{(1-\frac{\alpha}{2}, n-t)} \sqrt{\frac{2CM_{EE}}{r}}.$$

Para usar el procedimiento *DMS*, simplemente se comparan las diferencias observadas entre cada par de promedios con el valor correspondiente de la *DMS*.

Sí $|\bar{Y}_{.j} - \bar{Y}_{.j'}| > DMS$, se concluye que las medias poblacionales μ_j y $\mu_{j'}$ son estadísticamente diferentes.

6.4.2. Ejemplo en R para DMS

Siguiendo con el ejemplo anterior 6.2.7., se puede aplicar con la función `LSD.test`, esta se obtiene instalando y cargando la librería `agricolae`. Cuando se llevó a cabo el ANOVA que se observó diferencias significativas entre los tratamientos, por tanto, se lleva a cabo las comparaciones múltiples.

- Se utiliza la función `LSD.test` para llevar a cabo el método de diferencia mínima significativa

```

1 > LSD.test(ANOVA, "Porcentaje", console=TRUE)
2
3 Study: ANOVA ~ "Porcentaje"
4
5 LSD t Test for Resistencia
6
7 Mean Square Error: 8.06
8
9 Porcentaje, means and individual (95 %) CI
10
11 Resistencia      std      r      LCL      UCL      Min Max
12 15 9.8           3.346640 5 7.151566 12.44843 7 15
13 20 15.4          3.130495 5 12.751566 18.04843 12 18
14 25 17.6          2.073644 5 14.951566 20.24843 14 19
15 30 21.6          2.607681 5 18.951566 24.24843 19 25
16 35 10.8          2.863564 5 8.151566 13.44843 7 15
17
18 Alpha: 0.05 ; DF Error: 20
19 Critical Value of t: 2.085963
20

```

```

21 least Significant Difference: 3.745452
22
23 Treatments with the same letter are not significantly
    different.
24
25 Resistencia groups
26 30 21.6 a
27 25 17.6 b
28 20 15.4 b
29 35 10.8 c
30 15 9.8 c

```

De acuerdo a los resultados obtenidos con el método de la diferencia mínima significativa, se observa que entre los tratamientos 30, 25 y 35 existen diferencias significativas, pero entre los tratamientos 25 y 20, no existen diferencias significativas.

La información sobre la función `LSD.test` se obtiene de la página web (*LSD.Test function*, s.f.)

6.4.3. La prueba de intervalos múltiples de Duncan

La prueba de intervalos múltiples de Duncan es una opción útil cuando se requiere realizar comparaciones múltiples y se busca identificar qué grupos o tratamientos específicos presentan diferencias significativas. Sin embargo, es importante tener en cuenta que esta prueba asume igualdad de varianzas entre los grupos y debe aplicarse con precaución en situaciones en las que esta suposición no se cumpla. Se comienza calculando el error estándar para una media de tratamiento dado por:

$$S_{\bar{Y}.j} = \sqrt{\frac{CM_{EE}}{r}}. \quad (4)$$

Luego se ordena de manera descendente los t promedios de los tratamientos.

Los valores tabulados para la prueba de Duncan, a un nivel de significancia α , $Q_\alpha(p, gl_{EE})$ donde p representa el número de medias entre dos tratamientos y gl_{EE} el número de grados de libertad del error experimental, se obtienen de la siguiente forma: para cada comparación entre dos medias de tratamiento se debe calcular este valor, por ejemplo, al comparar el tratamiento con la

media mayor contra el resto de tratamientos, se procede así: al comparar esta media de tratamiento con la media más pequeña obtenida, se tiene $p = t$ medias de tratamiento entre las medias comparadas, así se busca el valor $Q_\alpha(t, gl_{EE})$, cuando se va a comparar la media más grande con la segunda media más pequeña, entonces hay $p = t - 1$ medias entre las comparadas; entonces ahora el valor de la tabla se busca con $Q_\alpha(t - 1, gl_{EE})$, este proceso se lleva a cabo hasta llegar a comparar la media más grande con la segunda media más grande, cuyo valor en la tabla será ahora $Q_\alpha(2, gl_{EE})$. Con el resto de comparaciones la metodología es similar, así cuando usted compara la segunda media más grande con la media más pequeña, existen $p = t - 1$ medias entre estas dos, por lo tanto el valor de la tabla será $Q_\alpha(t - 1, gl_{EE})$. Se puede seguir el procedimiento hasta comparar las $t(t - 1)/2$ medias de tratamientos en el experimento (Martínez y Robles, 2013, p.83).

Para cada comparación entre medias de tratamiento se calcula el valor $R_p = Q_\alpha(p, gl_{EE})S_{\bar{Y}_{.j}}$ para $p = 2, 3 \dots t$, y la hipótesis nula de igualdad de medias entre los tratamientos comparados se rechaza sí.

$$|\bar{Y}_{.j} - \bar{Y}_{.j'}| > R_p.$$

En este procedimiento se debe tener en cuenta que si dos medias no difieren significativamente, entonces cualquier otra media entre estas dos no debe diferir con las comparadas inicialmente (Martínez y Robles, 2013, p.87).

En muchos experimentos el número de repeticiones por tratamiento resulta ser de diferente tamaño r_j , aun así se puede aplicar la prueba de Duncan, aunque con menos precisión, reemplazando r por la media armonica r_h dada por (Martínez y Martínez, 1997).

$$r_h = \frac{t}{\sum_{j=1}^t 1/r_j}.$$

6.4.4. Ejemplo en R para la prueba de intervalos múltiples de Duncan

De acuerdo al ejemplo que se viene trabajando, en este caso ya no se debe instalar ni cargarla librería `agricolae`, ya que se realizó en el procedimiento anterior.

- Para realizar la prueba de Duncan se utiliza la función `duncan.test`.

```

1 > duncan.test(ANOVA,"Porcentaje",console=TRUE)
2
3 Study: ANOVA ~ "Porcentaje"
4
5 Duncan's new multiple range test
6 for Resistencia
7
8 Mean Square Error: 8.06
9
10 Porcentaje, means
11
12      Resistencia  std      r  Min Max
13 15          9.8    3.346640 5    7  15
14 20         15.4    3.130495 5   12  18
15 25         17.6    2.073644 5   14  19
16 30         21.6    2.607681 5   19  25
17 35         10.8    2.863564 5    7  15
18
19 Alpha: 0.05 ; DF Error: 20
20
21 Critical Range
22          2      3      4      5
23 3.745452 3.931466 4.049682 4.132249
24
25 Means with the same letter are not significantly
26      different.
27
28      Resistencia groups
29 30 21.6      a
30 25 17.6      b
31 20 15.4      b
32 35 10.8      c
33 15  9.8      c

```

Las medias se agrupan y se les asigna una letra. Aquellas medias que comparten la misma letra no son significativamente diferentes entre sí. En este caso, las medias del nivel "30" (21.6) se agrupan con la letra **a**, las del nivel "25" (17.6) y "20" (15.4) se agrupan con la letra **b**, y las del nivel "35" (10.8) y "15" (9.8) se agrupan con la letra **c**. En conclusión el test de comparación múltiple de Duncan, se encontraron diferencias significativas entre los diferentes niveles del factor Porcentaje en términos de la variable Resistencia. Las medias se agruparon en tres grupos distintos, identificados por las letras

a, b y c.

La información sobre la función `duncan.test` se obtiene de la página web (*duncan.test function*, s.f.)

6.4.5. Prueba honesta de Tukey

La prueba honesta de Tukey, también conocida como prueba de Tukey HSD (por sus siglas en inglés, Honest Significant Difference), es un procedimiento estadístico utilizado para realizar comparaciones múltiples entre las medias de varios grupos o tratamientos en un estudio experimental.

El objetivo de la prueba de Tukey HSD es determinar si existen diferencias significativas entre las medias de los grupos y, en caso afirmativo, identificar qué grupos presentan diferencias significativas.

Su procedimiento requiere el uso de la cantidad $q_\alpha(t, gl_{EE})$ para determinar el valor crítico de todas las comparaciones por pares, independientemente de las medias en un grupo. Así la prueba de Tukey declara dos medias significativamente diferentes si el valor absoluto de sus diferencias muestrales excede el valor (Martínez y Robles, 2013, p.88). Esto implica que.

$$|\bar{Y}_{.j} - \bar{Y}_{.j'}| > T_\alpha = q_\alpha(t, gl_{EE}) S_{\bar{Y}_{.j}}, \quad (5)$$

donde $S_{\bar{Y}_{.j}}$ está definida como $S_{\bar{Y}_{.j}} = \sqrt{\frac{CM_{EE}}{r}}$, debe notarse que en todas las comparaciones sólo se usa el valor crítico T_α de la tabla de Tukey.

6.4.6. Ejemplo en R para la prueba honesta de Tukey

De acuerdo al ejemplo 6.2.7., que se viene trabajando, se utiliza la librería `agricolae`.

- Para realizar la prueba de Tukey se utiliza la función `HSD.test`.

```
1 > TukeyHSD(ANOVA)
2 Tukey multiple comparisons of means
3 95% family-wise confidence level
4
```

```

5 Fit: aov(formula = Repeticiones ~ Porcentaje, data = Base
6         _1)
7 $Porcentaje
8         diff          lwr          upr          p adj
9 20-15    5.6    0.2270417 10.9729583 0.0385024
10 25-15    7.8    2.4270417 13.1729583 0.0025948
11 30-15   11.8    6.4270417 17.1729583 0.0000190
12 35-15    1.0   -4.3729583  6.3729583 0.9797709
13 25-20    2.2   -3.1729583  7.5729583 0.7372438
14 30-20    6.2    0.8270417 11.5729583 0.0188936
15 35-20   -4.6   -9.9729583  0.7729583 0.1162970
16 30-25    4.0   -1.3729583  9.3729583 0.2101089
17 35-25   -6.8  -12.1729583 -1.4270417 0.0090646
18 35-30 -10.8  -16.1729583 -5.4270417 0.0000624

```

Por ejemplo, si observamos la primera fila: 20-15, indica la comparación entre los niveles 20 y 15 del factor "Porcentaje". La diferencia media es de 5.6 y el intervalo de confianza para esta diferencia está entre 0.2270417 y 10.9729583. El valor p ajustado es de 0.0385024, lo que sugiere una diferencia estadísticamente significativa entre estos dos niveles.

La información sobre la función TukeyHSD se obtiene de la página web (*TukeyHSD function*, s.f.).

6.4.7. Prueba de Scheffé

Una ventaja de esta prueba es su robustez frente a la violación de los supuestos del ANAVA, además que no necesita igual número de repeticiones por tratamiento. Se pueden llevar a cabo cualquiera de las posibles comparaciones entre medias de tratamientos (Martínez y Robles, 2013, p.89).

En esta prueba la hipótesis nula es:

$$H_0 : \mu_j = \mu_{j'}, \text{ para toda } j \neq j' \text{ y se rechaza sí}$$

$$|\bar{Y}_{.j} - \bar{Y}_{.j'}| > \sqrt{(t-1) F_{(\alpha, t-1, v)}} \sqrt{CM_{EE} \left(\frac{1}{r_j} + \frac{1}{r_{j'}} \right)}, \quad (6)$$

donde: $v = gl_{EE}$.

6.4.8. Ejemplo en R para la prueba de Scheffé

Se tiene en cuenta el ejemplo 6.2.7., para realizar esta prueba también se utiliza la librería `agricolae`.

- Para realizar la prueba de Scheffé se utiliza la función `scheffe.test`.

```
1 > scheffe.test(ANOVA,"Porcentaje",console=TRUE)
2
3 Study: ANOVA ~ "Porcentaje"
4
5 Scheffe Test for Resistencia
6
7 Mean Square Error : 8.06
8
9 Porcentaje, means
10
11 Resistencia std r Min Max
12 15 9.8 3.346640 5 7 15
13 20 15.4 3.130495 5 12 18
14 25 17.6 2.073644 5 14 19
15 30 21.6 2.607681 5 19 25
16 35 10.8 2.863564 5 7 15
17
18 Alpha: 0.05 ; DF Error: 20
19 Critical Value of F: 2.866081
20
21 Minimum Significant Difference: 6.079555
22
23 Means with the same letter are not significantly
    different.
24
25 Resistencia groups
26 30 21.6 a
27 25 17.6 ab
28 20 15.4 bc
29 35 10.8 c
30 15 9.8 c
```

De acuerdo a los resultados obtenidos utilizando la prueba de Scheffé, el tratamiento 30 y 35 existen diferencias significativas, mientras que entre tratamientos 35 y 15 no existen diferencias significativas.

La información sobre la función `scheffe.test` se obtiene de la página web (*scheffe.test function*, s.f.)

6.4.9. Prueba de Bonferroni

Se recomendada cuando la F del ANAVA resulta no significativo, pero, también se puede usar en caso contrario, es una prueba medianamente conservadora, que se debe usar sólo para un número p no muy grande de comparaciones, puesto de lo contrario se incrementa el nivel de significancia (Martínez y Robles, 2013, p.89).

La hipótesis nula

$$H_0 : \mu_j = \mu_{j'}, \text{ para } j \neq j'$$

se rechaza, al nivel de significancia α , si

$$|\bar{Y}_{.j} - \bar{Y}_{.j'}| > t_{(\alpha/2p, gl_{EE})} \sqrt{CM_{EE} \left(\frac{1}{r_j} + \frac{1}{r_{j'}} \right)}, \quad (7)$$

donde p es el número de comparaciones que se piensan llevar a cabo y $t_{(\alpha/2p, v)}$ es el valor tabulado de la estadística de Bonferroni al nivel de significancia α (Martínez y Robles, 2013, p.90).

6.4.10. Ejemplo en R para la prueba de Bonferroni

Se tiene en cuenta el ejemplo 6.2.7., para realizar esta prueba se utiliza la librería `agricolae`.

- Para realizar la prueba de Bonferroni se utiliza la función `LSD.test`.

```
1 > LSD.test(ANOVA, "Porcentaje", p.adj= "bon", console=TRUE)
2
3 Study: ANOVA ~ "Porcentaje"
4
5 LSD t Test for Resistencia
6 P value adjustment method: bonferroni
7
8 Mean Square Error: 8.06
9
10 Porcentaje, means and individual (95 %) CI
11
12 Resistencia      std r      LCL      UCL Min Max
13 15              9.8 3.346640 5 7.151566 12.44843 7 15
14 20             15.4 3.130495 5 12.751566 18.04843 12 18
```

```

15 25          17.6 2.073644 5 14.951566 20.24843 14 19
16 30          21.6 2.607681 5 18.951566 24.24843 19 25
17 35          10.8 2.863564 5  8.151566 13.44843  7 15
18
19 Alpha: 0.05 ; DF Error: 20
20 Critical Value of t: 3.153401
21
22 Minimum Significant Difference: 5.662089
23
24 Treatments with the same letter are not significantly
   different.
25
26 Resistencia groups
27 30          21.6      a
28 25          17.6     ab
29 20          15.4     bc
30 35          10.8      c
31 15           9.8      c

```

Los resultados obtenidos utilizando la prueba de Bonferroni, se observa los mismos resultados obtenidos al realizar la Scheffé.

La información sobre la función `LSD.test` se obtiene de la página web (López y González, s.f.).

6.4.11. Comparación de medias de tratamientos con un control (prueba de Dunnett)

En varios experimentos uno de los tratamientos es un control o testigo y el investigador puede estar interesado en comparar este control con el resto de tratamientos, es decir llevar a cabo $t - 1$ comparaciones de medias de tratamientos (Martínez y Robles, 2013, p.83). Si se supone que el tratamiento t es un control y se desea probar las hipótesis (Martínez y Robles, 2013, p.83).

$$H_0: \mu_j = \mu_t \text{ contra } H_1: \mu_j \neq \mu_t,$$

para $j = 1, 2, \dots, t - 1$. El procedimiento de Dunnett es una modificación de la prueba t . Para cada hipótesis se calculan las diferencias que se observan en las medias muestrales.

$$|\bar{Y}_{.j} - \bar{Y}_{.t}|, \text{ para } j = 1, 2, \dots, t - 1.$$

La hipótesis nula $H_0: \mu_j = \mu_t$ se rechaza a un nivel de significancia α , si

$$|\bar{Y}_{.j} - \bar{Y}_{.t}| > d_\alpha(t-1, v) \sqrt{CM_{EE} \left(\frac{1}{r_j} + \frac{1}{r_t} \right)}, \quad (8)$$

donde la constante $d_\alpha(t-1, v)$ se encuentra en la Tabla de Dunnett. Siendo $v = gl_{EE}$ y α el nivel de significancia asociado a las $t-1$ pruebas.

6.4.12. Ejemplo en R para la prueba de Dunnett

Se tiene en cuenta el ejemplo 6.2.7., para realizar esta prueba se instala y carga la librería DescTools.

- Para realizar la prueba de Dunnett se utiliza la función `DunnettTest`.

```

1 > DunnettTest(Resistencia, Porcentaje)
2
3   Dunnett's test for comparing several treatments with a
4     control:
5     95% family-wise confidence level
6
7 $ '15'
8      diff      lwr.ci      upr.ci      pval
9 20-15   5.6   0.8373325  10.362668  0.0186 *
10 25-15   7.8   3.0373325  12.562668  0.0011 **
11 30-15  11.8   7.0373325  16.562668  6.9e-06 ***
12 35-15   1.0  -3.7626675   5.762668  0.9469
13
14 ---
15 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
16                 ' ' 1

```

En el análisis anterior se presentan diferencias significativas entre las medias de los tratamientos 20 con 15, 25 con 15 y 30 con 15, ya que el p_{valor} es inferior al compararlo con un $\alpha = 0,05$.

La información sobre la función `DunnettTest` se obtiene de la página web (*How to Perform Dunnett's test in R*, s.f.).

De acuerdo a la descripción aporta por R, la función `(ea1)` puede realizar análisis de varianza y otros análisis complementarios importantes para diseños simples, con datos balanceados y no balanceados. Aporta pruebas como las de Kruskal-Wallis y Friedman

6.4.13. Función (ea1) en R

Esta información fue obtenida de la página web (*Ea2 function*, s.f.). Otra forma de obtenerla es utilizar el comando `help(ea1)` en R.

El autor de la función (ea1) es Emmanuel Arnhold, realiza un análisis de la varianza en diseños simples, este se encuentra en la librería de R llamada (*easynova*).

- La función (ea1) está compuesto por:

```
1 > ea1(data, design = 1, alpha = 0.05, list = FALSE, p.adjust  
      =1, plot=2)
```

- Características

data: los datos son un data.frame.

design:

- 1 = diseño completamente aleatorio
- 2 = diseño de bloques aleatorios
- 3 = diseño de cuadrados latinos
- 4 = cuadrados latinos múltiples
- 5 = análisis con una covariable (diseño completamente aleatorio)
- 6 = análisis con una covariable (diseño de bloques aleatorios)
- 7 = bloques incompletos de tipo I y II
- 8 = bloques incompletos tipo III o bloques aumentados
- 9 = bloques incompletos tipo III en experimentos con animales
- 10 = entramado (análisis intra-bloque)
- 11 = entramado (análisis interbloques)
- 12 = diseño switchback
- 13 = diseño de bloques invertidos
- 14 = prueba de suma de rangos de Kruskal-Wallis
- 15 = prueba de suma de rangos de Friedman

alpha: nivel de significación para las comparaciones múltiples.

`list`: FALSE = una sola variable de respuesta TRUE = respuesta multivariante

`p.adjust`:1="none"; 2="holm"; 3="hochberg"; 4="hommel"; 5="bonferroni"; 6="BH"; 7="BY"; 8="fdr".

`plot`: 1 = gráfico de caja para los residuos; 2 = residuos estandarizados frente a los datos de la secuencia; 3 = residuos estandarizados frente a cuantiles teóricos.

La función (`ea1`) aporta un análisis de rendimientos de la varianza, medias (medias ajustadas), test de comparación múltiple (tukey, snk, duncan, t y scott knott) y análisis de residuos. También devuelve el análisis de kruskal-Wallis y Friedman.

6.4.14. Ejemplo en R para un diseño completamente aleatorio usando la función (`ea1`).

Utilizando los datos que se vienen trabajando, la función (`ea1`) permite obtener un análisis mucho más rápido, nos aporta con una sola línea de código el análisis de varianza, comparaciones de media, comparaciones múltiples, análisis de los residuales y pruebas de validación.

Los pasos a seguir para obtener esos resultados son los siguientes:

- Utilizando la función (`ea1`) se lleva a cabo un análisis con diseño completamente aleatorio. Las variables que se trabajan en esta función de acuerdo con este ejemplo son: la base de datos llamada `Base_0` y (`design=1`) ya que de acuerdo a la descripción previamente dada significa que el análisis que se va a realizar es de un DCA.

```
1 > ea1(Base_0, design=1, plot=2, list=FALSE)
```

1. Se obtiene el análisis de varianza

```
1 $ 'Analysis of variance '  
2   df type I SS mean square F value  p>F  
3 treatments  4   475.76    118.94 14.7568 <0.001  
4 Residuals  20   161.20     8.06    -      -
```

- Se obtiene los 5 promedios de los tratamientos acompañados y se realizan las comparaciones de medias utilizando las pruebas de Tukey, Newman–Keuls method, Duncan, t y Scott Knott.

```

1 $Means
2   treatment  mean  standard.error  tukey  snk  duncan  t
3 1          T4 21.6          1.2696    a    a      a a
4 2          T3 17.6          1.2696   ab    b      b b
5 3          T2 15.4          1.2696   bc    b      b b
6 4          T5 10.8          1.2696   cd    c      c c
7 5          T1  9.8          1.2696    d    c      c c
8 scott_knott
9 a
10 b
11 b
12 c
13 c

```

de acuerdo a los resultados de las comparaciones de medias, el tratamiento T2 se encuentra en un grupo diferente al T4 y T3, indicado por la letra "b" en la prueba tukey. Esto implica que hay una diferencia estadísticamente significativa entre el T2 y los tratamientos anteriores. El promedio del T2 es 15.4, lo que es más bajo en comparación con el T4 y T3.

```

1 $'Multiple comparison test '
2   pair contrast  p(tukey)  p(snk)  p(duncan)  p(t)
3 1  T4 - T3      4.0    0.2101  0.0375    0.0375  0.0375
4 2  T4 - T2      6.2    0.0189  0.0068    0.0034  0.0025
5 3  T4 - T5     10.8    0.0001  0.0000    0.0000  0.0000
6 4  T4 - T1     11.8    0.0000  0.0000    0.0000  0.0000
7 5  T3 - T2      2.2    0.7372  0.2347    0.2347  0.2347
8 6  T3 - T5      6.8    0.0091  0.0032    0.0016  0.0012
9 7  T3 - T1      7.8    0.0026  0.0017    0.0006  0.0003
10 8 T2 - T5      4.6    0.1163  0.0186    0.0186  0.0186
11 9 T2 - T1      5.6    0.0385  0.0143    0.0072  0.0054
12 10 T5 - T1     1.0    0.9798  0.5837    0.5837  0.5837
13
14 $'Residual analysis '
15 $'Residual analysis '$'residual analysis '
16   values
17 p.value Shapiro-Wilk test    0.1818
18 p.value Bartlett test       0.9198
19 coefficient of variation (%) 18.8800
20 first value most discrepant  3.0000

```

```

21 second value most discrepant 24.0000
22 third value most discrepant 21.0000

```

Los resultados de la prueba de comparación múltiple indican que todas las diferencias entre los tratamientos son estadísticamente significativas. El análisis residual sugiere que los residuos pueden seguir una distribución normal y que hay homogeneidad de varianza entre los grupos.

6.5. Contrastes

Los contrastes entre las medias de los tratamientos contribuyen a responder preguntas específicas sobre el experimento que se lleva cabo. Un contraste entre medias se define como:

Definición: Un contraste es toda combinación lineal de medias de los tratamientos, donde la suma algebraica de sus coeficientes es igual a cero. En general una combinación lineal o contraste es de la forma: (Martínez y Robles, 2013, p.91)

$$Z_1 = c_{11}\mu_1 + c_{12}\mu_2 + \cdots + c_{1t}\mu_t = \sum_{j=1}^t c_{1j}\mu_j \quad (9)$$

donde se cumple que $\sum_{j=1}^t \frac{c_{1j}}{r_j} = 0$.

En caso de que los tratamientos tienen el mismo número de repeticiones, la condición anterior se reduce a:

$$\sum_{j=1}^t c_{1j} = 0. \quad (10)$$

Las hipótesis que involucran grupos de medias de los tratamientos a partir del concepto de acuerdo al contraste es:

$$H_0 : Z_1 = \sum_{j=1}^t c_{1j}\mu_j = 0 \quad \text{contra} \quad H_1 : Z_1 = \sum_{j=1}^t c_{1j}\mu_j \neq 0.$$

donde Z_1 es el contraste que representa una comparación de medias.

Un estimador del contraste dado en la ecuación (9) teniendo en cuenta las medias de los tratamientos es:

$$\widehat{Z}_1 = c_{11}\bar{Y}_{.1} + c_{12}\bar{Y}_{.2} + \cdots + c_{1t}\bar{Y}_{.t} = \sum_{j=1}^t c_{1j}\bar{Y}_{.j}, \quad (11)$$

donde los $\bar{Y}_{.j}$ son los promedios procedentes de muestras aleatorias de tamaño r_j . En el caso que las muestras provengan de una población con distribución normal, por el teorema del límite central se tiene:

$$Y_{.j} \sim N(\mu_j, \sigma_j^2) \quad y \quad \bar{Y}_{.j} \sim N\left(\mu_j, \frac{\sigma_j^2}{r_j}\right).$$

De acuerdo al aporte anterior se obtiene los siguientes resultados:

$$E(\widehat{Z}_k) = E\left[\sum_{j=1}^t c_{kj}\bar{Y}_{.j}\right] = \sum_{j=1}^t c_{kj}E(\bar{Y}_{.j}) = \sum_{j=1}^t c_{kj}\mu_j.$$

y

$$V(\widehat{Z}_k) = V\left[\sum_{j=1}^t c_{kj}\bar{Y}_{.j}\right] = \sum_{j=1}^t c_{kj}^2 V(\bar{Y}_{.j}) = \sum_{j=1}^t c_{kj}^2 \frac{\sigma_j^2}{r_j} = \sigma^2 \sum_{j=1}^t \frac{c_{kj}^2}{r_j},$$

la estadística es:

$$t = \frac{\widehat{Z}_k - Z_k}{\sigma \sqrt{\sum_{j=1}^t \frac{c_{kj}^2}{r_j}}} \sim t_{g l e \epsilon}.$$

la prueba de hipótesis:

$$H_0 : Z_1 = 0 \quad \text{vs} \quad H_1 : Z_1 \neq 0$$

En la toma de decisiones se puede tomar α como nivel de significancia.

■ **Suma de cuadrados de un contraste:**

De acuerdo a la teoría aportada por Martínez y Robles, la suma de cuadrados de un contraste para promedios o totales viene dada por:

$$SC_{\widehat{Z}_k} = \frac{\left(\widehat{Z}_k\right)^2}{\sum_{j=1}^t \frac{c_{kj}^2}{r_j}} \quad \text{para promedios}$$

$$SC_{\widehat{Z}_k} = \frac{(\widehat{Z}_k)^2}{\sum_{j=1}^t r_j c_j^2} \quad \text{para totales}$$

y como cada contraste tiene un grado de libertad, se deduce que

$$CM_{\widehat{Z}_k} = \frac{SC_{\widehat{Z}_k}}{1} = SC_{\widehat{Z}_k}.$$

Con este resultado se llega a que la prueba de hipótesis

$$H_0 : Z_k = 0 \quad \text{contra} \quad H_1 : Z_k \neq 0$$

tenga un segundo estadístico de prueba, que es más utilizado, que el anterior, es dado por

$$F = \frac{CM_{\widehat{Z}_k}}{CM_{EE}} \sim F_{1, gl_{EE}}.$$

de tal forma que si al nivel de significancia α , $F_c > F_{\alpha, 1, gl_{EE}}$ entonces H_0 se rechaza.

■ **Contrastes ortogonales:**

Se establecen dos contrastes:

$$Z_1 = c_{11}\mu_1 + c_{12}\mu_2 + \cdots + c_{1t}\mu_t \quad \text{y} \quad Z_2 = c_{21}\mu_1 + c_{22}\mu_2 + \cdots + c_{2t}\mu_t$$

se dicen ortogonales sí,

$$\sum_{j=1}^t \frac{c_{1j}c_{2j}}{r_j} = 0$$

que para tratamientos con igual número de repeticiones se reduce a:

$$\sum_{j=1}^t c_{1j}c_{2j} = 0.$$

Se puede probar que dados t tratamientos, se pueden obtener a lo más $t - 1$ contrastes ortogonales cada uno con 1 grado de libertad y además que:

$$SC_{TTO} = SC_{\hat{z}_1} + SC_{\hat{z}_2} + \dots + SC_{\hat{z}_{t-1}},$$

puesto que los contrastes que se analizan son independientes dos a dos, lo que conlleva a que la variación que se deduce de uno de los contrastes no es posible volverla a tener en otro contraste.

De acuerdo a la descripción aportada por R la siguiente función aporta una prueba contrastes, devuelve una matriz que contiene coeficientes estimados, errores estándar, valores t y valores p bilaterales.

6.5.1. Función (`fit.contrast`) en R

Esta información fue obtenida de la página web (*fit.contrast function*, s.f.). Otra forma de obtenerla es utilizar el comando `help(fit.contrast)` en R.

La función (`fit.contrast`), calcula y prueba contrastes arbitrarios para objetos de regresión. Este se encuentra en la librería de R llamada (`gmodels`).

- La función (`fit.contrast`) está compuesta por:

```
1 > fit.contrast(model, varname, coeff, ... )
```

- Características

`model`: objeto de regresión (`lm`, `glm`, `aov`, `lme`) para el cual se calcularán los contrastes.

`varname`: nombre de la variable.

`coeff`: vector o matriz que especifica contrastes (uno por fila).

`showall`: devolver todos los coeficientes de regresión. Si es VERDADERO, se devolverán todos los coeficientes del modelo. Si es FALSO (valor predefinido), solo se devolverán los coeficientes correspondientes al contraste especificado.

`conf.int`: valor numérico en (0,1) o NULL. Si se especifica un valor numérico, se calcularán los intervalos de confianza con `conf.int` de probabilidad de cobertura nominal. Si es NULL, no se calcularán los intervalos de confianza.

`df`: indica si se debe devolver una columna que contenga los grados de libertad.

... argumentos opcionales proporcionados por los métodos.

Esta función calcula el(los) contraste(s) especificado(s) reajustando el modelo con los argumentos adecuados. Un contraste de la forma `c(1,0,0,-1)` compararía la media del primer grupo con la media del cuarto grupo.

Devuelve una matriz que contiene los coeficientes estimados, los errores estándar, los valores `t` y los valores `p`. Si `df` es TRUE, se incluye una columna adicional que contiene los grados de libertad. Si se especifica `conf.int` también se devuelven los límites de confianza inferior y superior.

6.5.2. Ejemplo en R para contrastes con un DCA usando la función (`fit.contrast`).

Se tienen en cuenta los resultados aportados por ejemplo 6.2.7., donde el ANOVA muestra que existen diferencias estadísticamente significativas entre los tratamientos. Se instala y luego se carga la librería `gmodels`, esta nos permite utilizar la función `fit.contrast`. Las comparaciones deseadas son:

- En términos ortogonales se escribe de esta forma:

a) $\mu_1 = \frac{\mu_2 + \mu_3 + \mu_4 + \mu_5}{2}$

b) $\mu_2 = \mu_3$

c) $\mu_4 = \mu_5$

d) $\mu_2 + \mu_3 = \mu_4 + \mu_5$

- Llevemos ahora a cabo las pruebas de hipótesis

a) T1 contra el resto

$$H_0 : Z1 = 2\mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5 = 0$$

$$H_1 : Z1 = 2\mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5 \neq 0$$

b) T2 contra T3

$$H_0 : Z2 = \mu_2 - \mu_3 = 0$$

$$H_1 : Z2 = \mu_2 - \mu_3 \neq 0$$

c) T4 contra T5

$$H_0 : Z3 = \mu_4 - \mu_5 = 0$$

$$H_1 : Z3 = \mu_4 - \mu_5 \neq 0$$

c) T2, T3 contra T4, T5

$$H_0 : Z4 = \mu_2 + \mu_3 - \mu_4 - \mu_5 = 0$$

$$H_1 : Z4 = \mu_2 + \mu_3 - \mu_4 - \mu_5 \neq 0$$

Ahora los pasos a seguir son:

1. Se crea una matriz de contrastes teniendo en cuenta las hipótesis planteadas.

```
1 > MatrizContraste=rbind("T1 vs resto"=c(2,-1,-1,-1,-1),
2 "T2 vs T3"=c(0,1,-1,0,0),
3 "T4 vs T5"=c(0,0,0,1,-1),
4 "T2,T3 vs T4,T5"=c(0,1,1,-1,-1))
```

2. Para realizar la prueba de contrastes se utiliza la función `fit.contrast`. Donde se observa la interacción entre los contrastes.

```
1 > Contrastes=fit.contrast(ANOVA, Porcentaje,
2 MatrizContraste, conf=0.95)
3 > Contrastes
4 Estimate Std. Error t value
5 PorcentajeT1 vs resto -17.46667 3.785352 -4.6142782
6 PorcentajeT2 vs T3 -2.20000 1.795550 -1.2252513
7 PorcentajeT4 vs T5 10.80000 1.795550 6.0148699
8 PorcentajeT2,T3 vs T4,T5 0.60000 2.539291 0.2362864
9 Pr(>|t|)
10 1.676859e-04
11 2.347148e-01
12 7.011202e-06
13 8.156135e-01
14
15 lower CI upper CI
16 PorcentajeT1 vs resto -25.362772 -9.570561
```

17	PorcentajeT2 vs T3	-5.945452	1.545452
18	PorcentajeT4 vs T5	7.054548	14.545452
19	PorcentajeT2 ,T3 vs T4 ,T5	-4.696869	5.896869

Los resultados de los contrastes indican que hay diferencias estadísticamente significativas entre el tratamiento 1 y el resto de los tratamientos, así como entre el tratamiento 4 y el tratamiento 5. No se encontraron diferencias significativas entre el tratamiento 2 y el tratamiento 3 ni entre la combinación de los tratamientos 2 y 3 con la combinación de los tratamientos 4 y 5.

6.6. Polinomios ortogonales

Se recomienda utilizar este análisis en situaciones donde los niveles del factor son equidistantes o igualmente espaciados, ya que puede simbolizarse mucho el ajuste de modelos polinomiales por el método de mínimos cuadrados. (Montgomery, 2012). Para un factor X con tres niveles, los efectos del factor están formados por dos componentes: lineal y cuadrático.

Cambio en A	Incremento en la respuesta
0 - 1	$\bar{a}_1 - \bar{a}_0$
1 - 2	$\bar{a}_2 - \bar{a}_1$

La tendencia lineal presenta un incremento en la variable respuesta el cual es igual a la diferencia al pasar de un nivel a otro, este efecto se mide con el promedio de los incrementos al pasar de \bar{a}_0 a \bar{a}_2 , puesto que de existir la linealidad, el incremento al pasar de \bar{a}_0 a \bar{a}_1 debe de ser igual al de pasar de \bar{a}_1 a \bar{a}_2 , así: (Martínez y Robles, 2013, p.97)

$$A_l = \frac{1}{2} [(\bar{a}_1 - \bar{a}_0) + (\bar{a}_2 - \bar{a}_1)] = \frac{1}{2} (\bar{a}_2 - \bar{a}_0).$$

en caso de que se presente una tendencia cuadrática es debido a los incrementos no proporcionales. En este caso el incremento que presenta. \bar{a}_0 a \bar{a}_1 es diferente al de pasar de \bar{a}_1 a \bar{a}_2 , de existir esa tendencia la diferencia entre los incrementos sería diferente de cero por esto el efecto cuadrático es la diferencia de los incrementos: (Martínez y Robles, 2013, p.97).

$$A_C = \frac{1}{2} [(\bar{a}_2 - \bar{a}_1) - (\bar{a}_1 - \bar{a}_0)] = \frac{1}{2} (\bar{a}_2 - 2\bar{a}_1 + \bar{a}_0).$$

A_L como A_C representan un contraste, por esto, para determinar la significancia, o presencia, de cada efecto se plantean las siguientes hipótesis nulas con respecto a los contrastes anteriores: (Martínez y Robles, 2013, p.97)

$$H_0 : A_L = 0 \Leftrightarrow \frac{1}{2} (\mu_2 - \mu_0) = 0 \Leftrightarrow \mu_0 - \mu_2 = 0$$

$$H_0 : A_c = 0 \Leftrightarrow \frac{1}{2} (\mu_2 - 2\mu_1 + \mu_0) = 0 \Leftrightarrow \mu_0 - 2\mu_1 + \mu_2 = 0.$$

De acuerdo a lo anterior los coeficientes ortogonales para el efecto lineal son: 1, 0, -1 y para el efecto cuadrático son 1, -2, 1.

Fisher estableció en 1919 que si t niveles del factor se analizan en un experimento, es posible extraer efectos polinomiales hasta de orden $t - 1$.

El modelo estadístico del polinomio es:

$$Y = \alpha_0 + \alpha_1 p_1(x) + \alpha_2 p_2(x) + \alpha_3 p_3(x) + \dots + \alpha_{t-1} p_{t-1}(x) + \epsilon$$

donde Y es la variable respuesta y X corresponde a los niveles del factor. La hipótesis que se van a utilizar para determinar la significancia de cada término del polinomio o efecto son:

$$H_0 : \alpha_i = 0 \text{ contra } H_1 : \alpha_i \neq 0.$$

Las pruebas de estas hipótesis son análogas a las realizadas sobre un conjunto de contrastes ortogonales, donde los C_{kj} ya no son seleccionados según la situación, sino que están dados en la tabla de polinomios ortogonales, para cada uno de los efectos ya sea lineal, cuadrático, cúbico, etc. Así, es suficiente encontrar las sumas de cuadrados de los $t - 1$ polinomios ortogonales $Z_k = \sum_{j=1}^t C_{kj} \mu_j$ en forma similar a como ocurre para los $t - 1$ contrastes ortogonales. Así, se cumple que: (Martínez y Robles, 2013, p.99).

$$\begin{aligned} SC_{TTO} &= SC_{E. \text{ Lineal}} + SC_{E. \text{ Cuadrático}} + SC_{E. \text{ Cúbico}} + SC_{E. \text{ Cuartico}} + \dots + SC_{E. t-1} \\ &= \sum_{K=1}^{t-1} SC_{E. K}, \end{aligned}$$

donde, para $K = 1, 2, \dots, t - 1$, la $SC_{E.K}$ se obtiene calculando la SC del contraste

$$Z_k = \sum_{j=1}^t c_{kj} \mu_j.$$

Las estimaciones por mínimos cuadrados de los parámetros del modelo polinomial son:

$$\hat{\alpha}_j = \frac{\sum_{j=1}^t \bar{Y}_{.j} P_j(Y)}{\sum_{j=1}^t [P_j(Y)]^2} \quad \text{para} \quad j = 0, 1, 2, \dots, t - 1,$$

es decir,

$$\hat{\alpha}_0 = \frac{\sum_{j=1}^t C_{0j} \bar{Y}_{.j}}{\sum_{j=1}^t C_{0j}^2} = \bar{Y}_{..}, \quad \hat{\alpha}_1 = \frac{\sum_{j=1}^t C_{1j} \bar{Y}_{.j}}{\sum_{j=1}^t C_{1j}^2}, \dots, \hat{\alpha}_k = \frac{\sum_{j=1}^t C_{kj} \bar{Y}_{.j}}{\sum_{j=1}^t C_{kj}^2},$$

donde los coeficientes C_{kj} son dados en la Tabla de polígonos ortogonales para el efecto lineal ($P_1(x)$), cuadrático ($P_2(x)$), cúbico ($P_3(x)$), etc.

La regresión polinomio: según (Gutiérrez y De la Vara, 2012), existen otros modelos de regresión que solo incluyen una variable independiente y que se aplican cuando se espera o se observa que la relación entre X y Y no es modelada por una línea recta. Son aquellos que pueden ser linealizados mediante la transformación ya sea de X , Y o ambas, algunos de estos modelos son exponencial, logarítmico, multiplicativo, inverso de Y , inverso de X , Y cuadrática y X cuadrática.

La descripción aportada por R lleva a cabo un análisis sobre el diseño completamente aleatorizado de un factor.

6.6.1. Función (crd) en R

Esta información fue obtenida de la página web (*crd functio*, *RDocumentation*, s.f). Otra forma de obtenerla es utilizar el comando `help(crd)` en R.

Los autores de la función (crd) son Eric B Ferreira, Nogueira y Piscitelli.

- La función (`crd`) está compuesta por:

```
1 > crd(treat, resp, quali = TRUE, mcomp = "tukey", nl = FALSE,  
2 hvar = "bartlett", sigT = 0.05, sigF = 0.05, unfold = NULL)
```

- Características

treat: Vector numérico o complejo que contiene los tratamientos.

resp: Vector numérico o complejo que contiene la variable de respuesta.

quali: Lógica. Si es TRUE (predeterminado), los tratamientos se asumen cualitativos, si es FALSE, cuantitativo.

mcomp: Permite elegir la prueba de comparación múltiple; el valor predeterminado es la prueba de Tukey, sin embargo, las opciones son: la prueba LSD ('lsd'), la prueba LSD con Bonferroni ('lsdb'), la prueba de Duncan ('duncan'), la prueba de Student-Newman Keuls ('snk'), la prueba de Scott-Knot ('sk'), la prueba de Calinski y Corsten ('ccF') y arranque la prueba de comparación múltiple ('ccboot').

nl: Lógica. Si es FALSE (predeterminado) se ajustan los modelos de regresión lineal. Si es TRUE, se ajustan los modelos de regresión no lineal.

hvar: Permite elegir la prueba de homogeneidad de varianzas; el valor predeterminado es la prueba de Bartlett, sin embargo, existen otras opciones: test de Levene ('levene'), test de Samiuddin ('samiuddin'), prueba de O'Neill y Mathews ('oneillmathews') y la prueba de Layard ('layard').

sigT: La significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5%.

sigF: La significación que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5%.

unfold: Dice lo que se debe hacer después del ANOVA. Si es NULL (predeterminado), recomendado se realizan pruebas; si es '0', solo se realiza ANOVA;

si '1', los efectos simples.

Esta función proporciona el análisis de varianza (ANOVA) para un Diseño Completamente al Azar (CRD, por sus siglas en inglés). Además, incluye la prueba de normalidad de Shapiro-Wilk para los residuos del modelo. En caso de que los tratamientos sean cuantitativos, se ajustan modelos de regresión y se presentan los resultados. Si los tratamientos son cualitativos, se realizan pruebas de comparación múltiple. Para obtener más información sobre el uso de esta función, se puede utilizar el comando `help(crd)`.

6.6.2. Ejemplo en R para un diseño completamente aleatorizado de un factor usando la función `(crd)`

Se utilizan los datos del ejemplo 6.2.7. Se instala y luego se carga la librería `ExpDes`, esta nos permite utilizar la función `crd`. Esta función aporta un análisis completo compuesto por el análisis de la varianza, Prueba de normalidad, Prueba de homogeneidad de varianzas.

La función `(crd)` lleva a cabo un análisis con diseño completamente aleatorio. Los argumentos de la función son: `(quali)` es `TRUE` ya que los tratamientos se asumen cualitativos, `(mcomp)` permite elegir la prueba de comparación múltiple, en este caso se utiliza `tukey`, `(sigT)` es la significación que se utiliza para la prueba de comparación múltiple; el valor predeterminado es 5% y `(sigF)` es la significación que se utilizara para la prueba F de ANOVA; el valor predeterminado es 5%

```
1 > crd(Porcentaje, Resistencia, quali = F, mcomp="lsdb",  
2 sigT = 0.05, sigF = 0.05)
```

■ Tabla de análisis de la varianza

```
1 -----  
2 Analysis of Variance Table  
3 -----  
4           DF      SS      MS      Fc      Pr>Fc  
5 Treatment   4 475.76 118.94 14.757 9.1279e-06  
6 Residuals  20 161.20   8.06  
7 Total      24 636.96  
8 -----  
9 CV = 18.88 %
```

Los resultados previos obtenidos de la tabla ANAVA permiten concluir que el p_{valor} es inferior al compararlo con un $\alpha = 0,05$, por tanto, existen diferencias significativas entre tratamientos.

- Prueba de normalidad de Shapiro-Wilk

```
1 -----
2 Shapiro-Wilk normality test
3 p-value: 0.1817575
4 According to Shapiro-Wilk normality test at 5%
5 of significance,
6 residuals can be considered normal.
7 -----
```

Según la prueba de normalidad de Shapiro-Wilk al 5% de significación, los residuos pueden considerarse normales.

- Prueba de homogeneidad de varianzas

```
1 -----
2 Homogeneity of variances test
3 p-value: 0.9197662
4 According to the test of bartlett at 5% of
5 significance,
6 residuals can be considered homocedastic.
7 -----
```

Según la prueba de homogeneidad de varianzas al 5% de significación, las varianzas pueden considerarse homocedasticas.

- Análisis de los modelos lineal, cuadrático y cubico

```
1 -----
2 Adjustment of polynomial models of regression
3 -----
4 Linear Model
5 =====
6 Estimate Standard.Error tc p.value
7 -----
8 b0 10.9400 2.0862 5.2439 0.00004
9 b1 0.1640 0.0803 2.0424 0.0545
10 -----
11
12 R2 of linear model
13 -----
14 0.070666
```

```

15 -----
16
17 Analysis of Variance of linear model
18 =====
19              DF      SS      MS      Fc      p.value
20 -----
21 Linear Effect 1  33.6200  33.6200  4.17  0.05452
22 Lack of fit   3  442.1400 147.3800 18.29  1e-05
23 Residuals    20 161.2000  8.0600
24 -----
25
26
27 Quadratic Model
28 =====
29      Estimate Standard.Error   tc   p.value
30 -----
31 b0 -39.9886      8.0786      -4.9500 0.0001
32 b1  4.5926      0.6834      6.7203  0
33 b2 -0.0886      0.0136      -6.5255  0
34 -----
35
36 R2 of quadratic model
37 -----
38 0.792068
39 -----
40
41 Analysis of Variance of quadratic model
42 =====
43              DF      SS      MS      Fc      p.value
44 -----
45 Linear Effect 1  33.6200  33.6200  4.17  0.05452
46 Quadratic Effect 1  343.2143 343.2143 42.58  0
47 Lack of fit   2  98.9257  49.4629  6.14  0.00835
48 Residuals    20 161.2000  8.0600
49 -----
50
51
52 Cubic Model
53 =====
54      Estimate Standard.Error   tc   p.value
55 -----
56 b0 62.6114      37.0268      1.6910  0.1064
57 b1 -9.0114      4.8397      -1.8620 0.0774
58 b2  0.4814      0.2012      2.3927  0.0266
59 b3 -0.0076      0.0027      -2.8394 0.0101

```

```

60 -----
61
62 R2 of cubic model
63 -----
64 0.928649
65 -----
66
67 Analysis of Variance of cubic model
68 =====
69              DF      SS      MS      Fc      p.value
70 -----
71 Linear Effect      1    33.6200   33.6200   4.17   0.05452
72 Quadratic Effect  1   343.2143  343.2143  42.58     0
73 Cubic Effect       1    64.9800   64.9800   8.06   0.01013
74 Lack of fit        1    33.9457   33.9457   4.21   0.05347
75 Residuals          20   161.2000    8.0600
76 -----

```

Teniendo en cuenta el coeficiente de correlación (R^2), el modelo que mejor se ajusta es el cúbico el más cercano a 1. Entonces se tiene:

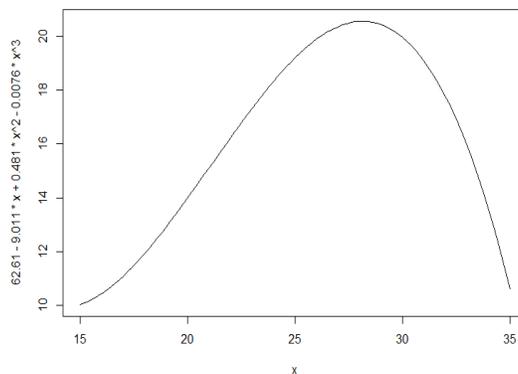
$$\hat{y} = 62,61 - 9,011x + 0,481x^2 - 0,0076x^3$$

■ Gráfico modelo estimado.

```

1 > curve(62.61-9.011*x+0.481*x^2-0.0076*x^3,15,35)

```



```

1 > f = function(x, b0, b1, b2, b3){b0 + b1*x + b2*x2 + b3*
    x3}

```

- Se escogen los parámetros del modelo que resulta significativo

- Para el mínimo

```
1 optimize(f, c(15,35), b0 = 62.61, b1 = 9 .01, b2 =  
    0.481,  
2 b3 = 0 .0076, maximum = FALSE)  
3 minimum  
4 [1] 15.00007  
5 objective  
6 [1] 10.11102
```

- Para el máximo

```
1 optimize(f , c(15, 35), b0 = 62.61, b1 = 9 .01, b2 =  
    0.481,  
2 b3 = 0 .0076, maximum = TRUE)  
3 maximum  
4 [1] 28.22621  
5 objective  
6 [1] 20.89361
```

La resistencia a la atención de un material en función del porcentaje de algodón, el valor máximo es de 28,22% y el mínimo es 15%.

6.7. Transformaciones

Cuando se lleva a cabo un análisis de varianza de los residuales estimados y estos no se ajustan a una distribución normal, también puede ocurrir que los residuales no tengan varianzas homogéneas. Si se presenta esta falta de normalidad la relación entre la media y la varianza de las poblaciones estudiadas, ocasionan heterogeneidad entre las varianzas muestrales.

El uso de las transformaciones consiste en cambiar la escala de las observaciones con el objetivo de que cumplan mejor los supuestos del modelo lineal para que ofrezcan inferencias validas del análisis de varianza. (Montgomery, 2012)

Las observaciones necesitan un cambio de expresión debido a:

1. Son muy asimétricas.

2. Tiene muchos puntos atípicos, sobre todo en las colas.
3. Después de ajustar un modelo a los datos se obtiene residuales muy grandes.

La solución para estas situaciones propuestas consiste en aplicar funciones matemáticas conocidas.

Según Peña (1987) las razones más fuertes para transformar son:

1. Promover simetría en los datos.
2. Promover variación estable en diferentes muestras.
3. Promover una relación lineal entre dos variables.
4. Facilitar la interpretación de una manera natural.

Una relación muy importante entre los datos originales y los transformados es la siguiente, si Y_i son los datos originales y $z_i = T(Y_i)$ los transformados, se cumple la siguiente propiedad: (Martínez y Robles, 2013, p.118)

$$\bar{z} = T(\bar{Y}) \quad \text{y} \quad S_z^2 = [T(\bar{Y})]^2 \times S_y^2.$$

Lo que quiere indicar que, si se tiene varias poblaciones, entonces la transformación afecta de igual forma la media y la varianza de cada una de ellas. (Martínez y Robles, 2013, p.118)

6.7.1. Transformaciones más frecuentes para homogenizar varianzas

1. Raíz cuadrada (\sqrt{Y}): se utiliza en datos de conteos, por ejemplo, número de insectos por parcela y número de malezas por unidad experimental, entre otros. Cuando los datos tiene ceros es recomendable utilizar la transformación $\sqrt{Y + 0,5}$, esta se utiliza también cuando los datos presentan porcentajes en intervalos de 0 a 20%. (Martínez y Robles, 2013, p.118).

2. *Logarítmica* ($\ln(Y)$) : es útil para transformar distribuciones con sesgo positivo, se aplica para los datos presentan diferentes órdenes de magnitud, también se recomienda cuando las varianzas y las medias son proporcionales, en caso de que las observación tengan ceros se debe utilizar $\ln(Y + 1)$. (Martínez y Robles, 2013, p.119).
3. *Angular o Arcoseno* ($\text{sen}^{-1}(Y)$) : esta transformaciones se utiliza cuando los datos de proporciones o porcentajes, donde expanden los porcentajes pequeños y comprimiendo los altos. (Martínez y Robles, 2013, p.119).
4. *Potencia* (Y^p) : esta transformación se recomienda cuando al hacer un gráfico entre las varianzas de las muestras contra las medias de las mismas, o una función de éstas, se revela una relación sistemática o dependencia entre las varianzas y la medias, es decir $\sigma \propto \mu^b$. (Martínez y Robles, 2013, p.119).

El tipo de transformación a llevar a cabo de acuerdo a la relación que existe entre σ y μ .

Tabla 5: Tipos de transformaciones según la relación entre σ y μ

Relación funcional	P	Transformación	Nombre
$\sigma \propto \mu^{-1}$	2	y^2	Cuadrática
$\sigma \propto \mu^0$	1	y	Idéntica
$\sigma \propto \mu^{1/2}$	1/2	\sqrt{y}	Raíz Cuadrada
$\sigma \propto \mu^1$	0	$\ln(y)$	Logarítmica
$\sigma \propto \mu^{3/2}$	-1/2	$1/\sqrt{y}$	Raíz Inversa
$\sigma \propto \mu^2$	-1	$1/y$	Inversa

Es claro que para llegar a un valor de la potencia puede ser necesario realizar varias transformaciones sobre las medias muestrales a fin de conseguir una curva de ajuste entre σ y μ .

6.7.2. Ejemplo en R de transformaciones

En el laboratorio del programa de ingeniería de alimentos, se llevó a cabo un ensayo en el que se investigaron las variables de cohesividad y tratamiento. El

tratamiento utilizado en el estudio se compuso de 6 niveles diferentes. Estas variables fueron seleccionadas como aspectos clave para examinar y comprender las propiedades de los alimentos en estudio. El objetivo del ensayo era analizar la relación entre la cohesividad de los alimentos y los distintos niveles de tratamiento aplicados.

1. Se visualizan los datos

```

1 > Base_2<-read.csv2("C:/Users/HpE2/Downloads/Datos120521.
  csv"
2 , header=T)
3 > TTO=as.factor(TTO)
4 > head(Base_2)
5   TTO APAR COLOR TEXTGRAS DUREZA FIRMEZA COHESIVI HUMEDAD
6 1    1   5.5   5.7   6.9   4.3   5.1   1.5   2.4
7 2    1   5.0   1.9   7.4   5.7   5.1   1.8   4.2
8 3    1   4.3   1.8   6.5   4.2   3.5   0.5   4.1
9 4    1   2.5   0.7   2.7   0.4   1.2   0.7   2.1
10 5    1   2.9   0.5   4.3   5.4   4.1   4.7   0.5
11 6    1   1.4   0.4   2.1   0.8   0.4   1.1   0.8

```

2. Se realiza un análisis de varianza para la variable cohesividad (COHESIVI), los resultados son los siguientes.

```

1 > mod=av(COHESIVI~TTO); mod
2 Call:
3   av(formula = COHESIVI ~ TTO)
4
5 Terms:
6
7           TTO Residuals
8 Sum of Squares      8.58667 155.46833
9 Deg. of Freedom           5         66
10
11 Residual standard error: 1.53479
12 Estimated effects may be unbalanced

```

3. Prueba de normalidad de los residuales del modelo.

```

1 > e=resid(mod)
2 > shapiro.test(e)
3
4   Shapiro-Wilk normality test
5
6 data:  e
7 W = 0.91174, p-value = 9.344e-05

```

Como el p_{valor} es inferior al compararlo con un $\alpha = 0,05$, entonces los residuales no se distribuyen de forma normal.

4. Prueba de Bartlett.

```
1 > bartlett.test(e~TTOS)
2
3   Bartlett test of homogeneity of variances
4
5 data:  COHESIVI and TTOS
6 Bartlett's K-squared = 1.8288, df = 5, p-value = 0.8723
```

Con un nivel de significancia del 0.05, como el $p_{valor}=0.8723$ es mayor que el nivel de significancia. Esto indica que no hay suficiente evidencia para afirmar que existen diferencias significativas entre las varianzas de las muestras. Por lo tanto, se asume la homogeneidad de varianzas entre las muestras evaluadas en el estudio.

5. Prueba de Levene.

```
1 > leveneTest(e~TTOS) ## Prueba de homogeneidad de levene
2 Levene's Test for Homogeneity of Variance (center =
3   median)
4   Df F value Pr(>F)
5 group 5  0.5028 0.7731
6      66
```

En este caso, como el p_{valor} es mayor que el nivel de significancia típicamente utilizado (por ejemplo, $\alpha = 0,05$), no se encontró evidencia suficiente para rechazar la hipótesis nula. Por lo tanto, se puede concluir que no hay diferencias significativas en las varianzas entre los grupos según el factor TTOS.

```
1 > skewness(e, na.rm=FALSE)
2 [1] 1.006607
```

La asimetría de los residuales del modelo es 1.006607. Indica que sugiere una asimetría positiva en la distribución.

6. Se crea un modelo con las variables de interés y se aplica la **transformación de raíz cuadrada**.

```
1 > mod1=aov(sqrt(COHESIVI)~TTO)
2 > summary(mod1)
3      Df Sum Sq Mean Sq F value Pr(>F)
```

```

4 TT0          5  1.234  0.2468  0.954  0.452
5 Residuals   66 17.074  0.2587

```

6.1. Prueba de Shapiro-Wilk.

```

1 > e1=resid(mod1)
2 > shapiro.test(e1)
3
4 Shapiro-Wilk normality test
5
6 data:  e1
7 W = 0.97053, p-value = 0.08906

```

6.2. Prueba de Bartlett.

```

1 > bartlett.test(e1~TTOS)
2
3 Bartlett test of homogeneity of variances
4
5 data:  e1 by TTOS
6 Bartlett's K-squared = 3.5192, df = 5, p-value = 0.6205

```

6.3. Prueba de Levene.

```

1 > leveneTest(e1~TTOS)
2 Levene's Test for Homogeneity of Variance (center =
  median)
3      Df F value Pr(>F)
4 group 5  1.3579 0.2516
5      66

```

Al llevar a cabo la transformación de raíz cuadrada se observa que se corrige la normalidad teniendo en cuenta los resultados de la prueba de Shapiro-Wilk y las pruebas de Bartlett y Levene arrojan que las varianzas son homogéneas.

7. Ahora se utiliza el mismo modelo pero en este caso se le aplica la **transformación logarítmica**.

```

1 > mod2=aov(log(COHESIVI)~TT0)
2 > summary(mod1)
3      Df Sum Sq Mean Sq F value Pr(>F)
4 TTOS    5   3.58  0.7155  1.182  0.327
5 Residuals 66  39.95  0.6053

```

7.1. Prueba de Shapiro-Wilk.

```

1 > e2=resid(mod2)
2 > shapiro.test(e2)
3
4 Shapiro-Wilk normality test
5
6 data: e2
7 W = 0.98273, p-value = 0.4272

```

7.2. Prueba de Bartlett.

```

1 > bartlett.test(e2~TTOS)
2
3 Bartlett test of homogeneity of variances
4
5 data: e2 by TTOS
6 Bartlett's K-squared = 9.0942, df = 5, p-value = 0.1054

```

7.3. Prueba de Levene.

```

1 > leveneTest(e2~TTOS)
2 Levene's Test for Homogeneity of Variance (center =
3 median)
4 Df F value Pr(>F)
5 group 5 2.2987 0.05484 .
6 66

```

Al utilizar la transformación logarítmica, obtenemos por medio del test de Shapiro-Wilk que los residuales si se distribuyen de manera normal, las pruebas de Bartlett y Levene arrojan que las varianzas son homogéneas.

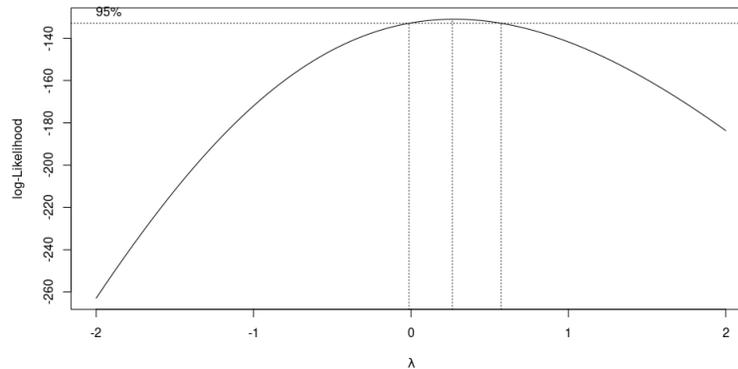
8. Transformación de Box-cox.

Se lleva a cabo un modelo ajustado con las variables cohesividad (COHESIVI) y tratamiento (TTOS).

```

1 > b=boxcox(COHESIVI~TTOS)

```



La línea vertical punteada central representa el parámetro estimado mientras que las otras dos representan su intervalo de confianza al 95 %.

Se encuentra el λ óptima para la transformación de Box-Cox. Luego utilizando la función `cbind` para crear un vector llamado `bc` que contenga los valores de lambda y verosimilitud, obtenemos un objeto de clase matriz, por último, el vector llamado `bc` se ordena de mayor a menor y así se obtiene el vector `orbc`.

Como son muchos datos se visualizan los seis primeros.

```

1 > lambda=b$x
2 > verosimili=b$y
3 > bc=cbind(lambda,verosimili)
4 > orbc=bc[order(-verosimili),]
5 > orbc
6           lambda verosimili
7 [1,]  0.26262626  -131.0004
8 [2,]  0.30303030  -131.0137
9 [3,]  0.22222222  -131.0607
10 [4,]  0.34343434  -131.0998
11 [5,]  0.18181818  -131.1957
12 [6,]  0.38383838  -131.2574

```

El λ que maximiza la función de verosimilitud es 0.2626.

```

1 > COHESIVI1=BoxCox( COHESIVI ,lambda=orbc [1,1])
2 > mod3=aoV(COHESIVI1 ~ TTOS)
3 > summary(mod3)
4           Df Sum Sq Mean Sq F value Pr(>F)

```

```

5 TTOS          5    4.11  0.8214    1.065  0.388
6 Residuals    66   50.91  0.7714

```

8.1. Prueba de Shapiro-Wilk.

```

1 > e3=resid(mod3)
2 > shapiro.test(e3)
3
4 Shapiro-Wilk normality test
5
6 data:  e3
7 W = 0.98384, p-value = 0.4849

```

Teniendo en cuenta la transformación Box Cox, obtenemos por medio del test de Shapir-Wilk que los residuales si se distribuyen de manera normal.

8.2. Prueba de Bartlett.

```

1 > bartlett.test(e3~TTOS)
2
3 Bartlett test of homogeneity of variances
4
5 data:  e3 by TTOS
6 Bartlett's K-squared = 5.6055, df = 5, p-value = 0.3465

```

8.3. Prueba de Levene.

```

1 > leveneTest(e3~TTOS)
2 Levene's Test for Homogeneity of Variance (center = median)
3   Df F value Pr(>F)
4 group 5  2.2987 0.05484.
5      66

```

La transformación realizada corrige la normalidad y las varianzas son homogéneas.

6.8. Diseño en bloques completamente aleatorizados (DBCA).

En este diseño experimental, se utiliza el término "bloque" para referirse a un grupo de unidades experimentales que comparten características similares que podrían afectar la respuesta al aplicar los tratamientos. Esta agrupación

se realiza con el propósito de controlar factores confusos y mejorar la precisión de los resultados obtenidos en el estudio. (Melo, López y Melo, 2007).

El diseño de bloques completamente al azar es el más sencillo de este tipo de diseños. Se utiliza para controlar y reducir el error experimental, ya que las unidades experimentales se estratifican en bloques de unidades homogéneas. Para llevar a cabo este procedimiento, se asigna al azar un número igual de unidades experimentales en cada bloque. Se considera que este diseño permite realizar comparaciones más precisas entre los tratamientos dentro del conjunto homogéneo de unidades experimentales en cada bloque. (Kuehl, 2001)

El diseño busca determinar diferencias entre los tratamientos, en donde cada uno de estos se aplica a distintas unidades experimentales en cada uno de los bloques. (Melo, López y Melo, 2007)

6.8.1. Aleatorización en el diseño en bloques al azar

De acuerdo a Martínez y Robles los tratamientos se deben aleatorizar dentro de cada uno de los bloques, en forma independiente. Por ejemplo, si el ensayo en campo consta de cinco tratamientos (t_1, t_2, t_3, t_4, t_5) y tres bloques, una de las posibles distribuciones de los tratamientos a las parcelas pueden ser:

Bloque I					Bloque II				
t_2	t_1	t_5	t_4	t_3	t_1	t_2	t_3	t_4	t_5
Bloque III									
t_3	t_2	t_1	t_4	t_5					

6.8.2. Modelo estadístico

El modelo estadístico para un experimento en un DBCA viene dado por la expresión:

$$Y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij} \begin{cases} i = 1, 2, \dots, r \\ j = 1, 2, \dots, t \end{cases} \quad (12)$$

donde $i = 1, 2, \dots, r$ y $j = 1, 2, \dots, t$

- Y_{ij} : es la observación de la variable de estudio Y en el i – *ésimo* bloque para el j – *ésimo* tratamiento.
- μ : es la media general común a todos los tratamientos.
- β_i : es el efecto del bloque i – *ésimo*.
- τ_j : es el efecto del tratamiento j – *ésimo*.

ϵ_{ij} es el error experimental en la unidad experimental i – *ésima* para el tratamiento j – *ésimo*.

6.8.3. Pruebas de hipótesis y análisis de varianza

El objetivo se concentra en los tratamientos y sus efectos sobre la respuesta.

Las hipótesis que se deben tener en cuenta son:

1. Para los tratamientos:

$$H_0 : \mu_{.1} = \mu_{.2} = \dots = \mu_{.t} \text{ contra } H_1 : \mu_{.j} \neq \mu_{.j'} \text{ para algún } j \neq j'.$$

2. Para los Bloques:

$$H_0 : \mu_{1.} = \mu_{2.} = \dots = \mu_{r.} \text{ contra } H_1 : \mu_{i.} \neq \mu_{i'.} \text{ para algún } i \neq i'.$$

Dado que $\mu_j = \mu + \tau_j$, entonces la primera hipótesis también se puede escribir como

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_t \text{ contra } H_1 : \tau_j \neq \tau_{j'} \text{ para algún } j \neq j'.$$

La siguiente Tabla 6, representa el análisis de varianza, la cual sirve de apoyo para corroborar las hipótesis 1 y 2 enumeradas anteriormente. donde,

$$SC_{Total} = \sum_{i=1}^r \sum_{j=1}^t Y_{ij}^2 - \frac{Y_{..}^2}{n}; \quad SC_{Bloques} = \sum_{i=1}^r Y_{i.}^2 - \frac{Y_{..}^2}{n}$$

$$SC_{TTO} = \sum_{j=1}^t Y_{.j}^2 - \frac{Y_{..}^2}{n} \quad \text{y} \quad SC_{EE} = SC_{Total} - SC_{Bloques} - SC_{TTO}.$$

Tabla 6: ANAVA para un DBCA

F.V.	gl	SC	CM	F_c	F_{Tabla}
Bloques	$r - 1$	$SC_{Bloques}$	$\frac{SC_{Bloques}}{r-1}$	$\frac{CM_{Bloques}}{CM_{EE}}$	$F_{\alpha, (r-1), (r-1)(t-1)}$
TTO	$t - 1$	SC_{TTO}	$\frac{SC_{TTO}}{t-1}$	$\frac{CM_{TTO}}{CM_{EE}}$	$F_{\alpha, t-1, (r-1)(t-1)}$
EE	$(r - 1)(t - 1)$	SC_{EE}	$\frac{SC_{EE}}{(r-1)(t-1)}$		
Total	$n - 1$	SC_{Total}			

además,

$$Y_{..} = \sum_{i=1}^r \sum_{j=1}^t Y_{ij}^2, \quad n = \sum_{j=1}^t r = tr, \quad gl_{Bloques} = r - 1, \quad gl_{TTO} = t - 1 \quad y$$

$$gl_{EE} = (r - 1)(t - 1).$$

Se debe tener en cuenta que $\frac{CM_{TTO}}{CM_{EE}}$, sigue siendo el cuadrado medio del tratamiento y cuadrado medio del error. La implementación de las pruebas de comparación múltiple en este diseño se hace en forma similar al de DCA, simplemente que ahora se utilizarán los gl_{EE} y el CM_{EE} (Martínez y Robles, 2013, p.111).

6.8.4. Ejemplo en R para un diseño en bloque completamente aleatorio usando la función (ea1).

En una empresa lechera se tienen varios silos para almacenar leche (cirternas de 60000L). Un aspecto crítico para que se conserve la leche es la temperatura de almacenamiento. Se sospecha que en algunos silos hay problemas, por ello, durante cinco días se decide registrar la temperatura a cierta hora crítica. Obviamente la temperatura de un día a otro es una fuente de variabilidad que podría impactar la variabilidad total de la variable temperatura de la leche. Por estas razones se tuvo que bloquear por día de medición (Gutiérrez y Dela Vara, 2012).

Tabla 7: Temperatura de la leche

Dia/Silo	A	B	C	D	E
Lunes	4.0	5.0	4.5	2.5	4.0
Martes	4.0	6.0	4.0	4.0	4.0
Miércoles	5.0	2.0	3.5	6.5	3.5
Jueves	0.5	4.0	2.0	4.5	2.0
Viernes	3.0	4.0	3.0	4.0	4.0

Es necesario instalar y cargar la librería (`easyanova`). Ahora los pasos a seguir son:

1. Se carga la base de datos a R y se observan los 6 primeros elementos.

```

1 > Base_DBCA <- read_excel("C:/Users/Hp E2/Desktop
2 /MONOGRAFIA-TODO/Base_DBCA.xlsx")
3
4 > data(Base_DBCA);head(Base_DBCA )
5   Silos      Dias Temp
6 1      A      Lunes  4.0
7 2      A      Martes  4.0
8 3      A Miercoles  5.0
9 4      A      Jueves  0.5
10 5     A      Viernes  3.0
11 6     B      Lunes  5.0

```

2. Utilizando la función (`ea1`) se lleva a cabo un análisis con diseño en bloque completamente aleatorio. Los argumentos que se usan en esta función de acuerdo con este ejemplo son: la base de datos llamada `Base` y (`design=2`) ya que de acuerdo a la descripción previamente dada significa que el análisis que se va a realizar es de un DBCA.

```

1 > r2=ea1(Base,design=2);r2
2
3 $ 'Analysis of variance '
4   df type III SS mean square F value  p>F
5 treatments  4      4.46      1.115  0.6904  0.6092
6 blocks      4      9.76      2.440  1.5108  0.246
7 residuals  16     25.84      1.615      -      -

```

Teniendo en cuenta los resultados previos obtenidos de la tabla ANAVA con un $\alpha = 0,05$, no se rechaza H_0 ; todas las medias son iguales y se

concluye que no hay una diferencia significativa entre los silos, es decir funcionan correctamente.

3. Ajuste de medias

```

1 $'Adjusted means '
2   treatment adjusted.mean standard.error tukey snk duncan
3 1          D             4.3           0.5683    a    a    a
4 2          B             4.2           0.5683    a    a    a
5 3          E             3.5           0.5683    a    a    a
6 4          C             3.4           0.5683    a    a    a
7 5          A             3.3           0.5683    a    a    a
8   t    scott_knott
9 1 a             a
10 2 a             a
11 3 a             a
12 4 a             a
13 5 a             a

```

Las pruebas de Tukey, snk, Duncan, t y Scott Knott muestran que los tratamientos no presentan diferencias entre ellos.

4. Comparaciones múltiples

```

1 $'Multiple comparison test '
2   pair contrast p(tukey) p(snk) p(duncan) p(t)
3 1  D - B       0.1    0.9999 0.9025    0.9025 0.9025
4 2  D - E       0.8    0.8536 0.5902    0.3598 0.3344
5 3  D - C       0.9    0.7940 0.6830    0.3182 0.2793
6 4  D - A       1.0    0.7268 0.7268    0.2770 0.2313
7 5  B - E       0.7    0.9034 0.3966    0.3966 0.3966
8 6  B - C       0.8    0.8536 0.5902    0.3598 0.3344
9 7  B - A       0.9    0.7940 0.6830    0.3182 0.2793
10 8  E - C       0.1    0.9999 0.9025    0.9025 0.9025
11 9  E - A       0.2    0.9991 0.9665    0.8170 0.8066
12 10 C - A       0.1    0.9999 0.9025    0.9025 0.9025

```

5. Análisis de residuales

```

1 $'Residual analysis '
2 $'Residual analysis '$'residual analysis '
3   values
4 p.value Shapiro-Wilk test    0.2759
5 p.value Bartlett test       0.4075
6 coefficient of variation (%) 3.0500
7 first value most discrepant 7.0000

```

```
8 second value most discrepant 3.0000
9 third value most discrepant 11.0000
```

De acuerdo a los datos anteriores se concluye que los residuales pueden considerarse normales, y la prueba de Bartlett muestra que las varianzas son homogéneas.

6.8.5. Función `(rbd)` en R

Esta información fue obtenida de la página web (Ferreira, Cavalcanti, Alves, Maintainer, 2022). Otra forma de obtenerla es utilizar el comando `help(rbd)` en R.

Los autores de la función `(rbd)` Eric B Ferreira, Denismar Alves Nogueira y Portya Piscitelli Cavalcanti, analiza los experimentos en diseños de bloques aleatorios equilibrados bajo un único factor, considerando un modelo fijo. este se encuentra en la librería de R llamada (`ExpDes`).

- La función `(rbd)` está compuesta por:

```
1 > rbd(treat, block, resp, quali = TRUE, mcomp = "tukey", nl =
  FALSE,
2 hvar = "oneillmathews", sigT = 0.05, sigF = 0.05, unfold = NULL)
```

- Características

`treat`: vector numérico o complejo que contiene los tratamientos.

`block`: vector numérico o complejo que contiene los bloques.

`resp`: vector numérico o complejo que contiene la variable de respuesta.

`quali`: lógica. Si es `TRUE` (por defecto), los tratamientos se asumen cualitativos, si es `FALSE`, cuantitativos.

`mcomp`: permite elegir la prueba de comparación múltiple; el valor por defecto es la prueba de Tukey, sin embargo, las opciones son: la prueba de LSD (`'lsd'`), la prueba de LSD con protección de Bonferroni (`'lsdb'`), la prueba de

Duncan ('duncan'), la prueba de Student-NewmanKeuls ('snk'), la prueba de Scott-Knot ('sk'), la prueba de Calinski y Corsten ('ccF') y la prueba de comparaciones múltiples bootstrap ('ccboot').

`nl`: lógica. Si es FALSE (por defecto) se ajustan los modelos de regresión lineal. SI ES TRUE, se ajustan los modelos de regresión no lineal.

`hvar`: permite elegir la prueba de homogeneidad de varianzas; por defecto es la prueba de O'Neill y Mathews ('oneillmathews'), pero hay otras opciones: la prueba de Han ('han'), y la prueba de Anscombe y Tukey ('anscombetukey').

`sigT`: la significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 %.

`sigF`: la significación que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5 %.

`unfold`: dice lo que debe hacerse después del ANOVA. Si es NULL (por defecto), se realizan las pruebas recomendadas; si es '0', sólo se realiza el ANOVA; si es '1', se prueban los efectos simples son probados.

Esta función aporta el ANOVA del `rbd`, la prueba de normalidad de Shapiro-Wilk para los residuos del modelo, los modelos de regresión ajustados (cuando los tratamientos son cuantitativos) y/o las pruebas de (cuando los tratamientos son cuantitativos) y/o las pruebas de comparación múltiple (cuando los tratamientos son cualitativos).

6.8.6. Ejemplo en R para un diseño en bloque completamente aleatorio usando la función (`rbd`).

Para este ejemplo se tienen en cuenta los datos del ejemplo anterior. Es necesario instalar y cargar la librería (`ExpDes`).

Utilizando la función (`rbd`) se lleva a cabo un análisis con diseño en bloque completamente aleatorio. Los argumentos que se usan en esta función de acuerdo con este ejemplo son: (`quali`) es TRUE ya que los tratamientos se asumen cualitativos, (`mcomp`) permite elegir la prueba de comparación múltiple, en este caso se utiliza `tukey`, (`sigT`) es la significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 % y

(sigF) es la significación que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5%. El comando completo es el siguiente:

```
1 > comp=rbd(Silos, Dias, Temp, quali = T, mcomp="tukey",
2 sigT = 0.05, sigF = 0.05);comp
```

Los resultados obtenidos son los siguientes

- Tabla de análisis de la varianza

```
1 -----
2 Analysis of Variance Table
3 -----
4           DF      SS      MS      Fc      Pr>Fc
5 Treatment  4   4.46  1.115  0.6904  0.60921
6 Block      4   9.76  2.440  1.5108  0.24602
7 Residuals 16  25.84  1.615
8 Total      24  40.06
9 -----
10 CV = 33.98 %
```

El análisis de varianza (ANOVA) muestra que no se encontraron diferencias significativas entre los tratamientos y los bloques en el estudio. Tanto el p_{valor} para el tratamiento como para el bloque son mayores que el nivel de significancia establecido (0.05), lo que indica que las diferencias observadas podrían deberse al azar y no a un efecto real de los tratamientos o bloques. Además, el coeficiente de variación (CV) del 33.98% indica que existe una variabilidad relativamente alta en los datos. En resumen, no se encontraron diferencias estadísticamente significativas entre los tratamientos y los bloques en este estudio.

- Prueba de normalidad de Shapiro-Wilk

```
1 -----
2 Shapiro-Wilk normality test
3 p-value:  0.152785
4 According to Shapiro-Wilk normality test at 5%
5 of significance, residuals can be considered normal.
6 -----
```

Según la prueba de normalidad de Shapiro-Wilk al 5% de significación, los residuos pueden considerarse normales.

- Prueba de homogeneidad de varianzas

```

1 -----
2 Homogeneity of variances test
3 p-value: 0.1391133
4 According to the test of oneillmathews at 5%
5 of significance, the variances can be considered
6 homocedastic.
7 -----

```

Según la prueba de homogeneidad de varianzas al 5% de significación, las varianzas pueden considerarse homocedásticas.

■ Prueba F

```

1 -----
2 According to the F test, the means can not be
3 considered distinct.
4
5 Levels Means
6 1      A    3.3
7 2      B    4.2
8 3      C    3.4
9 4      D    4.3
10 5      E    3.5
11 -----

```

Según la prueba F, no hay una diferencia estadísticamente significativa entre las medias de los diferentes niveles. Esto significa que no se encontró evidencia suficiente para concluir que las medias de los niveles A, B, C, D y E son distintas entre sí.

6.8.7. Ejemplo con un factor cuantitativo en R para un diseño en bloque completamente aleatorio usando la función `ea1`.

Lozano *et al* (2004), estudiaron el comportamiento agronómico del cultivo del maíz HS-36 para un programa de nutrición foliar complementario, bajo un diseño en bloques completamente al azar con 4 repeticiones o bloques. Los tratamientos que componen los programas de nutrición son los siguientes:

1: Testigo o control (nutrición edáfica).

2: Nutrición edáfica + aplicaciones de nutrientes foliares a los 12 días (1 *lt/ha*) y a los 18 y 25 días 2 *lt/ha*, en cada aplicación.

3: Nutrición edáfica + aplicaciones de nutrientes foliares a los 18 días (2 *lt/ha*), en cada aplicación.

4: Nutrición edáfica + aplicaciones de nutrientes foliares a los 25 días (2 *lt/ha*), en cada aplicación.

En cada aplicación para la variable longitud de la mazorca los datos recolectados se presentan en la Tabla 8.

Tabla 8: Datos sobre el cultivo del maíz HS-36

Bloque/TTO	1	2	3	4
1	1456,58	1212,81	1388,62	1343,46
2	1236,75	1494,88	1353,49	1442,87
3	1010,53	1484,1	1511,39	1382,97
4	1334,64	1640,54	2199,12	1797,86

1. Se carga la base de datos a R y se observan los 6 primeros elementos.

```
1 > Base_DBCAC <- read_excel("C:/Users/Hp E2/Desktop
2 /MONOGRAFIA-TODO/Base_DBCAC.xlsx")
3
4 > data(Base_DBCAC); head(Base_DBCAC)
5 Bloque  Trat  y
6 1      1  1456.58
7 1      2  1212.81
8 1      3  1388.62
9 1      4  1343.46
10 2     1  1236.75
11 2     2  1494.8
```

2. Utilizando la función (ea1) se lleva a cabo un análisis con diseño en bloque completamente aleatorio.

```
1 > ea1(Base_DBCAC, design=2)
```

```
1 $'Analysis of variance'
2
3      df type III SS mean square F value  p>F
4 treatments  3  443424.6  147808.21  3.69 0.0557
5 blocks      3  258182.2   86060.73  2.1485 0.1641
6 residuals  9  360508.5   40056.50  -      -
```

Para los tratamientos, el valor F es de 3.69 y el p_{valor} es de 0.0557, lo que indica que existe cierta evidencia para rechazar la hipótesis nula de que no hay diferencias significativas entre los tratamientos, aunque el resultado no es estadísticamente significativo a un nivel de significancia del 0.05. Para los bloques, el valor F es de 2.1485 y el p_{valor} es de 0.1641, lo que indica que no hay suficiente evidencia para rechazar la hipótesis nula de que no hay diferencias significativas entre los bloques.

```

1 $ 'Adjusted means '
2   treatment  adjusted.mean  standard.error  tukey  snk  duncan  t
3 1           4           1743.040      100.0706    a    a      a  a
4 2           2           1381.977      100.0706    a    b      b  b
5 3           1           1350.368      100.0706    a   ab      b  b
6 4           3           1347.247      100.0706    a   ab      b  b
7 scott_knott
8 a
9 b
10 b
11 b

```

Las pruebas de Tukey, muestra que los tratamientos no presentan diferencias entre ellos a diferencia de las pruebas snk, Duncan, t y Scott Knott las cuales muestran diferencias entre los tratamientos 2 y 4.

```

1 $ 'Multiple comparison test '
2   pair contrast  p(tukey)  p(snk)  p(duncan)  p(t)
3 1 4 - 2 361.0625  0.1174 0.0311  0.0311 0.0311
4 2 4 - 1 392.6725  0.0842 0.0514  0.0260 0.0216
5 3 4 - 3 395.7925  0.0815 0.0815  0.0279 0.0208
6 4 2 - 1  31.6100  0.9958 0.8282  0.8282 0.8282
7 5 2 - 3  34.7300  0.9944 0.9675  0.8196 0.8116
8 6 1 - 3   3.1200  1.0000 0.9829  0.9829 0.9829

```

```

1 $ 'Residual analysis '
2 $ 'Residual analysis ' $ 'residual analysis '
3                                     values
4 p.value Shapiro-Wilk test          0.2441
5 p.value Bartlett test              0.1492
6 coefficient of variation (%)      13.7500
7 first value most discrepant       1.0000
8 second value most discrepant      15.0000
9 third value most discrepant       13.0000

```

De acuerdo a los datos anteriores se concluye que los residuales pueden considerarse normales, y la prueba de Bartlett muestra que las varianzas son homogéneas.

6.8.8. Ejemplo con un factor cuantitativo en R para un diseño en bloque completamente aleatorio usando la función rbd.

Para ilustrar el procedimiento analicemos el trabajo preliminar llevado a cabo por Cardona (2003), donde se estudió el efecto de la aplicación de dosis de calcio, en tres dosis de aplicación, en la calidad de los frutos de tomate. La variable de interés fue la cantidad de calcio en dicho fruto. Las dosis de calcio estudiadas fueron de 0; 0,05; 0,1 y 0,15 el experimento se llevó a cabo bajo un diseño en bloque completamente al azar (DBCA) con 4 repeticiones. Los resultados obtenidos por unidad experimental son presentados en la Tabla 9.

Tabla 9: Datos para longitud de mazorca de maíz H5-36

Concentración	Repetición			
	1	2	3	4
0	0,172	0,204	0,244	0,148
0,05	0,212	0,136	0,268	0,204
0,1	0,252	0,272	0,212	0,228
0,15	0,424	0,468	0,524	0,452

1. Se carga la base de datos a R y se observan los 6 primeros elementos.

```
1 > Base_DBCAC_2 <- read_excel("C:/Users/Hp E2/Desktop
2 /MONOGRAFIA-TODO/Base_DBCAC_2.xlsx")
3
4 > data(Base_DBCAC); head(Base_DBCAC)
5 Bloque  Trat  y
6 0       1  0,172
7 0       2  0,172
8 0       3  0,244
9 0       4  0,148
10 0,05   1  0,212
11 0,05   2  0,136
```

2. Utilizando la función (rbd) se lleva a cabo un análisis con diseño en bloque completamente aleatorio.

```
1 > rbd(Bloque, Trat, y, quali = TRUE, mcomp = "tukey", nl = FALSE,
2 hvar = "oneillmathews", sigT = 0.05, sigF = 0.05, unfold = NULL)
```

Los resultados obtenidos son los siguientes

- Tabla de análisis de la varianza

```

1 -----
2 Analysis of Variance Table
3 -----
4           DF          SS          MS          Fc          Pr>Fc
5 Treatment   3 0.204795  0.068265  44.908  0.00001
6 Block       3 0.007699  0.002566   1.688  0.23835
7 Residuals   9 0.013681  0.001520
8 Total       15 0.226175
9 -----
10 CV = 14.22 %

```

Hay diferencias significativas entre los tratamientos, ya que el p_{valor} es muy pequeño (0.00001), lo que indica que la probabilidad de obtener diferencias tan grandes entre los tratamientos debido al azar es muy baja. Sin embargo, no hay diferencias significativas entre los bloques, ya que el p_{valor} (0.23835) es mayor que el nivel de significancia típico de 0.05. El coeficiente de variación (CV) del 14,22 % indica la variabilidad relativa de los datos en relación con la media.

- Prueba de normalidad de Shapiro-Wilk

```

1 -----
2 Shapiro-Wilk normality test
3 p-value:  0.05396234
4 According to Shapiro-Wilk normality test at 5% of
5 significance, residuals can be considered normal.
6 -----

```

Según la prueba de normalidad de Shapiro-Wilk al 5 % de significación, los residuos pueden considerarse normales.

- Prueba de homogeneidad de varianzas

```

1 -----
2 Homogeneity of variances test
3 p-value:  0.4965848
4 According to the test of oneillmathews at 5% of
5 significance, the variances can be considered
6 homocedastic.
7 -----

```

Según la prueba de homogeneidad de varianzas al 5 % de significación, las varianzas pueden considerarse homocedásticas.

■ Prueba F

```
1 Tukey's test
2 -----
3 Groups Treatments Means
4 a          0.15      0.467
5 b           0.1      0.241
6 b          0.05      0.205
7 b           0        0.184
8 -----
```

Esto indica que el tratamiento con un valor de 0.15 tiene una media significativamente diferente de los tratamientos del grupo *b*. Sin embargo, no se encontraron diferencias significativas entre los tratamientos dentro del grupo *b*.

6.9. Muestreo en unidades experimentales

Cuando se reúnen los datos del estudio, es necesario obtener muestras aleatorias de subunidades de las unidades experimentales, la unidad observa es una submuestra. Cuando se toma la información de unidades experimentales se realiza una selección aleatoria de algunas de las submuestras que componen la unidad experimental (Martínez y Robles, 2013, p.130).

Martínez y Robles proponen que para medir una variable en una unidad experimental se puede usar:

- Toda la unidad experimental, en cuyo caso se tiene un solo error en el modelo.
- Muestrear dentro de la unidad experimental, lo cual ocasiona otra fuente de error en el modelo. La diferencia entre el valor muestral y el de la unidad experimental completa constituye el error de muestreo, por lo cual es importante utilizar una buena técnica de selección de las muestras.

6.9.1. Muestreo de unidades experimentales en un DCA

Si se tienen s submuestras en cada una de las r unidades experimentales para los t tratamientos, el modelo estadístico para un diseño completamente

al azar (DCA) es: “ Martínez y Robles, 2013, p.131”

$$Y_{ijk} = \mu + \tau_j + \epsilon_{ij} + \theta_{ijk}, \quad (13)$$

para, $i = 1, 2, \dots, r$, $j = 1, 2, \dots, t$; y $k = 1, 2, \dots, s$
donde

- r : es el número de repeticiones del tratamiento j -ésimo.
- s : es el número de submuestras en la unidad experimental i - *ésima* bajo el tratamiento j - *ésimo*.
- Y_{ijk} : es la observación de la submuestra k - *ésima* en la i - *ésima* unidad experimental para el tratamiento j - *ésimo*.
- μ : es la media general común a todos los tratamientos.
- τ_j : es el efecto del tratamiento j -*ésimo* (para efectos fijos o aleatorios).
- ϵ_{ij} : es el error experimental de la i - *ésima* unidad experimental para el j - *ésimo* tratamiento.
- θ_{ijk} : es el error de la submuestra k - *ésima* en la i - *ésima* unidad experimental para el tratamiento j - *ésimo*.
- Los supuestos para los dos terminos de error son: $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ y $\theta_{ijk} \stackrel{iid}{\sim} N(0, \sigma_\theta^2)$.

6.9.2. Descomposición de la suma de cuadrados total DCA

El modelo estadístico, la desviación total tiene la siguiente descomposición:

$$Y_{ijk} - \bar{Y}_{...} = (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{.j.}) + (Y_{ijk} - \bar{Y}_{ij.})$$

cada componente de esta descomposición corresponde a:

- $\bar{Y}_{.j.} - \bar{Y}_{...}$ es la desviación del tratamiento j -ésimo.
- $\bar{Y}_{ij.} - \bar{Y}_{.j.}$ es el error experimental de la i -ésimo unidad experimental para el j -ésimo tratamiento.

- $Y_{ijk} - \bar{Y}_{ij.}$ es el error de muestreo de la k -ésimo submuestra en la i -ésima unidad experimental del j -ésimo tratamiento, donde $Y_{ij.} = \sum_{k=1}^s Y_{ijk}$ es el total de las s submuestras en la i -ésima unidad experimental bajo el tratamiento j -ésimo, $Y_{.j.} = \sum_{i=1}^r \sum_{k=1}^s Y_{ijk}$ es el total del tratamiento j -ésimo y $Y_{...} = \sum_{j=1}^t Y_{.j.} = \sum_{i=1}^r \sum_{j=1}^t \sum_{k=1}^s Y_{ijk}$ es el gran total.

El promedio por unidad experimental, por tratamiento y general, respectivamente, vienen dados por las siguientes expresiones:

$$\bar{Y}_{ij.} = \frac{Y_{ij.}}{s}; \bar{Y}_{.j.} = \frac{Y_{.j.}}{rs} \text{ y } \bar{Y}_{...} = \frac{Y_{...}}{rst}$$

En este diseño la descomposición de la suma de cuadrados total se puede escribir como:

$$\begin{aligned} SC_{Total} &= SC_{Tratamiento} + SC_{Error\ Experimental} + SC_{Error\ de\ Muestreo} \\ &= SC_{TTO} + SC_{EE} + SC_{EM}, \end{aligned}$$

donde,

$$SC_{Total} = \sum_{i=1}^r \sum_{j=1}^t \sum_{k=1}^s Y_{ijk}^2 - FC, \text{ siendo } FC = \frac{Y_{...}^2}{rst},$$

$$SC_{TTO} = \frac{1}{rs} \sum_{j=1}^t Y_{.j.}^2 - FC$$

$$SC_{EE} = \frac{1}{s} \sum_{i=1}^r \sum_{j=1}^t Y_{ij.}^2 - \frac{1}{rs} \sum_{j=1}^t Y_{.j.}^2 = \frac{1}{s} \sum_{i=1}^r \sum_{j=1}^t Y_{ij.}^2 - FC - SC_{TTO}$$

$$SC_{EM} = SC_{Total} - SC_{TTO} - SC_{EE}.$$

Así mismo, los grados de libertad para el error experimental y de muestreo son:

$$gl_{EE} = \sum_{j=1}^t (r_j - 1) = \sum_{j=1}^t (r - 1) = t(r - 1) \text{ y } gl_{EM} = \sum_{i=1}^r \sum_{j=1}^t (s - 1) = rt(s - 1).$$

La Tabla 10 de análisis de varianza para probar la hipótesis usual en esta estructura de diseño, es la siguiente

Tabla 10: ANAVA Para un DCA con submuestreo

F.V.	gl	SC	CM	$E[CM]$	F_c
TTO	$t - 1$	SC_{TTO}	CM_{TTO}	$\sigma_\theta^2 + \sigma_\varepsilon^2 + \frac{rs}{t-1} \sum_{j=1}^t \tau_j^2$	$\frac{CM_{TTO}}{CM_{EE}}$
EE	$t(r - 1)$	SC_{EE}	CM_{EE}	$\sigma_\theta^2 + \sigma_\varepsilon^2$	
EM	$tr(s - 1)$	SC_{EM}	CM_{EM}	σ_θ^2	
Total	$rst - 1$				

6.9.3. Función (Sub.DCA)

Como R no aporta una función que vaya dirigida a realizar el análisis de submuestreo con un DCA, se programó una función llamada (Sub.DCA).

- La función (Sub.DCA) está compuesto por:

```
1 > Sub.DCA(treat, SUBM, VaInts, quali = TRUE, sigF = 0.05, sigT = 0.05)
```

- Características

treat: son los tratamientos.

SUBM: las submuestras.

VaInts: es la variable de interés.

quali: lógica. Si es TRUE (por defecto), los tratamientos se asumen cualitativos, si es FALSE, cuantitativos.

sigT: la significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 %.

sigF: la significación que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5 %.

Esta función aporta el análisis de varianza, prueba de normalidad Shapiro-Wilk, prueba de homogeneidad de varianza de Bartlett y las comparaciones de medias. Ver en anexo.

6.9.4. Ejemplo en R de muestreo de unidades experimentales en un DCA

Un investigador de suelos estudió el crecimiento de plantas de cebada en tres niveles diferentes de salinidad en un medio controlado. Tenía dos contenedores replica de cada tratamiento, en un DCA y midió tres plantas de cada replica. Las masas secas de las plantas, en gramos, son las presentadas en la Tabla 11 (Martínez y Robles, 2013, p.132).

Tabla 11: Masa seca de plantas de cebada

Salinidad	Contenedor	<i>Peso (gr)</i>
Control	1	11, 29; 11, 08; 11, 10
	2	7, 37; 6, 55; 8, 50
6 barras	1	5, 64; 5, 98; 5, 69
	2	4, 20; 3, 34; 4, 21
12 barras	1	4, 83; 4, 77; 5, 66
	2	3, 28; 2, 61; 2, 69

Pruebe la hipótesis de que no hay diferencias entre las medias de los niveles salinos. Los pasos a seguir son los siguientes.

Para formar la base de datos se debe formar una secuencia que inicie de 1 hasta $r \times s$ donde r es el total de réplicas y s son las submuestras donde cada elemento de la secuencia se repite s veces, en este ejemplo se tiene $r \times s = 6$ entonces se debe armar la secuencia de 1, 2, 3, 4, 5 y 6; donde cada elemento de esta secuencia se repite $s=3$ veces. Tal y como se observa en la columna SUBM de la tabla siguiente.

1. La función `read.csv2`, permite cargar la base de datos, se hace un paréntesis en este paso ya que se requiere ordenar de forma secuencial como se observa a continuación.

```
1 > datos=read.csv2(file.choose())
2 > datos
3   Trat SUBM Contenedor  Peso
4 1     0     1           1 11.29
5 2     0     1           1 11.08
6 3     0     1           1 11.10
7 4     0     2           2  7.37
```

```

8 5      0      2      2 6.55
9 6      0      2      2 8.50
10 7     6      3      1 5.64
11 8     6      3      1 5.98
12 9     6      3      1 5.69
13 10    6      4      2 4.20
14 11    6      4      2 3.34
15 12    6      4      2 4.21
16 13   12     5      1 4.83
17 14   12     5      1 4.77
18 15   12     5      1 5.66
19 16   12     6      2 3.28
20 17   12     6      2 2.61
21 18   12     6      2 2.69

```

- Se emplea la función (Sub.DCA), empleando las variables de estudio. Solo se trabaja con las variables Trat, SUBM y Peso

```

1 > Sub.DCA(Trat,SUBM,Peso,quali = TRUE,sigF = 0.05,sigT =
0.05)

```

Esta fue programada para aportar los siguientes resultados:

- Análisis de varianza

```

1 Analysis of Variance Table
2 -----
3
4          DF          SS          MS          Fc          Pr>Fc
5 Treatment      2  98.572  49.286  4.4888  0.12535
6 Residuals      3  32.940  10.980
7 Sampling error 12   3.273   0.273
8 Total          17 134.785  60.539
9 -----
10 CV = 54.83 %

```

Los resultados previos obtenidos de la tabla ANAVA permiten concluir que el $p_{valor} = 0,12535$ es superior al compararlo con un $\alpha = 0,05$, por tanto, no existen diferencias estadísticas entre las medias de masa seca para los niveles de salinidad.

- Supuestos del modelo

```

1 Shapiro-Wilk normality test
2 p-value: 0.7798462
3 According to Shapiro-Wilk normality test at 5% of
4 significance, residuals can be considered normal.
5 -----
6 Bartlett homogeneity test
7 p-value: 0.3673749
8 According to Bartlett homogeneity test at 5% of
9 significance, it is considered that the residuals
10 have constant variance.
11 -----

```

■ Comparaciones de medias

```

1 According to the F test, the means can not be considered
2 distinct.
3 -----
4
5      Levels      Means
6 1         0 9.315000
7 2         12 3.973333
8 3          6 4.843333

```

Al realizar la prueba F, se ha determinado que las medias no son estadísticamente diferentes entre sí. Esto significa que no hay evidencia suficiente para concluir que las medias de los tres niveles son distintas. En otras palabras, no hay diferencias significativas entre los grupos representados por los niveles 0, 12 y 6 en relación a la variable que se está analizando.

6.9.5. Muestreo de unidades experimentales en un DBCA

Para el diseño sea en bloques completos al azar, con una estructura de tratamiento simple, el modelo que se ajusta a la situación para una variable respuesta Y es: (Martínez y Robles, 2013, p.133).

$$Y_{ijk} = \mu + \beta_i + \tau_j + \epsilon_{ij} + \theta_{ijk}; \quad (14)$$

con: $i = 1, 2, \dots, r$; $j = 1, 2, \dots, t$; y $k = 1, 2, \dots, s$, donde

- r : es el número de repeticiones del tratamiento j -ésimo.

- s : es el número de submuestras en el i –ésimo bloque en el tratamiento j .
- Y_{ijk} : es la observación de la submuestra k –ésima en el i –ésimo bloque para el tratamiento j –ésimo.
- μ : es la media general común a todos los tratamientos.
- β_i : es el efecto del bloque i –ésimo.
- τ_j : es el efecto del tratamiento j –ésimo (para efectos fijos o aleatorios).
- ϵ_{ij} : es el error experimental en el i –ésimo bloque para el j –ésimo tratamiento.
- θ_{ijk} : es el error de la submuestra k –ésima en el i –ésimo bloque para el tratamiento j –ésimo.

Los supuestos para los dos terminos de error son los mismos del DCA:

$$\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2) \text{ y } \theta_{ijk} \stackrel{iid}{\sim} N(0, \sigma_\theta^2).$$

La descomposición de la suma de cuadrados total sigue el mismo patrón de los casos anteriores, así se tiene que:

$$SC_{Total} = SC_{Bloques} + SC_{TTO} + SC_{EE} + SC_{EM}, \quad (15)$$

donde,

$$SC_{Total} = \sum_{i=1}^r \sum_{j=1}^t \sum_{k=1}^s Y_{ijk}^2 - FC, \quad (16)$$

con $FC = \frac{Y_{...}^2}{rst}$ y $Y_{...} = \sum_{i=1}^r \sum_{j=1}^t \sum_{k=1}^s Y_{ijk}$,

$$SC_{Bloques} = \frac{1}{st} \sum_{i=1}^r Y_{i..}^2 - FC,$$

$$SC_{TTO} = \frac{1}{sr} \sum_{j=1}^t Y_{.j.}^2 - FC,$$

$$SC_{EE} = \frac{1}{s} \sum_{i=1}^r \sum_{j=1}^t Y_{ij}^2 - FC - SC_B - SC_{TTO}$$

$$SC_{EM} = SC_{Total} - SC_{Bloques} - SC_{TTO} - SC_{EE}.$$

El ANAVA se muestra en la Tabla 12.

Tabla 12: ANAVA para un DBCA con muestreo en las unidades experimentales

F.V.	gl	SC	CM	$E[CM]$	F_c
Bloques	$r - 1$	$SC_{Bloques}$	$CM_{Bloques}$	$\sigma_\theta^2 + s\sigma_\rho^2 + \frac{ts}{r-1} \sum \beta_i^2$	$\frac{CM_{Bloques}}{CM_{EE}}$
TTO	$t - 1$	SC_{TTO}	CM_{TTO}	$\sigma_\theta^2 + s\sigma_\rho^2 + \frac{rs}{t-1} \sum \tau_j^2$	$\frac{CM_{TTO}}{CM_{EE}}$
EE	$(r - 1)(t - 1)$	SC_{EE}	CM_{EE}	$\sigma_\theta^2 + s\sigma_\rho^2$	
EM	$rt(s - 1)$	SC_{EM}	CM_{CM}	σ_θ^2	
TOTAL	$rts - 1$				

6.9.6. Función (Sub.DBCA)

Como R no aporta una función que vaya dirigida a realizar el análisis de submuestreo con un DBCA, se programó una función llamada (Sub.DBCA).

- La función (Sub.DBCA) está compuesta por:

```

1 > Sub.DBCA(treat, SUBM, block, VaInts, quali=T,
2 sigF=0.05, sigT=0.05)

```

- Características

treat: son los tratamientos.

SUBM: las submuestras.

block: contiene los bloques.

`VaInts`: es la variable de interés.

`quali`: Lógica. Si es `TRUE` (por defecto), los tratamientos se asumen cualitativos, si es `FALSE`, cuantitativos.

Esta función aporta el análisis de varianza, prueba de normalidad Shapiro-Wilk, prueba de homogeneidad de varianza de Bartlett y las comparaciones de medias.

6.9.7. Ejemplo en R de muestreo de unidades experimentales en un DBCA

Supongamos que en el ejemplo anterior se lleva a cabo con un diseño en bloque, en este caso la variable `Contenedor` es el factor de bloqueo. Los pasos a seguir son:

1. Cargar los datos como se realizó en el paso anterior

```
1 > Base=read.csv2(file.choose())
2 > Base
3   Trat SUBM Contenedor  Peso
4 1     0     1           1 11.29
5 2     0     1           1 11.08
6 3     0     1           1 11.10
7 4     0     2           2  7.37
8 5     0     2           2  6.55
9 6     0     2           2  8.50
10 7     6     3           1  5.64
11 8     6     3           1  5.98
12 9     6     3           1  5.69
13 10    6     4           2  4.20
14 11    6     4           2  3.34
15 12    6     4           2  4.21
16 13   12     5           1  4.83
17 14   12     5           1  4.77
18 15   12     5           1  5.66
19 16   12     6           2  3.28
20 17   12     6           2  2.61
21 18   12     6           2  2.69
```

2. Empleando la función `Sub.DBCA` se obtienen los siguientes resultados:

```
1 > Sub.DBCA(Trat, SUBM, Contenedor, Peso, quali=T,
2 sigF=0.05, sigT=0.05)
```

Esta fue programada para aportar los siguientes resultados:

■ Análisis de varianza

```
1 Analysis of Variance Table
2 -----
3           DF      SS      MS      Fc      Pr>Fc
4 Treatment    2  98.572  49.286  35.129  0.028
5 block        1  30.135  30.135  21.479  0.044
6 Residuals    2   2.805   1.403
7 Sampling error 12   3.273   0.273
8 Total        17 134.785  81.096
9 -----
10 CV = 19.6 %
```

Los resultados previos obtenidos de la tabla ANAVA permiten concluir que el p_{valor} es inferior al compararlo con un $\alpha = 0,05$, por tanto, existen diferencias significativas entre las medias de masa seca correspondiente a los tres niveles de salinidad del medio controlado.

■ Supuestos del modelo

```
1 Shapiro-Wilk normality test
2 p-value: 0.7798462
3 According to Shapiro-Wilk normality test at 5% of
4 significance, residuals can be considered normal.
5 -----
6 Bartlett homogeneity test
7 p-value: 0.3673749
8 According to Bartlett homogeneity test at 5% of
9 significance, it is considered that the residuals
10 have constant variance.
11 -----
```

■ Comparaciones de medias

```
1 Tukey's test
2 -----
3 Groups Treatments Means
4 a          0      9.315
5 b          6      4.843333
```

6	b	12	3.973333
7	-----		

Mediante la prueba de Tukey se observa que existen diferencias entre los tratamientos, ya que la prueba arroja dos grupos diferentes.

6.10. Diseño Cuadrado Latino

El Cuadrado Latino es un diseño experimental que utiliza la aleatorización para distribuir los tratamientos de manera equilibrada. En este diseño, los tratamientos se organizan en una matriz de tal manera que en la primera fila y columna se encuentran ordenados alfabéticamente.

Cuando se utiliza el Cuadrado Latino, se selecciona un patrón específico de distribución entre varios patrones disponibles. Luego, se procede a aleatorizar las filas y columnas manteniendo la distribución original del diseño. El objetivo es preservar la estructura inicial durante la aleatorización, lo que proporciona un marco sólido para llevar a cabo experimentos y analizar los efectos de los tratamientos de manera controlada y sistemática (Martínez y Robles, 2013, p.138).

Una forma común de obtener un patrón de Cuadrado Latino es asignar los tratamientos en orden alfabético a la primera fila y continuar asignando los tratamientos a las filas restantes en orden alfabético, comenzando con la siguiente letra del alfabeto. Este enfoque garantiza una distribución equilibrada de los tratamientos en el diseño.

En resumen, la aleatorización en el diseño de Cuadrado Latino implica la selección de un patrón inicial y la distribución aleatoria de filas y columnas, manteniendo la estructura del diseño original. Este enfoque proporciona una forma sistemática y controlada de llevar a cabo experimentos y analizar los efectos de los tratamientos. (Martínez y Robles, 2013, p.138).

6.10.1. Aleatorización en un Diseño Cuadrado Latino

En el diseño de Cuadrado Latino, se utiliza la aleatorización para seleccionar un patrón específico entre diferentes opciones disponibles. Un patrón típico

de Cuadrado Latino se caracteriza por tener los tratamientos distribuidos en orden alfabético en la primera fila y columna.

Para cada tamaño de Cuadrado Latino, existen varios patrones que se pueden utilizar como base para la aleatorización. Una vez elegido el patrón de Cuadrado Latino, se procede a distribuir aleatoriamente las filas y columnas, asegurando que se conserve la distribución original del diseño. El objetivo principal es preservar la estructura del diseño inicial durante todo el proceso de aleatorización (Martínez y Robles, 2013, p.138).

Una forma común de obtener un patrón de Cuadrado Latino es asignar los tratamientos en orden alfabético a la primera fila y luego continuar asignando los tratamientos a las filas restantes en orden alfabético, comenzando con la siguiente letra del alfabeto. Este enfoque garantiza una distribución equilibrada de los tratamientos en el diseño.

En resumen, la aleatorización en el diseño de Cuadrado Latino implica seleccionar un patrón inicial y luego realizar una distribución aleatoria de filas y columnas, manteniendo la estructura original del diseño. Este enfoque proporciona un marco sólido para llevar a cabo experimentos y analizar los efectos de los tratamientos de manera controlada y sistemática (Martínez y Robles, 2013, p.138).

Para $t = 5$, tratamientos, se tiene el cuadrado latino patrón presentado en la Tabla 13. (Martínez y Robles, 2013, p.138)

Tabla 13: Cuadrado latino patrón

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

Posteriormente se sortean al azar las columnas o las hileras, por ejemplo sortear en el cuadrado latino anterior, las columnas en el orden 3, 5, 2, 1, 4. entonces se obtiene el cuadrado latino dado en la Tabla 14:

Tabla 14: Cuadrado latino patrón aleatorizado

C	E	B	A	D
D	A	C	B	E
E	B	D	C	A
A	C	E	D	B
B	D	A	E	C

Tabla 15: Diseño cuadrado latino

B	D	A	E	C
C	E	B	A	D
D	A	C	B	E
A	C	E	D	B
E	B	D	C	A

Ahora se deben sortear las filas o hileras en un orden determinado, por ejemplo en el orden 5, 1, 2, 4, 3, obteniéndose finalmente el cuadrado latino a implementar es presentado en la Tabla 15.

Bajo esa estructura de diseño, se asignan los tratamientos a los cuadrados según la letra correspondiente.

6.10.2. Modelo Estadístico e Hipótesis

En el caso de un ensayo en un DCL, el modelo estadístico es el siguiente:

$$Y_{ij(k)} = \mu + \alpha_i + \beta_j + \tau_{(k)} + \epsilon_{ij(k)}, \quad (17)$$

con: $i = j = k = 1, 2, \dots, t$, donde,

- $Y_{ij(k)}$: es la observacion de la variable Y en la i –ésima fila y j –ésima columna para el tratamiento k –ésimo.
- μ : es la media general común a todos los tratamientos.
- α_i : es el efecto de la i –ésima fila.
- β_j : es el efecto de la j –ésima columna.
- $\tau_{(k)}$: es el efecto del tratamiento k –ésimo.

- $\epsilon_{ij(k)}$: es el error experimental del k –ésimo tratamiento en la i –ésima fila y j –ésima columna.

Los supuestos asumidos en el modelo para este término de error es que ellos son independientes, están normalmente distribuidos, con media cero y tienen varianza homogénea σ^2 , es decir: $\epsilon_{ij(k)} \stackrel{iid}{\sim} N(0, \sigma^2)$.

6.10.3. Análisis de Varianza

Para este diseño, la variación total se descompone en: filas, columnas, tratamientos y error, es decir se tiene que: (Martínez y Robles, 2013, p.140)

$$\begin{aligned} SC_{Total} &= SC_{Filas} + SC_{Columnas} + SC_{Tratamiento} + SC_{Error\ experimental} \\ &= SC_{Fila} + SC_{Col} + SC_{TTO} + SC_{EE}. \end{aligned} \quad (18)$$

El esquema general del ANAVA en un DCL, para efectos fijos y aleatorios se presenta en la Tabla 16

Tabla 16: ANAVA con cuadrados medios esperados para un DCL

F.V.	gl	SC	CM	$E(CM)$ Efectos fijos	$E(CM)$ Efectos aleatorios
Filas	$(t - 1)$	SC_{Fila}	CM_{fila}	$\sigma^2 + t\sigma_f^2$	$\sigma^2 + t\sigma_f^2$
Columnas	$(t - 1)$	SC_{Col}	CM_{Col}	$\sigma^2 + t\sigma_c^2$	$\sigma^2 + t\sigma_c^2$
TTO	$(t - 1)$	SC_{TTO}	CM_{TTO}	$\sigma^2 + \frac{1}{t-1} \sum_k \tau_k^2$	$\sigma^2 + t\sigma_k^2$
EE	$(t - 1)(t - 2)$	SC_{EE}	CM_{EE}	σ^2	σ^2
Total	$t^2 - 1$	SC_{Total}			

donde,

$$SC_{Total} = \sum_{i=1}^t \sum_{j=1}^t Y_{ij(k)}^2 - FC, \quad (19)$$

con

$$FC = \frac{\left(\sum_{i=1}^t \sum_{j=1}^t Y_{ij(k)} \right)^2}{t^2}, \quad (20)$$

$$SC_{Fila} = \frac{1}{t} \sum_{i=1}^t \sum_{j=1}^t Y_{i..}^2 - FC, \quad (21)$$

$$SC_{Col} = \frac{1}{t} \sum_{j=1}^t Y_{.j.}^2 - FC, \quad (22)$$

$$SC_{TTO} = \frac{1}{t} \sum_{k=1}^t Y_{..k}^2 - FC, \quad (23)$$

y

$$SC_{EE} = SC_{Total} - SC_{Fila} - SC_{Col} - SC_{TTO}. \quad (24)$$

El análisis de varianza resumida para llevar a cabo las pruebas de hipótesis de filas, columnas y tratamientos se presenta en la Tabla 17.

Tabla 17: Tabla resumida para un ANAVA de un DCL

F.V.	gl	SC	CM	F_c
Filas	$(t - 1)$	SC_{Fila}	CM_{Fila}	$\frac{CM_{Fila}}{CM_{EE}}$
Columnas	$(t - 1)$	SC_{Col}	CM_{Col}	$\frac{CM_{Col}}{CM_{EE}}$
TTO	$(t - 1)$	SC_{TTO}	CM_{TTO}	$\frac{CM_{TTO}}{CM_{EE}}$
Error	$(t - 1)(t - 2)$	SC_{EE}	CM_{EE}	
Total	$(t^2 - 1)$	SC_{Total}		

6.10.4. Función (latsd) en R

Esta información fue obtenida de la página web (Ferreira, Cavalcanti y Nogueira, 2022). Otra forma de obtenerla es utilizar el comando `help(latsd)` en R.

Los autores de la función (`latsd`) son Eric B Ferreira, Denismar Alves Nogueira y Portya Piscitelli Cavalcanti, el cual lleva a cabo un Diseño Cuadrado Latino, considerando un modelo fijo este se encuentra en la librería de R llamada (`ExpDes`).

- La función (`latsd`) está compuesto por:

```
1 > latsd(treat, row, column, resp, quali = TRUE, mcomp = "tukey",
2 sigT = 0.05, sigF = 0.05, unfold = NULL)
```

- Características

treat: Vector numérico o complejo que contiene los tratamientos.

row : Vector numérico o complejo que contiene las filas.

column : Vector numérico o complejo que contiene las columnas.

resp: Vector numérico o complejo que contiene la variable de respuesta.

quali: Lógico. Si es TRUE (por defecto), se asume que los tratamientos son cualitativos. Si es FALSE, se asume que son cuantitativos.

mcomp: Permite elegir la prueba de comparación múltiple; el valor predeterminado es la prueba de Tukey, sin embargo, las opciones son: la prueba LSD ('lsd'), la prueba LSD con protección de Bonferroni ('lsdb'), la prueba de Duncan ('duncan'), la prueba de Student-Newman-Keuls ('snk'), la prueba de Scott-Knott ('sk'), la prueba de Calinski y Corsten ('ccF') y la prueba de comparación múltiple mediante bootstrap ('ccboot').

sigT: La significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 %.

sigF: La significación que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5 %.

unfold: Dice lo que se debe hacer después del ANOVA. Si es NULL (predeterminado), recomendado se realizan pruebas; si es '0', solo se realiza ANOVA; si '1', los efectos simples.

Esta función contiene el ANOVA de la prueba LSD, la prueba de normalidad de Shapiro-Wilk para los residuos del modelo, los modelos de regresión ajustados (cuando los tratamientos son cuantitativos) y/o las pruebas de comparación múltiple (cuando los tratamientos son cualitativos).. Para más información utilizar el comando `help(latsd)`.

6.10.5. Ejemplo en R para un diseño cuadrado latino usando la función (latsd)

Se llevó a cabo un experimento para estudiar el efecto de las raciones de alimento, diferenciadas por fuente de proteína, en la producción diaria promedio de leche de las vacas. Se utilizaron 5 Raciones por experimento. Se utilizó un cuadrado latino 5×5 en el cual las filas representan diferentes vacas y las columnas distintos periodos de lactancia ¿Puede concluirse que las diferentes raciones con diferentes fuentes de proteína tienen un efecto en la producción diaria (*kg*) de las vacas?.

Los datos son dados en la Tabla 18.

Tabla 18: Producción promedio de leche de vaca

Periodos/Animal	Vaca 1	Vaca 2	Vaca 3	Vaca 4	Vaca 5
1	A(33,1)	B(34,4)	C(26,4)	D(34,6)	E(33,9)
2	C(30,7)	D(28,7)	E(24,9)	A(28,8)	B(28,0)
3	D(28,7)	E(28,8)	A(20,0)	B(31,9)	C(22,7)
4	E(31,4)	A(22,3)	B(18,7)	C(31,0)	D(21,3)
5	B(28,9)	C(22,3)	D(15,8)	E(30,9)	A(19,0)

Para llevar a cabo la solución del problema se tiene en cuenta los siguientes pasos:

1. Se carga la base de datos a R y se observan los 6 primeros elementos.

```
1 > dcl <- read_excel("C:/Users/Hp E2/Desktop/dcl.xlsx")
2 > head(dcl)
3   trat vaca periodo resp
4 1 A      1         1  33.1
5 2 C      1         2  30.7
6 3 D      1         3  28.7
7 4 E      1         4  31.4
8 5 B      1         5  28.9
9 6 B      2         1  34.4
```

2. Ahora se utiliza la función (latsd)

```
1 > latsd(dcl$trat,dcl$vaca,dcl$periodo,dcl$resp, quali =
      TRUE,
2 mcomp = "snk",sigT = 0.05, sigF = 0.05, unfold=NULL)
```

Los resultados obtenidos son los siguientes:

■ Analisis de varianza

```
1 -----
2 Analysis of Variance Table
3 -----
4           DF      SS      MS      Fc      Pr>Fc
5 Treatment  4   89.26  22.316  12.549  3.0136e-04
6 Row        4  353.12  88.281  49.644  2.2800e-07
7 Column     4  245.82  61.456  34.559  1.6930e-06
8 Residuals 12   21.34   1.778
9 Total      24  709.55
10 -----
11 CV = 4.92 %
```

En resumen, los resultados indican que los factores de tratamiento, fila y columna tienen efectos significativos en la variable de respuesta, mientras que los residuos explican una cantidad relativamente pequeña de la variabilidad en los datos.

■ Prueba de Shapiro-Wilk

```
1 -----
2 Shapiro-Wilk normality test
3 p-value:  0.367643
4 According to Shapiro-Wilk normality test at 5% of
5 significance, residuals can be considered normal.
6 -----
```

Según la prueba de normalidad de Shapiro-Wilk al 5% de significación, los residuos pueden considerarse normales.

■ Prueba de diferencia de medias

```
1 Student-Newman-Keuls's test (SNK)
2 -----
3 Groups  Treatments  Means
4 a       E          29.98
5 ab      B          28.38
6 bc      C          26.62
7 c       D          25.82
8 c       A          24.64
9 -----
```

Respondiendo la pregunta propuesta en el inciso los tratamientos son el 2 y el 5, ya que tienen la mayor producción de leche con respecto a los otros.

6.10.6. Juego de Cuadrados Latinos

Los Cuadrados Latinos de gran tamaño, como aquellos con nueve o más tratamientos, son poco comunes debido a la dificultad de obtener y manejar un número elevado de unidades experimentales en el campo. En estos casos, la disponibilidad de recursos suele ser limitada.

Por otro lado, cuando se trabaja con pocos tratamientos, como por ejemplo tres, los grados de libertad del error no son suficientes para obtener resultados precisos. En el caso de un Cuadrado Latino con tres tratamientos ($t = 3$), los grados de libertad del error se reducen a solo dos, lo que limita la precisión de los resultados obtenidos.

Para abordar esta limitación, se puede planificar un experimento en el que se ejecuten múltiples Cuadrados Latinos del mismo orden, es decir, se replica el diseño. Esta replicación permite aumentar los grados de libertad del error y mejorar la precisión de los resultados.

El análisis de varianza (ANAVA) se utiliza para analizar los datos obtenidos en estos diseños y realizar pruebas de hipótesis. La Tabla 19 muestra un ejemplo de ANAVA, que incluye el cálculo de las sumas de cuadrados (SC), las medias de cuadrados (CM) y las estadísticas de prueba correspondientes.

En resumen, los Cuadrados Latinos de gran tamaño son poco utilizados debido a la dificultad de manejar un alto número de unidades experimentales. Cuando se trabaja con pocos tratamientos, la replicación del diseño es necesaria para obtener resultados confiables. El ANAVA es una herramienta clave para analizar los datos y probar las hipótesis en estos diseños.

donde, $FC = \frac{Y^2}{st^2}$, $i = 1, 2, \dots, s$ es el número de cuadrados, $j = k = l = 1, 2, \dots, t$, son número de filas, columnas, o tratamientos respectivamente.

6.10.7. Ejemplo en R para un diseño cuadrado cuadrado latino replicado usando la función (ea2).

(Martínez y Martínez, 1997). Se usaron 6 animales en un experimento con tres tratamientos, cada animal recibió cada tratamiento en secuencia durante 3 períodos; al final, en cada período se determinó la glucosa en la sangre de

Tabla 19: ANAVA para un cuadrado latino replicado s veces

F.V.	gl	SC
Cuadrado	$s - 1$	$\sum_{i=1}^s \sum \frac{Y_{i..}^2}{t^2} - \frac{Y^2}{st^2}$
Filas(Cuadrado)	$s(t - 1)$	$\sum_{i=1}^s \sum_{j=1}^t \frac{Y_{ij.}^2}{t} - FC - SC_{cuadrado}$
Columna(Cuadrados)	$s(t - 1)$	$\sum_{i=1}^s \sum_{k=1}^t \frac{Y_{i.k.}^2}{t} - FC - SC_{cuadrado}$
TTO	$t - 1$	$\sum_{l=1}^t \frac{Y_{...l}^2}{st} - FC$
EE	$(t - 1)(st - s - 1)$	Por diferencia
Total	$st^2 - 1$	$\sum_{i=1}^s \sum_{j=1}^t \sum_{k=1}^t Y_{ijk(l)}^2 - FC$

cada animal. La glucosa en mg/100 ml y los tratamientos, A, B y C para cada animal son como se presentan en las Tablas 20 y 21.

Tabla 20: Glucosa en la sangre de seis animales

Periodo	Animal 1	Animal 2	Animal 3	Total
1	67 (A)	78 (B)	69 (C)	214
2	76 (B)	69 (C)	67 (A)	212
3	65 (C)	68 (A)	85 (B)	218
Total	208	215	221	644

Tabla 21: Glucosa en la sangre de seis animales

Periodo	Animal1	Animal 2	Animal 3	Total
1	80 (A)	71 (C)	84 (B)	235
2	56 (C)	75 (B)	71 (A)	202
3	83 (B)	75 (A)	67 (C)	225
Total	219	221	222	662

Las medias de los tratamientos fueron: 71(A), 80(B), y 66(C).

El diseño es un cuadrado latino replicado dos veces. La primera matriz corresponde al primer cuadrado latino y la segunda matriz es el segundo cuadrado latino. Cada animal recibió los tres tratamientos durante los tres periodos. Todos los tratamientos aparecen en cada periodo.

Los pasos a seguir para obtener esos resultados son los siguientes:

1. Se carga la base de datos y se observan los 6 primeros elementos.

```
1 > dclr <- read_excel("C:/Users/Hp E2/Desktop/dcl1.xlsx")
2 > head(dclr)
3   trat  replica animal periodo  resp
4 1 A          1      1         1    67
5 2 B          1      1         2    76
6 3 C          1      1         3    65
7 4 B          1      2         1    78
8 5 C          1      2         2    69
9 6 A          1      2         3    68
```

2. Utilizando la función (`ea2`) se lleva a cabo un análisis con diseño cuadrado latino replicado. Las variables que se trabajan en esta función de acuerdo con este ejemplo son: la base de datos llamada `dclr` y (`design=13`) ya que de acuerdo a la descripción previamente dada significa que el análisis que se va a realizar un diseño cuadrado latino replicado.

```
1 > ea2(dclr, design =13)
```

- Se obtiene el análisis de varianza

```
1 $'Analysis of variance'
2      df type III SS mean square F value p>F
3 treatments      2 601.44    300.72    30.75 0.003
4 squares          1 18.00     18.00     1.840 0.246
5 treatments:squares  2 93.00     46.50     4.755 0.087
6 squares:rows      4 29.77      7.44     0.761 0.601
7 squares:columns   4 197.11    49.27     5.039 0.073
8 residuals        4 39.11      9.77      -     -
```

Los resultados aportados por el análisis de varianza muestran que, la interacción entre squares y filas no se encuentra evidencia suficiente para indicar una interacción significativa entre los squares y las filas en la variable de interés. El p_{valor} asociado es 0.601, lo cual sugiere que las diferencias observadas no se deben a una interacción entre estos dos factores. La interacción entre squares y columnas, aunque no se alcanza significancia estadística (p_{valor} de 0.073), se observa una tendencia hacia una interacción entre los squares y las columnas en la variable

de interés. Esto sugiere que los efectos de los squares pueden depender de los niveles de las columnas, pero se requiere más evidencia para confirmar esta relación.

- Ajuste de medias y comparaciones múltiples (Tratamientos).

```

1 $'Adjusted means (treatments)'
2   treatment adjusted.mean standard.error tukey snk duncan
3 1          B          80.1667          1.2766      a   a     a
4 2          A          71.3333          1.2766      b   b     b
5 3          C          66.1667          1.2766      b   c     c
6 t   scott_knott
7 a   a
8 b   b
9 c   c
10
11 $'Multiple comparison test (treatments)'
12   pair contrast p(tukey) p(snk) p(duncan)  p(t)
13 1 B - A      8.8334   0.0176 0.0081    0.0081 0.0081
14 2 B - C     14.0000   0.0033 0.0033    0.0017 0.0015
15 3 A - C      5.1666   0.0952 0.0458    0.0458 0.0458

```

El análisis de medias ajustadas revela que el tratamiento B tiene la media ajustada más alta con 80.1667, seguido por el tratamiento A con 71.3333 y el tratamiento C con 66.1667. Las pruebas de comparación múltiple confirman que el tratamiento B es significativamente diferente de los tratamientos A y C, mientras que no se encontró una diferencia significativa entre los tratamientos A y C. Estos resultados indican que el tratamiento B tiene un efecto positivo en la variable medida en comparación con los otros tratamientos.

- Ajuste de medias y comparaciones múltiples (squares).

```

1 $'Adjusted means (squares)'
2   square adjusted.mean standard.error tukey snk
3 1      2          73.55556          1.0423      a   a
4 2      1          71.55556          1.0423      a   a
5 duncan t   scott_knott
6 a      a   a
7 a      a   a
8
9 $'Multiple comparison test (squares)'
10   pair contrast p(tukey) p(snk) p(duncan)  p(t)
11 1 2 - 1          2    0.2464 0.2464    0.2464 0.2464

```

En resumen, el análisis de medias ajustadas para los niveles de squares indica que no se encontró una diferencia significativa entre los niveles 1 y 2. Ambos niveles tienen medias ajustadas similares y no se puede establecer una superioridad estadística entre ellos. Esto sugiere que el factor squares no tiene un impacto significativo en la variable medida.

- Ajuste de medias y comparaciones múltiples de los tratamientos en niveles de squares.
- Ajuste de medias del tratamiento 1 en cada nivel de squares

```

1 $'Adjusted means (treatments in levels of squares) '$'
   treatments in 1'
2   interaction adjusted.mean standard.error tukey snk
3 2           B.1           79.6667           2.576   a   a
4 3           C.1           67.6667           2.576   a   b
5 1           A.1           67.3333           2.576   a  ab
6 duncan t scott_knott
7 a       a  a
8 b       b  b
9 b       b  b

```

- Ajuste de medias del tratamiento 2 en cada nivel de squares

```

1 $'Adjusted means (treatments in levels of squares) '$'
   treatments in 2'
2   interaction adjusted.mean standard.error tukey snk
3 5           B.2           80.6667           2.576   a   a
4 4           A.2           75.3333           2.576  ab   a
5 6           C.2           64.6667           2.576   b   b
6 duncan t scott_knott
7 a       a  a
8 a       a  a
9 b       b  b

```

- Comparaciones múltiples del tratamiento 1 en cada nivel de squares

```

1 $'Multiple comparison test (treatments in levels of
   squares)
2 '$'treatments in 1'
3   pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 B.1 - C.1 12.0000 0.0635 0.0301 0.0301 0.0301
5 2 B.1 - A.1 12.3334 0.0585 0.0585 0.0297 0.0276
6 3 C.1 - A.1 0.3334 0.9954 0.9315 0.9315 0.9315

```

- Comparaciones múltiples del tratamiento 2 en cada nivel de squares

```

1 $'Multiple comparison test (treatments in levels of
   squares)
2 '$'treatments in 2'
3      pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 B.2 - A.2    5.3334   0.3964 0.2170    0.2170 0.2170
5 2 B.2 - C.2   16.0000   0.0254 0.0254    0.0128 0.0118
6 3 A.2 - C.2   10.6666   0.0894 0.0429    0.0429 0.0429

```

El análisis de medias ajustadas para los tratamientos en diferentes niveles de squares muestra las siguientes conclusiones:

Para el nivel de squares 1, se encontraron diferencias significativas entre los tratamientos B.1 y C.1, así como entre los tratamientos B.1 y A.1. Sin embargo, no se encontraron diferencias significativas entre los tratamientos C.1 y A.1.

Para el nivel de squares 2, se encontró una diferencia significativa entre los tratamientos B.2 y A.2, pero no se encontraron diferencias significativas entre los tratamientos B.2 y C.2, ni entre los tratamientos A.2 y C.2.

Estos resultados indican que la interacción entre los tratamientos y los niveles de squares tiene un impacto significativo en la variable medida. Los tratamientos B.1 y B.2 muestran las medias ajustadas más altas en sus respectivos niveles de squares.

- Ajuste de medias y comparaciones múltiples de los squares en niveles de tratamientos.
- Ajuste de medias del square A en cada nivel del tratamiento

```

1 $'Adjusted means (squares in levels treatments) '$'squares
   in A'
2      interaction adjusted.mean standard.error tukey snk
3 4          A.2          75.3333          2.576    a    a
4 1          A.1          67.3333          2.576    a    a
   duncan t scott_knott
5 a      a    a
6 a      a    b

```

- Ajuste de medias del square B en cada nivel del tratamiento

```

1 $'Adjusted means (squares in levels treatments) '$squares
   in B'
2   interaction adjusted.mean standard.error tukey snk
3 5          B.2          80.6667          2.576   a   a
4 2          B.1          79.6667          2.576   a   a
5 duncan t scott_knott
6 a      a a
7 a      a a

```

- Ajuste de medias del square C en cada nivel del tratamiento

```

1 $'Adjusted means (squares in levels treatments) '$squares
   in C'
2   interaction adjusted.mean standard.error tukey snk
3 3          C.1          67.6667          2.576   a   a
4 6          C.2          64.6667          2.576   a   a
5 duncan t scott_knott
6 a      a a
7 a      a a

```

- Comparaciones múltiples del square A en cada nivel del tratamiento

```

1 $'Multiple comparison test (squares in levels treatments)
2 '$squares in A'
3   pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 A.2 - A.1      8    0.0931 0.0931    0.0931 0.0931

```

- Comparaciones múltiples del square B en cada nivel del tratamiento

```

1 $'Multiple comparison test (squares in levels treatments)
2 '$squares in B'
3   pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 B.2 - B.1      1    0.7973 0.7973    0.7973 0.7973

```

- Comparaciones múltiples del square C en cada nivel del tratamiento

```

1 $'Multiple comparison test (squares in levels treatments)
2 '$squares in C'
3   pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 C.1 - C.2      3    0.4565 0.4565    0.4565 0.4565

```

Estos resultados indican que no hay diferencias significativas entre los niveles de squares dentro de cada tratamiento.

■ Análisis de los residuales

```
1 $ 'Residual analysis '  
2  
3 p.value Shapiro-Wilk test 0.0269  
4 p.value Bartlett test (treatments) 1.0000  
5 p.value Bartlett test (squares) 0.7784  
6 p.value Bartlett test (interaction) 1.0000  
7 coefficient of variation (%) 4.3100  
8 first value most discrepant 9.0000  
9 second value most discrepant 5.0000  
10 third value most discrepant 1.0000
```

El análisis de residuos revela que los residuos no siguen una distribución normal, lo cual indica que los datos no cumplen con la suposición de normalidad. Además, las pruebas de homogeneidad de varianzas no muestran diferencias significativas entre los grupos de tratamientos y los niveles de squares. Esto sugiere que la variabilidad entre los grupos es similar.

El coeficiente de variación es del 4.31 %, lo cual indica una variabilidad moderada en los datos. Sin embargo, se identificaron algunos valores atípicos en los residuos, lo que sugiere que estos puntos podrían tener un impacto considerable en los resultados del análisis.

6.11. Análisis de covarianza

Según (Montgomery, 2012), el análisis de covarianza es útil para mejorar la precisión de un experimento. Si suponemos que en un experimento la variable respuesta y está relacionada linealmente con la variable independiente x , asimismo, el experimento no puede controlar la variable x , pero puede medirla al mismo tiempo que la variable respuesta, esta variable x se conoce como *covariable* o *variable concomitante*.

6.11.1. Análisis de covarianza en un DCA

Si se supone una relación lineal ente la respuesta y covariable el modelo que se propone es:

$$y_{ij} = \mu + \tau_i + \beta_i(x_{ij} - \bar{x}_{..}) + \epsilon_{ij} \quad i = 1, \dots, t; j = 1, \dots, r$$

Donde y_{ij} , es la j -ésima observación de la variable de respuesta tomada bajo el i -ésimo tratamiento o nivel del factor único, x_{ij} , es el valor de la covariable o variable concomitante correspondiente a y_{ij} , $\bar{x}_{..}$, es la media de las x_{ij} , μ es la media global, τ_i es el efecto del tratamiento i -ésimo y β_i es el coeficiente de regresión lineal en el i -ésimo tratamiento.

Se supone en el modelo anterior que los coeficientes de cada tratamiento son idénticos, si es cierto $\beta_i = \beta$ para i es común para los tratamientos y para su estimación se combina la información en las muestras de los tratamientos, donde al final se conserva una estimación ponderada de β .

La hipótesis de interés es:

$$H_0 = \beta_1 = \beta_2 = \dots = \beta_t = \beta.$$

En el caso de que se rechaza H_0 se concluye que al menos una pendiente no tiene la misma tendencia.

Tabla 22: Análisis de covariancia para un experimento de un factor con una covariable

		SC			Ajustado para la regresión		
Fuente de variación	gl	x	xy	y	y	gl	MC
Tratamiento	$a - 1$	T_{xx}	T_{xy}	T_{yy}			
Error	$a(n - 1)$	E_{xx}	E_{xy}	E_{yy}	SCE	$a(n - 1) - 1$	MC_E
Total	$an - 1$	S_{xx}	S_{xy}	S_{yy}	SCT	$an - 2$	
Tratamientos aj.					$SCT - SCE$	$a - 1$	$\frac{SCT - SCE}{a - 1}$

Sea

- $S_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{an}$
- $S_{xx} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - \frac{x_{..}^2}{an}$
- $S_{xy} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) = \sum_{i=1}^a \sum_{j=1}^n x_{ij}y_{ij} - \frac{(x_{..})(y_{..})}{an}$
- $T_{yy} = \sum_{i=1}^a (y_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^a \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$

- $T_{xx} = \sum_{i=1}^a (x_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^a \frac{x_{i.}^2}{n} - \frac{x_{..}^2}{an}$
- $T_{yy} = \sum_{i=1}^a (x_{i.} - \bar{x}_{..})(y_{i.} - \bar{y}_{..}) = \sum_{i=1}^a \frac{(x_{i.})(y_{i.})}{n} - \frac{(x_{..})(y_{..})}{an}$
- $E_{yy} = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = S_{yy} - T_{yy}$
- $E_{xx} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2 = S_{xx} - T_{xx}$
- $E_{xy} = \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}) = S_{xy} - T_{xy}$

En general, $S = T + E$, donde S , T y E sirven para representar las sumas de cuadrados y de productos cruzados del total, los tratamientos y el error, respectivamente. A diferencia de las sumas de cuadrados para x y y que son positivas, las sumas de productos cruzados (xy) pueden ser negativas (Montgomery, 2013, p.658).

Ahora, se muestra como se adapta la respuesta al efecto de la covariable aplicando el análisis de covariancia. Considerando el modelo $y_{ij} = \mu + \tau_i + \beta_i(x_{ij} - \bar{x}_{i.}) + \epsilon_{ij}$. Los estimadores de mínimos cuadrados de μ , τ_i y β_i son $\hat{\mu} = \bar{y}_{..}$, $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \beta_i(\bar{x}_{i.} - \bar{x}_{..})$ (Montgomery, 2000, p.658).

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}}$$

La suma de cuadrados de error del modelo es

$$SCE = E_{yy} - (E_{yy})^2/E_{xx}$$

con $a(n-1) - 1$ grados de libertad. La varianza del error experimental estima mediante

$$MC_E = \frac{SCE}{a(n-1) - 1}$$

En caso de que no existan efectos de tratamiento, los estimadores de mínimos cuadrados para μ y β son $\hat{\mu} = \bar{y}_{..}$ y $\hat{\beta} = S_{xy}/S_{xx}$. La suma de cuadrados de error de este modelo reducido es

$$SCT = S_{yy} - (S_{xy})^2/S_{xx}$$

con $an - 2$ grados de libertad. El termino $(S_{xy})^2$ es la reducción de la suma de cuadrados de y obtenida por medio de la regresión lineal de y sobre x . En consecuencia, para probar $H_0 : \tau_i = 0$, se calcula

$$F_0 = \frac{SCT - SCE/(a-1)}{SCE/(a(n-1)-1)}$$

Si la hipótesis nula es verdad, tiene una distribución $F_{a-1, a(n-1)-1}$, entonces se rechaza la hipótesis nula si $F_0 > F_{\alpha, a-1, a(n-1)-1}$.

6.11.2. Análisis de covarianza en un DBCA

El modelo estadístico de bloques con covariable es:

$$y_{ij} = \mu + \tau_i + \gamma_j + \beta(x_{ij} - \bar{x}_{.}) + \epsilon_{ij}$$

Donde $i = 1, 2, \dots, t$, t es el número de tratamiento y $j = 1, 2, \dots, r$ y r es el número de bloques. En este diseño se tienen los mismos supuestos del DCA, es decir los valores de x son fijos, sin error e independiente de los tratamientos, la regresión de y con respecto a x es lineal e independiente con respecto a los bloques y tratamientos, por último, los residuales deben presentar una distribución normal estándar.

Tabla 23: Análisis de covarianza para el diseño aleatorizado por bloques completos.

Fuente de variación	gl	SC			Ajustado para la regresión		
		x	xy	y	y	gl	MC
Bloques	$b-1$	R_{xx}	R_{xy}	R_{yy}			
Tratamiento	$a-1$	T_{xx}	T_{xy}	T_{yy}			
Error	$(a-1)(b-1)$	E_{xx}	E_{xy}	E_{yy}	SCE	$ab-a-b$	MC_E
Total	$ab-1$	S_{xx}	S_{xy}	S_{yy}		$ab-2$	
Tratamientos aj.					$SCT - SCE$	$a-1$	$\frac{SCT-SCE}{a-1}$

Sea

- $R_{xx} = \sum \frac{x_{.j}^2}{a} - \frac{x_{..}^2}{ab}$
- $R_{yy} = \sum \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$
- $R_{xy} = \sum \frac{x_{.j}y_{.j}}{a} - \frac{x_{..}y_{..}}{ab}$
- $T_{xx} = \sum \frac{x_{i.}^2}{b} - \frac{x_{..}^2}{ab}$

- $T_{yy} = \sum \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$
- $T_{xx} = \sum \frac{x_{i.}y_{i.}}{b} - \frac{x_{..}y_{..}}{ab}$
- $S_{xx} = \sum \sum x_{ij}^2 - \frac{x_{..}^2}{ab}$
- $S_{yy} = \sum \sum y_{ij}^2 - \frac{y_{..}^2}{ab}$
- $S_{xx} = \sum \sum x_{ij}y_{ij} - \frac{x_{..}y_{..}}{ab}$
- $E_{xx} = S_{xx} - R_{xx} - T_{xx}$
- $E_{yy} = S_{yy} - R_{yy} - T_{yy}$
- $E_{xy} = S_{xy} - R_{xy} - T_{xy}$

Con el propósito de eliminar los efectos de los bloques, la suma de cuadrados de tratamientos ajustados se calcula mediante una diferencia entre la suma de cuadrados del error ajustada y la suma de cuadrados ajustada por los tratamientos más error, ahora se puede observar el coeficiente de regresión β se estima mediante la siguiente ecuación (Montgomery, 2000, p.526).

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}}$$

La estadística apropiada para probar la hipótesis donde $H_0 = 0$ es:

$$F_0 = \frac{(E_{xy})^2/E_{xx}}{MS_E}$$

La hipótesis nula es verdadera, F_0 tiene una distribución $F_{1,ab-a-b}$, por lo tanto, $H_0 = 0$ se rechaza si $F_0 > F_{1,ab-a-b}$.

6.11.3. Ejemplo en R de análisis con covarianza con un DCA.

El siguiente ejemplo se lleva a cabo utilizando la librería (`easynova`), como ya se viene trabajando anteriormente se carga la librería para acceder a la función `ea1`. Ahora veamos el ejemplo propuesto.

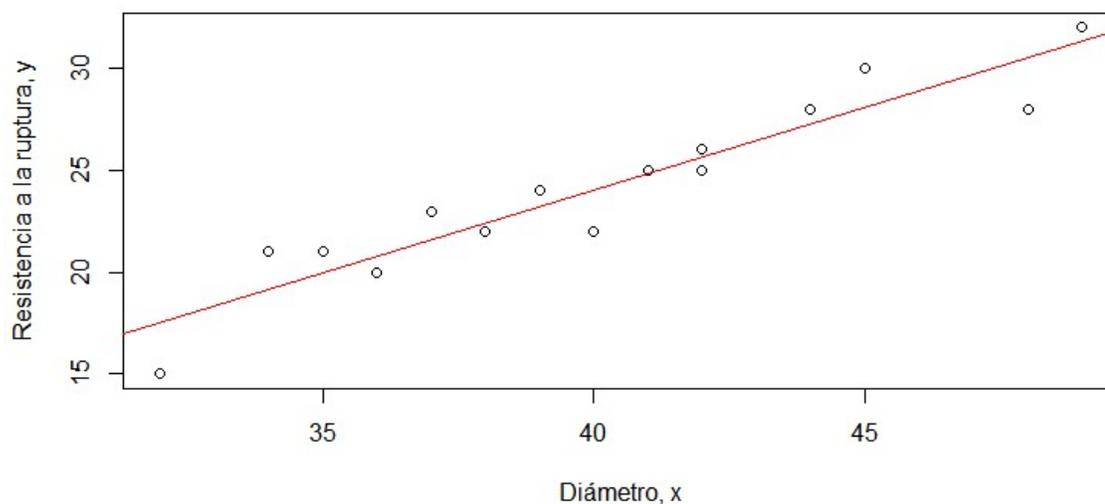
Se usan tres máquinas distintas para producir fibras monofilas para compañía textil. El ingeniero de proceso está interesado en determinar si existe diferencia en la resistencia a la ruptura de la fibra producida por las tres

máquinas. Sin embargo, la resistencia de una fibra depende del grosor de la misma, siendo generalmente más resistentes las fibras de mayor grosor. Se selecciona una muestra aleatoria de cinco fragmentos de fibra por cada máquina. La resistencia de cada fibra (y) y el grosor (x) correspondiente, se presentan en la Tabla 24 (Montgomery, 2000, p.656).

Tabla 24: Datos de resistencia a la ruptura.

Máquina A		Máquina B		Máquina C	
Diámetro	Resistencia	Diámetro	Resistencia	Diámetro	Resistencia
36	20	40	22	35	21
41	25	48	28	37	23
39	24	38	22	42	26
42	25	45	30	34	21
49	32	44	28	32	15

Figura 1: Resistencia a la ruptura (y) contra diámetro de fibra (x).



Se observa una fuerte relación lineal entre la resistencia a la ruptura y el espesor de la fibra por tanto es apropiado eliminar el efecto que tiene el espesor sobre la resistencia mediante un análisis de covariancia.

- Las hipótesis a verificar son:

$$1) H_0 : \beta = 0 \quad \text{contra} \quad H_1 : \beta \neq 0.$$

Los pasos a seguir para llevar este ejemplo a R son los siguientes:

1. Se carga la base datos y por cuestiones estéticas se visualizan los primeros seis elementos.

```
1 > COV_DCA <- read_excel("C:/Users/Hp E2/Desktop/
2 COVARIABLES-DCA.xlsx")
3 > head(COV_DCA)
4 # A tibble: 15 x 3
5   M     y     x
6 1 A     36    20
7 2 A     41    25
8 3 A     39    24
9 4 A     42    25
10 5 A     49    32
11 6 B     40    22
```

2. Utilizando la función (`ea1`) se lleva a cabo un análisis con covariable. Los argumentos que se usan de acuerdo con este ejemplo son: la base de datos llamada `COV_DCA` y (`design=5`) ya que de acuerdo a la descripción previamente dada significa que el análisis que se va a realizar es de un DCA.

```
1 > ea1(COV_DCA, design = 5)
```

Se obtienen los siguientes resultados:

■ Análisis de variación

```
1 > $'Analysis of variance'
2           df type I SS mean square F value    p>F
3 covariate   1  231.8140   231.8140  95.1904 <0.001
4 treatments  2    3.1314    1.5657  0.6429 0.5444
5 Residuals  11   26.7879    2.4353    -      -
```

■ Ajuste de medias

```
1 $'Adjusted means'
2 treatment adjusted.mean standard.error tukey snk duncan
3 1          C           24.8608         0.6979    a    a    a
4 2          A           24.0781         0.6979    a    a    a
5 3          B           23.4610         0.6979    a    a    a
6 t scott_knott
7 1 a           a
8 2 a           a
9 3 a           a
```

■ Comparaciones múltiples

```
1 $'Multiple comparison test '  
2   pair contrast p(tukey) p(snk) p(duncan)   p(t)  
3 1 C - A      0.7827    0.7148 0.4445    0.4445 0.4445  
4 2 C - B      1.3998    0.3655 0.3655    0.2034 0.1838  
5 3 A - B      0.6171    0.8096 0.5446    0.5446 0.5446
```

■ Análisis de residuales

```
1 $'Residual analysis '  
2 $'Residual analysis '$'residual analysis '  
3                                     values  
4 p.value Shapiro-Wilk test          0.6797  
5 p.value Bartlett test              0.9631  
6 coefficient of variation (%)       6.4700  
7 first value most discrepant       15.0000  
8 second value most discrepant      7.0000  
9 third value most discrepant       9.0000
```

Para el análisis de variable con un DCA, esta función ofrece, un análisis de varianza de las covariables, tratamientos y residuales, se realizan las comparaciones de medias utilizando las pruebas de Tukey, Newman–Keuls method, Duncan, t y Scott Knott. En las comparaciones múltiples de los tratamientos, se lleva a cabo los contrastes entre cada uno de ellos, también se observan los p_{valor} de las pruebas Tukey, Newman–Keuls method, Duncan y t. Para el análisis de los residuales se obtiene los p_{valor} de las pruebas Shapiro-Wilk test, Bartlett test, coeficiente de variación (%), primero, segundo y tercer valor más discrepante. El análisis que aporta este ejemplo en específico, plantea como hipótesis que existe diferencia en el nivel medio de la resistencia a la ruptura de la fibra, por tanto, el revisar los resultados se obtiene que no hay diferencias en la resistencia a la ruptura de la fibra producida por las tres máquinas.

6.11.4. Ejemplo en R de análisis con covarianza con un DBCA.

Supongamos que estás realizando un experimento para evaluar el efecto de diferentes tipos de fertilizantes en el crecimiento de plantas. Tienes cuatro tipos de fertilizantes (A, B, C, D) y quieres determinar si hay una relación entre el tipo de fertilizante utilizado y el crecimiento de las plantas. Decides utilizar un diseño en bloque completamente al azar para controlar la variabilidad entre los bloques. Tienes cuatro bloques de tierra diferentes en los que

plantarás las semillas de manera aleatoria. En cada bloque, eliges al azar una parcela para aplicar uno de los cuatro tipos de fertilizantes. Luego, mides la altura de las plantas después de un mes de crecimiento. Supongamos que además de evaluar el efecto de los diferentes tipos de fertilizantes en el crecimiento de las plantas, también quieres estudiar si la ubicación geográfica de los bloques tiene alguna influencia en el crecimiento. Agregaremos la variable Ubicación con dos niveles cualitativos: Norte y Sur. Los datos actualizados se mostrarán a continuación, con un total de 24 datos. Los datos se presentan en la Tabla 25.

Tabla 25: Datos fertilizantes

Fertilizante	Bloque	Ubicación	Altura
A	1	Norte	20
B	1	Sur	18
C	1	Norte	21
D	1	Sur	19
A	2	Norte	22
B	2	Sur	20
C	2	Norte	23
D	2	Sur	21
A	3	Norte	19
B	3	Sur	17
C	3	Norte	20
D	3	Sur	18
A	4	Norte	21
B	4	Sur	19
C	4	Norte	22
D	4	Sur	17
A	5	Norte	21
B	5	Sur	40
C	5	Norte	22
D	5	Sur	20
A	6	Norte	15
B	6	Sur	23
C	6	Norte	18
D	6	Sur	20

Recordar cargar y llamar la librería (`easyanova`), al igual que en el ejemplo anterior se utiliza la función (`ea1`), con un pequeño cambio y es que se utiliza la etiqueta (`design=6`). Los pasos a seguir para llevar a cabo este ejemplo son:

Los pasos a seguir para llevar este ejemplo a R son los siguientes:

1. Se carga la base datos y por cuestiones estéticas se visualizan los 6 primeros elementos.

```

1 > COV_DBCA <- read.csv("C:/Users/Hp E2/Desktop/COV_DBCA.
  csv"
2 , sep=";")
3 > head(COV_DBCA)
4   Fertilizante Bloque Ubicacion altura
5 1             A      1     Norte    20
6 2             B      1       Sur    18
7 3             C      1     Norte    21
8 4             D      1       Sur    19
9 5             A      2     Norte    22
10 6            B      2       Sur    20

```

2. Ahora se lleva a cabo un análisis con covariable para un DBCA.

```

1 > ea1(COV_DBCA,design = 6,alpha = 0.05, list = FALSE,
2 p.adjust=8, plot=2)

```

Se obtienen los siguientes resultados:

- Analisis de varianza

```

1 $'Analysis of variance'
2   df type I SS mean square F value  p>F
3 covariate  1   7.5571    7.5571  0.3375 0.5681
4 treatments 3  48.3333   16.1111  0.7195 0.5526
5 Residuals 19 425.4429   22.3917    -    -

```

El análisis de varianza (ANOVA) realizado revela que tanto la covariable en este caso (`Ubicacion`) como los tratamientos no presentan efectos significativos en la variable de respuesta. Tanto la covariable como los tratamientos no tienen una relación significativa con la variable medida en este estudio. En otras palabras, no se encontraron diferencias significativas entre los niveles de la covariable ni entre los tratamientos en cuanto a su influencia en la variable de interés.

■ Ajuste de medias

```

1 $ 'Adjusted means '
2   treatment  adjusted.mean  standard.error  tukey  snk  duncan
3 1           B           22.8333         1.9318      a    a    a
4 2           C           21.0000         1.9318      a    a    a
5 3           A           19.6667         1.9318      a    a    a
6 4           D           19.1667         1.9318      a    a    a
7 t.adjust.fdr  scott_knott
8 a              a
9 a              a
10 a             a
11 a             a

```

Los ajustes de medias revelan que no hay diferencias significativas entre los tratamientos en términos de sus medias ajustadas. Los tratamientos B y C tienen una media ajustada de 22.8333 y 21.0000, respectivamente, mientras que los tratamientos A y D tienen medias ajustadas de 19.6667 y 19.1667, respectivamente. Sin embargo, debido a la falta de significancia en el análisis de varianza, estos resultados deben interpretarse con precaución. Los diferentes métodos de ajuste de medias utilizados (Tukey, SNK y Duncan) también arrojan la misma conclusión de que no hay diferencias significativas entre los tratamientos.

■ Comparaciones múltiples

```

1 $ 'Multiple comparison test '
2   pair contrast  p(tukey)  p(snk)  p(duncan)  p(t) adjust.fdr
3 1 B - C      1.8333    0.9067  0.5103     0.5103    0.7573
4 2 B - A      3.1666    0.6588  0.4910     0.2866    0.7573
5 3 B - D      3.6666    0.5488  0.5488     0.2330    0.7573
6 4 C - A      1.3333    0.9608  0.6311     0.6311    0.7573
7 5 C - D      1.8333    0.9067  0.7828     0.5340    0.7573
8 6 A - D      0.5000    0.9977  0.8567     0.8567    0.8567

```

Las pruebas de comparación múltiple muestran que no hay diferencias significativas entre los pares de tratamientos en términos de medias ajustadas. Los valores p para las comparaciones entre los tratamientos B y C, B y A, B y D, C y A, C y D, y A y D son todos superiores al nivel de significancia de 0.05, lo que indica que no hay diferencias estadísticamente significativas entre estos pares de tratamientos. Esto respalda la conclusión anterior de que no hay diferencias significativas entre los tratamientos en términos de sus efectos en el crecimiento de las plantas.

- Análisis de residuales

```
1 $'Residual analysis '  
2 $'Residual analysis '$'residual analysis '  
3 values  
4 p.value Shapiro-Wilk test 0.0000  
5 p.value Bartlett test 0.0002  
6 coefficient of variation (%) 22.9000  
7 first value most discrepant 18.0000  
8 second value most discrepant 10.0000  
9 third value most discrepant 21.0000
```

El p_{valor} del test de Shapiro-Wilk es 0.0000, lo que indica que los residuos no siguen una distribución normal. El p_{valor} del test de Bartlett es 0.0002, lo que indica que la varianza de los residuos no es constante. Estos resultados sugieren que el modelo de ANCOVA podría no ser apropiado para los datos, ya que los residuos no cumplen con los supuestos de normalidad y homogeneidad de varianza. Es posible que se requieran métodos estadísticos alternativos o una transformación de los datos para obtener resultados más confiables.

6.12. Experimentos factoriales

De acuerdo a Martínez y Robles los experimentos factoriales permiten examinar las *interacciones* entre los distintos niveles de los factores estudiados, para posteriormente determinar cuál combinación de niveles de éstos optimizan la respuesta.

6.12.1. Conceptos básicos

Un factor es un conjunto de tratamientos que se pueden aplicar a las unidades experimentales, estos pueden ser cualitativos, como razas, variedades, métodos de aplicación de un producto, etc, o cuantitativos como dosis de un medicamento, raciones en una dieta, dosis de un fertilizante, distancias de siembra, tiempo de aplicación de un producto, temperatura, etc (Martínez y Robles, 2013, p.147)

Un nivel de un factor es un tratamiento particular del conjunto de tratamientos que constituyen el factor. Si el factor es cuantitativo, los niveles están formados por las diferentes cantidades del mismo; por ejemplo, para

el factor Temperatura: los niveles pueden ser 20°C, 25°C y 30°C. Si el factor es cualitativo los niveles los constituyen las diferentes categorías del mismo, es decir, si el factor es método de siembra, los niveles pueden ser: método tradicional y método tecnificado, en este caso tendría dos niveles (Martínez y Robles, 2013, p.147).

6.12.2. Efectos de un factor

Se conoce como efecto simple al efecto de otros factores. En el caso de que se tengan dos factores y estos se promedian esos efectos simples se obtiene un efecto principal. El efecto principal de un factor es una comparación entre las respuestas esperadas para los diferentes niveles de un factor (Martínez y Robles, 2013, p.148).

6.12.3. Experimento factorial bajo un DCA

El modelo estadístico para un experimento factorial bajo un DCA, con efectos fijos de dos factores A y B , con a y b niveles respectivamente es:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk} \quad (25)$$

con $i = 1, 2, 3, \dots, r$, $j = 1, 2, 3, \dots, a$ y $k = 1, 2, 3, \dots, b$, donde

- Y_{ijk} : es la respuesta en la i -ésima unidad experimental para el j -ésimo nivel del factor A y el k -ésimo nivel del factor B .
- μ : es la media general común a todos los tratamientos.
- α_j : es el efecto principal del j -ésimo nivel del factor A .
- β_k : es el efecto principal del k -ésimo nivel del factor B .
- $(\alpha\beta)_{jk}$: es el efecto del j -ésimo nivel del factor A con el k -ésimo nivel del factor B .
- ϵ_{ijk} : es el error aleatorio en la i -ésima unidad experimental para el j -ésimo nivel del factor A y el k -ésimo nivel del factor B .

Para estimar los parámetros del modelo se debe suponer que

$$\sum_{j=1}^a \alpha_j = \sum_{k=1}^b \beta_k = \sum_{j=1}^a \sum_{k=1}^b (\alpha\beta)_{jk} = 0$$

y además se supone que $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$.

Ya sea para un DCA o un DBCA, la información de cada unidad experimental se puede resumir como se ilustra en la Tabla 26.

Tabla 26: Distribución de observaciones en un factorial 2×2

TTO	Repeticiones					Total de TTO
a_1b_1	Y_{111}	Y_{211}	Y_{311}	\cdots	Y_{r11}	$Y_{\cdot 11}$
a_2b_1	Y_{121}	Y_{221}	Y_{321}	\cdots	Y_{r21}	$Y_{\cdot 21}$
a_1b_2	Y_{112}	Y_{212}	Y_{312}	\cdots	Y_{r12}	$Y_{\cdot 12}$
a_2b_2	Y_{122}	Y_{222}	Y_{322}	\cdots	Y_{r22}	$Y_{\cdot 22}$
Total de repeticiones	$Y_{1\cdot}$	$Y_{2\cdot}$	$Y_{3\cdot}$	\cdots	$Y_{r\cdot}$	$Y_{\cdot\cdot}$

Con los totales por tratamiento dados en la Tabla 27

Tabla 27: Totales para los factores A y B

A/B	b_1	b_2	$Y_{\cdot j}$
a_1	$Y_{\cdot 11}$	$Y_{\cdot 12}$	$Y_{\cdot 1}$
a_2	$Y_{\cdot 21}$	$Y_{\cdot 22}$	$Y_{\cdot 2}$
$Y_{\cdot k}$	$Y_{\cdot \cdot 1}$	$Y_{\cdot \cdot 2}$	$Y_{\cdot \cdot}$

6.12.4. Análisis de varianza

Se tienen dos estructuras, la de diseño y la de tratamiento, entonces tomando un DCA tenemos que

$$SC_{Total} = SC_{TTO} + SC_{EE}$$

y por la estructura de tratamiento se tiene que

$$SC_{TTO} = SC_A + SC_B + SC_{AB}.$$

Entonces la combinación de las dos estructuras arroja la siguiente descomposición de la suma de cuadrados total:

$$SC_{Total} = SC_A + SC_B + SC_{AB} + SC_{EE}$$

Así, las hipótesis a verificar son:

$$1) H_0 : (\alpha\beta)_{jk} = 0 \text{ contra } H_1 : (\alpha\beta)_{jk} \neq 0$$

$$2) H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de los tratamientos es diferente de cero

$$3) H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de las covariables es diferente de cero

Las estadísticas de prueba para corroborar estas hipótesis quedan expresadas como se muestran en la Tabla 28 del análisis de varianza respectivo, donde

Tabla 28: ANAVA para un experimento factorial 2×2 bajo un DCA

F.V.	gl	SC	CM	F_c	F_{tabla}
TTO	$ab - 1$	SC_{TTO}	CM_{TTO}	$\frac{CM_{TTO}}{CM_{EE}}$	$F_{\alpha, ab-1, ab(r-1)}$
A	$a - 1$	SC_A	CM_A	$\frac{CM_A}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
B	$b - 1$	SC_B	CM_B	$\frac{CM_B}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
AB	$(a - 1)(b - 1)$	SC_{AB}	CM_{AB}	$\frac{CM_{AB}}{CM_{EE}}$	$F_{\alpha, (a-1)(b-1), ab(r-1)}$
EE	$ab(r - 1)$	SC_{EE}	CM_{EE}		
Total	$abr - 1$	SC_{Total}			

$$SC_{Total} = \sum_{i=1}^r \sum_{j=1}^a \sum_{k=1}^b Y_{ijk}^2 - FC,$$

con $FC = \frac{Y_{...}^2}{rab}$ y $Y_{...} = \sum_{i=1}^r \sum_{j=1}^a \sum_{k=1}^b Y_{ijk}$,

$$SC_{TTO} = \sum_{j=1}^a \sum_{k=1}^b \frac{Y_{.jk}^2}{r} - FC,$$

$$SC_A = \sum_{j=1}^a \frac{Y_{.j}^2}{rb} - FC,$$

$$SC_B = \sum_{k=1}^b \frac{Y_{.k}^2}{ra} - FC,$$

$$SC_{AB} = SC_{TTO} - SC_A - SC_B,$$

$$SC_{EE} = SC_{Total} - SC_{TTO}$$

6.12.5. Errores estándar para medias de efectos principales e interacción

Los errores estándar para las medias de efectos principales vienen dados por:

1) Para medias de los niveles del factor A :

$$S_{\bar{Y}.j} = \sqrt{CM_{EE}/rb}$$

2) Para medias de los niveles del factor B :

$$S_{\bar{Y}.k} = \sqrt{CM_{EE}/ra}$$

3) Mientras que el error estándar para una media de la interacción viene dado por:

$$S_{\bar{Y}.jk} = \sqrt{CM_{EE}/r}$$

4) Error estándar de la diferencia entre dos medias de A :

$$S_{d_A} = \sqrt{\frac{2CM_{EE}}{rb}}$$

5) Error estándar de la diferencia entre dos medias de B :

$$S_{d_B} = \sqrt{\frac{2CM_{EE}}{ra}}$$

6) Error estándar de la diferencia entre dos medias de AB :

$$S_{d_{AB}} = \sqrt{\frac{2CM_{EE}}{r}}$$

1. La descomposición de la interacción:

Para cada factor se descompone la suma de cuadrado por cada nivel del otro factor.

Para el factor A en cada nivel de B definimos:

$$SCA : B_k = \frac{1}{r} \sum_{j=1}^a Y_{.jk}^2 - \frac{Y_{..k}^2}{ar}$$

$$CMA : B_k = \frac{SCA : B_k}{a - 1}$$

$$F_c = \frac{CMA : B_k}{CM_{EE}}$$

Tabla 29: ANAVA interacción de A dentro los niveles de B

$F.V.$	gl	SC	CM	F_c	F_{tabla}
TTO	$ab - 1$	SC_{TTO}	CM_{TTO}	$\frac{CM_{TTO}}{CM_{EE}}$	$F_{\alpha, ab-1, ab(r-1)}$
B	$b - 1$	SC_B	CM_B	$\frac{CM_B}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
$A : B_1$	$b - 1$	$SC_{A:B_1}$	$CM_{A:B_1}$	$\frac{CM_{A:B_1}}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
$A : B_2$	$b - 1$	$SC_{A:B_2}$	$CM_{A:B_2}$	$\frac{CM_{A:B_2}}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$A : B_b$	$b - 1$	$SC_{A:B_b}$	$CM_{A:B_b}$	$\frac{CM_{A:B_b}}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
EE	$ab(r - 1)$	SC_{EE}	CM_{EE}		

Ahora para el factor B en cada nivel de B definimos:

$$SCB : A_j = \frac{1}{r} \sum_{k=1}^B Y_{.jk}^2 - \frac{Y_{.j.}^2}{br}$$

$$CMB : A_j = \frac{SCB : A_j}{b - 1}$$

$$F_c = \frac{CMA : A_j}{CMEE}$$

Tabla 30: ANAVA interacción de B dentro los niveles de A

<i>F.V.</i>	<i>gl</i>	<i>SC</i>	<i>CM</i>	<i>F_c</i>	<i>F_{tabla}</i>
<i>TTO</i>	$ab - 1$	SC_{TTO}	CM_{TTO}	$\frac{CM_{TTO}}{CM_{EE}}$	$F_{\alpha, ab-1, ab(r-1)}$
<i>A</i>	$a - 1$	SC_A	CM_A	$\frac{CM_A}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
$B : A_1$	$a - 1$	$SC_{B:A_1}$	$CM_{B:A_1}$	$\frac{CM_{B:A_1}}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
$B : A_2$	$a - 1$	$SC_{B:A_2}$	$CM_{B:A_2}$	$\frac{CM_{B:A_2}}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$B : A_a$	$a - 1$	$SC_{B:A_a}$	$CM_{B:A_b}$	$\frac{CM_{B:A_a}}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
<i>EE</i>	$ab(r - 1)$	SC_{EE}	CM_{EE}		

2. Pruebas horizontales y verticales

Para analizar la interacción se pueden realizar pruebas verticales u horizontales. Estas consisten en analizar un factor a cada nivel del otro factor. Por ejemplo, podemos llevar a cabo pruebas verticales si analizamos las diferencias entre los niveles del factor A para cada nivel del factor B, con una cualesquiera de las pruebas de comparación múltiple estudiada anteriormente. En forma similar, si analizamos las diferencias entre los niveles del factor B a cada nivel del factor A, entonces estamos realizando pruebas horizontales.

Analicemos las diferencias del factor A cada nivel del factor B (pruebas verticales); las hipótesis para tener en cuenta son:

$$H_0 : \mu_j = \mu_{j'} \quad vs \quad H_0 : \mu_j \neq \mu_{j'}$$

donde $j, j' = 1, 2, 3, \dots, a$.

Analicemos las diferencias del factor B a cada nivel del factor A (pruebas horizontales); las hipótesis para tener en cuenta son:

$$H_0 : \mu_k = \mu_{k'} \quad vs \quad H_0 : \mu_k \neq \mu_{k'}$$

donde $k, k' = 1, 2, 3, \dots, b$.

6.12.6. Función (fat2.crd) en R

Esta información fue obtenida de la página web ([Fat2.Crd function - RDocumentation], s.f.) . Otra forma de obtenerla es utilizar el comando `help(fat2.crd)` en R.

Esta función lleva a cabo análisis de experimentos en Diseño Completamente Aleatorizado balanceado en esquema factorial doble, considerando un modelo fijo este se encuentra en la librería de R llamada (`ExpDes`).

Los autores de esta función son Eric B Ferreira, Denismar Alves Nogueira y Portya Piscitelli Cavalcanti.

- La función (`fat2.crd`) está compuesto por:

```
1 > fat2.crd(factor1, factor2, resp, quali = c(TRUE, TRUE),
2 mcomp = "tukey", fac.names = c("F1", "F2"), sigT = 0.05,
3 sigF = 0.05, unfold = NULL)
```

- Características

`data`: los datos son un formato de `data.frame`.

`factor1`: vector numérico o complejo que contiene los niveles del factor 1.

`factor2`: vector numérico o complejo que contiene los niveles del factor 2.

`resp`: vector numérico o complejo que contiene la variable de respuesta.

`quali`: lógica. Si es VERDADERO (predeterminado), los tratamientos se asumen cualitativos, si es FALSO, cuantitativos.

`mcomp`: permite elegir la prueba de comparación múltiple; por defecto es el test de Tukey, sin embargo, las opciones son: el test LSD (`'lsd'`), el test LSD con protección Bonferroni (`'lsdb'`), el test de Duncan (`'duncan'`), el test de Student- Newman-Keuls (`'snk'`), la prueba de Scott-Knott (`'sk'`), la prueba de Calinski y Corsten (`'ccF'`) y la prueba de comparaciones múltiples bootstrap (`'ceboot'`).

`fac.nombres`: permite etiquetar los factores 1 y 2.

`sigT`: la significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 %.

`sigF`: la significación que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5 %.

`unfold`: dice lo que se debe hacer después del ANOVA. Si es NULL (predeterminado), se realizan las pruebas recomendadas; si es '0', solo se realiza ANOVA; si '1', se prueban los efectos simples; si es '2', se despliega la doble interacción.

Esta función aporta un análisis de la varianza del referido CRD, la prueba de normalidad de Shapiro-Wilk para los residuales del modelo, los modelos de regresión ajustados (cuando los tratamientos son cuantitativos) y/o las pruebas de comparación múltiple (cuando los tratamientos son cualitativos).

6.12.7. Ejemplo en R para análisis para factores cualitativo × cuantitativo.

Se tienen tres métodos de elaboración de un producto (m1, m2 y m3) y dosis de un componente 0, 2, 4, 6, 8 y 10. El diseño experimental utilizado fue completamente al azar con 4 repeticiones. La información presentada en la Tabla 31 corresponde a la variable dependiente Y.

Parecido al proceso anterior, se debe tener en cuenta que esta vez se utiliza la librería `ExpDes`. Ahora se utiliza la función `fat2.crd`, las variables que se utilizan de acuerdo con este ejemplo son: `Dosis`, `Metodo` y `log(y1)`.

```
1 > fat2.crd( Dosis, Metodo, log(y1), quali = c(FALSE, TRUE),
2 mcomp = "tukey", fac.names = c("Dosis", "M todo"), sigT =
   0.05,
3 sigF = 0.05)
```

Se obtienen los siguientes resultados:

```
1 -----
2 Legend:
3 FACTOR 1:  Dosis
4 FACTOR 2:  Metodo
5 -----
```

Tabla 31: Métodos y dosis de un componente

Dosis	Método	rep 1	rep 2	rep 3	rep 4
0	m1	30.35	11.38	13.09	58.93
2	m1	29.79	16.5	16.83	18.37
4	m1	56.2	48.4	40.76	49.7
6	m1	23.36	28.7	20.95	23.87
8	m1	21.67	20.62	37.29	21.23
10	m1	35.4	45.9	24.37	21.06
0	m2	57.89	91.47	37.97	84.33
2	m2	66.54	80.9	28.29	83.58
4	m2	42.92	51.24	60.5	66.87
6	m2	27	35.7	34.6	22.74
8	m2	67.7	72.08	17.32	61.67
10	m2	26.99	27.9	10.50	13.86
0	m3	54.10	67.69	85.44	34.55
2	m3	40.79	68.9	93	57.8
4	m3	87.01	59.86	62.47	61.75
6	m3	40.50	73.73	99.87	98,49
8	m3	84.73	70.5	37.28	44.89
10	m3	42.36	37.37	38.54	40.36

■ Análisis de varianza

```

1 -----
2 Analysis of Variance Table
3 -----
4           DF      SS MS      Fc      Pr>Fc
5 Dosis           5  3.1012  4  4.0756  0.0032739
6 Metodo          2  7.1712  5 23.5607  0.0000000
7 Dosis*Metodo    10  4.6927  3  3.0836  0.0036060
8 Residuals       54  8.2180  2
9 Total           71 23.1831  1
10 -----
11 CV = 10.55 %

```

Los resultados indican que tanto el factor Dosis como el factor Método tienen un efecto significativo en la variable respuesta, al igual que su interacción. Esto se basa en los valores bajos de los p_{valor} asociados a los estadísticos F, que indican una evidencia de que al menos uno de los

niveles de los factores o su interacción influyen en la variable respuesta.

- Prueba de Shapiro-Wilk

```
1 -----
2 Shapiro-Wilk normality test
3 p-value: 0.06990718
4 According to Shapiro-Wilk normality test at 5% of
5   significance,
6 residuals can be considered normal.
6 -----
```

El p_{valor} resultante es 0.06990718. Según la prueba de normalidad de Shapiro-Wilk con un nivel de significancia del 5%, no tenemos suficiente evidencia para rechazar la hipótesis nula de que los residuos siguen una distribución normal. Por lo tanto, podemos considerar que los residuos se distribuyen aproximadamente de manera normal.

Descomposición de la interacción: en R este procedimiento se realiza de la siguiente manera.

- Análisis de varianza de la interacción de la dosis dentro de cada nivel de método.

```
1 Significant interaction: analyzing the interaction
2 -----
3 Analyzing Dosis inside of each level of Metodo
4 -----
5 Analysis of Variance Table
6 -----
7           DF      SS      MS      Fc  Pr.Fc
8 Metodo           2  7.17115  3.58558 23.5607    0
9 Dosis:Metodo 1    5  2.02370  0.40474  2.6595  0.032
10 Dosis:Metodo 2    5  4.85431  0.97086  6.3795 1e-04
11 Dosis:Metodo 3    5  0.91595  0.18319  1.2037 0.3199
12 Residuals       54  8.21796  0.15218
13 Total           71 23.18307
14 -----
```

La dosis influenciada es preferida apenas en el método 2. Lo cual explica la interacción.

- Ajuste de modelos polinómicos de regresión de la dosis dentro del nivel 1 de Método. A continuación, se presentan los modelos lineal, cuadrático y cubico, estos tienen su respectivo R^2 y análisis de varianza.

```

1 -----
2 Dosis  inside of the level  1  of  Metodo
3 -----
4 Adjustment of polynomial models of regression
5 -----

```

- Estimaciones del modelo lineal el R^2 y análisis de varianza.

```

1 Linear Model
2 =====
3 Estimate Standard.Error tc p.value
4 -----
5 b0  3.1978      0.1412      22.6523      0
6 b1  0.0194      0.0233       0.8334     0.4083
7 -----
8
9 R2 of linear model
10 -----
11 1
12 -----
13 0.052238
14 -----
15
16 Analysis of Variance of linear model
17 =====
18 DF SS MS Fc p.value
19 -----
20 Linear Effect 1 0.1057 0.1057 0.69 0.40826
21 Lack of fit 4 1.9180 0.4795 3.15 0.0212
22 Residuals 54 8.2180 0.1522
23 -----

```

El coeficiente de determinación (R^2) del modelo lineal es 0.052238, lo que sugiere que el modelo explica aproximadamente el 5.2% de la variabilidad de la respuesta medida. El análisis de varianza del modelo lineal muestra que el efecto lineal es no significativo ($p = 0,40826$). Sin embargo, se observa una falta de ajuste significativa ($p = 0,0212$), lo que indica que el modelo no ajusta perfectamente los datos observados.

- Estimaciones del modelo cuadrático el R^2 y análisis de varianza.

```

1 Quadratic Model
2 =====
3 Estimate Standard.Error tc p.value
4 -----

```

```

5 b0  3.0929      0.1768      17.4952      0
6 b1  0.0981      0.0831      1.1805      0.2430
7 b2 -0.0079      0.0080      -0.9863     0.3284
8 -----
9
10 R2 of quadratic model
11 -----
12      1
13 -----
14 0.125400
15 -----
16
17 Analysis of Variance of quadratic model
18 =====
19              DF      SS      MS      Fc      p.value
20 -----
21 Linear Effect      1    0.1057  0.1057  0.69  0.40826
22 Quadratic Effect  1    0.1481  0.1481  0.97  0.32836
23 Lack of fit       3    1.7699  0.5900  3.88  0.01392
24 Residuals        54    8.2180  0.1522
25 -----

```

El coeficiente de determinación (R^2) del modelo cuadrático es 0.125400, lo que sugiere que el modelo explica aproximadamente el 12.5 % de la variabilidad de la respuesta medida. El análisis de varianza del modelo cuadrático muestra que tanto el efecto lineal como el efecto cuadrático no son significativos ($p = 0,40826$ y $p = 0,32836$, respectivamente). Sin embargo, se observa una falta de ajuste significativa ($p = 0,01392$), lo que indica que el modelo no ajusta perfectamente los datos observados.

- Estimaciones del modelo cubico el R^2 y análisis de varianza.

```

1 Cubic Model
2 =====
3      Estimate Standard.Error      tc      p.value
4 -----
5 b0  3.0166      0.1911      15.7815      0
6 b1  0.2724      0.1856      1.4673      0.1481
7 b2 -0.0556      0.0461      -1.2046     0.2336
8 b3  0.0032      0.0030      1.0498      0.2985
9 -----
10
11 R2 of cubic model
12 -----
13      1

```

```

14 -----
15 0.208272
16 -----
17
18 Analysis of Variance of cubic model
19 =====
20              DF      SS      MS      Fc      p.value
21 -----
22 Linear Effect      1    0.1057  0.1057  0.69  0.40826
23 Quadratic Effect  1    0.1481  0.1481  0.97  0.32836
24 Cubic Effect      1    0.1677  0.1677  1.1   0.2985
25 Lack of fit      2    1.6022  0.8011  5.26  0.00815
26 Residuals       54    8.2180  0.1522
27 -----

```

El coeficiente de determinación (R^2) del modelo cúbico es 0.208272, lo que sugiere que el modelo explica aproximadamente el 20.8% de la variabilidad de la respuesta medida. El análisis de varianza del modelo cúbico muestra que tanto el efecto lineal, el efecto cuadrático y el efecto cúbico no son significativos ($p = 0,40826$, $p = 0,32836$ y $p = 0,2985$, respectivamente). Sin embargo, se observa una falta de ajuste significativa ($p = 0,00815$), lo que indica que el modelo no ajusta perfectamente los datos observados.

- Ajuste de modelos polinómicos de regresión de la dosis dentro del nivel 2 de método. A continuación, se presentan los modelos lineal, cuadrático y cúbico, estos tienen su respectivo R^2 y análisis de varianza.

```

1 -----
2 Dosis inside of the level 2 of Metodo
3 -----
4 Adjustment of polynomial models of regression
5 -----

```

- Estimaciones del modelo lineal el R^2 y análisis de varianza.

```

1 Linear Model
2 =====
3      Estimate Standard.Error      tc      p.value
4 -----
5 b0  4.2760      0.1412      30.2895      0
6 b1 -0.1084      0.0233      -4.6492  0.00002
7 -----
8

```

```

9 R2 of linear model
10 -----
11      2
12 -----
13 0.677636
14 -----
15
16 Analysis of Variance of linear model
17 =====
18              DF      SS      MS      Fc      p.value
19 -----
20 Linear Effect 1   3.2895  3.2895  21.61   2e-05
21 Lack of fit   4   1.5649  0.3912   2.57   0.04808
22 Residuals    54  8.2180  0.1522
23 -----

```

El coeficiente de determinación (R^2) del modelo lineal es 0.677636, lo que sugiere que el modelo explica aproximadamente el 67.8% de la variabilidad de la respuesta medida. El análisis de varianza del modelo lineal muestra que el efecto lineal es altamente significativo ($p = 2e - 05$). Sin embargo, se observa una falta de ajuste significativa ($p = 0,04808$), lo que indica que el modelo no ajusta perfectamente los datos observados.

- Estimaciones del modelo cuadrático el R^2 y análisis de varianza.

```

1 -----
2 Quadratic Model
3 =====
4      Estimate Standard.Error    tc    p.value
5 -----
6 b0  4.1460      0.1768      23.4525    0
7 b1 -0.0109      0.0831      -0.1312  0.8961
8 b2 -0.0097      0.0080      -1.2214  0.2272
9 -----
10
11 R2 of quadratic model
12 -----
13      2
14 -----
15 0.724404
16 -----
17
18 Analysis of Variance of quadratic model
19 =====

```

	DF	SS	MS	Fc	p.value
Linear Effect	1	3.2895	3.2895	21.61	2e-05
Quadratic Effect	1	0.2270	0.2270	1.49	0.22724
Lack of fit	3	1.3378	0.4459	2.93	0.04171
Residuals	54	8.2180	0.1522		

El coeficiente de determinación (R^2) del modelo cuadrático es 0.724404, lo que sugiere que el modelo explica aproximadamente el 72.4% de la variabilidad de la respuesta medida. El análisis de varianza del modelo cuadrático muestra que el efecto lineal es altamente significativo ($p = 2e - 05$), mientras que el efecto cuadrático no es significativo ($p = 0,22724$). Además, se observa una falta de ajuste significativa ($p = 0,04171$), lo que indica que el modelo no ajusta perfectamente los datos observados.

- Estimaciones del modelo cubico el R^2 y análisis de varianza.

```

1 -----
2 Cubic Model
3 =====
4 Estimate Standard.Error tc p.value
5 -----
6 b0 4.2093 0.1911 22.0213 0
7 b1 -0.1554 0.1856 -0.8368 0.4064
8 b2 0.0298 0.0461 0.6458 0.5212
9 b3 -0.0026 0.0030 -0.8702 0.3880
10 -----
11
12 R2 of cubic model
13 -----
14 2
15 -----
16 0.748146
17 -----
18 Analysis of Variance of cubic model
19 =====
20 DF SS MS Fc p.value
21 -----
22 Linear Effect 1 3.2895 3.2895 21.61 2e-05
23 Quadratic Effect 1 0.2270 0.2270 1.49 0.22724
24 Cubic Effect 1 0.1153 0.1153 0.76 0.38803
25 Lack of fit 2 1.2226 0.6113 4.02 0.02364

```

```
26 Residuals          54  8.2180  0.1522
```

```
27 -----
```

El coeficiente de determinación (R^2) del modelo cúbico es 0.748146, lo que sugiere que el modelo explica aproximadamente el 74.8% de la variabilidad de la respuesta medida. El análisis de varianza del modelo cúbico muestra que el efecto lineal es altamente significativo ($p = 2e - 05$), mientras que los efectos cuadrático y cúbico no son significativos ($p = 0,22724$ y $p = 0,38803$, respectivamente). Además, se observa una falta de ajuste significativa ($p = 0,02364$), lo que indica que el modelo no ajusta perfectamente los datos observados.

- Comparaciones de medias

```
1 -----
2           Levels      Means
3 1           0  4.048999
4 2          10  3.679152
5 3           2  4.132670
6 4           4  4.203953
7 5           6  4.298884
8 6           8  4.029439
9 -----
```

Estos resultados sugieren que hay diferencias significativas en las medias entre al menos algunos de los niveles. Sin embargo, es importante realizar pruebas estadísticas adecuadas, como análisis de varianza (ANOVA) o pruebas de comparación múltiple, para evaluar la significancia estadística de estas diferencias y determinar qué niveles difieren significativamente entre sí.

- Análisis del método dentro de cada nivel de Dosis

```
1 -----
2 Analyzing Metodo inside of each level of Dosis
3 -----
4 Analysis of Variance Table
5 -----
6           DF          SS          MS          Fc  Pr.Fc
7 Dosis          5  3.10122  0.62024  4.0756  0.0033
8 M todo:Dosis  0  2  2.59710  1.29855  8.5327  6e-04
9 M todo:Dosis  2  2  3.39909  1.69954  11.1677  1e-04
10 M todo:Dosis  4  2  0.21389  0.10695  0.7027  0.4997
11 M todo:Dosis  6  2  2.83694  1.41847  9.3207  3e-04
```

```

12 M todo:Dosis 8 2 1.56886 0.78443 5.1545 0.0089
13 M todo:Dosis 10 2 1.24802 0.62401 4.1003 0.022
14 Residuals 54 8.21796 0.15218
15 Total 71 23.18307

```

El factor Dosis es estadísticamente significativo, con un valor de F_c de 4.0756 y un p_{valor} de 0.0033. Esto indica que hay diferencias significativas en las medias entre al menos algunos de los niveles de Dosis. Los efectos de Método:Dosis también son estadísticamente significativos para algunos niveles de Dosis. Los niveles 0, 2, 6 y 8 de Dosis muestran diferencias significativas en las medias con respecto a Método, ya que tienen valores de F_c superiores a 4 y valores de p_{valor} inferiores a 0.01. Sin embargo, los niveles 4 y 10 de Dosis no muestran diferencias significativas en las medias con respecto a Método, ya que sus valores de F_c son bajos y sus valores de p_{valor} son más altos (0.4997 y 0.022, respectivamente)

El método influenciada es preferida en la dosis 0, 2, 6,8 y 10.

- Método dentro del nivel 2 de Dosis

```

1 -----
2 Metodo inside of the level 2 of Dosis
3 -----
4 Tukey's test
5 -----
6 Groups Treatments Means
7 a 3 4.13267
8 a 2 4.089832
9 b 1 2.982854
10 -----

```

Según la prueba de Tukey, los grupos 3 y 2 tienen medias que no difieren significativamente entre sí, ya que comparten la misma letra *a*. Sin embargo, el grupo 1 tiene una media significativamente más baja que los grupos 3 y 2, ya que se le asigna la letra *b*. Esto indica que dentro del nivel 2 de Dosis, hay diferencias significativas en las medias entre el grupo 1 y los grupos 3 y 2, pero no hay una diferencia significativa entre los grupos 3 y 2.

- Comparaciones de medias

```

1 -----

```

```

2      Levels      Means
3 1          1  3.880531
4 2          2  4.000313
5 3          3  4.203953
6 -----

```

Según estos resultados, se puede observar que el nivel 3 tiene la media más alta, seguido por el nivel 2 y luego el nivel 1. Estas diferencias en las medias sugieren que puede haber variaciones significativas en los valores promedio entre los diferentes niveles.

- Método dentro del nivel 6 de dosis

```

1 -----
2 Metodo  inside of the level  6  of  Dosis
3 -----
4 Tukey's test
5 -----
6 Groups Treatments Means
7 a          3    4.298884
8 b          2    3.384742
9 b          1    3.180671

```

Según estos resultados, se puede concluir que el tratamiento 3, con una media de 4.298884, es significativamente diferente de los tratamientos 1 y 2, que tienen medias de 3.180671 y 3.384742 respectivamente. Sin embargo, no hay una diferencia significativa entre los tratamientos 1 y 2, ya que ambos pertenecen al grupo *b*. Estas conclusiones se basan en el análisis de Tukey, que compara todas las combinaciones de tratamientos y determina si existen diferencias significativas entre ellos.

- Método dentro del nivel 8 de dosis

```

1 -----
2 Metodo  inside of the level  8  of  Dosis
3 -----
4 Tukey's test
5 -----
6 Groups Treatments Means
7 a      3    4.029439
8 a      2    3.866631
9 b      1    3.194083
10 -----

```

Según estos resultados, se puede concluir que los tratamientos 3 y 2, con medias de 4.029439 y 3.866631 respectivamente, son significativamente diferentes del tratamiento 1, que tiene una media de 3.194083. Sin embargo, no hay una diferencia significativa entre los tratamientos 3 y 2, ya que ambos pertenecen al grupo *a*. Estas conclusiones se basan en el análisis de Tukey, que compara todas las combinaciones de tratamientos y determina si existen diferencias significativas entre ellos.

- Método dentro del nivel 10 de dosis

```
1 -----
2 Metodo inside of the level 10 of Dosis
3 -----
4 Tukey's test
5 -----
6 Groups Treatments Means
7 a          3      3.679152
8 ab         1      3.408476
9 b          2      2.901119
10 -----
```

Según estos resultados, se puede concluir que el tratamiento 3, con una media de 3.679152, es significativamente diferente del tratamiento 2, que tiene una media de 2.901119. Además, el tratamiento 1, con una media de 3.408476, muestra una diferencia intermedia y se encuentra entre los grupos *a* y *b*. Estas conclusiones se basan en el análisis de Tukey, que compara todas las combinaciones de tratamientos y determina si existen diferencias significativas entre ellos.

Un aporte adicional de la anterior función consiste en llevar a cabo un análisis de varianza sobre la descomposición de las interacciones, a continuación, se observa la función (ea2) pero esta no lleva cabo el proceso de la descomposición de las interacciones.

6.12.8. Función (ea2) en R

Esta información fue obtenida de la página web (easynova, s.f.). Otra forma de obtenerla es utilizar el comando `help(ea2)` en R.

El autor de la función (`ea2`) Emmanuel Arnhold, el cual lleva cabo un análisis de la varianza en parcelas factoriales y divididas este se encuentra en la librería de R llamada (`easynova`).

- La función (`ea2`) está compuesta por:

```
1 > ea2(data, design = 1, alpha = 0.05, cov = 4, list = FALSE,  
2 p.adjust=1, plot=2)
```

- Características

`data`: los datos son un `data.frame`.

`design`:

- 1 = doble factorial en diseño completamente aleatorio
- 2 = doble factorial en diseño de bloques aleatorios
- 3 = doble factorial en diseño de cuadrados latinos
- 4 = parcela dividida en diseño completamente aleatorio
- 5 = parcela dividida en diseño de bloques aleatorios
- 6 = parcela dividida en diseño de cuadrado latino
- 7 = triple factorial en un diseño completamente aleatorizado
- 8 = factorial triple en diseño de bloques aleatorizados
- 9 = factorial doble en diseño de parcela dividida (completamente aleatorizado)
- 10 = doble factorial en parcela dividida (bloque aleatorio)
- 11 = análisis agrupado de experimentos con bloques jerárquicos
- 12 = análisis agrupado de réplicas de cuadrados latinos (filas jerárquicas)
- 13 = análisis combinado de réplicas de cuadrados latinos (filas y columnas jerárquicas)

`cov`: para diseños de parcela dividida

- 1 = Autorregresivo
- 2 = Autorregresivo heterogéneo
- 3 = Proceso Autorregresivo Continuo
- 4 = Simetría compuesta
- 5 = No estructurado

`alpha`: nivel de significación para las comparaciones múltiples.

`list`: FALSE = una sola variable de respuesta TRUE = respuesta multivariable.

`p.adjust`: 1="none"; 2="holm"; 3="hochberg"; 4="hommel"; 5="bonferroni"; 6="BH"; 7="BY"; 8="fdr".

`plot`: 1 = gráfico de caja para los residuos; 2 = residuos estandarizados frente a los datos de la secuencia; 3 = residuos estandarizados frente a cuantiles teóricos.

Esta función aporta un análisis de la varianza, las medias (medias ajustadas), la prueba de comparación múltiple (tukey, snk, duncan t y scott knott) y análisis de residuos.

6.12.9. Ejemplo en R para un experimento factorial con un DCA.

Viloria *et al.*, (2004) llevaron a cabo un experimento bajo un diseño completamente al azar, con 4 repeticiones, para evaluar el efecto de bioherbicidas de 4 extractos de pasto Jonson y mentolada sobre el crecimiento de 4 cultivos de ciclo corto. Los extractos y cultivos utilizados son mostrados en la Tabla 32.

Tabla 32: Factores y niveles estudiados

Cultivos	Extractos
Algodón(<i>c1</i>)	Mentolada (<i>e1</i>)
Arroz (<i>c2</i>)	Pasto (<i>e2</i>)
Frijol (<i>c3</i>)	Testigo(agua) (<i>e3</i>)
Maíz (<i>c4</i>)	Tween (<i>e4</i>)

Los resultados finales de altura de planta se muestran en la Tabla 33. Así, las hipótesis a verificar son:

$$1) H_0 : (\alpha\beta)_{jk} = 0 \text{ contra } H_1 : (\alpha\beta)_{jk} \neq 0$$

$$2) H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_j = 0$$

Tabla 33: Datos para altura de plantas

Cultivo(C)	extracto(E)	rep 1	rep 2	rep 3	rep 4
c1	e1	10	9	8,6	10
c1	e2	8	7,8	4,5	7,2
c1	e3	18	24,8	21,5	20,5
c1	e4	9,7	11	9	18,2
c2	e1	9	12,4	10,4	13
c2	e2	0	7,5	5,7	6
c2	e3	10,5	17,5	18	0
c2	e4	0	9	11	0
c3	e1	7,6	8	11,6	9
c3	e2	6	6	8	7,6
c3	e3	19,5	18	18,8	21,2
c3	e4	11	6	11,1	9
c4	e1	10	12	9,8	9
c4	e2	5,4	4,6	7,5	6,4
c4	e3	23	19	22	20
c4	e4	8,5	8,5	10	10,5

contra

H_1 : Al menos uno de los coeficientes de los tratamientos es diferente de cero

$$3) H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de las covariables es diferente de cero

Para llevar a cabo este ejemplo en R , se requiere la librería (**easyanova**). Los pasos a seguir son:

1. Se observan los primeros 6 valores de la base de datos llamada **base_DCA**. Se debe tener en cuenta que el factor **factor_1**, que representa el cultivo, ha sido renombrado de c1, c2, c3 y c4 a 1, 2, 3 y 4 respectivamente. Asimismo, el factor **factor_2**, que representa el extracto, también ha sido renombrado de e1, e2, e3 y e4 a 1, 2, 3 y 4 respectivamente.

```

1 > base_DCA=read.csv2(file.choose(),sep=";",dec=",");base_
  DCA
2 > head(base_DCA)
3   factor_1 factor_2 Rep    Y
4 1         1         1   1 10.0
5 2         1         1   2   9.0
6 3         1         1   3   8.6
7 4         1         1   4 10.0
8 5         1         2   1   8.0
9 6         1         2   2   7.8

```

2. Utilizando la función (ea2) se lleva a cabo un análisis factorial. Los argumentos que se utilizan en esta función de acuerdo con este ejemplo son: la base de datos llamada `base_DCA`, `design=1` ya que de acuerdo a la descripción previamente dada significa que el análisis que se va a realizar es de un DCA, `alpha = 0.05` es el nivel de significación para las comparaciones múltiples, `cov = 4` significa simetría compuesta y `list = FALSE` representa una sola variable respuesta.

```

1 > ea2(base_DCA[-3], design=1,alpha = 0.05, cov = 4, list
  = FALSE)

```

Se obtienen los siguientes resultados:

- Análisis de varianza general

```

1 $'Analysis of variance'
2           df type III SS mean square F value p>F
3 factor_1      3  166.7125  55.5708 5.3174 0.003
4 factor_2      3 1308.9188 436.3063 41.7489 <0.001
5 factor_1:factor_2  9  208.4912  23.1657 2.2167 0.037
6 residuals    48  501.6350  10.4507    -    -

```

El $p_{valor} = 0,037$ que aporta la interacción de los factores 1 y 2 es inferior al compararlo con un $\alpha = 0,05$, se concluye que hay diferencia significativa, es decir, que existe una interacción significativa entre los factores, cuando un factor está al nivel del otro el comportamiento es diferente cuando cambia el nivel.

- Analizando significancia de interacción - A dentro de niveles de B.

1. A dentro del nivel b_1 :

$$SCA : B_1 = \frac{1}{4}(37,6^2 + 44,8^2 + 36^2 + 40,8^2) - \frac{159,4^2}{16} = 10,9475$$

2. A dentro del nivel b_2 :

$$SCA : B_2 = \frac{1}{4}(27,5^2 + 19,2^2 + 27,6^2 + 23,9^2) - \frac{98,2^2}{16} = 11,7625$$

3. A dentro del nivel b_3 :

$$SCA : B_3 = \frac{1}{4}(84,3^2 + 46^2 + 77,5^2 + 84^2) - \frac{292,3^2}{16} = 252,3669$$

4. A dentro del nivel b_4 :

$$SCA : B_4 = \frac{1}{4}(47,9^2 + 20^2 + 37,1^2 + 37,5^2) - \frac{142,5^2}{16} = 100,1269$$

Entonces el ANAVA parcial respectivo es:

Tabla 34: ANAVA Parcial de A dentro los niveles de B

F.V.	gl	SC	CM	Fc
TTO	15	1684.12	112.27	10.74**
B	3	1308.91	436.30	41.75**
A : B1	3	10.9475	3.6491	0.3492NS
A : B2	3	11.7625	3.9208	0.3751NS
A : B3	3	252.3669	84.1223	8.0499**
A : B4	3	100.1269	33.3756	3.1938*
EE	48	501.63	10.45	

Se concluye que para los niveles B1 y B2 no hay diferencia significativa entre los niveles de A, en tanto que para los niveles B3 y B4 existen diferencia significativa entre los niveles de A, esto explica la interacción de A con B.

- Analizando significancia de interacción - B dentro de niveles de A.

1. B dentro del nivel a_1 :

$$SCB : A_1 = \frac{1}{4}(37,6^2 + 27,5^2 + 84,8^2 + 47,9^2) - \frac{197,8^2}{16} = 468,5625$$

2. B dentro del nivel a_2 :

$$SCB : A_2 = \frac{1}{4}(44,8^2 + 19,2^2 + 46^2 + 20^2) - \frac{130^2}{16} = 166,67$$

3. B dentro del nivel a_3 :

$$SCB : A_3 = \frac{1}{4}(36^2 + 27,6^2 + 77,5^2 + 37,1^2) - \frac{178,4^2}{16} = 374,555$$

4. B dentro del nivel a_4 :

$$SCB : A_4 = \frac{1}{4}(40,8^2 + 23,9^2 + 84^2 + 37,5^2) - \frac{186,2^2}{16} = 507,6225$$

Entonces el ANAVA parcial respectivo es:

Tabla 35: ANAVA Parcial de B dentro los niveles de A

F.V.	gl	SC	CM	Fc
TTO	15	1684.12	112.27	10.74**
A	3	166.71	55.57	5.32**
B : A1	3	468.5625	156.187	14.9461*
B : A2	3	166.67	55.5566	5.3164*
B : A3	3	374.555	124.8518	11.9475*
B : A4	3	507.6225	169.2075	16.1921*
EE	48	501.63	10.45	

Se concluye que para todos los niveles del factor A hay diferencia significativa entre los niveles de B, esto explica la interacción de B con A.

Los errores estándar para las medias de tratamiento son:

$$\text{Para A: } S_{\bar{Y}.j} = \sqrt{CM_{EE}/rb} = \sqrt{\frac{10,46}{16}} = 0,8085$$

$$\text{Para B: } S_{\bar{Y}..k} = \sqrt{CM_{EE}/ra} = \sqrt{\frac{10,46}{16}} = 0,8085$$

y para una media de la interacción $A \times B$

$$S_{\bar{Y}.jk} = \sqrt{CM_{EE}/r} = \sqrt{\frac{10,46}{4}} = 1,6170$$

El siguiente proceso tiene sentido cuando la interacción no es significativa.

- Ajuste de medias y la comparación múltiple para el factor 1.

```

1 $'Adjusted means (factor 1)'
2   factor_1 adjusted.mean standard.error tukey snk duncan
3 1          1         12.3625         0.8082    a   a     a
4 2          4         11.6375         0.8082    a   a     a
5 3          3         11.1500         0.8082   ab   a     a
6 4          2          8.1250         0.8082    b   b     b
7 t   scott_knott
8 a     a
9 a     a
10 a    a
11 a    b
12
13
14 $'Multiple comparison test (factor 1)'
15   pair contrast p(tukey) p(snk) p(duncan)  p(t)
16 1 1 - 4      0.7250   0.9204 0.5289    0.5289 0.5289
17 2 1 - 3      1.2125   0.7147 0.5427    0.3238 0.2941
18 3 1 - 2      4.2375   0.0030 0.0030    0.0010 0.0005
19 4 4 - 3      0.4875   0.9737 0.6716    0.6716 0.6716
20 5 4 - 2      3.5125   0.0177 0.0096    0.0048 0.0035
21 6 3 - 2      3.0250   0.0518 0.0110    0.0110 0.0110

```

De acuerdo con el análisis anterior donde se tienen en cuenta los tratamientos del factor 1, la prueba tukey, encuentra diferencias significativas ente los tratamientos 1 y 2, para las pruebas snk, duncan, t y scott_knott los tratamientos 1, 3 y 4 no presentan diferencias significativas entre ellos.

■ Ajuste de medias y la comparación múltiple para el factor 2.

```

1 $'Adjusted means (factor 2)'
2   factor_2 adjusted.mean standard.error tukey snk duncan
3 1          3         18.2688         0.8082    a   a     a
4 2          1          9.9625         0.8082    b   b     b
5 3          4          8.9062         0.8082   bc   b     b
6 4          2          6.1375         0.8082    c   c     c
7 t   scott_knott
8 a     a
9 b     b
10 b    b
11 c     c
12
13 $'Multiple comparison test (factor 2)'
14   pair contrast p(tukey) p(snk) p(duncan)  p(t)
15 1 3 - 1      8.3063   0.0000 0.0000    0.0000 0.0000

```

16	2	3	-	4	9.3626	0.0000	0.0000	0.0000	0.0000
17	3	3	-	2	12.1313	0.0000	0.0000	0.0000	0.0000
18	4	1	-	4	1.0563	0.7921	0.3600	0.3600	0.3600
19	5	1	-	2	3.8250	0.0084	0.0045	0.0022	0.0016
20	6	4	-	2	2.7687	0.0864	0.0192	0.0192	0.0192

De acuerdo con el análisis anterior donde se tienen en cuenta los tratamientos del factor 2, la prueba tukey, encuentra diferencias significativas ente los tratamientos 3 y 2, para las pruebas snk, duncan, t y scott_knott los tratamientos 2 y 3 no presentan diferencias significativas entre ellos.

Analicemos las diferencias del factor A a cada nivel del factor B (pruebas verticales); en este caso estaríamos corroborando a cada nivel de B (b_1, b_2, b_3, b_4) la hipótesis nula.

$$H_0 : \mu_j = \mu_{j'} \quad vs \quad H_0 : \mu_j \neq \mu_{j'} \quad \text{donde } j, j' = 1, 2, 3, 4$$

- Significancia de interacción del factor 1 dentro de niveles del factor 2 (pruebas verticales).
- Ajuste de medias del factor 1 dentro del nivel 1 del factor 2.

```

1 $'Adjusted means (factor 1 in levels of factor 2)'$'
   factor_1 in 1'
2   treatment adjusted.mean standard.error tukey snk duncan
3 2           2.1           11.20           1.6164   a   a   a
4 4           4.1           10.20           1.6164   a   a   a
5 1           1.1            9.40           1.6164   a   a   a
6 3           3.1            9.05           1.6164   a   a   a
7 t scott_knott
8 a   a
9 a   a
10 a  a
11 a  a

```

- Ajuste de medias del factor 1 dentro del nivel 2 del factor 2.

```

1 $'Adjusted means (factor 1 in levels of factor 2)'$'
   factor_1 in 2'
2   treatment adjusted.mean standard.error tukey snk duncan
3 7           3.2            6.900           1.6164   a   a   a
4 5           1.2            6.875           1.6164   a   a   a

```

```

5 8      4.2      5.975      1.6164      a      a      a
6 6      2.2      4.800      1.6164      a      a      a
7 t scott_knott
8 a      a
9 a      a
10 a     a
11 a     a

```

- Ajuste de medias del factor 1 dentro del nivel 3 del factor 2.

```

1 $'Adjusted means (factor 1 in levels of factor 2) '$'
  factor_1 in 3'
2   treatment adjusted.mean standard.error tukey snk
3 9          1.3          21.200          1.6164      a      a
4 12          4.3          21.000          1.6164      a      a
5 11          3.3          19.375          1.6164      a      a
6 10          2.3          11.500          1.6164      b      b
7 duncan t scott_knott
8 a      a      a
9 a      a      a
10 a     a      a
11 b     b      b

```

- Ajuste de medias del factor 1 dentro del nivel 4 del factor 2.

```

1 $'Adjusted means (factor 1 in levels of factor 2) '$'
  factor_1 in 4'
2   treatment adjusted.mean standard.error tukey snk
3 13          1.4          11.975          1.6164      a      a
4 16          4.4          9.375          1.6164      ab     ab
5 15          3.4          9.275          1.6164      ab     ab
6 14          2.4          5.000          1.6164      b      b
7 duncan t scott_knott
8 a      a      a
9 ab     ab     a
10 ab    ab     a
11 b     b      b

```

- Comparaciones múltiples del factor 1 dentro del nivel 1 del factor 2.

```

1 $'Multiple comparison test (factor 1 in levels of factor
  2)'
2 $'factor_1 in 1'
3   pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 2.1 - 4.1      1.00    0.9717 0.6637      0.6637 0.6637
5 2 2.1 - 1.1      1.80    0.8598 0.7125      0.4638 0.4349

```

6	3	2.1	-	3.1	2.15	0.7832	0.7832	0.3993	0.3517
7	4	4.1	-	1.1	0.80	0.9851	0.7279	0.7279	0.7279
8	5	4.1	-	3.1	1.15	0.9580	0.8702	0.6397	0.6172
9	6	1.1	-	3.1	0.35	0.9987	0.8790	0.8790	0.8790

- Comparaciones múltiples del factor 1 dentro del nivel 2 del factor 2.

```

1 $'Multiple comparison test (factor 1 in levels of factor
2 2)'
3 $'factor_1 in 2'
4 pair contrast p(tukey) p(snk) p(duncan) p(t)
5 1 3.2 - 1.2 0.025 1.0000 0.9913 0.9913 0.9913
6 2 3.2 - 4.2 0.925 0.9773 0.9139 0.7065 0.6875
7 3 3.2 - 2.2 2.100 0.7950 0.7950 0.4104 0.3629
8 4 1.2 - 4.2 0.900 0.9791 0.6955 0.6955 0.6955
9 5 1.2 - 2.2 2.075 0.8008 0.6381 0.3984 0.3686
10 6 4.2 - 2.2 1.175 0.9553 0.6096 0.6096 0.6096

```

- Comparaciones múltiples del factor 1 dentro del nivel 3 del factor 2.

```

1 $'Multiple comparison test (factor 1 in levels of factor
2 2)'
3 $'factor_1 in 3'
4 pair contrast p(tukey) p(snk) p(duncan) p(t)
5 1 1.3 - 4.3 0.200 0.9998 0.9306 0.9306 0.9306
6 2 1.3 - 3.3 1.825 0.8548 0.7058 0.4576 0.4286
7 3 1.3 - 2.3 9.700 0.0006 0.0006 0.0002 0.0001
8 4 4.3 - 3.3 1.625 0.8923 0.4806 0.4806 0.4806
9 5 4.3 - 2.3 9.500 0.0007 0.0004 0.0002 0.0001
10 6 3.3 - 2.3 7.875 0.0064 0.0012 0.0012 0.0012

```

- Comparaciones múltiples del factor 1 dentro del nivel 4 del factor 2.

```

1 $'Multiple comparison test (factor 1 in levels of factor
2 2)'
3 $'factor_1 in 4'
4 pair contrast p(tukey) p(snk) p(duncan) p(t)
5 1 1.4 - 4.4 2.600 0.6685 0.2610 0.2610 0.2610
6 2 1.4 - 3.4 2.700 0.6415 0.4701 0.2720 0.2434
7 3 1.4 - 2.4 6.975 0.0188 0.0188 0.0063 0.0037
8 4 4.4 - 3.4 0.100 1.0000 0.9653 0.9653 0.9653
9 5 4.4 - 2.4 4.375 0.2359 0.1458 0.0758 0.0616
10 6 3.4 - 2.4 4.275 0.2544 0.0676 0.0676 0.0676

```

De acuerdo con el análisis anterior donde se analiza la interacción del factor 1 dentro de cada nivel del factor 2 se observa que: para los niveles

1 y 2 del factor 2 no se encuentran diferencias, en cambio para los niveles 3 y 4 si se observan diferencias entre los promedios.

Analicemos las diferencias del factor B a cada nivel del factor A (pruebas horizontales) en este caso estaríamos corroborando a cada nivel de A (a_1, a_2, a_3, a_4) la hipótesis nula.

$$H_0 : \mu_k = \mu_{k'} \quad vs \quad H_0 : \mu_k \neq \mu_{k'} \quad \text{donde } k, k' = 1, 2, 3, 4$$

- Significancia de interacción del factor 2 dentro de niveles del factor 1 (pruebas horizontales).
- Ajuste de medias del factor 2 dentro del nivel 1 del factor 1.

```

1 $'Adjusted means (factor 2 in levels of factor 1) '$'
2   factor_2 in 1'
3   treatment adjusted.mean standard.error tukey snk
4 9          1.3          21.200          1.6164    a    a
5 13          1.4          11.975          1.6164    b    b
6 1           1.1           9.400          1.6164    b    b
7 5           1.2           6.875          1.6164    b    b
8 duncan t      scott_knott
9 a         a         a
10 b        b        b
11 bc       bc       b
12 c        c        b

```

- Ajuste de medias del factor 2 dentro del nivel 2 del factor 1.

```

1 $'Adjusted means (factor 2 in levels of factor 1) '$'
2 $'factor_2 in 2'
3   treatment adjusted.mean standard.error tukey snk
4 10          2.3          11.5          1.6164    a    a
5 2           2.1          11.2          1.6164    a    a
6 14          2.4           5.0          1.6164    b    b
7 6           2.2           4.8          1.6164    b    b
8 duncan t      scott_knott
9 a         a         a
10 a        a         a
11 b        b         b
12 b        b         b

```

- Ajuste de medias del factor 2 dentro del nivel 3 del factor 1.

```

1 $'Adjusted means (factor 2 in levels of factor 1)'
2 $'factor_2 in 3'
3   treatment adjusted.mean standard.error tukey snk
4 11      3.3      19.375      1.6164      a   a
5 15      3.4       9.275      1.6164      b   b
6  3      3.1       9.050      1.6164      b   b
7  7      3.2       6.900      1.6164      b   b
8 duncan t scott_knott
9 a      a   a
10 b     b   b
11 b     b   b
12 b     b   b

```

- Ajuste de medias del factor 2 dentro del nivel 4 del factor 1.

```

1 $'Adjusted means (factor 2 in levels of factor 1)'
2 $'factor_2 in 4'
3   treatment adjusted.mean standard.error tukey snk
4 12      4.3      21.000      1.6164      a   a
5  4      4.1      10.200      1.6164      b   b
6 16      4.4       9.375      1.6164      b   b
7  8      4.2       5.975      1.6164      b   b
8 duncan t scott_knott
9 a      a   a
10 b     b   b
11 b     b   b
12 b     b   b

```

- Comparaciones múltiples del factor 2 dentro del nivel 1 del factor 1.

```

1 $'Multiple comparison test (factor 2 in levels of factor
2   1)'
3 $'factor_2 in 1'
4   pair contrast p(tukey) p(snk) p(duncan) p(t)
5 1 1.3 - 1.4    9.225    0.0011 0.0002    0.0002 0.0002
6 2 1.3 - 1.1   11.800    0.0000 0.0000    0.0000 0.0000
7 3 1.3 - 1.2   14.325    0.0000 0.0000    0.0000 0.0000
8 4 1.4 - 1.1    2.575    0.6752 0.2656    0.2656 0.2656
9 5 1.4 - 1.2    5.100    0.1295 0.0761    0.0388 0.0304
10 6 1.1 - 1.2    2.525    0.6885 0.2748    0.2748 0.2748

```

- Comparaciones múltiples del factor 2 dentro del nivel 2 del factor 1.

```

1 $'Multiple comparison test (factor 2 in levels of factor
2   1)'
3 $'factor_2 in 2'

```

	pair	contrast	p(tukey)	p(snk)	p(duncan)	p(t)
1	2.3 - 2.1	0.3	0.9992	0.8961	0.8961	0.8961
2	2.3 - 2.4	6.5	0.0321	0.0176	0.0089	0.0065
3	2.3 - 2.2	6.7	0.0257	0.0257	0.0086	0.0052
4	2.1 - 2.4	6.2	0.0443	0.0092	0.0092	0.0092
5	2.1 - 2.2	6.4	0.0357	0.0197	0.0099	0.0073
6	2.4 - 2.2	0.2	0.9998	0.9306	0.9306	0.9306

- Comparaciones múltiples del factor 2 dentro del nivel 3 del factor 1.

```

1 $'Multiple comparison test (factor 2 in levels of factor
  1) '
2 $'factor_2 in 3'
3      pair contrast p(tukey) p(snk) p(duncan)  p(t)
4 1 3.3 - 3.4   10.100   0.0003 0.0001    0.0001 0.0001
5 2 3.3 - 3.1   10.325   0.0002 0.0001    0.0001 0.0000
6 3 3.3 - 3.2   12.475   0.0000 0.0000    0.0000 0.0000
7 4 3.4 - 3.1    0.225   0.9997 0.9220    0.9220 0.9220
8 5 3.4 - 3.2    2.375   0.7276 0.5562    0.3339 0.3040
9 6 3.1 - 3.2    2.150   0.7832 0.3517    0.3517 0.3517

```

- Comparaciones múltiples del factor 2 dentro del nivel 4 del factor 1.

```

1 $'Multiple comparison test (factor 2 in levels of factor
  1) '
2 $'factor_2 in 4'
3      pair contrast p(tukey) p(snk) p(duncan)  p(t)
4 1 4.3 - 4.1   10.800   0.0001 0.0000    0.0000 0.0000
5 2 4.3 - 4.4   11.625   0.0000 0.0000    0.0000 0.0000
6 3 4.3 - 4.2   15.025   0.0000 0.0000    0.0000 0.0000
7 4 4.1 - 4.4    0.825   0.9837 0.7198    0.7198 0.7198
8 5 4.1 - 4.2    4.225   0.2640 0.1651    0.0862 0.0707
9 6 4.4 - 4.2    3.400   0.4529 0.1435    0.1435 0.1435

```

De acuerdo con el análisis anterior donde se analizan la interacción del factor 2 dentro de cada nivel del factor 1 se observa que hay diferencias entre los promedios dentro de cada interacción.

- Análisis residuales.

```

1 $'Residual analysis '
2 $'Residual analysis '$'residual analysis '
3      values
4 p.value Shapiro-Wilk test      0.0004
5 p.value Bartlett test (factor_1) 0.0000

```

```

6 p.value Bartlett test (factor_2)      0.0001
7 p.value Bartlett test (treatments)   0.0004
8 coefficient of variation (%)          29.8800
9 first value most discrepant          28.0000
10 second value most discrepant        27.0000
11 third value most discrepant         16.0000

```

Los errores no se distribuyen de forma normal. La prueba de Bartlett muestra que para los factores 1 y 2 presentan varianzas homogéneas, de acuerdo a estos resultados se considera hacer una transformación, este proceso queda recomendado para el lector, la desviación estándar es el 29.8% del tamaño de la media.

6.12.10. Experimento factorial bajo un DBCA

En el caso de un experimento factorial con dos factores A y B , ambos de efectos fijos, con a y b niveles, respectivamente. El modelo estadístico para este caso es. (Martínez y Robles, 2013, p.170)

$$Y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk} \quad (26)$$

con $i = 1, 2, 3, \dots, r$, $j = 1, 2, 3, \dots, a$ y $k = 1, 2, 3, \dots, b$, donde

- Y_{ijk} : es la respuesta en la i – *ésima* unidad experimental para el j – *ésimo* nivel del factor A y el k – *ésimo* nivel del factor B .
- μ : es la media general común a todos los tratamientos.
- ρ_i : es el efecto del bloque i – *ésimo*.
- α_j : es el efecto principal del j – *ésimo* nivel del factor A .
- β_k : es el efecto principal del k – *ésimo* nivel del factor B .
- $(\alpha\beta)_{jk}$: es el efecto de interacción del j – *ésimo* nivel del factor A con el k – *ésimo* nivel del factor B .
- ϵ_{ijk} : es el error aleatorio en la i – *ésima* unidad experimental para el j – *ésimo* nivel del factor A y el k – *ésimo* nivel del factor B .

Los supuestos para los tratamientos son:

$$\sum_{j=1}^a \alpha_j = \sum_{k=1}^b \beta_k = \sum_{j=1}^a \sum_{k=1}^b (\alpha\beta)_{jk} = 0$$

Nuevamente se supone que $\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$.

El arreglo de la información, al igual que los totales por tratamiento se muestran en las Tablas 36 y 37 respectivamente.

Tabla 36: Arreglo de la información para un factorial 2^2 bajo DBCA

TTO	Bloques					Total de TTO
a_1b_1	Y_{111}	Y_{211}	Y_{311}	\cdots	Y_{r11}	$Y_{\cdot 11}$
a_1b_2	Y_{121}	Y_{221}	Y_{321}	\cdots	Y_{r21}	$Y_{\cdot 21}$
a_2b_1	Y_{112}	Y_{212}	Y_{312}	\cdots	Y_{r12}	$Y_{\cdot 12}$
a_2b_2	Y_{122}	Y_{222}	Y_{322}	\cdots	Y_{r22}	$Y_{\cdot 22}$
Total de bloques	$Y_{1\cdot}$	$Y_{2\cdot}$	$Y_{3\cdot}$	\cdots	$Y_{r\cdot}$	$Y_{\cdot\cdot}$

Tabla 37: Totales para los factores A y B en un DBCA

Niveles	Factor B				
Factor A	1	2	\cdots	b	Totales de A ($Y_{\cdot j}$)
1	$Y_{\cdot 11}$	$Y_{\cdot 12}$	\cdots	$Y_{\cdot 1b}$	$Y_{\cdot 1}$
2	$Y_{\cdot 21}$	$Y_{\cdot 22}$	\cdots	$Y_{\cdot 2b}$	$Y_{\cdot 2}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a	$Y_{\cdot a1}$	$Y_{\cdot a2}$	\cdots	$Y_{\cdot ab}$	$Y_{\cdot a}$
Totales de B ($Y_{\cdot k}$)	$Y_{\cdot 1}$	$Y_{\cdot 2}$	\cdots	$Y_{\cdot b}$	$Y_{\cdot\cdot}$

Tabla 38: ANAVA para un factorial $a \times b$ en un DBCA

F.V.	gl	SC	CM	F_c	F_{tabla}
Bloques	$r - 1$	$SC_{Bloques}$	$CM_{Bloques}$	$\frac{CM_{Bloques}}{CM_{EE}}$	$F_{\alpha, r-1, (ab-1)(r-1)}$
A	$a - 1$	SC_A	CM_A	$\frac{CM_A}{CM_{EE}}$	$F_{\alpha, a-1, (ab-1)(r-1)}$
B	$b - 1$	SC_B	CM_B	$\frac{CM_B}{CM_{EE}}$	$F_{\alpha, b-1, (ab-1)(r-1)}$
AB	$(a - 1)(b - 1)$	SC_{AB}	CM_{AB}	$\frac{CM_{AB}}{CM_{EE}}$	$F_{\alpha, (a-1)(b-1), (ab-1)(r-1)}$
EE	$(ab - 1)(r - 1)$	SC_{EE}	CM_{EE}		
Total	$abr - 1$	SC_{Total}			

6.12.11. Ejemplo en R para un experimento factorial con un DBCA

Para ilustrar el procedimiento se supone un DBCA, el trabajo de Cardona (2003) donde se estudió el efecto de la aplicación de dosis de calcio y la frecuencia de aplicación sobre la firmeza como medida de la calidad de los frutos de tomate. El experimento se llevó a cabo con 4 repeticiones y los niveles de cada factor y datos resultantes son dados en las Tablas 39 y 40, respectivamente.

Tabla 39: Factores para firmeza del fruto del tomate

Calcio (C)	Frecuencia (F)
$C_0(0,05 \%)$	$F_0(1)$
$C_1(0,10 \%)$	$F_1(2)$
$C_2(0,15 \%)$	$F_2(3)$

Tabla 40: Resultados de firmeza del fruto del tomate

Tratamientos	Repeticiones de Y_{ijk}			
C_0F_0	2,1	2,6	1,5	2,5
C_0F_1	1,65	2,15	2,55	2,15
C_0F_2	3,6	2,2	3	3,05
C_1F_0	3,8	2,45	2,75	3,45
C_1F_1	2,75	3,4	2,7	2,75
C_1F_2	2,4	2,1	2,45	2,75
C_2F_0	2,9	3,75	2,8	3,05
C_2F_1	2,9	3,6	2,05	1,8
C_2F_2	2,8	3,1	2,9	2,9

Así, las hipótesis a verificar son:

$$1) H_0 : (\alpha\beta)_{jk} = 0 \text{ contra } H_1 : (\alpha\beta)_{jk} \neq 0$$

$$2) H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de los tratamientos es diferente de cero

$$3) H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de las covariables es diferente de cero

Se requiere cargar la librería (`easynova`). Los pasos a seguir para llevar a cabo el análisis de un experimento factorial con un DBCA son:

1. Se carga la base de datos y se observan los primeros 6 elementos. Se debe tener en cuenta que de acuerdo al ejemplo el `factor_1` es el Calcio y el `factor_2` es el frecuencia.

```

1 > base_DBCA=read.csv2(file.choose(), sep=";", dec=",");base
  _DBCA
2 > head(base_DBCA)
3   factor_1 factor_2 Bloques1   y2
4 1    0.05      1         1    2.10

```

5	2	0.05	1	2	2.60
6	3	0.05	1	3	1.50
7	4	0.05	1	4	2.50
8	5	0.05	2	1	1.65
9	6	0.05	2	2	2.15

- Utilizando la función (ea2) se lleva a cabo un análisis factorial. Los argumentos que se usan en esta función de acuerdo con este ejemplo son: la base de datos llamada `base_DBCA` y (`design=2`) ya que de acuerdo a la descripción previamente dada significa que el análisis que se va a realizar es de un DBCA.

```
1 > ea2(DatCuanti, design=2, alpha = 0.05, cov = 4, list =
  FALSE)
```

Los resultados obtenidos son:

- Análisis de varianza general

```
1 $'Analysis of variance'
2           df type III SS mean square F value p>F
3 factor_1    2     1.4717    0.7358  2.9235 0.0731
4 factor_2    2     0.5067    0.2533  1.0065 0.3804
5 blocks      3     0.4474    0.1491  0.5926 0.6259
6 factor_1:factor_2  4     2.8404    0.7101  2.8213 0.0474
7 residuals   24     6.0407    0.2517    -    -
```

De acuerdo a la información aportada por el análisis de varianza, el `factor_1` y el `factor_2` no mostraron diferencias significativas entre los grupos, lo que sugiere que estos factores no tienen un efecto importante en la variable dependiente. Los bloques tampoco mostraron diferencias significativas, lo que indica que la variabilidad observada en la variable dependiente no se debe a las diferencias entre los bloques utilizados en el estudio. Sin embargo, se encontró una interacción significativa entre el `factor_1` y el `factor_2`. Esto implica que el efecto de uno de los factores puede depender del nivel del otro factor.

Al calcular las respectivas sumas de cuadrados se llega la Tabla 41.

Tabla 41: ANAVA para firmeza del fruto del tomate

F.V.	gl	SC	CM	Fc
Bloques	3	0.447	0.149	0.59NS
F	2	0.506	0.253	1.01NS
C:F1	2	2.375	1.1877	4.718*
C:F2	2	1.216	0.6081	2.416NS
C:F3	2	0.720	0.3602	1.431NS
EE	24	6.040	0.251	
Total	35	11.306		

Se concluye que para los niveles F2 y F3 no hay diferencia significativa entre los niveles de Calcio, en tanto que para al nivel F1 existe diferencia significativa entre el nivel de Calcio.

Tabla 42: ANAVA para firmeza del fruto del tomate

F.V.	gl	SC	CM	Fc
Bloques	3	0.447	0.149	0.59NS
C	2	1.471	0.735	2.92*(8%)
F:C1	2	1.765	0.8827	3.507*
F:C2	2	0.991	0.4956	1.969NS
F:C3	2	0.590	0.2952	1.172NS
EE	24	6.040	0.251	
Total	35	11.306		

Se concluye que para los niveles C2 y C3 no hay diferencia significativa entre los niveles de Frecuencia, en tanto que para al nivel C1 existe diferencia significativa entre el nivel de Frecuencia.

- Se realiza el ajuste de medias para cada uno de los factores y sus respectivas comparaciones múltiples en las que se tienen en cuenta pruebas tales como Tukey, Duncan, Scott Knott entre otras.
- Ajuste de medias y comparaciones de medias del factor 1.

1 \$ 'Adjusted means (factor 1) '

```

2 factor_1 adjusted.mean standard.error tukey snk duncan
3 1 0.15 2.8792 0.1448 a a a
4 2 0.1 2.8125 0.1448 a a ab
5 3 0.05 2.4208 0.1448 a a b
6 t scott_knott
7 a a
8 ab a
9 b a
10
11 $'Multiple comparison test (factor 1)'
12 pair contrast p(tukey) p(snk) p(duncan) p(t)
13 1 0.15 - 0.1 0.0667 0.9433 0.7475 0.7475 0.7475
14 2 0.15 - 0.05 0.4584 0.0849 0.0849 0.0434 0.0347
15 3 0.1 - 0.05 0.3917 0.1568 0.0678 0.0678 0.0678

```

Teniendo en cuenta los tratamientos del factor 1, no se encuentra diferencias significativas en las pruebas tukey, snk y scott_knott..

- Ajuste de medias y comparaciones de medias del factor 2.

```

1 $'Adjusted means (factor 2)'
2 factor_2 adjusted.mean standard.error tukey snk duncan
3 1 1 2.804167 0.1448 a a a
4 2 3 2.770833 0.1448 a a a
5 3 2 2.537500 0.1448 a a a
6 t scott_knott
7 a a
8 a a
9 a a
10
11 $'Multiple comparison test (factor 2)'
12 pair contrast p(tukey) p(snk) p(duncan) p(t)
13 1 1 - 3 0.0333 0.9855 0.8721 0.8721 0.8721
14 2 1 - 2 0.2667 0.4077 0.4077 0.2304 0.2052
15 3 3 - 2 0.2333 0.4999 0.2658 0.2658 0.2658

```

De acuerdo con el análisis anterior donde se tienen en cuenta los tratamientos del factor 2, la prueba tukey, snk, duncan, t y scott_knott, no encuentro diferencias significativas ente los tratamientos.

- Analizando significancia de interacción del factor 1 dentro de niveles del factor 2 (pruebas verticales). Se tiene en cuenta las siguientes hipótesis.

$$H_0 : \mu_j = \mu_{j'} \quad vs \quad H_0 : \mu_j \neq \mu_{j'} \quad \text{donde } j, j' = 1, 2, 3, 4$$

- Ajustes de medias del factor 1 dentro del nivel 1 del factor 2.

```

1 $'Adjusted means (factor 1 in levels of factor 2) '$'
   factor_1 in 1'
2   treatment adjusted.mean standard.error tukey snk duncan
3 3      0.15.1          3.1250          0.2508    a  a    a
4 2      0.1.1           3.1125          0.2508    a  a    a
5 1      0.05.1          2.1750          0.2508    b  b    b
6 t   scott_knott
7 a           a
8 a           a
9 b           b

```

- Ajustes de medias del factor 1 dentro del nivel 2 del factor 2.

```

1 $'Adjusted means (factor 1 in levels of factor 2) '$'
   factor_1 in 2'
2   treatment adjusted.mean standard.error tukey snk duncan
3 5      0.1.2           2.9000          0.2508    a  a    a
4 6      0.15.2          2.5875          0.2508    a  a   ab
5 4      0.05.2          2.1250          0.2508    a  a    b
6 t   scott_knott
7 a           a
8 ab          a
9 b           a

```

- Ajustes de medias del factor 1 dentro del nivel 3 del factor 2.

```

1 $'Adjusted means (factor 1 in levels of factor 2) '$'
   factor_1 in 3'
2   treatment adjusted.mean standard.error tukey snk duncan
3 7      0.05.3          2.9625          0.2508    a  a    a
4 9      0.15.3          2.9250          0.2508    a  a    a
5 8      0.1.3           2.4250          0.2508    a  a    a
6 t   scott_knott
7 a           a
8 a           a
9 a           a

```

- Comparaciones múltiples del factor 1 dentro del nivel 1 del factor 2.

```

1 $'Multiple comparison test (factor 1 in levels of factor
   2) '$'
2 $'Multiple comparison test (factor 1 in levels of factor
   2) '$'
3 $'factor_1 in 1'

```

	pair	contrast	p(tukey)	p(snk)	p(duncan)	p(t)
1	0.15.1-0.1.1	0.012	0.999	0.9722	0.9722	0.9722
2	0.15.1-0.05.1	0.950	0.033	0.0339	0.0171	0.0131
3	0.1.1-0.05.1	0.937	0.036	0.0142	0.0142	0.0142

- Comparaciones múltiples del factor 1 dentro del nivel 2 del factor 2.

```

1 $'Multiple comparison test (factor 1 in levels of factor
2 2) '
2 $'factor_1 in 2'
3 pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 0.1.2-0.15.2 0.3125 0.6572 0.3870 0.3870 0.3870
5 2 0.1.2-0.05.2 0.7750 0.0943 0.0943 0.0483 0.0389
6 3 0.15.2-0.05.2 0.4625 0.4067 0.2046 0.2046 0.2046

```

- Comparaciones múltiples del factor 1 dentro del nivel 3 del factor 2.

```

1 $'Multiple comparison test (factor 1 in levels of factor
2 2) '
2 $'factor_1 in 3'
3 pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 0.05.3-0.15.3 0.0375 0.9939 0.9167 0.9167 0.9167
5 2 0.05.3-0.1.3 0.5375 0.3017 0.3017 0.1643 0.1427
6 3 0.15.3-0.1.3 0.5000 0.3519 0.1715 0.1715 0.1715

```

La interacción del factor 1 dentro de cada nivel del factor 2, se encuentra diferencias entre los promedios en el nivel 1 y 2, pero en el nivel 3 no se encuentra diferencia.

- Analizando significancia de interacción del factor 2 dentro de niveles del factor 1 (pruebas horizontales). Se tiene en cuenta las siguientes hipótesis.

$$H_0 : \mu_{.k} = \mu_{.k'} \quad vs \quad H_0 : \mu_{.k} \neq \mu_{.k'} \quad \text{donde } k, k' = 1, 2, 3, 4$$

- Ajustes de medias del factor 2 dentro del nivel 0.05 del factor 1.

```

1 $'Adjusted means (factor 2 in levels of factor 1) '
2 $'factor_2 in 0.05 '
3 treatment adjusted.mean standard.error tukey snk duncan
4 7 0.05.3 2.9625 0.2508 a a a
5 1 0.05.1 2.1750 0.2508 a b b
6 4 0.05.2 2.1250 0.2508 a ab b

```

```

7 t scott_knott
8 a a
9 b b
10 b b

```

- Ajustes de medias del factor 2 dentro del nivel 0.1 del factor 1.

```

1 $'Adjusted means (factor 2 in levels of factor 1)'$'
   factor_2 in 0.1'
2   treatment adjusted.mean standard.error tukey snk duncan
3 2      0.1.1          3.1125         0.2508    a  a    a
4 5      0.1.2          2.9000         0.2508    a  a    a
5 8      0.1.3          2.4250         0.2508    a  a    a
6 t scott_knott
7 a a
8 a a
9 a a

```

- Ajustes de medias del factor 2 dentro del nivel 0.15 del factor 1.

```

1 $'Adjusted means (factor 2 in levels of factor 1)'$'
   factor_2 in 0.15'
2   treatment adjusted.mean standard.error tukey snk duncan
3 3      0.15.1          3.1250         0.2508    a  a    a
4 9      0.15.3          2.9250         0.2508    a  a    a
5 6      0.15.2          2.5875         0.2508    a  a    a
6 t scott_knott
7 a a
8 a a
9 a a

```

- Comparaciones múltiples del factor 2 dentro del nivel 0.05 del factor 1.

```

1 $'Multiple comparison test (factor 2 in levels of factor
   1)'
2 $'factor_2 in 0.05'
3   pair          contrast p(tukey) p(snk) p(duncan) p(t)
4 1 0.05.3-0.05.1  0.7875  0.0880  0.0361  0.0361  0.0361
5 2 0.05.3-0.05.2  0.8375  0.0663  0.0663  0.0337  0.0267
6 3 0.05.1-0.05.2  0.0500  0.9891  0.8891  0.8891  0.8891

```

- Comparaciones múltiples del factor 2 dentro del nivel 0.1 del factor 1.

```

1 $'Multiple comparison test (factor 2 in levels of factor
   1)'
2 $'factor_2 in 0.1'

```

```

3           pair contrast p(tukey) p(snk) p(duncan) p(t)
4 1 0.1.1-0.1.2  0.2125  0.8219  0.5547  0.5547  0.5547
5 2 0.1.1-0.1.3  0.6875  0.1498  0.1498  0.0779  0.0644
6 3 0.1.2-0.1.3  0.4750  0.3880  0.1930  0.1930  0.1930

```

- Comparaciones múltiples del factor 2 dentro del nivel 0.15 del factor 1.

```

1 $'Multiple comparison test (factor 2 in levels of factor
  1)'
2 $'factor_2 in 0.15'
3           pair      contrast p(tukey) p(snk) p(duncan) p(t)
4 1 0.15.1-0.15.3  0.2000  0.8404 0.5781  0.5781  0.5781
5 2 0.15.1-0.15.2  0.5375  0.3017 0.3017  0.1643  0.1427
6 3 0.15.3-0.15.2  0.3375  0.6138 0.3508  0.3508  0.3508

```

La interacción del factor 2 dentro de cada nivel del factor 1, se encuentra diferencias entre los promedios en el nivel 0.05, pero en el resto de los niveles del factor 1 no se encuentra diferencia.

- Análisis residual

```

1 $'Residual analysis'
2 $'Residual analysis'$'residual analysis'
3           values
4 p.value Shapiro-Wilk test 0.8824
5 p.value Bartlett test (factor_1) 0.8897
6 p.value Bartlett test (factor_2) 0.7392
7 p.value Bartlett test (treatments) 0.3949
8 coefficient of variation (%) 18.5500
9 first value most discrepant 30.0000
10 second value most discrepant 10.0000
11 third value most discrepant 32.0000

```

Los errores se distribuyen de forma normal. La prueba de Bartlett muestra que para los factores 1, 2 y la interacción de los tratamientos las varianzas homogéneas, de acuerdo a estos resultados se considera hacer una transformación, este proceso queda recomendado para el lector. El coeficiente de variación es 18.5%.

6.13. Experimentos con diferentes tamaños de parcelas

En muchos experimentos factoriales es necesario utilizar más de un tamaño de parcela para las unidades experimentales. Tal es el caso de factores que

exigen unidades experimentales grandes como los sistemas de labranza, y otras unidades más pequeñas, como variedad, dosis, etc. Otro caso importante es cuando se mide una unidad experimental en tiempos diferentes, en ese caso todos los tratamientos se observan en cada tiempo. Entre las estructuras más comunes con diferente tamaño de parcelas se encuentran (Kuehl, 2001, pg.469).

- Parcelas divididas.
- Parcelas subdivididas.
- Franjas o bloques divididos.
- Medidas repetidas.

6.13.1. Aleatorización

La aleatorización de los tratamientos asignados a las parcelas principales se lleva a cabo de acuerdo al diseño experimental que se implementó. Por ejemplo, si es un DBCA, entonces la asignación se llevará a cabo por cada bloque. Considérese un experimento donde se desea probar el factor A con tres niveles, el cual se asigna a las parcelas grandes en un DBCA y un factor B con cuatro niveles. Una posible distribución en campo podría ser la siguiente (Martínez y Robles, 2013, p.191).

Bloque I			Bloque II			Bloque III		
a_2	a_1	a_3	a_1	a_2	a_3	a_1	a_3	a_2
b_1	b_2	b_2	b_2	b_4	b_2	b_3	b_4	b_1
b_4	b_3	b_4	b_3	b_1	b_3	b_4	b_1	b_3
b_2	b_1	b_3	b_1	b_2	b_1	b_1	b_3	b_4
b_3	b_4	b_1	b_4	b_3	b_4	b_2	b_2	b_2

Otra manera de distribución donde se tienen dos tipos de errores el error (a) generado por las diferencias en las observaciones de las parcelas principales o unidades grandes y el error (b) que corresponde a las subparcelas, generado por la diferencia entre las observaciones de las parcelas pequeñas (Martínez y Robles, 2013).

Bloque I											
a_1				a_2				a_3			
b_1	b_2	b_4	b_3	b_2	b_4	b_1	b_3	b_3	b_4	b_1	b_2
Bloque II											
a_2				a_3				a_1			
b_4	b_2	b_1	b_3	b_2	b_1	b_3	b_4	b_3	b_1	b_2	b_4
Bloque III											
a_3				a_1				a_2			
b_3	b_1	b_4	b_2	b_1	b_2	b_4	b_3	b_2	b_1	b_3	b_1

Considerando que el investigador espera que el error (a) sea mayor que el error (b), pues las parcelas grandes usualmente presentan mayor heterogeneidad que las subparcelas. Además, la estimación del error en parcelas grandes es menos precisa que la del error (b) ya que el número de grados de libertad de las primeras es menor (Martínez y Robles, 2013, p.192).

6.13.2. Parcelas divididas en el espacio

Kuehl propone que para diseñar parcelas divididas se usa un modelo mixto que refleje las distintas varianzas del error experimental para las subparcelas y la parcela completa, esto incluye los efectos del error aleatorio por separado para ambas. El modelo estadístico para un diseño en bloques completos al azar, está dado por:

$$Y_{ijk} = \mu + \alpha_j + \rho_i + \varepsilon_{ij} + \beta_k + (\alpha\beta)_{jk} + \theta_{ijk} \quad (27)$$

con $i = 1, 2, 3, \dots, r$, $j = 1, 2, 3, \dots, a$ y $k = 1, 2, 3, \dots, b$, donde

- Y_{ijk} : es la respuesta en la i -ésima unidad experimental para el j -ésimo nivel del factor A y el k -ésimo nivel del factor B .
- μ : es la media general.
- ρ_i : es el efecto del bloque i -ésimo.
- α_j : es el efecto aditivo del nivel j -ésimo de A .
- ε_{ij} : es el error de las parcelas principales.

- β_k : es el efecto principal del k -ésimo nivel del factor B .
- $(\alpha\beta)_{jk}$: es el efecto aditivo de la combinación del nivel j -ésimo de A con el nivel k -ésimo de B .
- θ_{ijk} : es el error de las subparcelas.

Los supuestos para los errores, son los clásicos ya estudiados.

$$\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2); \theta_{ijk} \stackrel{iid}{\sim} N(0, \sigma_\theta^2)$$

El cuadrado medio del error para las parcelas principales, se estima por:

$$\sigma^2 + b\sigma_b^2$$

6.13.3. Análisis de varianza

Las sumas de cuadrados según el modelo definido están dadas por:

$$\begin{aligned} SC_{total}(subparc.) &= SC_{Bloques} + SC_A + SC_{Error}(parc.ppal) \\ &+ SC_B + SC_{AB} + SC_{Error}(Subparc.) \\ &= SC_{Bloques} + SC_A + SC_{E(a)} + SC_B + SC_{AB} + SC_{E(b)} \end{aligned} \quad (28)$$

$$\begin{aligned} SC_{(Parc.ppal)} &= SC_{Bloques} + SC_A + SC_{Error}(Parc.ppal) \\ &= SC_{Bloques} + SC_A + SC_{E(a)} \\ &= \sum_{i=1}^r \sum_{j=1}^a \frac{Y_{ij}^2}{b} - FC \end{aligned} \quad (29)$$

con $FC = \frac{Y_{\dots}^2}{rab}$

$$SC_{Total}(Subparc.) = \sum_{i=1}^r \sum_{j=1}^a \sum_{k=1}^b Y_{ijk}^2 - \frac{Y_{\dots}^2}{rab} \quad (30)$$

$$SC_{Bloques} = \sum_{i=1}^r \frac{Y_{i..}^2}{ab} - \frac{Y_{...}^2}{rab} \quad (31)$$

$$SC_A = \sum_{j=1}^a \frac{Y_{.j.}^2}{rb} - \frac{Y_{...}^2}{rab} \quad (32)$$

$$SC_{E(a)} = SC_{(Parc.ppal)} - SC_{Bloques} - SC_A \quad (33)$$

$$SC_B = \sum_{k=1}^b \frac{Y_{k..}^2}{ra} - \frac{Y_{...}^2}{rab} \quad (34)$$

$$SC_{AB} = \left[\sum_{j=1}^a \sum_{k=1}^b \frac{Y_{.jk}^2}{r} - \frac{Y_{...}^2}{rab} \right] - SC_A - SC_B \quad (35)$$

$$= SC_{TTO} - SC_A - SC_B \quad (36)$$

$$SC_{E(b)} = SC_{Total(Subparc.)} - SC_{(parc.ppal)} - SC_B - SC_{AB}. \quad (37)$$

Las hipótesis a probar son las usuales de un experimento factorial:

$$1) H_0 : (\alpha\beta)_{jk} = 0 \text{ contra } H_1 : (\alpha\beta)_{jk} \neq 0$$

$$2) H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de los tratamientos es diferente de cero

$$3) H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de las covariables es diferente de cero

Ahora, bajo un modelo de efectos fijos con bloques aleatorios, los cuadrados medios esperados para definir la estadística de prueba de cada hipótesis se muestran en la Tabla 43.

Tabla 43: Grados de libertad y cuadrados medios esperados

F.V.	gl	$E[CME]$
Bloques	$r - 1$	$\sigma_\theta^2 + b\sigma^2 + ab\sigma_\rho^2$
A	$a - 1$	$\sigma_\theta^2 + b\sigma^2 + \frac{rb}{a-1} \sum_{i=1}^a \alpha_j^2$
$E(a)$: Error parcela principal	$(r - 1)(a - 1)$	$\sigma_\theta^2 + b\sigma^2$
Total, parcela principal	$ra - 1$	
B	$b - 1$	$\sigma_\theta^2 + \frac{ra}{b-1} \sum_{i=1}^a \beta_k^2$
AB	$(a - 1)(b - 1)$	$\sigma_\theta^2 + \frac{r}{(a-1)(b-1)} \sum_{i=1}^a (\alpha\beta)_{jk}^2$
$E(b)$: Error Subparcela	$a(r - 1)(b - 1)$	σ_θ^2

Las siguientes Tablas 44 y 45 muestra el análisis de varianza correspondiente para llevar a cabo las pruebas de cada una de estas hipótesis anteriores.

Tabla 44: ANAVA para una estructura en parcelas divididas bajo un DCA

F.V.	gl	SC	CM	FC
A	$a - 1$	SC_A	CM_A	$\frac{CM_A}{CM_{E(a)}}$
$E(a)$: Error parcela principal	$(r - 1)(a - 1)$	$SC_{E(a)}$	$CM_{E(a)}$	
Total parcela principal	$ra - 1$	$SC_{Total}(PP)$		
B	$b - 1$	SC_B	CM_B	$\frac{CM_B}{CM_{E(b)}}$
AB	$(a - 1)(b - 1)$	SC_{AB}	CM_{AB}	$\frac{CM_{AB}}{CM_{E(b)}}$
$E(b)$: Error Subparcela	$a(r - 1)(b - 1)$	$SC_{E(a)}$	$CM_{E(a)}$	
Total	$(abr - 1)$	SC_{Total}		

Tabla 45: ANAVA para una estructura en parcelas divididas bajo un DBCA

F.V.	gl	SC	CM	FC
Bloques	$r - 1$	$SC_{Bloques}$	$CM_{Bloques}$	$\frac{CM_{Bloques}}{CM_{E(a)}}$
A	$a - 1$	SC_A	CM_A	$\frac{CM_A}{CM_{E(a)}}$
$E(a)$: Error parcela principal	$(r - 1)(a - 1)$	$SC_{E(a)}$	$CM_{E(a)}$	
Total parcela principal	$ra - 1$	$SC_{Total(PP)}$		
B	$b - 1$	SC_B	CM_B	$\frac{CM_B}{CM_{E(b)}}$
AB	$(a - 1)(b - 1)$	SC_{AB}	CM_{AB}	$\frac{CM_{AB}}{CM_{E(b)}}$
$E(b)$: Error Subparcela	$a(r - 1)(b - 1)$	$SC_{E(a)}$	$CM_{E(a)}$	
Total	$(abr - 1)$	SC_{Total}		

6.13.4. Errores estándar para la comparación de medias de tratamientos

Una vez se hayan determinado las diferencias significativas entre los tratamientos de los factores o las combinaciones de tratamiento de los factores involucrados, se procede a realizar comparaciones de medias para identificar específicamente dónde se encuentran esas diferencias. El análisis de las interacciones y los efectos principales se realiza de manera similar a lo estudiado en el capítulo de experimentos factoriales. Los errores estándar utilizados para comparar las medias de los tratamientos de las parcelas principales, subparcelas e interacción se presentan en la Tabla 46.

Tabla 46: Errores estándar en una estructura de parcelas divididas

Diferencia entre:	Comparación	Efecto	Error Estándar
Dos medias de A	$a_i - a_j$	$a_1 - a_2$	$\sqrt{\frac{2CM_{E(a)}}{rb}}$
Dos medias de B	$b_i - b_j$	$b_1 - b_2$	$\sqrt{\frac{2CM_{E(b)}}{ra}}$
Dos medias B al mismo nivel de A	$a_i b_j - a_i b_k$	$a_1 b_1 - a_2 b_1$	$\sqrt{\frac{2CM_{E(b)}}{r}}$
Dos medias A al mismo nivel de B	$a_i b_j - a_k b_j$	$a_1 b_1 - a_2 b_1$	$\sqrt{\frac{2[(b-1)CM_{E(b)} + CM_{E(a)}]}{rb}}$
Diferentes niveles de B (dos medias de TTO)	$a_i b_j - a_k b_l$	$a_1 b_2 - a_2 b_1$	$\sqrt{\frac{2[(b-1)CM_{E(b)} + CM_{E(a)}]}{rb}}$

En los primeros tres casos, los grados de libertad del error para llevar a cabo cualquier prueba de comparación son los dados por los gl del error que se utiliza. Para los últimos dos casos, el error estándar es combinación de los cuadrados medios de las parcelas grandes y las subparcelas; así que los grados de libertad para estos errores deben aproximarse a los grados de libertad adecuados utilizando el procedimiento dado por Satterthwaite (1946). Esta aproximación es como sigue:

$$gl = \frac{[(b-1)CM_{E(b)} + CM_{E(a)}]^2}{\frac{[(b-1)CM_{E(b)}]^2}{gl_{E(b)}} + \frac{[CM_{E(a)}]^2}{gl_{E(a)}}} \quad (38)$$

Con estos errores estándar se pueden llevar a cabo cualquier comparación de tratamiento utilizando las pruebas de Duncan, Tukey, DMS, Dunnett, etc, estudiadas anteriormente.

Los grados de libertad y los cuadrados medios esperados están dados en la Tabla 47.

Tabla 47: Grados de libertad y cuadrados medios esperados para un DBCA

F.V.	gl	$E[CME]$
Bloques	$r - 1$	$\sigma_\theta^2 + b\sigma^2 + ab\sigma_\rho^2$
A	$a - 1$	$\sigma_\theta^2 + b\sigma^2 + \frac{rb}{a-1} \sum_{i=1}^a \alpha_j^2$
$e(a)$: Error parcela principal	$(r - 1)(a - 1)$	$\sigma_\theta^2 + b\sigma^2$
<i>Bloques * A</i>	$ra - 1$	
B	$b - 1$	$\sigma_\theta^2 + \frac{ra}{b-1} \sum_{i=1}^a \beta_k^2$
AB	$(a - 1)(b - 1)$	$\sigma_\theta^2 + \frac{r}{(a-1)(b-1)} \sum_{i=1}^a (\alpha\beta)_{jk}^2$
$e(b)$: Error Subparcela	$a(r - 1)(b - 1)$	σ_θ^2

1. La descomposición de la interacción:

Para cada factor se descompone la suma de cuadrado por cada nivel del otro factor.

Para el factor A en cada nivel de B definimos:

$$SCA : B_k = \frac{1}{r} \sum_{j=1}^a Y_{.jk}^2 - \frac{Y_{..k}^2}{ar}$$

$$CMA : B_k = \frac{SCA : B_k}{a - 1}$$

$$F_c = \frac{CMA : B_k}{CMEE}$$

Tabla 48: ANAVA interacción de A dentro los niveles de B

<i>F.V.</i>	<i>gl</i>	<i>SC</i>	<i>CM</i>	<i>F_c</i>	<i>F_{tabla}</i>
<i>TTO</i>	$ab - 1$	SC_{TTO}	CM_{TTO}	$\frac{CM_{TTO}}{CM_{EE}}$	$F_{\alpha, ab-1, ab(r-1)}$
<i>B</i>	$b - 1$	SC_B	CM_B	$\frac{CM_B}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
$A : B_1$	$b - 1$	$SC_{A:B_1}$	$CM_{A:B_1}$	$\frac{CM_{A:B_1}}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
$A : B_2$	$b - 1$	$SC_{A:B_2}$	$CM_{A:B_2}$	$\frac{CM_{A:B_2}}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$A : B_b$	$b - 1$	$SC_{A:B_b}$	$CM_{A:B_b}$	$\frac{CM_{A:B_b}}{CM_{EE}}$	$F_{\alpha, b-1, ab(r-1)}$
<i>EE</i>	$ab(r - 1)$	SC_{EE}	CM_{EE}		

Ahora para el factor B en cada nivel de B definimos:

$$SCB : A_j = \frac{1}{r} \sum_{j=1}^B Y_{.jk}^2 - \frac{Y_{.j}^2}{br}$$

$$CMB : A_j = \frac{SCB : A_j}{b - 1}$$

$$F_c = \frac{CMA : A_j}{CM_{EE}}$$

Tabla 49: ANAVA interacción de B dentro los niveles de A

<i>F.V.</i>	<i>gl</i>	<i>SC</i>	<i>CM</i>	<i>F_c</i>	<i>F_{tabla}</i>
<i>TTO</i>	$ab - 1$	SC_{TTO}	CM_{TTO}	$\frac{CM_{TTO}}{CM_{EE}}$	$F_{\alpha, ab-1, ab(r-1)}$
<i>A</i>	$a - 1$	SC_A	CM_A	$\frac{CM_A}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
$B : A_1$	$a - 1$	$SC_{B:A_1}$	$CM_{B:A_1}$	$\frac{CM_{B:A_1}}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
$B : A_2$	$a - 1$	$SC_{B:A_2}$	$CM_{B:A_2}$	$\frac{CM_{B:A_2}}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$B : A_a$	$a - 1$	$SC_{B:A_a}$	$CM_{B:A_b}$	$\frac{CM_{B:A_a}}{CM_{EE}}$	$F_{\alpha, a-1, ab(r-1)}$
<i>EE</i>	$ab(r - 1)$	SC_{EE}	CM_{EE}		

2. Pruebas horizontales y verticales

Para analizar la interacción se pueden realizar pruebas verticales u horizontales. Estas consisten en analizar un factor a cada nivel del otro factor.

Las diferencias del factor A cada nivel del factor B (pruebas verticales); las hipótesis para tener en cuenta son:

$$H_0 : \mu_j = \mu_{j'} \quad vs \quad H_0 : \mu_j \neq \mu_{j'}$$

donde $j, j' = 1, 2, 3, 4$

Las diferencias del factor B a cada nivel del factor A (pruebas horizontales); las hipótesis para tener en cuenta son:

$$H_0 : \mu_k = \mu_{k'} \quad vs \quad H_0 : \mu_k \neq \mu_{k'}$$

donde $k, k' = 1, 2, 3, 4$

6.13.5. Función (split2.crd) en R

Los autores de la función (`split2.crd`) son Eric B Ferreira, Denismar Alves Nogueira y Portya Piscitelli Cavalcanti, analiza los experimentos que presenta un diseño completamente al azar. Este se encuentra en la librería de R llamada (`ExpDes`).

- La función (`plit2.crd`) está compuesto por:

```
1 > split2.crd( factor1, factor2, repet resp,  
2 quali = c(TRUE, TRUE), mcomp = "tukey", fac.names = c("F1", "F2")  
3 sigT = 0.05, sigF = 0.05, unfold = NULL)
```

- Características

`factor1` : vector numérico o complejo que contiene los niveles del factor 1.

`factor2` : vector numérico o complejo que contiene los niveles del factor 2.

`repet` : vector numérico o complejo que contiene las replicaciones.

`resp`: vector numérico o complejo que contiene la variable de respuesta.

`quali`: lógica. Si es TRUE (por defecto), los tratamientos se asumen cualitativos, si es FALSE, cuantitativos.

`mcomp`: permite elegir la prueba de comparación múltiple; el valor por defecto es la prueba de Tukey, sin embargo, las opciones son: la prueba de LSD ('lsd'), la prueba de LSD con protección de Bonferroni ('lsdb'), la prueba de Duncan ('duncan'), la prueba de Student-NewmanKeuls ('snk'), la prueba de Scott-Knot ('sk'), la prueba de Calinski y Corsten ('ccF') y la prueba de comparaciones múltiples bootstrap ('ccboot').

`fac.names`: permite etiquetar los factores 1 y 2.

`sigT`: la significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 %.

`sigF`: la significación que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5 %.

`unfold`: dice lo que debe hacerse después del ANOVA. Si es NULL (por defecto), se realizan las pruebas recomendadas; si es '0', sólo se realiza el ANOVA; si es '1', se prueban los efectos simples son probados.

Esta función aporta el ANOVA del referido CRD, la prueba de normalidad de Shapiro-Wilk para los residuos del modelo, los modelos de regresión ajustados (cuando los tratamientos son cuantitativos) y/o el pruebas de comparación múltiple (cuando los tratamientos son cualitativos).

6.13.6. Ejemplo en R de parcelas divididas para un DCA

En un proceso de fermentación del lactosuero en la elaboración del queso se experimenta con 2 tipo de lactosuero ácido (entero (1) y desproteínizado (2) y 3 niveles de inóculo (5, 10 y 15 %) con una estructura factorial 2x3 en parcelas divididas y un DCA con 4 repeticiones. Las parcelas grandes corresponden al tipo de lactosuero y las subparcelas a las dosis del inóculo. Una de las variables de interés es la acidez a las ocho horas de haber comenzado el experimento

Tabla 50: fermentación del lactosuero

Lactosuero	Rep	Inoculo	acidez
1	1	5	49.93
1	1	10	36.00
1	1	15	103.56
1	2	5	40.33
1	2	10	36.83
1	2	15	94.40
1	3	5	42.00
1	3	10	37.93
1	3	15	85.23
1	4	5	44.08
1	4	10	36.92
1	4	15	94.40
2	1	5	37.16
2	1	10	52.46
2	1	15	93.83
2	2	5	39.70
2	2	10	48.20
2	2	15	70.93
2	3	5	32.46
2	3	10	50.33
2	3	15	82.38
2	4	5	36.44
2	4	10	50.33
2	4	15	82.38

Así, las hipótesis a verificar son:

$$1) H_0 : (\alpha\beta)_{jk} = 0 \text{ contra } H_1 : (\alpha\beta)_{jk} \neq 0$$

$$2) H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de los tratamientos es diferente de cero

$$3) H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de las covariables es diferente de cero

Se requiere la librería (ExpDes) para realizar el procedimiento. Los pasos a seguir son los siguientes:

1. Los primeros 6 elementos de la base de datos son:

```
1 > data(base); head(base)
2 > base
3 Lactosuero Rep Inoculo acidez
4           1   1         5  49.93
5           1   1        10  36.00
6           1   1        15 103.56
7           1   2         5  40.33
8           1   2        10  36.83
9           1   2        15  94.40
```

2. Utilizando la función (split2.crd) se lleva a cabo un análisis de parcelas divididas con un DCA.

```
1 > split2.crd(Lactosuero, Inoculo, Rep, acidez, quali = c(TRUE
2 mcomp = "tukey", fac.names = c("Lactosuero", "Inoculo"))
```

Se obtienen los siguientes resultados:

■ Análisis de varianza

```
1 -----
2 Analysis of Variance Table
3 -----
4           DF      SS      MS      Fc      Pr>Fc
5 Lactosuero      1    26.1    26.1    0.727 0.42647
6 Error a         6   215.0    35.8
7 Inoculo         2 11550.0  5775.0 226.649 0.00000
8 Lactosuero*Inoculo  2   739.3   369.7  14.508 0.00063
9 Error b        12   305.8    25.5
10 Total          23 12836.1
11 -----
12 CV 1 = 10.42383 %
13 CV 2 = 8.7901 %
```

El p_{valor} que aporta la interacción de lactosuero e inoculo es inferior al compararlo con un $\alpha = 0,05$, se concluye que hay diferencia significativa, es decir que existe una interacción significativa entre lactosuero e inoculo.

- Prueba de normalidad

```

1 -----
2 Shapiro-Wilk normality test (Error b)
3 p-value: 0.1944154
4 According to Shapiro-Wilk normality test at 5%
5 of significance, residuals can be considered normal.
6 -----

```

La prueba de normalidad de Shapiro-Wilk al 5% de importancia, los residuos pueden considerarse normales.

- Análisis de la interacción de lactosuero dentro de cada nivel de inoculo.

```

1 -----
2 Significant interaction: analyzing the interaction
3 -----
4 Analyzing Lactosuero inside of each level of Inoculo
5 -----
6
7           DF      SS      MS
8 Lactosuero : Inoculo 5    1.00000 116.8920 116.89205
9 Lactosuero : Inoculo 10   1.00000 359.6562 359.65620
10 Lactosuero : Inoculo 15   1.00000 288.8406 288.84061
11 Pooled Error              17.50206 506.3399 28.93031
12 -----
13 Fc      p.value
14 4.040470 0.060093
15 12.431813 0.002500
16 9.984014 0.005566
17 NA      NA
18 -----

```

El lactosuero influenciado es preferido en el inoculo 10 y 15.

- Lactosuero dentro del Inóculo 5

```

1 -----
2 Lactosuero inside of Inoculo 5
3 -----
4 According to F test, the means of this factor are not
   distinct.

```

```

5 -----
6   Levels  Means
7 1         1 44.085
8 2         2 36.440
9 -----

```

Según la prueba F, las medias de este factor no son distintas.

- Lactosuero dentro del Inóculo 10

```

1 Lactosuero inside of Inoculo 10
2 -----
3 Tukey's test
4 -----
5 Groups Treatments Means
6 a     2     50.33
7 b     1     36.92
8 -----

```

- Lactosuero dentro del Inóculo 15

```

1 Lactosuero inside of Inoculo 15
2 -----
3 Tukey's test
4 -----
5 Groups Treatments Means
6 a     1     94.3975
7 b     2     82.38
8 -----

```

- Análisis de la interacción de lactosuero dentro de cada nivel de inoculo.

```

1 -----
2 Analyzing Inoculo inside of each level of Lactosuero
3 -----
4           DF          SS          MS          Fc
5 Inoculo : Lactosuero 1    2 7848.4639 3924.2319 154.01347
6 Inoculo : Lactosuero 2    2 4440.8243 2220.4121  87.14403
7 Error b                   12  305.7575   25.4798      NA
8 -----
9 p.value
10 0
11 0
12 NA
13 -----

```

El inoculo influenciado es preferido en el lactosuero 1 y 2.

- Inóculo dentro de Lactosuero 1.

```

1 Inoculo inside of Lactosuero 1
2 -----
3 Tukey's test
4 -----
5 Groups Treatments Means
6 a    15    94.3975
7 b     5    44.085
8 b    10    36.92
9 -----

```

- Inóculo dentro de Lactosuero 2.

```

1 Inoculo inside of Lactosuero 2
2 -----
3 Tukey's test
4 -----
5 Groups Treatments Means
6 a    15    82.38
7 b    10    50.33
8 c     5    36.44
9 -----

```

6.13.7. Función (split2.rbd) en R

Los autores de la función (split2.rbd) Eric B Ferreira, Denismar Alves Nogueira y Portya Piscitelli Cavalcanti, Analiza experimentos en esquema de parcela dividida en diseño de bloques aleatorios equilibrados, considerando un modelo fijo. Este se encuentra en la librería de R llamada (ExpDes).

- La función (split2.rbd) está compuesto por:

```

1 > split2.rbd( factor1, factor2, block, resp, quali = c(TRUE,
2   TRUE),
3   mcomp = "tukey", fac.names = c("F1", "F2"), sigT = 0.05,
4   sigF = 0.05, unfold = NULL )

```

- Características

factor1 : vector numérico o complejo que contiene los niveles del factor 1.

factor2 : vector numérico o complejo que contiene los niveles del factor 2.

block: vector numérico o complejo que contiene los bloques.

repet : vector numérico o complejo que contiene las repeticiones.

resp: vector numérico o complejo que contiene la variable de respuesta.

quali: lógica. Si es TRUE (por defecto), los tratamientos se asumen cualitativos, si es FALSE, cuantitativos.

mcomp: permite elegir la prueba de comparación múltiple; el valor por defecto es la prueba de Tukey, sin embargo, las opciones son: la prueba de LSD ('lsd'), la prueba de LSD con protección de Bonferroni ('lsdb'), la prueba de Duncan ('duncan'), la prueba de Student-NewmanKeuls ('snk'), la prueba de Scott-Knot ('sk'), la prueba de Calinski y Corsten ('ccF') y la prueba de comparaciones múltiples bootstrap ('ccboot').

fac.names: permite etiquetar los factores 1 y 2.

sigT: la significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 %.

sigF: la significación que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5 %.

unfold: dice lo que debe hacerse después del ANOVA. Si es NULL (por defecto), se realizan las pruebas recomendadas; si es '0', sólo se realiza el ANOVA; si es '1', se prueban los efectos simples son probados.

Esta función aporta el ANOVA de la referida RBD, la prueba de normalidad de Shapiro-Wilk para los residuos del modelo, los modelos de regresión ajustados (cuando los tratamientos son cuantitativos) y/o las pruebas de comparación múltiple (cuando los tratamientos son cualitativos).

6.13.8. Ejemplo en R de parcelas divididas para un DBCA

Jaller *et al.* (2001), evaluaron el efecto de la interacción de nitrógeno (N) y potasio (K) en una investigación experimental. El diseño utilizado fue de bloques al azar bajo una estructura de tratamiento factorial en parcelas divididas, donde las parcelas principales correspondían a 4 dosis de nitrógeno, 0, 100, 200 y 300 kg/ha , las subparcelas estaban conformadas por 3 niveles de fertilización potásica, 200, 600 y 1000 kg/ha fraccionadas en tres dosis de aplicación. Tabla 51.

Tabla 51: Peso del racimo

N	K	r_1	r_2	r_3
0	200	11	8.6	9
0	600	13	11.8	11
0	1000	11	13	10
Y_{i1}		35	33.4	30
100	200	10	10	10
100	600	11	12.9	15
100	1000	12	10	11
Y_{i2}		33	32.9	36
200	200	15	14.14	17.9
200	600	13	13	11
200	1000	11	11	14
Y_{i3}		39	38.14	4.9
300	200	11	11	11
300	600	14.69	15.36	14.8
300	1000	14.78	16.18	15.27
Y_{i4}		40.47	42.54	41.07

Así, las hipótesis a verificar son:

$$1) H_0 : (\alpha\beta)_{jk} = 0 \text{ contra } H_1 : (\alpha\beta)_{jk} \neq 0$$

$$2) H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de los tratamientos es diferente de cero

$$3) H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de las covariables es diferente de cero

Se requiere la librería (ExpDes) para realizar el procedimiento. Los pasos a seguir son los siguientes:

1. los primeros 6 elementos de la base de datos son:

```

1 > data(base_DBCA); head(base_DBCA)
2 > base_DBCA
3 Nitrogeno Potasio Block  PesoR
4 0          200      1      11
5 0          200      2      8,6
6 0          200      3      9
7 0          600      1      13
8 0          600      2     11,8
9 0          600      3      11

```

2. Utilizando la función (split2.rbd) se lleva a cabo un análisis de parcelas divididas con un DBCA.

```

1 > split2.rbd(Nitrogeno, Potasio, Block, PesoR, quali = c(TRUE
2 , FALSE),
3 mcomp = "tukey", fac.names = c("Nitrogeno", "Potasio"))

```

Se obtienen los siguientes resultados:

■ Análisis de varianza

```

1 -----
2 Analysis of Variance Table
3 -----
4              DF      SS      MS      Fc  Pr(>Fc)
5 Nitrogeno      3  54.926 18.3087  9.9568 0.009572 **
6 Block          2   0.429  0.2143  0.1165 0.891954
7 Error a        6  11.033  1.8388
8 Potasio        2  13.514  6.7569  4.1198 0.036038 *
9 Nitrogeno*Potasio  6  69.631 11.6052  7.0758 0.000813 ***
10 Error b       16  26.242  1.6401
11 Total         35 175.774
12 ---

```

```

13 Signif. codes:  0      ***    0.001    **    0.01    *
      0.05      .    0.1      1
14 -----
15 CV 1 = 10.98441 %
16 CV 2 = 10.37401 %

```

No se presenta interacción significativa por tanto se lleva a cabo el análisis de los efectos simples.

- A continuación se encuentra el ajuste de los modelos lineal, cuadrático y cubico, su respectivo R^2 y por último el análisis de varianza. Teniendo en cuenta la variable nitrógeno.
- Análisis del modelo lineal

```

1 -----
2 Nitrogeno
3 -----
4
5 Linear Model
6 =====
7      Estimate Standard.Error    tc    p.value
8 -----
9 b0 10.7587      0.3782      28.4487    0
10 b1  0.0106      0.0020      5.2317    0.0019
11 -----
12
13 R2 of linear model
14 -----
15 0.916307
16 -----
17
18 Analysis of Variance of linear model
19 =====
20      DF    SS      MS    Fc    p.value
21 -----
22 Linear Effect 1  50.3291  50.3291  27.37  0.00195
23 Lack of fit   2   4.5969   2.2985   1.25  0.35173
24 Residuals    6  11.0328   1.8388
25 -----

```

Los resultados indican que el factor Nitrógeno tiene un efecto significativo en la respuesta del modelo lineal, y el modelo en general muestra

una buena capacidad de ajuste, explicando aproximadamente el 91.6% de la variabilidad observada.

■ Análisis del modelo cuadrático

```

1 Quadratic Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0 10.7737      0.4406      24.4543    0
6 b1  0.0101      0.0071      1.4312  0.2023
7 b2 0.000002     0.00002     0.0664  0.9492
8 -----
9
10 R2 of quadratic model
11 -----
12 0.916454
13 -----
14
15 Analysis of Variance of quadratic model
16 =====
17                DF    SS      MS      Fc    p.value
18 -----
19 Linear Effect    1  50.3291 50.3291 27.37 0.00195
20 Quadratic Effect 1  0.0081 0.0081  0    0.94924
21 Lack of fit     1  4.5888 4.5888  2.5  0.16525
22 Residuals       6 11.0328 1.8388
23 -----

```

los resultados indican que el modelo cuadrático para el factor Nitrógeno no muestra un efecto significativo para el término lineal ni el término cuadrático. Sin embargo, el modelo en general muestra una buena capacidad de ajuste, con un coeficiente de determinación (R^2) de aproximadamente 91.6%.

■ Análisis del modelo cubico

```

1 Cubic Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0 10.9333      0.4520      24.1883    0
6 b1 -0.0149      0.0173     -0.8585  0.4236
7 b2  0.0002      0.0001      1.5722  0.1669
8 b3 -0.000001     0          -1.5797  0.1652

```

```

9 -----
10
11 R2 of cubic model
12 -
13 1
14 -
15
16 Analysis of Variance of cubic model
17 =====
18                DF      SS      MS      Fc      p.value
19 -----
20 Linear Effect    1    50.3291  50.3291  27.37  0.00195
21 Quadratic Effect 1     0.0081   0.0081    0    0.94924
22 Cubic Effect    1     4.5888   4.5888    2.5   0.16525
23 Lack of fit     0         0         0         0     1
24 Residuals       6    11.0328  1.8388
25 -----

```

El coeficiente de determinación (R^2) para el modelo cúbico es igual a 1, lo que indica que el modelo explica la variabilidad total de los datos.

En cuanto al análisis de varianza, se encuentra que el efecto lineal del factor Nitrógeno es significativo, mientras que el efecto cuadrático y cúbico no lo son. Además, se observa que no hay falta de ajuste en el modelo ya que el p_{valor} para la prueba de falta de ajuste es igual a 1.

- A continuación se encuentra el ajuste de los modelos lineal y cuadrático, su respectivo R^2 y por último el análisis de varianza. Teniendo en cuenta la variable potasio.
- Análisis del modelo lineal

```

1 -----
2 Potasio
3 -----
4
5 Linear Model
6 =====
7      Estimate Standard.Error    tc      p.value
8 -----
9 b0  11.6831      0.4464      26.1688    0
10 b1   0.0011      0.0006      1.6879   0.1108
11 -----
12

```

```

13 R2 of linear model
14 -----
15 0.345781
16 -----
17
18 Analysis of Variance of linear model
19 =====
20                DF    SS      MS      Fc    p.value
21 -----
22 Linear Effect  1    4.6728  4.6728  2.85  0.11082
23 Lack of fit    1    8.8410  8.8410  5.39  0.03377
24 Residuals     16   26.2419  1.6401
25 -----

```

- Análisis del modelo cuadrático

```

1 Quadratic Model
2 =====
3      Estimate Standard.Error    tc    p.value
4 -----
5 b0  10.0186      0.8446      11.8626    0
6 b1   0.0090      0.0035      2.5989  0.0194
7 b2  -0.00001      0      -2.3217  0.0338
8 -----
9
10 R2 of quadratic model
11 -
12 1
13 -
14
15 Analysis of Variance of quadratic model
16 =====
17                DF    SS      MS      Fc    p.value
18 -----
19 Linear Effect    1    4.6728  4.6728  2.85  0.11082
20 Quadratic Effect 1    8.8410  8.8410  5.39  0.03377
21 Lack of fit      0     0      0      0     1
22 Residuals       16   26.2419  1.6401
23 -----

```

Teniendo en cuenta los resultados anteriores el coeficiente de correlación (R^2), el modelo que mejor se ajusta es el cuadrático ya que es igual a uno.

- Interacción, analizando nitrógeno dentro de cada nivel de potasio.

```

1 Significant interaction: analyzing the interaction
2 -----
3 Analyzing Nitrogeno inside of each level of Potasio
4 -----
5           DF          SS          MS
6 Nitrogeno : Potasio 200    3.00000  71.48640  23.828800
7 Nitrogeno : Potasio 600    3.00000  16.13229   5.377431
8 Nitrogeno : Potasio 1000   3.00000  36.93823  12.312742
9 Pooled Error              21.20059  36.17556   1.706347
10 -----
11      Fc      p.value
12 13.964803  0.000030
13  3.151428  0.046153
14  7.215849  0.001621
15 NA        NA
16 -----

```

El nitrógeno influye con preferencia al potasio con niveles de 200 y 1000. Lo cual explica la interacción.

- A continuación se encuentra el ajuste de los modelos lineal, cuadrático y cubico, su respectivo R^2 y por último el análisis de varianza. Para el Nitrógeno dentro de Potasio 200.
- Análisis del modelo lineal

```

1 Linear Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0 10.0413      0.6310      15.9136    0
6 b1  0.0101      0.0034       2.9886  0.0069
7 -----
8
9 R2 of linear model
10 -----
11 0.213201
12 -----
13
14 Analysis of Variance of linear model
15 =====
16           DF          SS          MS          Fc          p.value
17 -----
18 Linear Effect      1      15.2410  15.2410   8.93   0.00695
19 Lack of fit        2      56.2454  28.1227  16.48   5e-05

```

```

20 Residuals      21.2006 36.1756 1.7064
21 -----

```

■ Análisis del modelo cuadrático

```

1 Quadratic Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0  8.7547      0.7351      11.9098    0
6 b1  0.0487      0.0118      4.1238  0.0005
7 b2 -0.0001      0.00004     -3.4121  0.0026
8 -----
9
10 R2 of quadratic model
11 -----
12 0.491102
13 -----
14
15 Analysis of Variance of quadratic model
16 =====
17                                DF      SS      MS      Fc    p.value
18 -----
19 Linear Effect                1      15.2410 15.2410  8.93  0.00695
20 Quadratic Effect              1      19.8661 19.8661 11.64  0.0026
21 Lack of fit                   1      36.3793 36.3793 21.32  0.00015
22 Residuals                    21.2006 36.1756 1.7064
23 -----

```

■ Análisis del modelo cubico

```

1 Cubic Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0  9.5333      0.7542      12.6407    0
6 b1 -0.0733      0.0289      -2.5334  0.0192
7 b2  0.0010      0.0003      4.0638  0.0006
8 b3 -0.000003     0          -4.6174  0.0001
9 -----
10
11 R2 of cubic model
12 -
13 1
14 -
15

```

```

16 Analysis of Variance of cubic model
17 =====
18                DF      SS      MS      Fc      p.value
19 -----
20 Linear Effect      1      15.2410  15.2410  8.93  0.00695
21 Quadratic Effect   1      19.8661  19.8661  11.64  0.0026
22 Cubic Effect       1      36.3793  36.3793  21.32  0.00015
23 Lack of fit        0         0         0         0         1
24 Residuals          21.2006  36.1756  1.7064
25 -----

```

Teniendo en cuenta los resultados anteriores el coeficiente de correlación (R^2), el modelo que mejor se ajusta es el cúbico ya que es igual a uno.

- A continuación se encuentra el ajuste de los modelos lineal, cuadrático y cubico, su respectivo R^2 y por último el análisis de varianza. Para el Nitrógeno dentro de Potasio 600.
- Análisis del modelo lineal

```

1
2 Linear Model
3 =====
4      Estimate Standard.Error   tc      p.value
5 -----
6 b0  11.7833         0.6310     18.6744     0
7 b1   0.0084         0.0034     2.4955    0.0209
8 -----
9
10 R2 of linear model
11 -----
12 0.658681
13 -----
14
15 Analysis of Variance of linear model
16 =====
17                DF      SS      MS      Fc      p.value
18 -----
19 Linear Effect      1      10.6260  10.6260  6.23  0.0209
20 Lack of fit        2       5.5062   2.7531   1.61  0.22272
21 Residuals          21.2006  36.1756  1.7064
22 -----

```

- Análisis del modelo cuadrático

```

1 Quadratic Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0 12.1792      0.7351      16.5685    0
6 b1 -0.0035      0.0118      -0.2930 0.7724
7 b2 0.00004      0.00004      1.0497 0.3057
8 -----
9
10 R2 of quadratic model
11 -----
12 0.775231
13 -----
14
15 Analysis of Variance of quadratic model
16 =====
17                                DF      SS      MS      Fc  p.value
18 -----
19 Linear Effect          1      10.6260 10.6260 6.23 0.0209
20 Quadratic Effect      1       1.8802  1.8802  1.1 0.30567
21 Lack of fit           1       3.6260  3.6260  2.13 0.15956
22 Residuals             21.2006 36.1756  1.7064
23 -----

```

■ Análisis del modelo cubico

```

1 Cubic Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0 11.9333      0.7542      15.8230    0
6 b1  0.0351      0.0289      1.2114 0.2391
7 b2 -0.0003      0.0003      -1.2871 0.2120
8 b3 0.000001      0           1.4578 0.1596
9 -----
10
11 R2 of cubic model
12 -
13 1
14 -
15
16 Analysis of Variance of cubic model
17 =====
18                                DF      SS      MS      Fc  p.value
19 -----
20 Linear Effect          1      10.6260 10.6260 6.23 0.0209

```

21	Quadratic Effect	1	1.8802	1.8802	1.1	0.30567
22	Cubic Effect	1	3.6260	3.6260	2.13	0.15956
23	Lack of fit	0	0	0	0	1
24	Residuals	21.2006	36.1756	1.7064		
25	-----					

Teniendo en cuenta los resultados anteriores el coeficiente de correlación (R^2), el modelo que mejor se ajusta es el cúbico ya que es igual a uno.

- A continuación se encuentra el ajuste de los modelos lineal, cuadrático y cubico, su respectivo R^2 y por último el análisis de varianza. Para el Nitrógeno dentro de Potasio 1000.
- Análisis del modelo lineal

```

1 Linear Model
2 =====
3 Estimate Standard.Error tc p.value
4 -----
5 b0 10.4513 0.6310 16.5634 0
6 b1 0.0132 0.0034 3.9226 0.0008
7 -----
8
9 R2 of linear model
10 -----
11 0.710780
12 -----
13
14 Analysis of Variance of linear model
15 =====
16 DF SS MS Fc p.value
17 -----
18 Linear Effect 1 26.2549 26.2549 15.39 0.00077
19 Lack of fit 2 10.6833 5.3416 3.13 0.06438
20 Residuals 21.2006 36.1756 1.7064
21 -----

```

El coeficiente de determinación (R^2) para el modelo lineal es 0.710780, lo que indica que el modelo explica el 71.08 % de la variabilidad total de los datos.

En el análisis de varianza, se encuentra que el efecto lineal del factor es significativo, ya que el p_{valor} para el efecto lineal es 0.00077. Sin embargo, se observa falta de ajuste en el modelo, ya que el p_{valor} para

la prueba de falta de ajuste es 0.06438. Esto sugiere que el modelo lineal puede no capturar completamente la variabilidad de los datos y que podría ser necesario considerar términos adicionales en el modelo.

- Análisis del modelo cuadrático

```

1 Quadratic Model
2 =====
3 Estimate Standard.Error tc p.value
4 -----
5 b0 11.3872 0.7351 15.4910 0
6 b1 -0.0148 0.0118 -1.2575 0.2222
7 b2 0.0001 0.00004 2.4817 0.0215
8 -----
9
10 R2 of quadratic model
11 -----
12 0.995293
13 -----
14
15 Analysis of Variance of quadratic model
16 =====
17 DF SS MS Fc p.value
18 -----
19 Linear Effect 1 26.2549 26.2549 15.39 0.00077
20 Quadratic Effect 1 10.5094 10.5094 6.16 0.02153
21 Lack of fit 1 0.1739 0.1739 0.1 0.75268
22 Residuals 21.2006 36.1756 1.7064
23 -----

```

El coeficiente de determinación (R^2) para el modelo cuadrático es 0.995293, lo que indica que el modelo explica el 99.53 % de la variabilidad total de los datos.

En el análisis de varianza, se encuentra que tanto el efecto lineal como el efecto cuadrático del factor son significativos, ya que los p_{valor} para ambos efectos son inferiores a 0.05. Sin embargo, se observa falta de ajuste en el modelo, ya que el p_{valor} para la prueba de falta de ajuste es 0.75268. Esto sugiere que el modelo cuadrático puede no capturar completamente la variabilidad de los datos y que podría ser necesario considerar términos adicionales en el modelo.

- Análisis del modelo cubico

```

1 Cubic Model
2 =====
3      Estimate Standard.Error    tc    p.value
4 -----
5 b0 11.3333      0.7542      15.0274    0
6 b1 -0.0064      0.0289      -0.2216  0.8268
7 b2 0.00001      0.0003      0.0502   0.9604
8 b3 0.000000      0           0.3192   0.7527
9 -----
10
11 R2 of cubic model
12 -
13 1
14 -
15
16 Analysis of Variance of cubic model
17 =====
18                DF      SS      MS      Fc      p.value
19 -----
20 Linear Effect      1      26.2549  26.2549  15.39  0.00077
21 Quadratic Effect   1      10.5094  10.5094   6.16  0.02153
22 Cubic Effect       1       0.1739   0.1739   0.1   0.75268
23 Lack of fit        0       0         0         0     1
24 Residuals          21.2006  36.1756  1.7064
25 -----

```

El coeficiente de determinación (R^2) para el modelo cúbico es 1, lo que indica que el modelo explica el 100% de la variabilidad total de los datos. Sin embargo, es importante tener en cuenta que un (R^2) de 1 puede ser indicativo de sobreajuste en el modelo.

En el análisis de varianza, se encuentra que tanto el efecto lineal como el efecto cuadrático del factor son significativos, ya que los p_{valor} para ambos efectos son inferiores a 0.05. Sin embargo, el efecto cúbico no es significativo, ya que el p_{valor} es 0.75268. Además, se observa falta de ajuste en el modelo, ya que el p_{valor} para la prueba de falta de ajuste es 1. Esto sugiere que el modelo cúbico puede no capturar completamente la variabilidad de los datos y que podría ser necesario considerar términos adicionales en el modelo.

- Análisis de varianza

```

1 -----
2 Analyzing Potasio inside of each level of Nitrogeno

```

```

3 -----
4           DF          SS          MS          Fc
5 Potasio : Nitrogeno 0      2  9.36000  4.680000  2.853453
6 Potasio : Nitrogeno 100   2 13.66889  6.834444  4.167044
7 Potasio : Nitrogeno 200   2 24.85369 12.426844  7.576799
8 Potasio : Nitrogeno 300   2 35.26220 17.631100 10.749897
9 Error b                    16 26.24189  1.640118      NA
10 -----
11 p.value
12 0.087132
13 0.034933
14 0.004841
15 0.001098
16 NA
17 -----
18      Levels      Means
19 1   1000    11.333333
20 2    200     9.533333
21 3    600    11.933333

```

El potasio influye con preferencia al nitrógeno con niveles de 200 y 300. Lo cual explica la interacción. Según la prueba F, las medias de este factor no son diferentes.

- A continuación se encuentra el ajuste de los modelos lineal y cuadrático, su respectivo R^2 y por último el análisis de varianza. Para el Potasio dentro de Nitrógeno 100.
- Análisis del modelo lineal

```

1 Linear Model
2 =====
3      Estimate Standard.Error   tc   p.value
4 -----
5 b0 10.5722      0.8929      11.8402   0
6 b1  0.0013      0.0013      0.9563  0.3531
7 -----
8
9 R2 of linear model
10 -----
11 0.109738
12 -----
13
14 Analysis of Variance of linear model
15 =====

```

	DF	SS	MS	Fc	p.value
Linear Effect	1	1.5000	1.5000	0.91	0.35313
Lack of fit	1	12.1689	12.1689	7.42	0.01502
Residuals	16	26.2419	1.6401		

El coeficiente de determinación (R^2) para el modelo lineal es 0.109738, lo que indica que el modelo explica aproximadamente el 10.97% de la variabilidad total de los datos.

En el análisis de varianza, se encuentra que el efecto lineal del factor no es significativo, ya que el p_{valor} para el efecto lineal es 0.35313. Sin embargo, se observa falta de ajuste en el modelo, ya que el p_{valor} para la prueba de falta de ajuste es 0.01502. Esto sugiere que el modelo lineal puede no capturar completamente la variabilidad de los datos y que podría ser necesario considerar términos adicionales en el modelo o explorar otras relaciones no lineales.

■ Análisis del modelo cuadrático

```

1 Quadratic Model
2 =====
3 Estimate Standard.Error tc p.value
4 -----
5 b0 6.6667 1.6891 3.9468 0.0012
6 b1 0.0198 0.0069 2.8555 0.0114
7 b2 -0.00002 0.00001 -2.7239 0.0150
8 -----
9
10 R2 of quadratic model
11 -
12 1
13 -
14
15 Analysis of Variance of quadratic model
16 =====
17 DF SS MS Fc p.value
18 -----
19 Linear Effect 1 1.5000 1.5000 0.91 0.35313
20 Quadratic Effect 1 12.1689 12.1689 7.42 0.01502
21 Lack of fit 0 0 0 0 1
22 Residuals 16 26.2419 1.6401
23 -----

```

El coeficiente de determinación (R^2) para el modelo cuadrático es 1, lo que indica que el modelo explica el 100 % de la variabilidad total de los datos. Esto sugiere que el modelo cuadrático ajusta perfectamente los datos observados.

En el análisis de varianza, se encuentra que tanto el efecto lineal como el efecto cuadrático son significativos. Los p_{valor} para el efecto lineal y el efecto cuadrático son 0.35313 y 0.01502, respectivamente. Esto indica que tanto el término lineal como el término cuadrático tienen un efecto significativo en la respuesta.

No se encuentra evidencia de falta de ajuste en el modelo, ya que el p_{valor} para la prueba de falta de ajuste es 1. Esto sugiere que el modelo cuadrático captura bien la variabilidad de los datos observados.

- A continuación se encuentra el ajuste de los modelos lineal y cuadrático, su respectivo R^2 y por último el análisis de varianza. Para el Potasio dentro de Nitrógeno 200.
- Análisis del modelo lineal

```

1 Linear Model
2 =====
3 Estimate Standard.Error tc p.value
4 -----
5 b0 16.0978 0.8929 18.0285 0
6 b1 -0.0046 0.0013 -3.5193 0.0029
7 -----
8
9 R2 of linear model
10 -----
11 0.817327
12 -----
13
14 Analysis of Variance of linear model
15 =====
16 DF SS MS Fc p.value
17 -----
18 Linear Effect 1 20.3136 20.3136 12.39 0.00285
19 Lack of fit 1 4.5401 4.5401 2.77 0.11562
20 Residuals 16 26.2419 1.6401
21 -----

```

- Análisis del modelo cuadrático

```

1 Quadratic Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0 18.4833      1.6891      10.9426    0
6 b1 -0.0159      0.0069      -2.2989 0.0353
7 b2 0.00001      0.00001     1.6638 0.1156
8 -----
9
10 R2 of quadratic model
11 -
12 1
13 -
14
15 Analysis of Variance of quadratic model
16 =====
17                DF    SS        MS        Fc    p.value
18 -----
19 Linear Effect    1  20.3136  20.3136  12.39  0.00285
20 Quadratic Effect 1   4.5401   4.5401   2.77  0.11562
21 Lack of fit      0    0         0         0     1
22 Residuals       16 26.2419  1.6401
23 -----

```

Teniendo en cuenta los resultados anteriores el coeficiente de correlación (R^2), el modelo que mejor se ajusta es el cuadrático ya que es igual a uno.

- A continuación se encuentra el ajuste de los modelos lineal, cuadrático y cubico, su respectivo R^2 y por último el análisis de varianza. Para el Potasio dentro de Nitrógeno 300.

- Análisis del modelo lineal

```

1 Linear Model
2 =====
3      Estimate Standard.Error   tc    p.value
4 -----
5 b0 10.4792      0.8929      11.7360    0
6 b1  0.0055      0.0013      4.2174 0.0006
7 -----
8
9 R2 of linear model
10 -----

```

```

11 0.827292
12 -----
13
14 Analysis of Variance of linear model
15 =====
16                DF    SS      MS      Fc    p.value
17 -----
18 Linear Effect  1    29.1721  29.1721  17.79  0.00065
19 Lack of fit    1     6.0900   6.0900   3.71   0.07193
20 Residuals     16    26.2419   1.6401
21 -----

```

■ Análisis del modelo cuadrático

```

1
2 Quadratic Model
3 =====
4      Estimate Standard.Error    tc    p.value
5 -----
6 b0    7.7162         1.6891     4.5682  0.0003
7 b1    0.0186         0.0069     2.6893  0.0161
8 b2   -0.00001      0.00001    -1.9270  0.0719
9 -----
10
11 R2 of quadratic model
12 -
13 1
14 -
15
16 Analysis of Variance of quadratic model
17 =====
18                DF    SS      MS      Fc    p.value
19 -----
20 Linear Effect    1    29.1721  29.1721  17.79  0.00065
21 Quadratic Effect 1     6.0900   6.0900   3.71   0.07193
22 Lack of fit      0     0         0         0       1
23 Residuals       16    26.2419   1.6401
24 -----

```

Teniendo en cuenta los resultados anteriores el coeficiente de correlación (R^2), el modelo que mejor se ajusta es el cuadrático ya que es igual a uno.

6.14. Medidas repetidas en el tiempo

Se lleva a cabo cuando una unidad experimental es observada en diferentes períodos de tiempo, después de aplicar cierto tratamiento. Si el factor estudiado consta de un sólo tratamiento, entonces en realidad se busca es la curva de tendencia de la variable de estudio Y , en el tiempo para dicho tratamiento.

En el caso de tener varios tratamientos, cuyo efecto se mide en diferentes períodos de tiempo, entonces el experimento es similar a cuando se lleva a cabo un diseño en parcelas divididas, con la diferencia que como el tiempo no se puede aleatorizar, entonces los períodos de tiempo, que corresponden a las subparcelas, no cumplen este proceso, por eso este diseño también es conocido como parcelas divididas en el tiempo.

El hecho de no tener aleatorización en las subparcelas, lleva a que deban cumplirse ciertos requerimientos para que éste diseño se pueda analizar como un diseño en parcelas divididas.

6.14.1. Supuestos del modelo

El modelo estadístico es:

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

El modelo de análisis de varianza usual supone que los errores son variables aleatorias independientes con distribución normal media cero y varianza constante. La independencia en las observaciones garantiza covarianza cero. En mediciones repetidas la aleatorización de los tratamientos no garantiza la independencia de las observaciones dado que el mismo sujeto o unidad experimental está siendo medida en diferentes tiempos, lo que lleva a pensar que existe cierto grado de asociación entre las observaciones, pero se espera que la correlación sea constante para cualquier aleatorización que se haga, porque como se dijo anteriormente, aleatorizar no garantiza la independencia entre medidas de la misma unidad experimental en tiempos distintos.

Gómez (1997) presenta una metodología para llevar a cabo el análisis en medidas repetidas; esta se basa en tomar los efectos de las subparcelas (periodos de tiempo) como aleatorios, esto ocasiona que el error de las subparcelas en

el diseño de parcelas divididas tenga dos componentes, una debida a la variación entre subparcelas y la otra debida a la variación de la interacción entre repeticiones y los periodos de tiempo, es decir que se debe introducir una nueva fuente de variación en la tabla ANAVA, constituida por la interacción de las repeticiones con el factor tiempo. Para un experimento en un diseño completamente al azar con r repeticiones donde se estudian los niveles del factor A , con a tratamientos, el ANAVA es dado en la Tabla 52 donde (T) es el tiempo y (Rep) representa las repeticiones.

Tabla 52: ANAVA para una estructura en parcelas divididas en un DCA

F.V.	gl	SC	CM	F_c
A	$a - 1$	SC_A	CM_A	$\frac{CM_A}{CM_{E^*(a)}}$
$E(a)$	$a(r - 1)$	$SC_{E(a)}$	$CM_{E(a)}$	
<i>Total parc.ppal</i>	$ar - 1$	$SC_{Total}(PP)$		
<i>Tiempo</i>	$t - 1$	SC_T	CM_T	$\frac{CM_T}{CM_{Rep \times T}}$
$A \times \textit{Tiempo}$	$(a - 1)(t - 1)$	SC_{AT}	CM_{AT}	$\frac{CM_{AT}}{CM_{E(t)}}$
$Rep \times T$	$(r - 1)(t - 1)$	$SC_{Rep \times T}$	$CM_{Rep \times T}$	
$E(t)$	$(a - 1)(r - 1)(t - 1)$	$SC_{E(t)}$	$CM_{E(t)}$	
<i>Total</i>	$art - 1$	SC_{Total}		

Siendo $CM_{E^*(a)} = CM_{E(a)} + CM_{AT} - CM_{E(t)}$ con grados de libertad según Satterthwhite,

$$gl_{E^*(A)} = \frac{(CM_{E(a)} + CM_{AT} - CM_{E(t)})^2}{\frac{(CM_{E(a)})^2}{gl_{E(a)}} + \frac{(CM_{AT})^2}{gl_{AT}} + \frac{(CM_{E(t)})^2}{gl_{E(t)}}} \quad (39)$$

Si la interacción entre los factores A y T no se considera significativa, entonces el cuadrado medio para llevar a cabo la prueba de hipótesis sobre el factor A es el $CM_{E(a)}$. Las sumas de cuadrados son:

$$SC_{Total} = \sum_{i=1}^r \sum_{j=1}^a \sum_{k=1}^t Y_{ijk}^2 - \frac{Y_{...}^2}{rat},$$

$$SC_{Total}(Parc.ppal) = \sum_{i=1}^r \sum_{j=1}^a \frac{Y_{ij.}^2}{b} - FC,$$

donde $FC = \frac{Y_{..}^2}{rat}$,

$$SC_{E(a)} = SC_{Total}(Parc.ppal) - SC_A,$$

$$SC_T = \sum_{k=1}^b \frac{Y_{..K}^2}{ra} - \frac{Y_{..}^2}{rat},$$

$$SC_{A \times T} = \left[\sum_{j=1}^a \sum_{k=1}^b \frac{Y_{.jk}^2}{r} - \frac{Y_{..}^2}{rat} \right] - SC_A - SC_T,$$

$$SC_{Rep} = \sum_{i=1}^r \frac{Y_{i..}^2}{at} - \frac{Y_{..}^2}{rat},$$

$$SC_{Rep \times T} = \left[\sum_{i=1}^r \sum_{k=1}^t \frac{Y_{i.k}^2}{r} - \frac{Y_{..}^2}{rat} \right] - SC_{Rep} - SC_T,$$

$$SC_{E(t)} = SC_{Total} - SC_{Total}(parc.ppal) - SC_T - SC_{A \times T} - SC_{Rep \times T}.$$

6.14.2. Errores estándar para comparaciones múltiples.

Para llevar a cabo comparaciones múltiples con este tipo de análisis, los errores estándar para cada situación son los siguientes:

1) Para comparar medias de los niveles de A , el error estándar es:

$$S_{\bar{Y}_{.j.}} = \sqrt{\frac{CM_{E(a)}}{rt}}.$$

2) Para comparar medias de los niveles de T , el error estándar es:

$$S_{\bar{Y}_{..k}} = \sqrt{\frac{CM_{Rep \times T}}{ar}}.$$

3) Para comparar medias de los niveles de A , en el tiempo k -ésimo, el error estándar es:

$$S_{\bar{Y}.j.} = \sqrt{\frac{CM_{E_{(ak)}}}{r}},$$

donde $CM_{E_{(ak)}}$ es el cuadrado medio del análisis de varianza en el tiempo k -ésimo.

4) Para comparar medias de los niveles de T , k y k' , para el tratamiento j -ésimo

$$S_{\bar{Y}..k} = \sqrt{\frac{CM_{(E_k)} + CM_{(E_{k'})}}{2r}},$$

donde $CM_{(E_k)}$ y $CM_{(E_{k'})}$ son los cuadrados medios de los errores al realizar un análisis de varianza en los tiempos k y k' .

6.14.3. Función (MRT.DCA)

Como R no aporta una función que vaya dirigida a realizar el análisis de medidas repetidas en el tiempo con un DCA, se programó una función llamada (MRT.DCA).

- La función (MRT.DCA) está compuesta por:

```
1 > MRT.DCA=function(Y,x1,x2,rep,quali,mcomp,sigF,sigT)
```

- Características

Y: variable respuesta.

x1: variable de interes 1.

x2: variable de interes 2.

quali: Lógica. Si es TRUE (por defecto), los tratamientos se asumen cualitativos, si es FALSE, cuantitativos.

mcomp: permite elegir la prueba de comparación múltiple; el valor predeterminado es la prueba de Tukey, sin embargo, las opciones son: la prueba LSD ('lsd'), la prueba LSD con Bonferroni ('lsdb'), el test de Duncan ('duncan'),

el test de Student-NewmanKeuls ('snk'), el test de Scott-Knott ('sk'), el test de Calinski y Corsten ('ccF') y prueba de comparación múltiple bootstrap ('ccboot').

sigT: la significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 %.

sigF: la significancia que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5 %.

Esta función aporta el análisis de varianza, analisis de la interaccion, prueba de normalidad Shapiro-Wilk, prueba de homogeneidad de varianza de Bartlett y las comparaciones de medias.

6.14.4. Ejemplo en R de medidas repetidas bajo un DCA

Argel *et al.*, (2006) llevaron a cabo un experimento para estudiar el área de cobertura (cm^2), de la maleza rabo ratón al suelo, en cuatro periodos diferentes, a los 28, 56, 84 y 112 días a cuatro densidades de planta, el diseño experimental fue completamente al azar con 4 repeticiones por tratamiento. Los resultados obtenidos son dados por la Tabla 53. (Martínez y Robles, 2013, p.210)

Tabla 53: Densidad de siembra y área de cobertura de la maleza rabo ratón

Días	Den	Rep	Cob	Días	Den	Rep	Cob
28	1	1	75	28	4	2	300
28	1	2	100	28	4	3	325
28	1	3	100	56	1	1	125
28	2	1	125	56	1	2	150
28	2	2	125	56	1	3	150
28	2	3	150	56	2	1	250
28	3	1	250	56	2	2	300
28	3	2	225	56	2	3	350
28	3	3	225	56	3	1	500
28	4	1	275	56	3	2	475
56	3	3	650	84	3	1	875
56	4	1	500	84	3	2	700
56	4	2	500	84	3	3	1050
56	4	3	525	84	4	1	800
84	1	1	200	84	4	2	800
84	1	2	250	84	4	3	875
84	1	3	275	112	1	1	300
84	2	1	325	112	1	2	400
84	2	2	400	112	1	3	425
84	2	3	450	112	2	1	500
112	2	2	500	112	3	3	1700
112	2	3	675	112	4	1	1150
112	3	1	1425	112	4	2	1175
112	3	2	1250	112	4	3	1325

Así, las hipótesis a verificar son:

$$1) H_0 : (\alpha\beta)_{jk} = 0 \text{ contra } H_1 : (\alpha\beta)_{jk} \neq 0$$

$$2) H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de los tratamientos es diferente de cero

$$3) H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de las covariables es diferente de cero

Para crear la base de datos de acuerdo a este ejemplo debe tener una secuencia teniendo en cuenta el total de repeticiones en este caso son 3, entonces si se tiene $r = 1$, la densidad es 1 y tiene su respectivo valor de días y de cobertura, si se tiene $r = 2$, la densidad es 1 y tiene su respectivo valor de días y de cobertura cuando se llegue a $r = 3$ y la densidad es 1, se comienza de nuevo con $r = 1$ y densidad 2, así sucesivamente.

Los pasos a seguir para solucionar este ejemplo en R son los siguientes.

1. Cargar la base de datos.

```
1 > datos1=read.csv2(file.choose());datos1
2   Dias Densidad Rep Cobertura
3 1 28     1     1    75
4 2 28     1     2   100
5 3 28     1     3   100
6 4 28     2     1   125
7 5 28     2     2   125
8 6 28     2     3   150
```

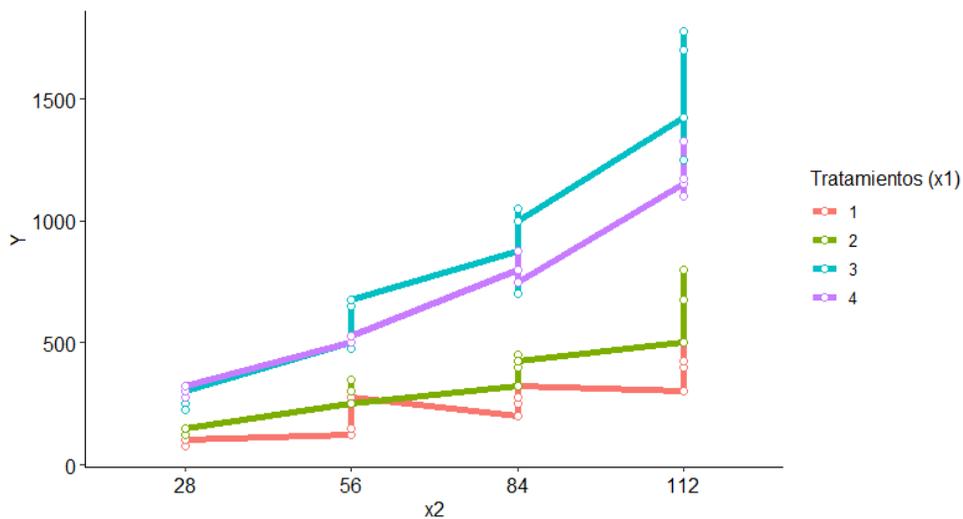
2. Se creó la función MRT.DCA, para llevar a cabo este análisis sobre las medidas repetidas en el tiempo con un DCA. Este paso es opcional aquí cambiamos el nombre de las variables originales por aquellas que en lo personal se consideran más cómodas, por ejemplo Y es la cobertura, x1 la densidad, x2 los días y rep las repeticiones.

```
1 > Y=datos1$Cobertura
2 > x1=as.factor(datos1$Densidad)
3 > x2=as.factor(datos1$Dias)
4 > rep=as.factor(datos1$Rep)
5 > MRT.DCA=function(Y,x1,x2,rep,quali,mcomp,sigF,sigT)
```

- Empleando la función (MRT.DCA) se obtienen los siguientes resultados.

```
1 > MRT.DCA=function(Y,x1,x2,rep,quali,mcomp,sigF,sigT)
```

- Gráfico para observar las variables de estudio.



El gráfico aporta una visualización sobre el comportamiento de la cobertura (Y), para distintos periodos (x2) con respecto a cada tratamiento (x1).

■ Análisis de varianza general

```

1 -----
2 Analysis of Variance Table
3 -----
4           GL          SC          CM          Fcal  valor.P
5 Treat          3  3654726.56  1218242.1875    7.5537  0.0042
6 E(a)          12   283359.38    23613.2812
7 Time           3  4822304.69  1607434.8958  206.3248    0
8 Treat x Time   9  1270429.69   141158.8542   40.4012    0
9 Rep x Time     9    70117.19    7790.7986
10 E(t)          27   94335.94    3493.9236
11 Total         63 10195273.44
12 -----
13 CV 1 = 29.054 %
14 CV 2 = 11.176 %

```

Los resultados indican que tanto el tratamiento como el tiempo tienen

efectos significativos en la variable de interés. Además, se observa una interacción significativa entre el tratamiento y el tiempo. Sin embargo, no se encuentra una interacción significativa entre la repetición y el tiempo.

- Supuestos del modelo general

```

1 -----
2 Shapiro-Wilk normality test
3
4   Shapiro-Wilk normality test
5
6 data:  eijk
7 W = 0.89164, p-value = 3.93e-05
8 -----
9 Bartlett homogeneity test
10
11   Bartlett test of homogeneity of variances
12
13 data:  eijk by Fk
14 Bartlett's K-squared = 7.227, df = 3, p-value = 0.065
15 -----

```

Los datos no siguen una distribución normal según el test de Shapiro-Wilk, pero no hay evidencia suficiente para rechazar la hipótesis de homogeneidad de varianza según el test de Bartlett. Esto implica que los datos no cumplen el supuesto de normalidad, pero sí cumplen el supuesto de homogeneidad de varianza para realizar análisis estadísticos adicionales.

- Análisis de varianza para cada periodo de tiempo con sus respectivos supuestos del modelo.
- Análisis de varianza por periodo 28 de tiempo, prueba de normalidad y prueba de homogeneidad de varianza.

```

1 -----
2 Analysis of Variance by period 28 of time
3 -----
4 Analysis of Variance Table
5
6 Response: y
7      Df Sum Sq Mean Sq F value    Pr(>F)
8 Treatments  3 115781    38594  70.571 6.823e-08 ***

```

```

9 Residuals 12 6562 547
10 -----
11 Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
12 0.05 '.' 0.1 1
13 -----
14 Shapiro-Wilk normality test
15 data: resid(mod)
16 W = 0.94923, p-value = 0.4775
17 -----
18 Bartlett test of homogeneity of variances
19
20 data: resid(mod) by Treatments
21 Bartlett's K-squared = 3.542, df = 3, p-value = 0.3154

```

Los resultados muestran que hay diferencias significativas entre los tratamientos, con un p_{valor} muy pequeño ($6.823e-08$). Esto indica que al menos uno de los tratamientos tiene un efecto significativo en la variable de respuesta.

La prueba de normalidad de Shapiro-Wilk se realizó en los residuos del modelo y arrojó un estadístico W de 0.94923, con un p_{valor} de 0.4775. Esto indica que no hay suficiente evidencia para rechazar la hipótesis nula de normalidad de los residuos.

La prueba de homogeneidad de varianzas de Bartlett también se realizó en los residuos del modelo, pero esta vez se agruparon por tratamientos. El estadístico K-squared de Bartlett fue de 3.542, con 3 grados de libertad y un p_{valor} de 0.3154. Esto indica que no hay suficiente evidencia para rechazar la hipótesis nula de homogeneidad de varianzas entre los tratamientos.

- Análisis de varianza por periodo 56 de tiempo, prueba de normalidad y prueba de homogeneidad de varianza.

```

1 -----
2 Analysis of Variance by period 56 of time
3 -----
4 Analysis of Variance Table
5
6 Response: y
7      Df Sum Sq Mean Sq F value    Pr(>F)
8 Treatments  3 423750   141250  32.286 5.008e-06 ***
9 Residuals  12  52500     4375

```

```

10 -----
11 Signif. codes: 0    ***    0.001    **    0.01    *
      0.05    .    0.1    1
12 -----
13 Shapiro-Wilk normality test
14
15 data:  resid(mod)
16 W = 0.94582, p-value = 0.4265
17 -----
18 Bartlett test of homogeneity of variances
19
20 data:  resid(mod) by Treatments
21 Bartlett's K-squared = 7.3154, df = 3, p-value =
22 0.0625

```

De acuerdo con el análisis de varianza realizado para el período 56 del tiempo, se obtuvo un valor de F de 32.286 con un p_{valor} muy pequeño (5.008e-06), lo que indica diferencias significativas entre los tratamientos. Los códigos de significancia muestran que el resultado es altamente significativo.

En relación a los supuestos, se llevó a cabo la prueba de normalidad de Shapiro-Wilk para los residuos del modelo. El valor de estadístico W fue de 0.94582, y el p_{valor} asociado fue de 0.4265. Estos resultados indican que no hay suficientes evidencias para rechazar la hipótesis nula de normalidad de los residuos.

Asimismo, se realizó la prueba de homogeneidad de varianzas de Bartlett, agrupando los residuos por tratamientos. El valor de K-squared de Bartlett fue de 7.3154, con 3 grados de libertad y un p_{valor} de 0.0625. Según estos resultados, no se encontraron suficientes evidencias para rechazar la hipótesis nula de homogeneidad de varianzas entre los tratamientos.

- Análisis de varianza por periodo 84 de tiempo, prueba de normalidad y prueba de homogeneidad de varianza.

```

1 -----
2 Analysis of Variance by period 84 of time
3 -----
4 Analysis of Variance Table
5
6 Response: y
7      Df  Sum Sq Mean Sq F value    Pr(>F)

```

```

8  Treatments  3 1160313  386771  47.45 6.256e-07 ***
9  Residuals  12  97812  8151
-----
10
11 Signif. codes:0    ***    0.001    **    0.01    *
    0.05    .    0.1    1
-----
12
13 Shapiro-Wilk normality test
14
15 data:  resid(mod)
16 W = 0.95327, p-value = 0.5432
-----
17
18 Bartlett test of homogeneity of variances
19
20 data:  resid(mod) by Treatments
21 Bartlett's K-squared = 5.6831, df = 3, p-value =
22 0.1281

```

- Análisis de varianza por periodo 112 de tiempo, prueba de normalidad y prueba de homogeneidad de varianza.

```

1  -----
2  Analysis of Variance by period 112 of time
3  -----
4  Analysis of Variance Table
5
6  Response: y
7      Df  Sum Sq Mean Sq F value  Pr(>F)
8  Treatments  3 3225313 1075104  44.344 9.07e-07 ***
9  Residuals  12  290937  24245
-----
10
11 Signif. codes:0    ***    0.001    **    0.01    *
    0.05    .    0.1    1
-----
12
13 Shapiro-Wilk normality test
14
15 data:  resid(mod)
16 W = 0.96682, p-value = 0.7848
-----
17
18 Bartlett test of homogeneity of variances
19
20 data:  resid(mod) by Treatments
21 Bartlett's K-squared = 3.8114, df = 3, p-value =
22 0.2826

```

El p_{valor} asociado al valor F se proporciona como 6.256e-07, lo cual es

extremadamente pequeño. Indica una fuerte evidencia en contra de la hipótesis nula de no haber diferencia entre los grupos de tratamiento.

La prueba de normalidad de Shapiro-Wilk muestra que los residuos del modelo siguen una distribución aproximadamente normal. El valor de la estadística de prueba (W) es 0.96682 y el p_{valor} asociado es 0.7848, lo cual indica que no hay suficiente evidencia para rechazar la hipótesis nula de normalidad. Esto sugiere que los residuos se ajustan razonablemente bien a una distribución normal.

La prueba de Bartlett para la homogeneidad de varianzas muestra que las varianzas de los residuos entre los diferentes grupos de tratamiento son aproximadamente iguales. El valor de la estadística de prueba (K-cuadrado de Bartlett) es 3.8114, con 3 grados de libertad y un p_{valor} asociado de 0.2826. Dado que el p_{valor} es mayor que el nivel de significancia común de 0.05, no hay suficiente evidencia para rechazar la hipótesis nula de homogeneidad de varianzas.

- Análisis de la interacción para cada tratamiento dentro de cada nivel de tiempo
- Tratamientos dentro del nivel 28 del tiempo

```
1 -----
2 treatments inside of the level 28 of time
3 -----
4
5 Tukey's test
6 -----
7 Groups Treatments Means
8 a 4 306.25
9 b 3 250
10 c 2 137.5
11 c 1 93.75
12 -----
```

Según los resultados de la prueba de Tukey, los grupos a y b no son estadísticamente diferentes entre sí, lo cual significa que no hay una diferencia significativa en las medias entre los tratamientos 4 y 3. Sin embargo, los grupos c y a, así como los grupos c y b, son estadísticamente diferentes entre sí, lo que indica que hay diferencias significativas en las medias entre los tratamientos 2 y 4, así como entre los tratamientos 2 y 3.

- Tratamientos dentro del nivel 56 del tiempo

```

1 -----
2 treatments  inside of the level  56  of  time
3 -----
4 -----
5 Tukey's test
6 -----
7 Groups  Treatments  Means
8   a      3           575
9   a      4           512.5
10  b      2           287.5
11  b      1           175
12 -----

```

- Tratamientos dentro del nivel 84 del tiempo

```

1 -----
2 treatments  inside of the level  84  of  time
3 -----
4 -----
5 Tukey's test
6 -----
7 Groups  Treatments  Means
8   a      3           906.25
9   a      4           806.25
10  b      2           400
11  b      1           262.5
12 -----

```

Según los resultados de la prueba de Tukey, para los niveles 56 y 88, los grupos a y b son estadísticamente diferentes entre sí, lo cual indica que hay diferencias significativas en las medias entre los tratamientos 3 y 2, así como entre los tratamientos 3 y 1, y entre los tratamientos 4 y 2, así como entre los tratamientos 4 y 1.

- Tratamientos dentro del nivel 112 del tiempo

```

1 -----
2 treatments  inside of the level  112  of  time
3 -----
4 -----
5 Tukey's test
6 -----
7 Groups  Treatments  Means
8   a      3           1537.5

```

9	b	4	1187.5
10	c	2	618.75
11	c	1	406.25
12	-----		

Según los resultados de la prueba de Tukey, los grupos a y b son estadísticamente diferentes entre sí, lo cual indica que hay diferencias significativas en las medias entre los tratamientos 3 y 4. Además, los grupos c y a, así como los grupos c y b, también son estadísticamente diferentes entre sí, lo que indica que hay diferencias significativas en las medias entre los tratamientos 2 y 3, así como entre los tratamientos 2 y 4, y entre los tratamientos 1 y 3, así como entre los tratamientos 1 y 4.

6.14.5. Medidas repetidas bajo un DBCA

Si el experimento es en bloques completos al azar entonces el ANAVA es dado en la Tabla 54.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

donde:

- Y_{ij} : es la observación en la i -ésima unidad experimental, en el j -ésimo bloque y en el k -ésimo momento de tiempo.
- μ : es la media global.
- α_i : es el efecto fijo del i -ésimo tratamiento.
- β_j : es el efecto fijo del j -ésimo bloque.
- γ_k : es el efecto fijo del k -ésimo momento de tiempo.
- $(\alpha\beta)_{ij}$: es la interacción entre el i -ésimo tratamiento y el j -ésimo bloque.
- $(\alpha\gamma)_{ik}$: es la interacción entre el i -ésimo tratamiento y el k -ésimo momento de tiempo.
- $(\beta\gamma)_{jk}$: es la interacción entre el j -ésimo bloque y el k -ésimo momento de tiempo.

- ϵ_{ij} : es el error aleatorio asociado a la i -ésima unidad experimental, el j -ésimo bloque y el k -ésimo momento de tiempo..

En este tipo de diseño, se utilizan bloques para controlar la variabilidad inherente a las condiciones del experimento. Los bloques se seleccionan de tal manera que las unidades experimentales dentro de cada bloque sean similares. Luego, los tratamientos se asignan aleatoriamente a las unidades experimentales dentro de cada bloque.

Tabla 54: Análisis de Varianza para un diseño en medidas repetidas bajo un DBCA

F.V.	gl	SC	CM	F_c
Bloques	$r - 1$	$SC_{Bloques}$	$CM_{Bloques}$	$\frac{CM_{Bloques}}{CM_{E^*(a)}}$
A	$a - 1$	SC_A	CM_A	$\frac{CM_A}{CM_{E^*(a)}}$
E(a)	$(a - 1)(r - 1)$	$SC_{E(a)}$	$CM_{E(a)}$	
Total parcela principal	$ar - 1$	$SC_{Total(PP)}$		
Tiempo(T)	$t - 1$	SC_T	CM_T	$\frac{CM_T}{CM_{Bloques*T}}$
$A \times Tiempo$	$(a - 1)(t - 1)$	SC_{AT}	CM_{AT}	$\frac{CM_{AT}}{CM_{E(t)}}$
$Bloques \times Tiempo$	$(r - 1)(t - 1)$	$SC_{Bloques \times T}$	$CM_{Bloques * T}$	
E(t)	$(a - 1)(r - 1)(t - 1)$	$SC_{E(t)}$	$CM_{E(t)}$	
Total	$art - 1$	SC_{Total}		

Donde

$$SC_{Total} = \sum_{i=1}^r \sum_{j=1}^a \sum_{k=1}^t Y_{ijk} - \frac{Y_{...}^2}{rat},$$

$$SC_{Total}(Parc.ppal) = \sum_{i=1}^r \sum_{j=1}^a \frac{Y_{ij.}^2}{t} - FC \text{ con } FC = \frac{Y_{...}^2}{rat},$$

$$SC_{Bloques} = \sum_{i=1}^r \frac{Y_{i..}}{at} - \frac{Y_{...}^2}{rat},$$

$$SC_A = \sum_{j=1}^a \frac{Y_{.j}}{rt} - \frac{Y_{...}^2}{rat},$$

$$SC_{E(a)} = SC_{E(a)} = SC_{Total}(Parc.ppal) - SC_{Bloques} - SC_A,$$

$$SC_T = \sum_{k=1}^t \frac{Y_{.K}}{ra} - \frac{Y_{...}^2}{rat},$$

$$SC_{A \times T} = \left[\sum_{j=1}^a \sum_{k=1}^b \frac{Y_{.jk}^2}{r} - \frac{Y_{...}^2}{rat} \right] - SC_A - SC_T,$$

$$SC_{Bloques \times T} = \left[\sum_{i=1}^r \sum_{k=1}^t \frac{Y_{i.k}^2}{r} - \frac{Y_{...}^2}{rat} \right] - SC_{Bloques} - SC_T,$$

$$SC_{E(t)} = SC_{Total} - SC_{Total}(Parc.ppal) \\ - SC_T - SC_{A \times T} - SC_{Bloques \times T}.$$

6.14.6. Función (MRT.DBCA)

Como R no aporta una función que vaya dirigida a realizar el análisis de medidas repetidas en el tiempo con un DBCA, se programó una función llamada (MRT.DBCA).

- La función (MRT.DBCA) está compuesto por:

```
1 > MRT.DBCA(Y,x1,x2,block,quali = T,mcomp="tukey",
2 sigF = 0.05,sigT = 0.05)
```

- Características

Y: variable respuesta.

x1: variable de interes 1.

x2: variable de interes 2.

block: vector numérico o complejo que contiene los bloques.

quali: lógica. Si es TRUE (por defecto), los tratamientos se asumen cualitativos, si es FALSE, cuantitativos.

mcomp: permite elegir la prueba de comparación múltiple; el valor predeterminado es la prueba de Tukey, sin embargo, las opciones son: la prueba LSD ('lsd'), la prueba LSD con Bonferroni ('lsdb'), el test de Duncan ('duncan'), el test de Student-NewmanKeuls ('snk'), el test de Scott-Knott ('sk') y prueba de comparación múltiple bootstrap ('ccboot').

sigT: la significación que se utilizará para la prueba de comparación múltiple; el valor predeterminado es 5 %.

sigF: la significancia que se utilizará para la prueba F de ANOVA; el valor predeterminado es 5 %.

Esta función aporta el análisis de varianza, analisis de la interaccion, prueba de normalidad Shapiro-Wilk, prueba de homogeneidad de varianza de Bartlett y las comparaciones de medias. Ver en anexo.

6.14.7. Ejemplo en R de medidas repetidas bajo un DBCA

Un investigador de suelos estudió la acidez en las plantas en siete niveles diferentes de pH en un medio controlado. Tenía cuatro contenedores, en un DBCA y lo midió cada hora. Los datos son los siguientes.

Figura 2: datos sobre la acidez en las plantas

Hora	Contenedor/TTO	Acidez						
		a_0	a_1	a_2	a_3	a_4	a_5	a_6
1	1	23,8	28,5	28	18,5	28,5	18,8	25,3
	2	20,3	38,4	32,2	20,9	18,2	27	24,2
	3	28	35,9	42,8	35,2	46,3	20,4	36,5
	4	20,5	34,3	41,2	30,6	38	35,1	38,3
2	1	26,8	31,7	31	26,5	30	21,6	29
	2	26	32	27	22,6	20,6	25,4	27,4
	3	26,5	37,2	46	25,6	38	27	30
	4	15,6	33,3	38	28	32	29,8	33,8
3	1	24,3	27,7	28,4	22,4	35,4	17,6	23,3
	2	20,8	28,6	29	22,8	17,6	25,8	25,2
	3	23,7	33,4	35,9	25,6	32,4	22,8	27,1
	4	19,8	28,4	32,9	24,5	29,2	31	34,8

Así, las hipótesis a verificar son:

$$1) H_0 : (\alpha\beta)_{jk} = 0 \text{ contra } H_1 : (\alpha\beta)_{jk} \neq 0$$

$$2) H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de los tratamientos es diferente de cero

$$3) H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_j = 0$$

contra

H_1 : Al menos uno de los coeficientes de las covariables es diferente de cero

Para formar la base de datos de acuerdo a este ejemplo debe tener en cuenta el periodo de tiempo en este caso es la hora, como se observa a continuación cuando la hora=1, se encuentran el bloque 1 con los dos tratamientos y su respectiva acidez.

Estos son los pasos para llevar a cabo este ejemplo en R.

1. Se cargan los datos

```

1 > datos1=read.csv2(file.choose());datos1
2     Periodo Bloques TTOS Acidez
3 1         1         1  a0   23.8
4 2         1         1  a1   28.5
5 3         1         1  a2   28.0
6 4         1         1  a3   18.5
7 5         1         1  a4   28.5
8 6         1         1  a5   18.8

```

2. Esta función (MRT.DBCA) fue creada para llevar a cabo el tema de medidas repetidas en el tiempo para un DBCA, aquí al igual que en el ejemplo anterior se cambian las variables, esto es completamente opcional.

```

1 > Y=datos1$Acidez
2 > x1=as.factor(datos1$TTOS)
3 > x2=as.factor(datos1$Periodo)
4 > block=as.factor(datos1$Bloques)

```

3. Empleando la función (MRT.DBCA) se obtienen los siguientes resultados.

```

1 > MRT.DBCA(Y,x1,x2,block,quali = T,mcomp="tukey",
2 sigF = 0.05,sigT = 0.05)

```

- Gráfico para observar las variables de estudio.

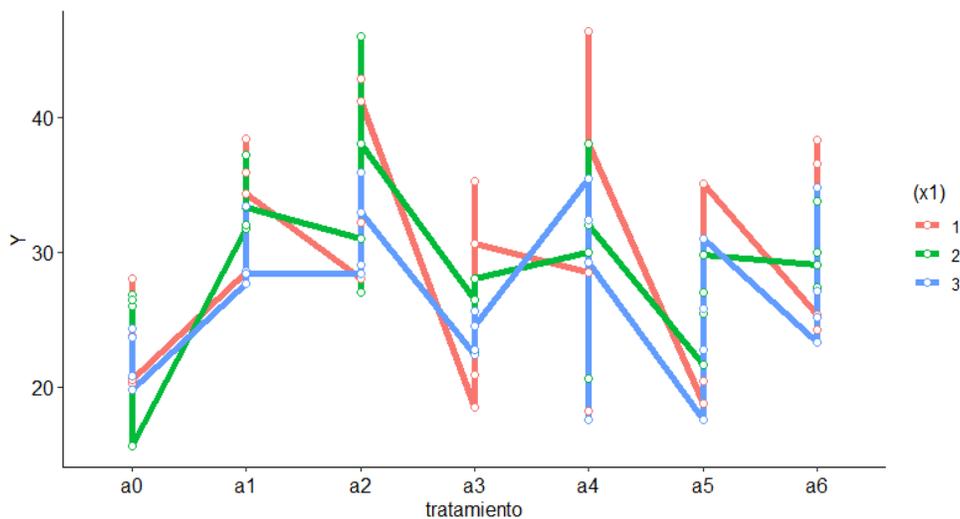
El gráfico aporta una visualización sobre el comportamiento de la acidez (Y), para cada uno de los tratamientos en las 3 horas.

- Análisis de varianza general

```

1 -----
2 Analysis of Variance Table
3 -----
4           GL          SC          CM      Fcal  valor.P
5 Block           3  745.2308  248.4103
6 Treat           6 1280.2967  213.3828  4.6532  0.0037
7 E(a)           21  962.9967   45.857
8 Time            2  145.2307   72.6154  2.617  0.1524
9 Treat x Time   12   39.1776    3.2648  0.3889  0.9822
10 Block x Time   6  166.4874   27.7479
11 E(t)           36  302.2176    8.3949
12 Total          83 3641.6375
13 -----

```



El análisis de varianza indica que el factor de tratamiento tiene un efecto significativo en los resultados, mientras que el factor de tiempo y la interacción entre tratamiento y tiempo no son significativos.

■ Supuestos del modelo general

```

1 -----
2 Shapiro-Wilk normality test
3
4 data:  eijk
5 W = 0.99539, p-value = 0.9927
6 -----
7 Bartlett homogeneity test
8
9 Bartlett test of homogeneity of variances
10
11 data:  eijk by Fk
12 Bartlett's K-squared = 5.2562, df = 6, p-value = 0.5114

```

Según los resultados de las pruebas, los datos analizados se ajustan a una distribución normal y las varianzas entre los grupos son similares. Esto sugiere que se cumplen las suposiciones necesarias para realizar análisis posteriores o pruebas estadísticas adicionales sobre estos datos.

- Análisis de varianza para cada periodo de tiempo con sus respectivos supuestos del modelo
- Análisis de varianza por periodo 1 de tiempo, prueba de normalidad y prueba de homogeneidad de varianza.

```

1 -----
2 Analysis of Variance by period 1 of time
3 -----
4 Analysis of Variance Table
5
6 Response: y
7           Df Sum Sq Mean Sq F value    Pr(>F)
8 Treatments  6 583.60   97.266   3.0124 0.032218 *
9 Block       3 618.68  206.228   6.3871 0.003882 **
10 Residuals  18 581.19   32.288
11 -----
12 Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
13                 0.05 '.' 0.1 ' ' 1
14 -----
15 Shapiro-Wilk normality test
16 data: resid(mod)
17 W = 0.95857, p-value = 0.3223
18 -----
19 Bartlett test of homogeneity of variances
20
21 data: resid(mod) by Treatments
22 Bartlett's K-squared = 8.6585, df = 6, p-value = 0.1937

```

Teniendo en cuenta los resultados del análisis de varianza por periodo 1 de tiempo, los tratamientos muestran una diferencia significativa en la variable de respuesta y con un p_{valor} de 0.032218. Esto indica que al menos uno de los tratamientos tiene un efecto estadísticamente significativo en la variable de respuesta. Los bloques también tienen un efecto significativo en la variable de respuesta, con un p_{valor} de 0.003882.

De acuerdo a la prueba de normalidad de Shapiro-Wilk el p_{valor} asociado fue 0.3223, lo que sugiere que los residuos siguen una distribución normal.

De acuerdo a la prueba de homogeneidad de varianzas de Bartlett el p_{valor} es 0.1937, indican que no se encontraron diferencias significativas en las varianzas entre los grupos de tratamientos.

- Análisis de varianza por periodo 2 de tiempo, prueba de normalidad y prueba de homogeneidad de varianza.

```

1 -----
2 Analysis of Variance by period 2 of time
3 -----
4 Analysis of Variance Table
5
6 Response: y
7      Df Sum Sq Mean Sq F value Pr(>F)
8 Treatments  6 452.78   75.464   3.7977 0.01275 *
9 Block       3 188.04   62.679   3.1543 0.05025 .
10 Residuals 18 357.68   19.871
11 -----
12 Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
13                0.05 '.' 0.1 ' ' 1
14 -----
15 Shapiro-Wilk normality test
16 data: resid(mod)
17 W = 0.98012, p-value = 0.8533
18 -----
19 Bartlett test of homogeneity of variances
20
21 data: resid(mod) by Treatments
22 Bartlett's K-squared = 7.6287, df = 6, p-value = 0.2666

```

Teniendo en cuenta los resultados del análisis de varianza por periodo 2 de tiempo, los tratamientos tienen un efecto significativo en la variable de respuesta para el período 2 de tiempo, con un p_{valor} de 0.01275. Además, los bloques muestran una tendencia a tener un efecto en la variable de respuesta, aunque no alcanza significancia estadística con un p_{valor} de 0.05025.

La prueba de normalidad de Shapiro-Wilk el p_{valor} asociado fue 0.8533, lo que sugiere que los residuos siguen una distribución normal.

La prueba de homogeneidad de varianzas de Bartlett el p_{valor} es 0.2666, indican que no se encontraron diferencias significativas en las varianzas entre los grupos de tratamientos.

- Análisis de varianza por periodo 3 de tiempo, prueba de normalidad y prueba de homogeneidad de varianza.

```

1 -----
2 Analysis of Variance by period 3 of time
3 -----
4 Analysis of Variance Table
5
6 Response: y
7           Df Sum Sq Mean Sq F value Pr(>F)
8 Treatments  6 283.09  47.182  2.6024 0.05391 .
9 Block       3 105.00  34.999  1.9304 0.16085
10 Residuals 18 326.35  18.130
11 -----
12 Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
13                 0.05 '.' 0.1 ' ' 1
14 -----
15 Shapiro-Wilk normality test
16 data: resid(mod)
17 W = 0.9902, p-value = 0.9939
18 -----
19 Bartlett test of homogeneity of variances
20
21 data: resid(mod) by Treatments
22 Bartlett's K-squared = 10.537, df = 6, p-value = 0.1038
23 -----

```

Los resultados indican que los tratamientos muestran una posible influencia en la variable de respuesta para el período 3 de tiempo, aunque no alcanza significancia estadística con un valor p de 0.05391. Los bloques no tienen un efecto significativo en la variable de respuesta.

La prueba de normalidad de Shapiro-Wilk el p_{valor} asociado fue 0.9939, lo que sugiere que los residuos siguen una distribución normal.

La prueba de homogeneidad de varianzas de Bartlett el p_{valor} es 0.1038, indican que no se encontraron diferencias significativas en las varianzas entre los grupos de tratamientos.

- Comparaciones de medias por periodo de tiempo

```

1 -----
2 Tukey's test
3 -----
4 Groups Treatments Means
5 a         a2      34.36667

```

6	ab	a1	32.45
7	bc	a4	30.51667
8	c	a6	29.575
9	d	a3	25.26667
10	d	a5	25.19167
11	e	a0	23.00833
12	-----		

Las letras que se utilizan para identificar los grupos indican que existen diferencias significativas entre las medias de esos tratamientos. Por ejemplo, el grupo a tiene la media más alta, seguido por el grupo ab, bc, c, d, y finalmente el grupo e con la media más baja.

6.15. Introducción a la metodología de superficie de respuesta

Según los autores Gutiérrez y De la Vara en su libro “Análisis y Diseño de Experimentos” (2012), el método de superficie de respuesta se utiliza cuando los niveles y respuestas obtenidos en un experimento no son suficientes. En este método, se lleva a cabo un proceso en el que se desplaza la región experimental en una dirección adecuada, permitiendo explorar de manera más detallada dicha región. El objetivo es encontrar las condiciones óptimas de operación del proceso.

Sin embargo, durante este proceso no solo se considera la región experimental, sino también la región de operabilidad. Esta última proporciona las condiciones en las que el proceso puede operar. Delimitar el tamaño de la región de operabilidad resulta complicado, ya que implica considerar múltiples factores simultáneamente. Es necesario establecer límites que abarquen todas las condiciones en las que el proceso pueda ser operado de manera efectiva.

El método de superficie de respuesta se utiliza para encontrar las condiciones óptimas de operación de un proceso, y durante este proceso también se debe considerar la región de operabilidad, que define las condiciones en las que el proceso puede ser operado de manera efectiva.

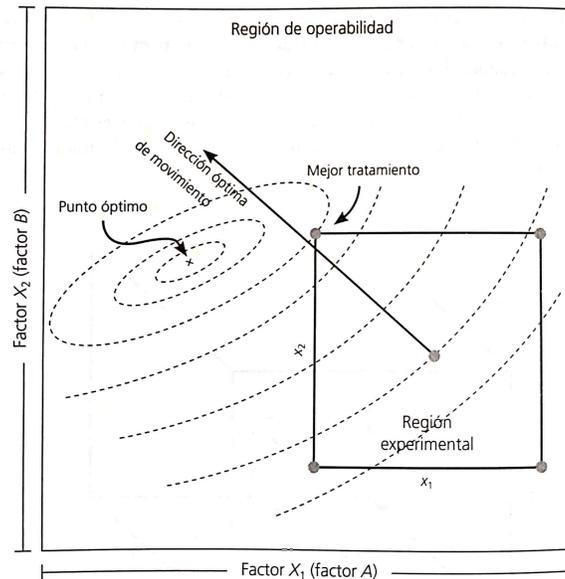
6.15.1. Mejor tratamiento y punto óptimo

En el contexto del diseño de experimentos, se distingue entre el “mejor tratamiento” y el “punto óptimo”. El “mejor tratamiento” se refiere a la combinación de niveles de los factores que se encuentra dentro de la región experimental del estudio. En esta región, se busca identificar la combinación de niveles que maximiza o minimiza la respuesta objetivo, considerando únicamente los niveles utilizados en el experimento.

Por otro lado, el “punto óptimo” se encuentra dentro de la región de operabilidad del proceso. Este proceso implica un enfoque más complejo, ya que requiere llevar a cabo experimentos de forma secuencial y utilizar diversas técnicas de análisis. El objetivo es determinar la combinación de niveles de los factores que maximice o minimice la respuesta objetivo, considerando restricciones adicionales y condiciones operativas del proceso.

Mientras que el “mejor tratamiento” se encuentra dentro de la región experimental y se enfoca en maximizar o minimizar la respuesta objetivo utilizando los niveles del experimento, el “punto óptimo” se encuentra dentro de la región de operabilidad y considera restricciones y condiciones operativas adicionales a través de experimentos secuenciales y técnicas de análisis más avanzadas.

Figura 3: Mejor tratamiento y punto óptimo, región experimental y región de operabilidad.



Fuente: Gutiérrez H y De la Vara R.,(2022), Mejor tratamiento y punto óptimo.[Figura]. Recuperado de Análisis y diseño de experimento edición 3.

Esta Figura representa el verdadero comportamiento de la respuesta y , se tiene un punto óptimo localizado en el centro de la elipse más pequeña, que por cierto está fuera de la región experimental inicial. La superficie de respuesta en la figura se puede imaginar como una montaña y la región experimental se ubica a un costado de la cima: cada curva de nivel representa puntos sobre la montaña que tienen la misma altura.

6.15.2. Elementos de la metodología de superficie de respuesta

Esta metodología tiene en cuenta tres aspectos los cuales son:

- Diseño
- Modelo
- Técnica de optimización

6.15.3. Diseños de superficie de respuesta

De acuerdo con (Gutiérrez y De la Vara, 2012), el diseño de experimentos se utiliza cuando el objetivo es ajustar un modelo para descubrir una superficie. Algunas propiedades importantes de la metodología de superficie de respuesta son:

- Que el diseño genere una distribución satisfactoria. Es decir que aporte información de interés en la región que se estudia.
- El diseño debe requerir un número mínimo de corridas experimentales, para evitar el gasto de recursos.
- El diseño debe permitir que otros diseños de orden mayor se construyan a partir de él. Para que en caso de que se detecte una curvatura se pueda explicar el comportamiento
- El experimento debe permitir la detección de la falta de ajuste, donde se necesitan repeticiones en el centro del diseño.
- En las repeticiones del punto central el diseño debe aportar un estimador puro de la varianza del error.
- La ortogonalidad y la rotabilidad son importantes porque aumentan la eficiencia del diseño.

6.15.4. Diseños para modelos de primer orden

Suponga que se desea utilizar el modelo de primer orden, es decir

$$Y = \beta_0 + \beta_i x_i + \beta_j x_j + \beta_{ij} x_i x_j + \varepsilon$$

para estudiar el comportamiento de cierta característica de calidad, que se supone depende de $k = 2$ factores de proceso. Un criterio importante para la selección del diseño de primer orden es que la varianza de la respuesta predicha $var[\hat{Y}(x)]$ en el punto $x' = (x_1, x_2, \dots, x_k)$ sea mínima, ya que es importante buscar la dirección óptima de movimiento a partir de los predichos del modelo, por tanto, los diseños ortogonales de primer orden son la única clase de diseños que minimizan la varianza de los coeficientes de regresión, los más utilizados son:

- Diseños factoriales 2^k :
- Diseños factorial fraccionados 2^{k-p}
- Diseños simplex:
- Diseño de Plackett – Burman

6.15.5. Diseños para modelos de segundo orden

El diseño de segundo orden, se utiliza cuando se quiere explorar una región que se espera sea más compleja o cuando se cree que el punto óptimo ya se encuentra dentro de la región experimental. El modelo de segundo orden está dado por la ecuación.

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{j=1}^k \beta_j x_j^2 + \sum_{i=1}^k \sum_{j=1}^k \beta_{ij} x_i x_j + \varepsilon,$$

donde $i < j$.

Se debe tener en cuenta que la estimación de los parámetros de un modelo polinómico de segundo grado se podrían utilizar factoriales 3^k , pero es necesario realizar un número alto de combinaciones. Existen otras clases de diseños que permiten ajustar este tipo de modelo con un número menor de combinaciones. Entre ellos están:

- Diseños Box-Behnken
- Diseño de composición central (DCC)
- Diseño de composición central con centros en las caras
- Diseño central compuesto pequeño

6.15.6. Técnicas de optimización

Cuando el modelo ya fue ajustado correctamente, se comienza a explorar la superficie para encontrar la combinación de niveles en los factores que dan por resultado el valor óptimo de la respuesta o para determinar la dirección óptima del movimiento. Se debe tener en cuenta que la técnica depende del modelo ajustado.

- **Análisis canónico:** Este se aplica cuando se tiene un modelo de segundo grado y se desea ampliar el estudio de la región experimental.
- **Análisis de codillera:** Es utilizado cuando el punto estacionario no es satisfactorio, en este se calcula el máximo y el mínimo de la superficie de respuesta sobre esferas concéntricas al centro del diseño empezando por una esfera de radio casi cero y posteriormente se abre la exploración incrementando el radio de la siguiente esfera.

6.15.7. Ejemplo en R de superficie de respuesta

Un ingeniero químico está interesado en determinar las condiciones de operación que maximizan el rendimiento de un proceso. Hay dos variables de control que influyen Tiempo de reacción y temperatura de reacción, el punto de operación actual es 35 minutos y 155°F que da un rendimiento del 40 % aproximadamente. Se hace un diseño experimental variando el tiempo (30 a 40 minutos) y la temperatura (150 a 160°F). Por simplicidad se codifican las variables en el intervalo (-1, 1). Si las variables codificadas son X_1 y X_2 y las variables naturales son ξ_1 y ξ_2 se tiene:(Montgomery, 2002, pg.431).

$$x_1 = \frac{\xi_1 - [\max(\xi_1) + \min(\xi_1)]/2}{[\max(\xi_1) - \min(\xi_1)]/2} = \frac{\xi_1 - (40 + 30)/2}{(40 - 30)/2} = \frac{\xi_1 - 35}{5}$$

$$x_2 = \frac{\xi_2 - [\max(\xi_2) + \min(\xi_2)]/2}{[\max(\xi_2) - \min(\xi_2)]/2} = \frac{\xi_2 - (160 + 150)/2}{(160 - 150)/2} = \frac{\xi_2 - 155}{5}$$

Para llevar a cabo este proceso en R se requieren las librerías (ggpubr), (magrittr) y (visreg), este ejemplo también se encuentra ilustrado en la página web RPubS. (Gil, 2019). Los pasos a seguir son los siguientes:

1. Importamos los datos

```
1 > datos<-read_excel(data = datos,sheet = 1);datos
2 > dA.c = coded.data(data = datos,x1 ~ (tiempo - 35)/5, x2
  ~
3 (temperatura - 155)/5);dA.c
```

```
1   Corridas tiempo temperatura bloque   y1
2 1           1     30           150     1 39.3
3 2           2     30           160     1 40.0
4 3           3     40           150     1 40.9
```

```

5 4      4      40      160      1 41.5
6 5      5      35      155      1 40.3
7 6      6      35      155      1 40.5
8 7      7      35      155      1 40.7
9 8      8      35      155      1 40.2
10 9     9      35      155      1 40.6
11
12 Data are stored in coded form using these coding formulas
13     ...
13 x1 ~ (tiempo - 35)/5
14 x2 ~ (temperatura - 155)/5

```

Se desea maximizar la variable respuesta y_1 , pero, no se sabe que valores toman los factores ya que se necesita maximizar. Por tanto, decidimos empezar aplicando un diseño factorial 2^k aumentado con cinco réplicas del punto central ($x_1 = 0$, $x_2 = 0$) para poder estimar el error experimental y la adecuación de un modelo.

2. Codificamos los factores

```

1 > as.data.frame(dA.c)

```

	Corridas	x1	x2	bloque	y1
2	1	-1	-1	1	39.3
3	2	-1	1	1	40.0
4	3	1	-1	1	40.9
5	4	1	1	1	41.5
6	5	0	0	1	40.3
7	6	0	0	1	40.5
8	7	0	0	1	40.7
9	8	0	0	1	40.2
10	9	0	0	1	40.6

Estos son los cinco puntos centrales se usan como réplicas para verificar la adecuación del modelo de primer orden.

3. Se ejecuta un modelo de primer orden con interacción. Si se quiere un modelo de segundo orden solo hay que cambiar FO() por SO() y eliminar el componente TWI()

```

1
2 > dA.c.rsm = rsm(y ~ FO(x1, x2) + TWI(x1, x2) , data = dA
3     .c)
4 > summary(dA.c.rsm)

```

```

5 Call:
6 rsm(formula = y ~ FO(x1, x2) + TWI(x1, x2), data = dA.c)
7
8           Estimate Std. Error  t value  Pr(>|t|)
9 (Intercept) 40.444444    0.062311 649.0693 1.648e-13 ***
10 x1           0.775000    0.093467   8.2917 0.0004166 ***
11 x2           0.325000    0.093467   3.4772 0.0177127 *
12 x1:x2       -0.025000    0.093467  -0.2675 0.7997870
13 ---
14 Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
15                 0.05 '.' 0.1 ' ' 1
16 Multiple R-squared:  0.9418, Adjusted R-squared:  0.9069
17 F-statistic: 26.97 on 3 and 5 DF,  p-value: 0.00163
18
19 Analysis of Variance Table
20
21 Response: y
22           Df  Sum Sq Mean Sq F value    Pr(>F)
23 FO(x1, x2)  2  2.82500  1.41250  40.4213 0.0008188
24 TWI(x1, x2)  1  0.00250  0.00250   0.0715 0.7997870
25 Residuals    5  0.17472  0.03494
26 Lack of fit  1  0.00272  0.00272   0.0633 0.8137408
27 Pure error   4  0.17200  0.04300
28
29 Stationary point of response surface:
30 x1 x2
31 13 31
32
33 Stationary point in original units:
34     tiempo temperatura
35     100             310
36
37 Eigenanalysis:
38 eigen() decomposition
39 $values
40 [1]  0.0125 -0.0125
41
42 $vectors
43           [,1]      [,2]
44 x1 -0.7071068 -0.7071068
45 x2  0.7071068 -0.7071068

```

La significancia estadística de los coeficientes estimados, y por tanto del efecto de los factores, viene determinada por los valores del estadístico

t del t-test y el p-valor asociado (p_{valor} es menor a 0,05 supone un efecto significativo). Como el ($p_{valor} = 0,941$), no es significativo se concluye que el modelo se aproxima bastante a los datos y resulta útil para modelarlos. Su ecuación en unidades codificadas sería:

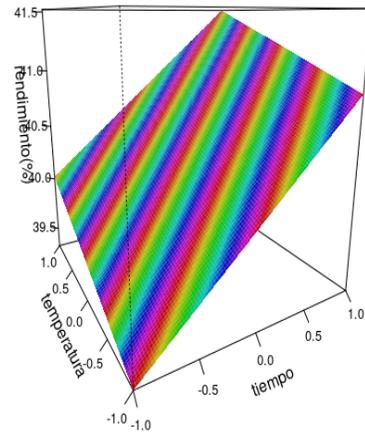
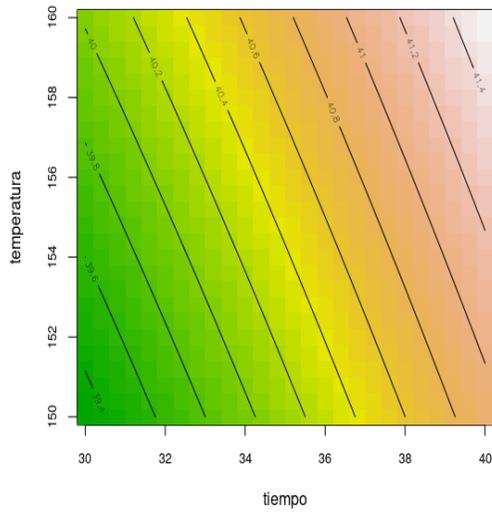
$$y = 40,444 + 0,775x_1 + 0,325x_2$$

Se debe tener en cuenta que, al tratarse de un buen ajuste con un modelo lineal, no nos encontramos en el entorno del máximo de la variable respuesta de la función que se está buscando.

4. Optimización del modelo

La optimización gráfica, es simple para superponer los dos gráficos de correspondientes a y_1 e y_2 , esto ayuda a inspeccionar la zona del espacio factorial, para escoger los posibles valores óptimos.

```
1 -----  
2 > par(mfrow=c(1,2))  
3 > contour(dA.c.rsm, ~x1+x2, image=T, cex.axis=0.8)  
4 > visreg2d(dA.c.rsm, "x1", "x2", plot.type="persp",  
5 xlab="Tiempo", ylab="\nHumedad", zlab="\nCarbono",  
6 cex.axis=0.8, col=rainbow(50), scale = "response")  
7 -----
```

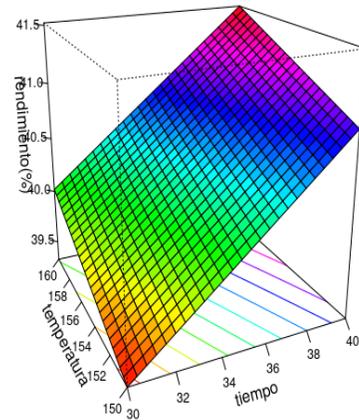
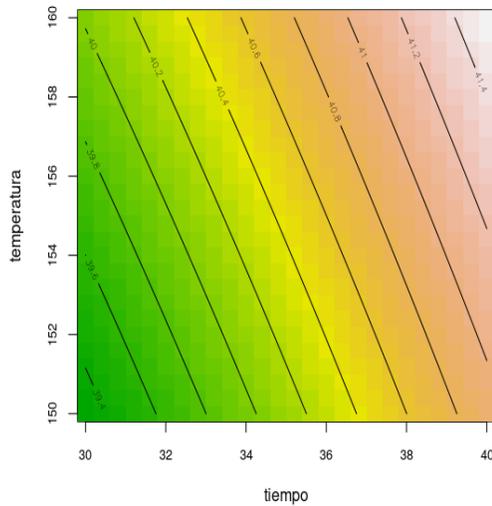


5. Contorno

```

1 > par(mfrow=c(1,2))
2 > contour(dA.c.rsm, ~x1+x2, image=T, cex.axis=0.8)
3 > persp(dA.c.rsm, ~x1+x2, zlab="rendimiento(%)",
4 col=rainbow(50),
5 contours=("colors"), cex.axis=0.8)
6
7 -----

```



6. Análisis de la segunda corrida

```

1 > segundacorrida <-read_excel(data = datos1,sheet = 1)
2 > segundacorrida <-coded.data(data = segundacorrida,
3     x1 ~ (tiempo - 85)/5, x2 ~ (temperatura - 175)/5)
4
5     Corrida tiempo temperatura bloque  y1  y2
6 1         1     80.0           170     1 76.5 2940
7 2         2     80.0           180     1 77.0 3470
8 3         3     90.0           170     1 78.0 3680
9 4         4     90.0           180     1 79.5 3890
10 5        5     85.0           175     1 79.9 3480
11 6        6     85.0           175     1 80.3 3200
12 7        7     85.0           175     1 80.0 3410
13 8        8     85.0           175     1 79.7 3290
14 9        9     85.0            17

```

7. Se codifican nuevamente los factores

```
1 > as.data.frame(segundacorrída)
2
3 Corrida      x1      x2 bloque    y1    y2
4 1          1 -1.00 -1.0      1 76.5 2940
5 2          2 -1.00  1.0      1 77.0 3470
6 3          3  1.00 -1.0      1 78.0 3680
7 4          4  1.00  1.0      1 79.5 3890
8 5          5  0.00  0.0      1 79.9 3480
9 6          6  0.00  0.0      1 80.3 3200
10 7          7  0.00  0.0      1 80.0 3410
11 8          8  0.00  0.0      1 79.7 3290
12 9          9  0.00  0.0      1 79.8 3500
13 10         11  1.42  0.0      1 78.4 3360
14 11         12 -1.42  0.0      1 75.6 3020
15 12         13  0.00  1.4      1 78.5 3630
16 13         14  0.00 -1.4      1 77.0 3150
```

8. Se ejecuta un modelo de segundo orden.

```
1 > modelo2.y1 <- rsm(y1 ~ S0(x1, x2), data =
2     segundacorrída)
3
4 > summary(modelo2.y1)
5
6 Call:
7 rsm(formula = y1 ~ S0(x1, x2), data = segundacorrída)
8
9      Estimate Std. Error  t value  Pr(>|t|)
10 (Intercept) 79.939060    0.120726 662.1530 < 2.2e-16 ***
11 x1           0.992929    0.095248  10.4247 1.625e-05 ***
12 x2           0.517677    0.095924   5.3968 0.001012 **
13 x1:x2        0.250000    0.134977   1.8522 0.106427
14 x1^2        -1.363382    0.101658 -13.4115 3.005e-06 ***
15 x2^2        -1.017172    0.103931  -9.7870 2.465e-05 ***
16 ---
17 Signif. codes:  0 '***' 0.001 '**' 0.01 '*'
18                 0.05 '.' 0.1
19
20 Multiple R-squared:  0.9823, Adjusted R-squared:  0.9696
21 F-statistic: 77.48 on 5 and 7 DF,  p-value: 5.66e-06
22
23 Analysis of Variance Table
24
25 Response: y1
26      Df  Sum Sq Mean Sq  F value  Pr(>F)
```

```

24 FO(x1, x2)    2 10.0421  5.0210  68.8994  2.484e-05
25 TWI(x1, x2)  1  0.2500  0.2500   3.4305   0.1064
26 PQ(x1, x2)   2 17.9409  8.9704 123.0937  3.514e-06
27 Residuals    7  0.5101  0.0729
28 Lack of fit  3  0.2981  0.0994   1.8750   0.2747
29 Pure error   4  0.2120  0.0530
30
31 Stationary point of response surface:
32      x1      x2
33 0.3918880 0.3026276
34
35 Stationary point in original units:
36      tiempo temperatura
37 86.95944  176.51314
38
39 Eigenanalysis:
40 eigen() decomposition
41 $values
42 [1] -0.9767582 -1.4037958
43
44 $vectors
45      [,1]      [,2]
46 x1 -0.3076329 -0.9515051
47 x2 -0.9515051  0.3076329

```

Son significativos (p -valor < 0,05) con un error estándar relativamente bajo, siendo el tiempo (x_1) el factor con un efecto más significativo sobre el rendimiento (mayor coeficiente absoluto de regresión). Su ecuación en unidades codificadas sería:

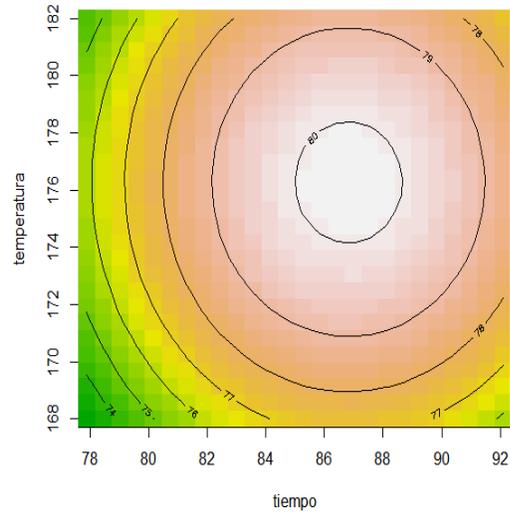
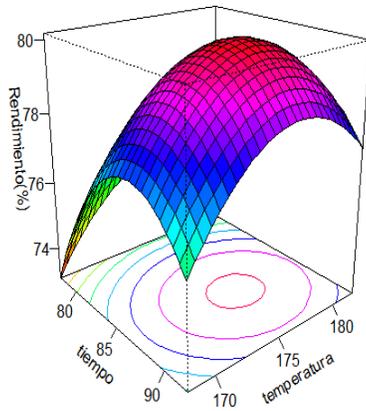
$$\hat{y} = 79,939060 + 0,992x_1 + 0,517x_2 - 1,363x_1^2 - 1,017x_2^2$$

9. Superficie 3D y grafico de contorno

```

1 > par(mfrow = c(1,2))
2 > persp(modelo2.y1.sinI, x2 ~ x1,
3       zlab = "Rendimiento(%)",
4       contours = list(z = "bottom", col = "colors"),
5       at = c(summary(modelo2.y1.sinI$canonical$xs)),
6       theta = 50,
7       phi = 20)
8
9
10 > contour(modelo2.y1.sinI, ~ x1 + x2, image = TRUE)

```



7. CÓDIGOS

1. Muestreo de unidades experimentales en un DCA

```

1 Sub.DCA=function (treat, repet, resp, quali = TRUE, mcomp
  = "tukey"
2 , sigT = 0.05, sigF = 0.05)
3 {
4   library(ExpDes)
5   Trat <- factor(treat)
6   rep=factor(repet)
7   anava <- aov(resp ~ Trat+(Trat:rep))
8   tab <- summary(anava)
9   colnames(tab[[1]]) <- c("DF", "SS", "MS", "Fc", "Pr>Fc"
  )
10  tab[[1]] <- rbind(tab[[1]], c(apply(tab[[1]], 2, sum)))
11  rownames(tab[[1]]) <- c("Tratamiento", "Residuals", "
  Sampling error",
12  "Total")
13  cv <- round(sqrt(tab[[1]][2,3])/mean(resp) * 100, 2)
14  tab[[1]][3, 4] = " "
15  tab[[1]][2,4]=" "
16  tab[[1]][2,5]=" "
17  tab[[1]][1,4]=tab[[1]][1,3]/tab[[1]][2,3]

```

```

18 tab[[1]][1,5]=pf(tab[[1]][1,3]/tab[[1]][2,3],tab
    [[1]][1,1],
19                 tab[[1]][2,1],lower.tail = F)
20 cat("-----\nAnalysis of Variance Table\n-----\n")
21 print(tab[[1]])
22 cat("-----\nCV =",
23     cv, "%\n")
24 pvalor.shapiro <- shapiro.test(anava$residuals)$p.value
25 cat("\n-----\nShapiro-Wilk normality test\n")
26 cat("p-value: ", pvalor.shapiro, "\n")
27 if (pvalor.shapiro < 0.05) {
28     cat("WARNING: at 5% of significance, residuals can
        not be
29     considered normal")
30 }
31 else {
32     cat("According to Shapiro-Wilk normality test at 5%
        of significance,
33     residuals can be considered normal.")
34 }
35 pvalor.bartlett <- bartlett.test(anava$residuals~Trat)$
    p.value
36 cat("\n-----\nBartlett homogeneity test\n")
37 cat("p-value: ", pvalor.bartlett, "\n")
38 if (pvalor.bartlett < 0.05) {
39     cat("WARNING: at 5% of significance,it is considered
        that
40     the residuals do not have constant variance")
41 }
42 else {
43     cat("According to Bartlett homogeneity test at 5% of
        significance,
44     it is considered that the residuals have constant
        variance.")
45 }
46
47 if (tab[[1]][1, 5] < sigF) {
48     if (quali == TRUE) {
49         if (mcomp == "tukey") {
50             tukey(resp, Trat, tab[[1]][2,1], tab[[1]][2,2],
                sigT)
51         }
52         if (mcomp == "duncan") {
53             duncan(resp, Trat, tab[[1]][2,1], tab[[1]][2,2],
                sigT)

```

```

54     }
55     if (mcomp == "lsd") {
56         lsd(resp, Trat, tab[[1]][2, 1], tab[[1]][2,2],
57             sigT)
58     }
59     if (mcomp == "lsdb") {
60         lsdb(resp, Trat, tab[[1]][2, 1], tab[[1]][2,2],
61             sigT)
62     }
63     if (mcomp == "sk") {
64         scottknott(resp, Trat, tab[[1]][2,1], tab
65             [[1]][2,2], sigT)
66     }
67     if (mcomp == "snk") {
68         snk(resp, Trat, tab[[1]][2, 1], tab[[1]][2,2],
69             sigT)
70     }
71     if (mcomp == "ccboot") {
72         ccboot(resp, Trat, tab[[1]][2, 1], tab[[1]][2,2],
73             sigT)
74     }
75     }
76     else {
77         reg.poly(resp, treat, tab[[1]][2,1], tab[[1]][2,2],
78             tab[[1]][1,1],
79             tab[[1]][1,2])
80     }
81     }
82     else {
83         cat("\nAccording to the F test, the means can not be
84             considered
85             distinct.\n")
86         mean.table <- tapply.stat(resp, treat, mean)
87         colnames(mean.table) <- c("Levels", "Means")
88         print(mean.table)
89         cat("-----\n")
90     }
91 }

```

2. Muestreo de unidades experimentales en un DBCA

```

1 Sub.DBCA=function (treat, repet,block,resp, quali = TRUE,
2     mcomp = "tukey",
3     sigT = 0.05,sigF = 0.05) {
4     library(ExpDes)
5     Trat <- factor(treat)

```

```

5 rep=factor(repet)
6 bloq=factor(block)
7 anava <- aov(resp ~ Trat+bloq+(Trat:rep))
8 tab <- summary(anava)
9 colnames(tab[[1]]) <- c("DF", "SS", "MS", "Fc", "Pr>Fc"
10 )
11 tab[[1]] <- rbind(tab[[1]], c(apply(tab[[1]], 2, sum)))
12 rownames(tab[[1]]) <- c("Treatment","block","Residuals
13 ",
14 "Sampling error", "Total")
15 cv <- round(sqrt(tab[[1]][3,3])/mean(resp) * 100, 2)
16 tab[[1]][3, 5] = " "
17 tab[[1]][3,4]=" "
18 tab[[1]][1,4]=tab[[1]][1,3]/tab[[1]][3,3]
19 tab[[1]][2,4]=tab[[1]][2,3]/tab[[1]][3,3]
20 tab[[1]][1,5]=pf(tab[[1]][1,3]/tab[[1]][3,3],tab
21 [[1]][1,1],
22 tab[[1]][3,1],lower.tail = F)
23 tab[[1]][1,5]=pf(tab[[1]][2,3]/tab[[1]][3,3],tab
24 [[1]][2,1],
25 tab[[1]][3,1],lower.tail = F)
26 cat("-----\nAnalysis of Variance Table\n-----\n")
27 print(tab[[1]])
28 cat("-----\nCV =",
29 cv, "%\n")
30 pvalor.shapiro <- shapiro.test(anava$residuals)$p.value
31 cat("\n-----\nShapiro-Wilk normality test\n")
32 cat("p-value: ", pvalor.shapiro, "\n")
33 if (pvalor.shapiro < 0.05) {
34 cat("WARNING: at 5% of significance, residuals
35 can not be considered normal")
36 }
37 else {
38 cat("According to Shapiro-Wilk normality test at 5%
39 of significance,
40 residuals can be considered normal.")
41 }
42 pvalor.bartlett <- bartlett.test(anava$residuals~Trat)$
43 p.value
44 cat("\n-----\nBartlett homogeneity test\n")
45 cat("p-value: ", pvalor.bartlett, "\n")
46 if (pvalor.bartlett < 0.05) {
47 cat("WARNING: at 5% of significance,it is considered
48 that the
49 residuals do not have constant variance")

```

```

43 }
44 else {
45   cat("According to Bartlett homogeneity test at 5% of
46     significance,
47     it is considered that the residuals have constant
48     variance.")
49 }
50
51 if (tab[[1]][1, 5] < sigF) {
52   if (quali == TRUE) {
53     if (mcomp == "tukey") {
54       tukey(resp, Trat, tab[[1]][3,1], tab[[1]][3,2],
55         sigT)
56     }
57     if (mcomp == "duncan") {
58       duncan(resp, Trat, tab[[1]][3,1], tab[[1]][3,2],
59         sigT)
60     }
61     if (mcomp == "lsd") {
62       lsd(resp, Trat, tab[[1]][3,1], tab[[1]][3,2],
63         sigT)
64     }
65     if (mcomp == "lsdb") {
66       lsdb(resp, Trat, tab[[1]][3,1], tab[[1]][3,2],
67         sigT)
68     }
69     if (mcomp == "sk") {
70       scottknott(resp, Trat, tab[[1]][3,1], tab
71         [[1]][3,2], sigT)
72     }
73     if (mcomp == "snk") {
74       snk(resp, Trat, tab[[1]][3,1], tab[[1]][3,2],
75         sigT)
76     }
77     if (mcomp == "ccboot") {
78       ccboot(resp, Trat, tab[[1]][3,1], tab[[1]][3,2],
79         sigT)
80     }
81   }
82   else {
83     reg.poly(resp, treat, tab[[1]][3, 1], tab
84       [[1]][3,2],
85       tab[[1]][1, 1], tab[[1]][1,2])
86   }
87 }

```

```

78 else {
79   cat("\nAccording to the F test, the means can not be
      considered
80   distinct.\n")
81   mean.table <- tapply.stat(resp, treat, mean)
82   colnames(mean.table) <- c("Levels", "Means")
83   print(mean.table)
84   cat("-----\n")
85 }
86 }

```

3. Medidas repetidas en el tiempo con un DCA

```

1 MRT.DCA=function(Y,x1,x2,rep,quali,mcomp,sigF,sigT){
2   library(ExpDes)
3   a=length(levels(x1))
4   t=length(levels(x2))
5   r=length(levels(rep))
6   N=a*t*r
7   nv1<-length(summary(x2))
8   lf1<-levels(x2)
9   Fc=(sum(Y)^2)/N
10  SCT=sum(Y^2)-Fc
11  SCA=(sum(tapply(Y,x1,sum)^2)/(t*r))-Fc
12  SCPP=(sum(tapply(Y,paste(x1,rep,sep=""),sum)^2)/t)-Fc
13  SCRep=(sum(tapply(Y,rep,sum)^2)/(t*a))-Fc
14  SCEa=SCPP-SCA
15  SCT=(sum(tapply(Y,x2,sum)^2)/(a*r))-Fc
16  SCAt=(sum(tapply(Y,paste(x1,x2,sep=""),sum)^2)/r)-Fc-
      SCA-SCT
17  SCTrep=(sum(tapply(Y,paste(rep,x2,sep=""),sum)^2)/a)-Fc-
      SCRep-SCT
18  SCEb=SCT-SCPP-SCTrep-SCAt-SCT
19  SC=c(SCA,SCEa,SCT,SCAt,SCTrep,SCEb,SCT)
20  GL=c(a-1,a*(r-1),t-1,(a-1)*(t-1),(t-1)*(r-1),(a-1)*(r-
      1)*(t-1),N-1)
21  CM=c((SC/GL)[1:6],NA)
22  CMEa=CM[2]+CM[4]-CM[6]
23  GLa=ceiling(((CMEa^2)/((CM[2]^2/(a*(r-1)))+(CM[4]^2/((a-
      1)*(r-1))))+
24  (CM[6]^2/((a-1)*(t-1)*(r-1))))))
25  Fcal=c(CM[1]/CM[2],NA,CM[3]/CM[5],CM[4]/CM[6],NA,NA,NA)
26  valor.P=1-c(pf(Fcal[1],a-1,a*(r-1)),NA,pf(Fcal[3],t-1,(
      t-1)*(r-1)),
27  pf(Fcal[4],(a-1)*(r-1),(a-1)*(r-1)*(t-1)),NA,NA,NA)
28  if(valor.P[4]<sigF){

```

```

29     Fcal[1]=CM[1]/CMEa
30     valor.P[1]=pf(CM[1]/CMEa,a-1,GLa,lower.tail = F)
31 }
32 Anova=data.frame(GL,SC,CM,Fcal,valor.P)
33 rownames(Anova)=c("Treat","E(a)","Time","Treat x Time",
34 "Rep x Time",
35 "E(t)","Total")
36 Anova<-round(Anova,4)
37 cv1 = round(sqrt(Anova[2,3])/mean(Y) * 100,3)
38 cv2 = round(sqrt(Anova[6,3])/mean(Y) * 100,3)
39 Anova[2,4:5] <- ""
40 Anova[5,4:5] <- ""
41 Anova[6,4:5] <- ""
42 Anova[7,3:5] <- ""
43 Fi<-rep
44 Fj<-x2
45 Fk<-x1
46 yijk<-Y
47 yb...<-mean(yijk)
48 ybi..<-tapply(yijk,Fi,mean)
49 yb.j.<-tapply(yijk,Fj,mean)
50
51 nfi<-levels(as.factor(Fi))
52 nfj<-levels(as.factor(Fj))
53 nfk<-levels(as.factor(Fk))
54
55 ybij.<-matrix(0,ncol=length(nfj),nrow=length(nfi))
56 for(i in 1:length(nfi)){
57   for(j in 1:length(nfj)){
58     ybij.[i,j]<-mean(yijk[which(Fi==nfi[i] & Fj==nfj[j]
59     ])]))
60   }
61 }
62
63 yb.jk<-matrix(0,ncol=length(nfk),nrow=length(nfj))
64 for(j in 1:length(nfj)){
65   for(k in 1:length(nfk)){
66     yb.jk[j,k]<-mean(yijk[which(Fj==nfj[j] & Fk==nfk[k]
67     ])]))
68   }
69 }
70
71 t=1
72 eijk<-numeric()

```

```

71 for(i in 1:length(nfi)){
72   for(j in 1:length(nfj)){
73     for(k in 1:length(nfk)){
74       eijk[t]<-yijk[which(Fi==nfi[i] & Fj==nfj[j] & Fk
75 ==nfk[k])]
76       +4*yb...-ybi..[i]-ybij.[i,j]-2*yb.j.[j]-yb.jk[j,k
77 ]
78       t<-t+1
79     }
80   }
81 }
82 i=1;j=1;k=1
83 sum(eijk)
84 Shapiro=shapiro.test(eijk)
85 Bartlett=bartlett.test(eijk~Fk)
86 ana.tiempo<-list()
87 Shapiro.tiempo<-list()
88 Bartlett.tiempo<-list()
89 for(i in 1:nv1){
90   y<-Y[x2 == lf1[i]]
91   Treatments<-x1[x2 == lf1[i]]
92   mod<-aov(y~Treatments)
93   Anava=anova(mod)
94   ana.tiempo[[i]]<-Anava
95   Shapiro.tiempo[[i]]<-shapiro.test(resid(mod))
96   Bartlett.tiempo[[i]]<-bartlett.test(resid(mod)~
97   Treatments)
98 }
99 cat("-----\nAnalysis of Variance Table by period\n
100 -----\n")
101 for(i in 1:nv1){
102   cat("-----\n")
103   cat("\n\n", "Analisis de varianza", " Para el periodo
104   ",
105   lf1[i], " of ", "time", "\n-----\n")
106   print(ana.tiempo[[i]])
107   print(Shapiro.tiempo[[i]])
108   print(Bartlett.tiempo[[i]])
109 }
110 cat("-----\nAnalysis of Variance Table\n-----\n")
111 print(Anova)
112 cat("-----\nCV 1 =",
113   cv1, "%\nCV 2 =", cv2, "%\n")

```

```

111   cat("\n-----\nShapiro-Wilk normality test\n")
112   print(Shapiro)
113   cat("\n-----\nBartlett homogeneity test\n")
114   print(Bartlett)
115
116   if ( valor.P[4] > sigF) {
117     cat("\nNot significant interaction: analyzing the
118         simple effect
119         \n-----\n")
120     if (quali==TRUE && valor.P[1] <= sigF) {
121       cat("-----\nTreat\n-----")
122       if (mcomp == "tukey") {
123         tukey(Y,x1,GL[2],CM[2],sigT)
124       }
125       if (mcomp == "duncan") {
126         duncan(Y,x1,GL[2],CM[2], sigT)
127       }
128       if (mcomp == "lsd") {
129         lsd(Y,x1,GL[2],CM[2], sigT)
130       }
131       if (mcomp == "lsdb") {
132         lsdb(Y,x1,GL[2],CM[2],sigT)
133       }
134       if (mcomp == "sk") {
135         scottknott(Y,x1,GL[2],CM[2], sigT)
136       }
137       if (mcomp == "snk") {
138         snk(Y,x1,GL[2],CM[2], sigT)
139       }
140       if (mcomp == "ccboot") {
141         ccboot(Y,x1,GL[2],CM[2], sigT)
142       }
143     }
144     if (quali==TRUE && valor.P[1] > sigF) {
145       cat("\n-----\nAccording to the F test, the means of
146         this treatments are statistical equal.\n-----\n")
147       mean.table <- tapply.stat(Y,x1,mean)
148       colnames(mean.table) <- c("Levels", "Means")
149       print(mean.table)
150       cat("-----")
151     }
152     if (quali== FALSE && valor.P[1] <= sigF) {
153       x1=as.numeric(x1)
154       cat("-----\nTreat\n-----")
155       reg.poly(Y,x1,GL[2],SC[2],GL[1],SC[1])

```

```

155 }
156 if (quali== FALSE && valor.P[1] > sigF) {
157   cat("\n-----\nAccording to the F test, the means of
158   this treatments are statistical equal.\n-----\n")
159   mean.table <- tapply.stat(Y,x1, mean)
160   colnames(mean.table) <- c("Levels", "Means")
161   print(mean.table)
162   cat("-----")
163 }
164 }
165 if (valor.P[4]<= sigF) {
166   cat("\n-----\nSignificant interaction: analyzing the
167   interaction
168   \n-----\n")
169   cat("\nAnalyzing ", "treatments", " inside of each
170   level of ",
171   "time", "\n-----\n")
172   for (i in 1:nv1) {
173     if (quali== TRUE) {
174       cat("\n-----\n", "treatments", " inside of the
175       level ",
176       lf1[i], " of ", "time", "\n-----")
177       if (mcomp == "tukey") {
178         mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]])
179         Anava=as.matrix(anova(mod))
180         tukey(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
181         [2,1],Anava[2,2],
182         sigT)
183       }
184       if (mcomp == "duncan") {
185         mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]])
186         Anava=as.matrix(anova(mod))
187         duncan(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
188         [2,1],Anava[2, 2],
189         sigT)
190       }
191       if (mcomp == "lsd") {
192         mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]])
193         Anava=as.matrix(anova(mod))
194         lsd(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava[2,1],
195         Anava[2, 2],
196         sigT)
197       }
198       if (mcomp == "lsdb") {
199         mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]])

```

```

194     Anava=as.matrix(anova(mod))
195     lsdb(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava[2,1],
Anava[2, 2],
196     sigT)
197   }
198   if (mcomp == "sk") {
199     mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]])
200     Anava=as.matrix(anova(mod))
201     scottknott(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
[2,1],Anava[2, 2],
202     sigT)
203   }
204   if (mcomp == "snk") {
205     mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]])
206     Anava=as.matrix(anova(mod))
207     snk(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava[2,1],
Anava[2, 2],
208     sigT)
209   }
210   if (mcomp == "ccboot") {
211     mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]])
212     Anava=as.matrix(anova(mod))
213     ccboot(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
[2,1],Anava[2, 2],
214     sigT)
215   }
216   }
217   else {
218     cat("\n\n", "treatments", " inside of the level ",
219         lf1[i], " of ", "time", "\n-----")
220     mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]])
221     Anava=as.matrix(anova(mod))
222     reg.poly(Y[x2 == lf1[i]],as.numeric(x1[x2 == lf1[i]
223     ])),
                Anava[2,1], Anava[2,2], Anava[1,1], Anava
[1,2])
224   }
225 }
226 }
227 }

```

4. Medidas repetidas en el tiempo con un DBCA

```

1 MRT.DBCA=function(Y,x1,x2,block,quali,mcomp,sigF,sigT){
2   library(ExpDes)
3   a=length(levels(x1))

```

```

4  t=length(levels(x2))
5  r=length(levels(block))
6  N=a*t*r
7  nv1<-length(summary(x2))
8  lf1<-levels(x2)
9  Fc=(sum(Y)^2)/N
10 SCT=sum(Y^2)-Fc
11 SCbloq=(sum(tapply(Y,block,sum)^2)/(t*a))-Fc
12 SCA=(sum(tapply(Y,x1,sum)^2)/(t*r))-Fc
13 SCPP=(sum(tapply(Y,paste(x1,block,sep=""),sum)^2)/t)-Fc
14 SCEa=SCPP-SCA-SCbloq
15 SCt=(sum(tapply(Y,x2,sum)^2)/(a*r))-Fc
16 SCAt=(sum(tapply(Y,paste(x1,x2,sep=""),sum)^2)/r)-Fc-
   SCA-SCt
17 SCTbloq=(sum(tapply(Y,paste(block,x2,sep=""),sum)^2)/a)
   -Fc-SCbloq-SCt
18 SCEb=SCT-SCPP-SCTbloq-SCAt-SCt
19 SC=c(SCbloq,SCA,SCEa,SCt,SCAt,SCTbloq,SCEb,SCT)
20 GL=c(r-1,a-1,a*(r-1),t-1,(a-1)*(t-1),(t-1)*(r-1),(a-1)*
   (r-1)*(t-1),N-1)
21 CM=c((SC/GL)[1:7],NA)
22 CMEa=CM[3]+CM[5]-CM[7]
23 GLa=ceiling(((CMEa^2)/((CM[3]^2/(a*(r-1)))+(CM[5]^2/((a
   -1)*(r-1))))+
24 (CM[7]^2/((a-1)*(t-1)*(r-1))))))
25 Fcal=c(NA,CM[2]/CM[3],NA,CM[4]/CM[6],CM[5]/CM[7],NA,NA,
   NA)
26 valor.P=1-c(NA,pf(Fcal[2],a-1,a*(r-1)),NA,
27 pf(Fcal[4],t-1,(t-1)*(r-1)),pf(Fcal[5],(a-1)*(r-1),(a
   -1)*(r-1)*(t-1)),
28 NA,NA,NA)
29 if(valor.P[5]<sigF){
30   Fcal[1]=CM[2]/CMEa
31   valor.P[2]=pf(CM[2]/CMEa,a-1,GLa,lower.tail = F)
32 }
33 Anova=data.frame(GL,SC,CM,Fcal,valor.P)
34 rownames(Anova)=c("Block","Treat","E(a)","Time","Treat
   x Time",
35 "Block x Time","E(t)","Total")
36 Anova<-round(Anova,4)
37 cv1 = round(sqrt(Anova[3,3])/mean(Y) * 100,3)
38 cv2 = round(sqrt(Anova[7,3])/mean(Y) * 100,3)
39 Anova[1,4:5]<-" "
40 Anova[3,4:5]<-" "
41 Anova[6,4:5]<-" "

```

```

42 Anova[7,4:5] <-""
43 Anova[8,3:5] <-""
44 Fi<-block
45 Fj<-x2
46 Fk<-x1
47 yijk<-Y
48 yb...<-mean(yijk)
49 ybi..<-tapply(yijk,Fi,mean)
50 yb.j.<-tapply(yijk,Fj,mean)
51
52 nfi<-levels(as.factor(Fi))
53 nfj<-levels(as.factor(Fj))
54 nfk<-levels(as.factor(Fk))
55
56 ybij.<-matrix(0,ncol=length(nfj),nrow=length(nfi))
57 for(i in 1:length(nfi)){
58   for(j in 1:length(nfj)){
59     ybij.[i,j]<-mean(yijk[which(Fi==nfi[i] & Fj==nfj[j]
60     )])
61   }
62 }
63
64 yb.jk<-matrix(0,ncol=length(nfk),nrow=length(nfj))
65 for(j in 1:length(nfj)){
66   for(k in 1:length(nfk)){
67     yb.jk[j,k]<-mean(yijk[which(Fj==nfj[j] & Fk==nfk[k]
68     )])
69   }
70 }
71
72 t=1
73 eijk<-numeric()
74 for(i in 1:length(nfi)){
75   for(j in 1:length(nfj)){
76     for(k in 1:length(nfk)){
77       eijk[t]<-yijk[which(Fi==nfi[i] & Fj==nfj[j] & Fk
78       ==nfk[k])]
79       +4*yb...-ybi..[i]-ybij.[i,j]-2*yb.j.[j]-yb.jk[j,k]
80     ]
81     t<-t+1
82   }
83 }
84 i=1;j=1;k=1

```

```

83 sum(eijk)
84 Shapiro=shapiro.test(eijk)
85 Bartlett=bartlett.test(eijk~Fk)
86 ana.tiempo<-list()
87 Shapiro.tiempo<-list()
88 Bartlett.tiempo<-list()
89 for(i in 1:nv1){
90   y<-Y[x2 == lf1[i]]
91   Treatments<-x1[x2 == lf1[i]]
92   Block<-block[x2 == lf1[i]]
93   mod<-aov(y~Treatments+Block)
94   Anava=anova(mod)
95   ana.tiempo[[i]]<-Anava
96   Shapiro.tiempo[[i]]<-shapiro.test(resid(mod))
97   Bartlett.tiempo[[i]]<-bartlett.test(resid(mod)~
    Treatments)
98
99 }
100 cat("\n-----\nAnalysis of Variance Table by period\n
    -----\n")
101 for(i in 1:nv1){
102   cat("\n-----\nAnalysis of Variance Table\n-----\n
    n")
103   print(Anova)
104   cat("\n\n", "Analisis de varianza", " Para el periodo
    ",
105       lf1[i], " of ", "time", "\n-----\n")
106   print(ana.tiempo[[i]])
107   print(Shapiro.tiempo[[i]])
108   print(Bartlett.tiempo[[i]])
109 }
110 cat("\n-----\nShapiro-Wilk normality test\n-----\n")
111 print(Shapiro)
112 cat("\n-----\nBartlett homogeneity test\n-----\n")
113 print(Bartlett)
114
115 if ( valor.P[5]> sigF) {
116   cat("\n-----\nNot significant interaction:
    analyzing
117   the simple effect\n-----\n")
118   if (quali==TRUE && valor.P[2] <= sigF) {
119     cat("\n-----\nTreat\n-----\n")
120     if (mcomp == "tukey") {
121       tukey(Y,x1,GL[3],CM[3],sigT)

```

```

122 }
123 if (mcomp == "duncan") {
124     duncan(Y,x1,GL[3],CM[3], sigT)
125 }
126 if (mcomp == "lsd") {
127     lsd(Y,x1,GL[3],CM[3], sigT)
128 }
129 if (mcomp == "lsdb") {
130     lsdb(Y,x1,GL[3],CM[3], sigT)
131 }
132 if (mcomp == "sk") {
133     scottknott(Y,x1,GL[3],CM[3], sigT)
134 }
135 if (mcomp == "snk") {
136     snk(Y,x1,GL[3],CM[3], sigT)
137 }
138 if (mcomp == "ccboot") {
139     ccboot(Y,x1,GL[3],CM[3], sigT)
140 }
141 }
142 if (quali==TRUE && valor.P[2] > sigF) {
143     cat("\n-----\nAccording to the F test, the means
of
144     this treatments are statistical equal.\n-----\n")
145     mean.table <- tapply.stat(Y,x1,mean)
146     colnames(mean.table) <- c("Levels", "Means")
147     print(mean.table)
148     cat("\n-----\n")
149 }
150 if (quali== FALSE && valor.P[2] <= sigF) {
151     x1=as.numeric(x1)
152     cat("\n-----\nTreat\n-----\n-")
153     reg.poly(Y,x1,GL[3],SC[3],GL[2],SC[2])
154 }
155 if (quali== FALSE && valor.P[2] > sigF) {
156     cat("\n-----\nAccording to the F test, the means
of
157     this treatments are statistical equal.\n-----\n")
158     mean.table <- tapply.stat(Y,x1, mean)
159     colnames(mean.table) <- c("Levels", "Means")
160     print(mean.table)
161     cat("\n-----\n")
162 }
163 }
164 if (valor.P[5]<= sigF) {

```

```

165     cat("\n-----\nSignificant interaction: analyzing
the interaction
166     \n-----\n")
167     cat("\nAnalyzing ", "treatments", " inside of each
level of ",
168         "time", "\n-----\n")
169     for (i in 1:nv1) {
170         if (quali== TRUE) {
171             cat("\n-----\n", "treatments", " inside of the
level ",
172                 lf1[i], " of ", "time", "\n-----\n")
173             if (mcomp == "tukey") {
174                 mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]]+block
[x2 == lf1[i]])
175                 Anava=as.matrix(anova(mod))
176                 tukey(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
[3,1],Anava[3,2],
177                     sigT)
178             }
179             if (mcomp == "duncan") {
180                 mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]]+block
[x2 == lf1[i]])
181                 Anava=as.matrix(anova(mod))
182                 duncan(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
[3,1],Anava[3, 2],
183                     sigT)
184             }
185             if (mcomp == "lsd") {
186                 mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]]+block
[x2 == lf1[i]])
187                 Anava=as.matrix(anova(mod))
188                 lsd(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
[3,1],Anava[3, 2],
189                     sigT)
190             }
191             if (mcomp == "lsdb") {
192                 mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]]+block
[x2 == lf1[i]])
193                 Anava=as.matrix(anova(mod))
194                 lsdb(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
[3,1],Anava[3, 2],
195                     sigT)
196             }
197             if (mcomp == "sk") {
198                 mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]]+block

```

```

[x2 == lf1[i]])
199     Anava=as.matrix(anova(mod))
200     scottknott(Y[x2 == lf1[i]],x1[x2 == lf1[i]],
Anava[3,1],
201     Anava[3, 2],sigT)
202   }
203   if (mcomp == "snk") {
204     mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]]+block
[x2 == lf1[i]])
205     Anava=as.matrix(anova(mod))
206     snk(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
[3,1],Anava[3, 2],
207     sigT)
208   }
209   if (mcomp == "ccboot") {
210     mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]]+block
[x2 == lf1[i]])
211     Anava=as.matrix(anova(mod))
212     ccboot(Y[x2 == lf1[i]],x1[x2 == lf1[i]],Anava
[3,1],Anava[3, 2],
213     sigT)
214   }
215   }
216   else {
217     cat("\n-----\n", "treatments", " inside of the
level ",
218     lf1[i], " of ", "time", "\n-----\n")
219     mod<-aov(Y[x2 == lf1[i]]~x1[x2 == lf1[i]]+block[
x2 == lf1[i]])
220     Anava=as.matrix(anova(mod))
221     reg.poly(Y[x2 == lf1[i]],as.numeric(x1[x2 == lf1[
i]]),
222     Anava[3,1],Anava[3,2],Anava[1,1],Anava[1,2])
223   }
224   }
225 }
226
227 }

```

8. BIBLIOGRAFÍA.

[1] Montgomery D. design and analysis of experiments, 8th edition (2012).

[2] (s/f). Redalyc.org. Recuperado el 15 de abril de 2023,
<https://www.redalyc.org/pdf/402/40223164022.pdf>

[3] Ciencia UANL. (2012). Diseños experimentales. Sistema de Información Científica Redalyc, Red de Revistas Científicas.
<https://www.redalyc.org/pdf/402/40223164022.pdf>

[4] Hinkelmann, K. and Kempthorne, O. Design and Anaysis of Experiments. John Wiley & Sons 1994.

[5] Gutiérrez H. y De la Vara R., Análisis y Diseño de experimentos 3ra edición. McGraw Hill. 2012.

[6] Martínez G. y Robles J, Métodos y Diseño estadístico de Experimentos, Universidad de Córdoba . 2013

[7] Martínez R. y Martínez N. Diseño de Experimentos. Fondo Nacional Universitario. 1997.

[8] Juan Carlos Correa, Estudio de potencia de pruebas de Homogeneidad de varianza, Universidad Nacional de Colombia.

[9] *Ea2 function - RDocumentation.* (s/f). Rdocumentation.org. Recuperado el 15 de abril de 2023, de <https://www.rdocumentation.org/packages/easyanova/versions/7.0/topics/ea2>

[10] *fit.contrast function - RDocumentation.* (s/f). Rdocumentation.org. Recuperado el 15 de abril de 2023, de <https://www.rdocumentation.org/packages/gmodels/versions/2.18.1.1/topics/fit.contrast>

[11] *crd function - RDocumentation.* (s/f). Rdocumentation.org. Recuperado el 15 de abril de 2023, de <https://www.rdocumentation.org/packages/ExpDes/versions>

/1.2.2/topics/crd

[12] López Pérez , L. A., Melo Martínez, O. O., Melo Martínez, S. E. (2007). *Diseño de Experimentos [Métodos y Aplicaciones]*. Bogotá: Universidad Nacional de Colombia. Bogotá: Universidad Nacional de Colombia.

[13] Kuehl , R. (2001). *Diseño de experimentos: principios estadísticos para el diseño y análisis de investigaciones* (2.^a ed.). Thomson Learning. Thomson Learning.

[14] Ferreira, E. B., Cavalcanti, P., Alves, D., Maintainer, N. (2022). *Package “ExpDes”*. <https://doi.org/10.4236/am.2014.519280>

[15] (S/f). Rdocumentation.org. Recuperado el 15 de abril de 2023, de <https://www.rdocumentation.org/packages/easyanova/versions/7.0/topics/ea2>

[16] R: *Fat2.Crd function - RDocumentation*. (s/f). Rdocumentation.org. Recuperado el 8 de junio de 2023, de <https://www.rdocumentation.org/packages/ExpDes/versions/1.2.2/topics/fat2.crd>

[17] Peña. S. D. *Estadística Modelos y Métodos*. Tomo I. Alianza Editorial.S.A. Madrid. 1987.

[18] RPubs - Metodología de superficie de respuesta (RSM). (2019, 27 septiembre). RPubs. Recuperado 11 de diciembre de 2021, de https://rpubs.com/Cristina_Gil/RSM.

[19] R: *Hartley’s maximum F-ratio test of homogeneity of variances*. (s/f). R-project.org. Recuperado el 15 de abril de 2023, de <https://search.r-project.org/CRAN/refmans/PMCMRplus/html/hartleyTest.html>

[20] RPubs - *Bartlett’s test for homogeneity*. (s/f). Rpubs.com. Recuperado el 15 de abril de 2023, de <https://rpubs.com/PAVelasquezVasconez/355232>

[21] RPubs - *Test de Levene*. (s/f). Rpubs.com. Recuperado el 15 de abril de 2023, de <https://rpubs.com/icaro/411288>

- [22] *LSD. Test function - RDocumentation.* (s/f). Rdocumentation.org. Recuperado el 15 de abril de 2023, de <https://www.rdocumentation.org/packages/agricolae/versions/1.3-5/topics/LSD.test>
- [23] *duncan.test function - RDocumentation.* (s/f). Rdocumentation.org. Recuperado el 15 de abril de 2023, de <https://www.rdocumentation.org/packages/agricolae/versions/1.3-5/topics/duncan.test>
- [24] *HSD. Test function - RDocumentation.* (s/f). Rdocumentation.org. Recuperado el 15 de abril de 2023, de <https://www.rdocumentation.org/packages/agricolae/versions/1.3-5/topics/HSD.test>
- [25] *scheffe.test function - RDocumentation.* (s/f). Rdocumentation.org. Recuperado el 15 de abril de 2023, de <https://www.rdocumentation.org/packages/agricolae/versions/1.3-5/topics/scheffe.test>
- [26] López, E., González, B., Agr. Isidro Chex, P. (s/f). *Diseño y Análisis de Experimentos usando Lenguaje R y R Studio.* Amazonaws.com. Recuperado el 15 de abril de 2023, de http://rstudio-pubs-static.s3.amazonaws.com/423211_e99d2b66206f4885abed23b68dc44ba4.html
- [27] *How to Perform Dunnett's test in R.* (2021, agosto 27). R-Bloggers. <https://www.r-bloggers.com/2021/08/how-to-perform-dunnetts-test-in-r/>
- [28] Cardona. C. Aplicación de dosis de calcio sobre la calidad o firmeza del fruto del tomate. Tesis de grado de Maestría. Universidad Nacional de Colombia. Bogotá. 2003
- [29] Lozano H, Elkin D, y Ogaza G, Jacobo. Respuesta del maíz HS-36 a un programa de nutrición foliar suplementario en suelos del valle del Sinú Medio. Tesis de grado, Universidad de Córdoba 2004.
- [30] Vilorio R. y Rhenal. R. Efecto bioherbicida de extracto de mentolada (*Stemodia Duratifolia* L.S.W) y pasto Jonson (*Soorghun Halepense* L. Pers) sobre especies de arvenses y cultivos en el valle Medio del Sinú. Tesis de grado, Universidad de Córdoba, 2004.

[31] Argel R, J... Evaluación de dos especies arvenses (*Callista cordipolia* (sw) E.S. Anders y Woods) canutillo y (*Geophila macropoda* Ruiz y Pav) oreja de ratón, como cobertura potencial en cultivos de banano en Turbo Antioquia. Tesis de grado, Universidad de Córdoba, 2006.

[32] Gómez H. Estadística Experimental Aplicada a las Ciencias Agrícolas. Universidad Nacional de Colombia. Sede Medellín. 1997.

[33] Ferreira, E. B., Cavalcanti, P., Alves, D., Maintainer, N. (2022). Package “ExpDes”. <https://doi.org/10.4236/am.2014.519280>