



Diseño de experimentos

Exámen de semana 01

Kevin Heberth Haquehua Apaza

08 de setiembre del 2025

Tabla de contenidos

DCA-Comparaciones múltiples-supuestos	1
EJERCICIO	1
Preguntas	2
Solución	2

DCA-Comparaciones múltiples-supuestos

EJERCICIO

Un grupo de ingenieros agrónomos realizaron una investigación experimental en el fundo Huancayo en el distrito de Huariaca, provincia Pasco, región Pasco, en condiciones de campo con el fin de determinar el efecto de cinco fertilizantes complejos en días de floración del cultivo de lino (*Linum usitatissimum*)

Descripción de los tratamientos (fertilizantes)

- T1: YaraVera AMIDAS
- T2: YaraMila COMPLEX
- T3: YaraTera KRISTA K
- T4: N20, P20, K20
- T5: N18, P15, K10

Cuadro 1: Datos de floración de cultivo de lino (en días).

Repetición	Tratamientos					Totales
	T1	T2	T3	T4	T5	
1	86	65	75	84	80	390
2	80	67	74	81	85	387
3	80	63	75	82	84	384
4	82	70	77	84	87	400
$\sum y_{.j}$	328	265	301	331	336	1561
$\bar{y}_{.j}$	82,00	66,25	75,25	82,75	84,00	
$\bar{y}_{..}$			78,05			

Fuente: (Alcántara Proa & Atencio Falcón, 2017)



Preguntas

- Plantee y describa el modelo que considere adecuado para analizar este caso.
- ¿Se puede afirmar que al menos uno de los días de floración difiere del resto al aplicar los distintos fertilizantes complejos?
- Verifique si el modelo propuesto cumple con los supuestos.
- Estime el coeficiente de variabilidad.
- ¿Cuáles son los fertilizantes que tienen mejor efecto en la floración media en el cultivo de lino?
- ¿El tiempo promedio de floración al aplicar el fertilizante $T3$ es superior al aplicar el fertilizante $T2$ en más de 2 días?
- Al aplicar los fertilizantes $T2$ y $T3$ (en forma conjunta), ¿el tiempo medio en hacer efecto la floración es menor que al aplicar los fertilizantes $T4$ y $T5$ (en forma conjunta)?

Indicaciones finales

- El documento debe utilizar un lenguaje académico.
- Subir el informe final-documento en pdf.
- Subir el Script y data comprimido (se descontaran puntos).
- Subir tu informe antes de las 13:00 pm (07-09-2025)

Solución

- Plantee y describa el modelo que considere adecuado para analizar este caso.**

Tomando en cuenta el contexto en el que solo se desea evaluar el efecto de cinco fertilizantes, el modelo adecuado a utilizar es un diseño completo al azar. En las cuales se tienen los tratamiento de los distintos fertilizantes, si evaluamos el estudio original indica que se utiliza el diseño bloque completo al azar, en donde las repeticiones vienen a ser los bloques.

El modelo aditivo lineal es el siguiente

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Donde:

- Y_{ij} Es la floración de (*Linum usitatissimum*) del i -ésimo tratamiento y j -ésima repetición.
- μ es el efecto de la media general en la floración de (*Linum usitatissimum*).
- τ_i es el efecto del i -ésimo tratamiento.
- ϵ_{ij} es el efecto del error experimental en el i -ésimo tratamiento y j -ésima repetición. para $i = \{1,2,3,4,5\}$ tratamientos y $j = \{1,2,3,4\}$ repeticiones. Además, el modelo asume que $\epsilon_{ij} \sim N(0, \sigma^2)$



b) ¿Se puede afirmar que al menos uno de los días de floración difiere del resto al aplicar los distintos fertilizantes complejos?

Empezemos desarrollando los datos del archivo Excel que también se adjuntará en el trabajo, para posteriormente hacer la lectura del archivo.

```
library(readxl)
library(here)

## here() starts at C:/Users/La Fuente Estadistic/Desktop/INLASER-LA FUENTE/General_Analytics/Maestria_
UNSAAC

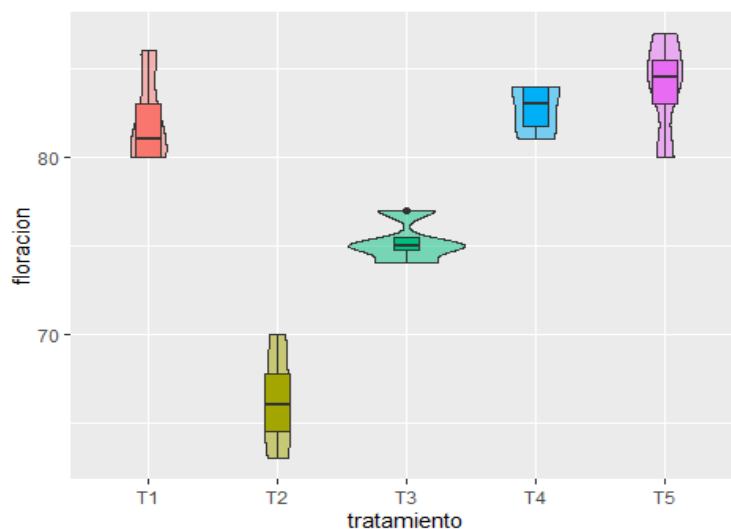
data <- read_excel(here("11 Diseno Experimentos/Examen_1/data_floracion.xlsx"))
head(data)

## # A tibble: 6 × 2
##   tratamiento floracion
##   <chr>      <dbl>
## 1 T1        86
## 2 T2        65
## 3 T3        75
## 4 T4        84
## 5 T5        80
## 6 T1        80

attach(data) # Para tomar Los datos
data$tratamiento = as.factor(tratamiento) # convierte tratamiento en factor
```

Empezemos viendo un gráfico de la dispersión y el efecto de los tratamientos en la floración del cultivo de (*Linum usitatissimum*)

```
library(ggplot2)
ggplot(data, aes(x = tratamiento, y = floracion, fill = tratamiento)) +
  geom_violin(alpha = 0.5) +
  theme(legend.position = "none") +
  geom_boxplot(width = 0.2)
```



Interpretación: Se observa que el YaraMila COMPLEX presenta la menor floración de (*Linum usitatissimum*), seguido del YaraTera KRISTA K. Seguidamente los tratamientos: YaraVera AMIDAS, (N20, P20, K20), (N18, P15, K10) presentan la mayor floración de (*Linum usitatissimum*) sin mostrar diferencias entre ellos.

Ahora definamos el modelo

```
mod.dca = lm(floracion ~ tratamiento) # modelo DCA
```

Y veamos el análisis de la varianza para afirmar que al menos uno de los días de floración difiere del resto al aplicar los distintos fertilizantes complejos.

```
summary(aov(mod.dca)) # Análisis de varianza
```

```
##      Df Sum Sq Mean Sq F value  Pr(>F)
## tratamiento  4  880.7   220.18   37.42 1.23e-07 ***
## Residuals   15   88.3    5.88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Efectivamente, se puede afirmar que al menos uno de los días de floración de (*Linum usitatissimum*) difiere del resto al aplicar los distintos fertilizantes complejos. Ya que existe evidencia suficiente para rechazar la hipótesis nula que indica que los días de floración de (*Linum usitatissimum*) no difieran en los tratamientos complejos.

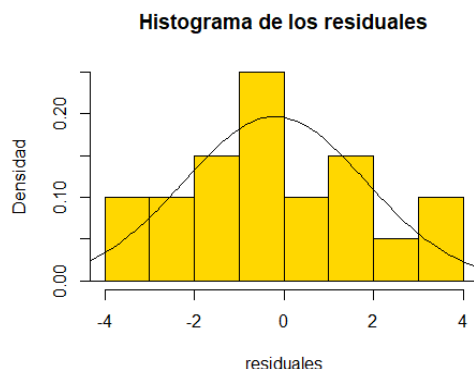
c) Verifique si el modelo propuesto cumple con los supuestos.

- **Normalidad de errores**

Nos planteamos las hipótesis H_0 : Los errores se distribuyen normalmente H_1 : Los errores no se distribuyen normalmente

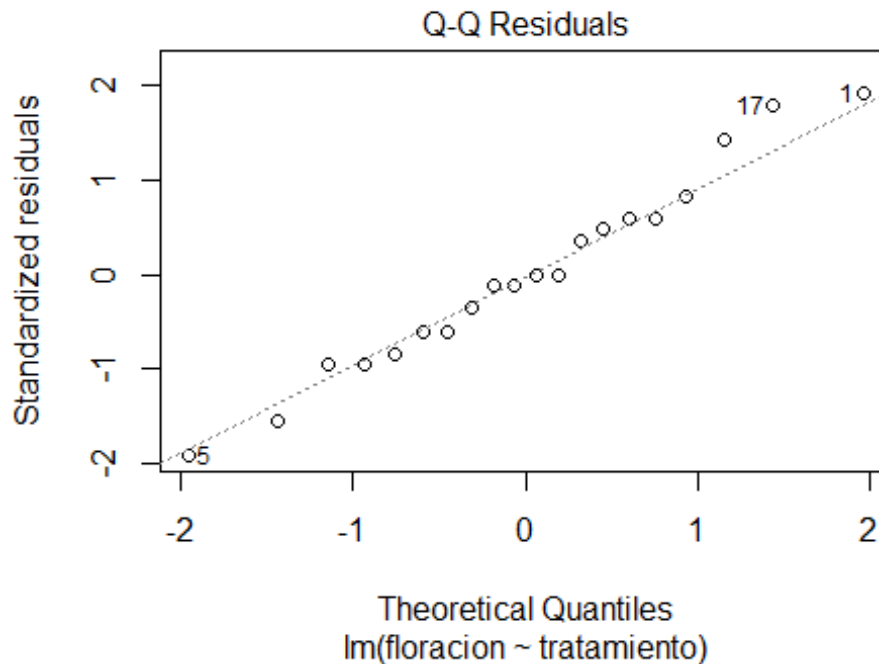
Veamos los residuales del modelo

```
library(agricolae)
residuales = residuals(mod.dca) # extrae residuales
histo      = hist(residuales, col="gold", xlab="residuales", main="Histograma de los residuales", ylab="Densidad", prob=1)
normal.freq(histo, frequency=3) # densidad
```



Se observa que los errores se distribuyen a una normal, veamos el QQ-plot

```
plot(mod.dca, which = 2)
```



A excepción de unos valores, la mayor parte se centra a la recta de residuales, veamos ahora la prueba de Shapiro-Wilk de normalidad de residuos

```
shapiro.test(residuales)

##
## Shapiro-Wilk normality test
##
## data: residuales
## W = 0.97749, p-value = 0.8977
```

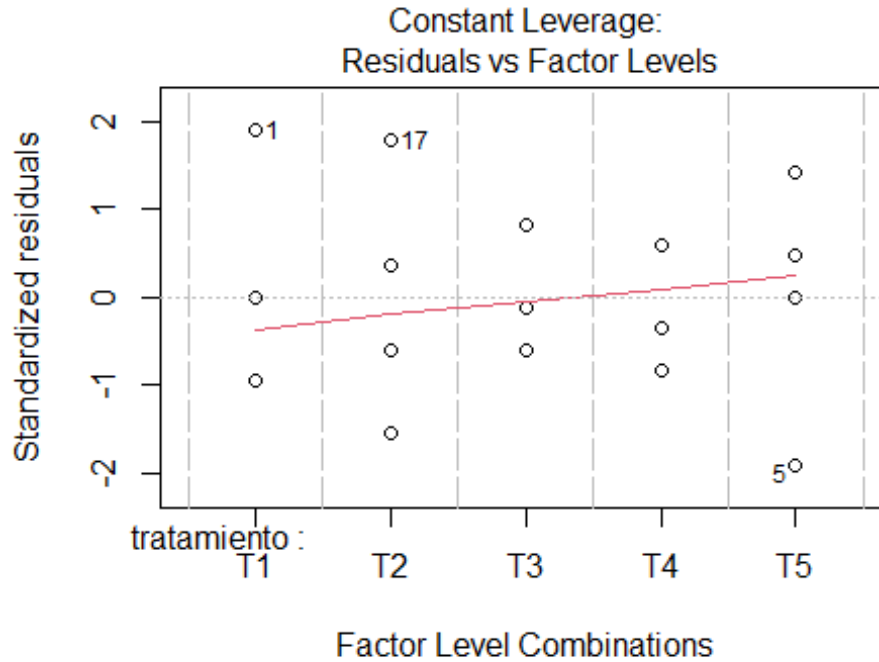
Como se tiene un pvalor > 0.05 no se rechaza H_0 lo cual indica que el modelo cumple el supuesto de normalidad de errores.

- **Homocedasticidad**

Nos planteamos las hipótesis: H_0 : Existe homogeneidad de varianzas H_1 : No existe homogeneidad de varianzas

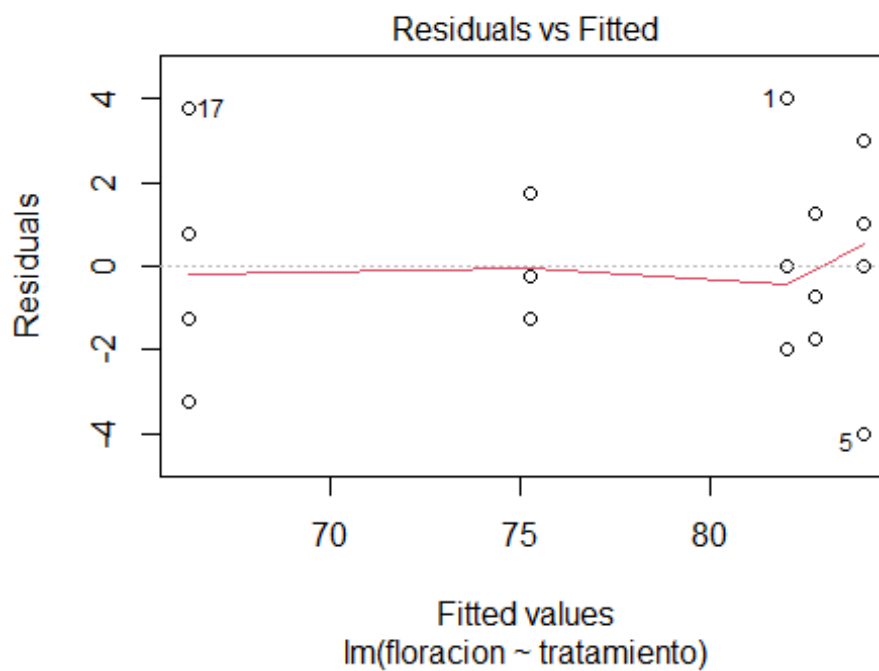
Veamos si la variabilidad es distinta en los factores

```
plot(mod.dca, which = 5)
```



Y si la variabilidad se incrementa con la media general

```
plot(mod.dca, which = 1)
```



Se observa que la varianza se mantiene constante en los residuos. Utilizemos la prueba de Breusch-Pagan para evaluar este supuesto

```
library(car)

## Loading required package: carData

ncvTest(mod.dca)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.06641474, Df = 1, p = 0.79663
```

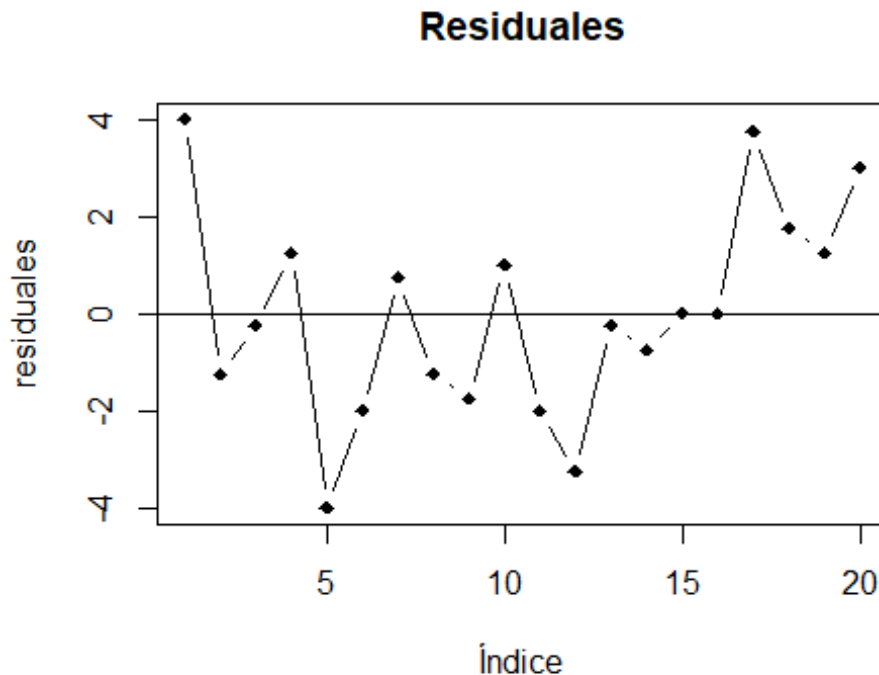
Como el pvalor > 0.05 no se rechaza H_0 en lo cual se indica que la varianza es constante en los residuos, por lo que se cumple el supuesto de homocedasticidad.

- **Independencia de errores**

Para ver la no autocorrelación de los errores nos planteamos las hipótesis H_0 : Existe independencia de errores H_1 : No existe independencia de errores

Veamos de forma gráfica la autocorrelación de los errores

```
plot(residuals(mod.dca), pch = 18,
     type = "b", ylab = "residuales",
     xlab = "Índice", main="Residuales")
abline(h=0)
```





No se observa un patrón de dependencia en los errores, realizemos la prueba de Durbin-Watson

```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

dwtest(mod.dca, alternative = c("two.sided"))

##
## Durbin-Watson test
##
## data: mod.dca
## DW = 1.3994, p-value = 0.3096
## alternative hypothesis: true autocorrelation is not 0
```

Como el pvalor > 0.05 no se rechaza H_0 lo cual indica que existe independencia de errores o no están autocorrelacionados de primer orden, lo cual indica que se cumple el supuesto de independencia de errores.

d) **Estime el coeficiente de variabilidad.**

Para cuantificar el grado de homogeneidad de los resultados del experimento veamos el coeficiente de variabilidad

```
# Coeficiente de variación
cv = cv.model(mod.dca)
cv

## [1] 3.107698
```

El cual nos da un valor de 3.11% el cual es menor al 10% en el que podemos indicar que hay homogeneidad en los resultados del experimento, esto indica que no hay factores adicionales que puedan influir en el estudio realizado.

e) **¿Cuáles son los fertilizantes que tienen mejor efecto en la floración media en el cultivo de lino?**

Veamos la media general

```
med_general = mean(floracion); med_general

## [1] 78.05
```

Ahora la media de los tratamientos con los efectos


```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':
##
##   recode

## The following objects are masked from 'package:stats':
##
##   filter, lag

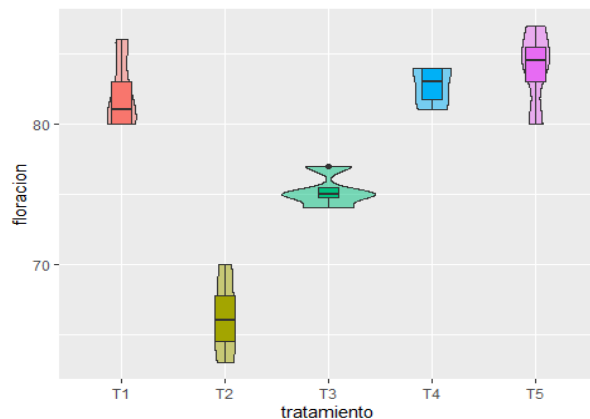
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

med_trat = aggregate(floracion~tratamiento, FUN = mean)
med_trat%>% rename(Media = floracion) %>% mutate(Efecto = Media-med_general)

## tratamiento Media Efecto
## 1      T1 82.00  3.95
## 2      T2 66.25 -11.80
## 3      T3 75.25  -2.80
## 4      T4 82.75  4.70
## 5      T5 84.00  5.95
```

Se observa que el tratamiento que tiene el menor efecto en la floración de (*Linum usitatissimum*) es YaraMila COMPLEX y el que presenta el mayor efecto es (N18, P15, K10). Asimismo por el gráfico visto anteriormente

```
ggplot(data, aes(x = tratamiento, y = floracion, fill = tratamiento)) +
  geom_violin(alpha = 0.5) +
  theme(legend.position = "none")+
  geom_boxplot(width = 0.2)
```



Se observa que los tratamientos que poseen mejores efectos para mejorar la floración de (*Linum usitatissimum*), son los tratamientos: YaraVera AMIDAS, (N20, P20, K20), (N18, P15, K10) aunque no se muestra diferencias entre ellos.

- f) **¿El tiempo promedio de floración al aplicar el fertilizante T_3 es superior al aplicar el fertilizante T_2 en más de 2 días?**

Formalizamos la expresión

$$\mu_3 > 2\mu_2$$

$$\mu_3 - 2\mu_2 > 0$$

Nos planteamos las hipótesis $H_0: \mu_3 - 2\mu_2 \leq 0$ $H_a: \mu_3 - 2\mu_2 > 0$

Y por medio de los contrastes ortogonales realizamos la comparación

```
library(gmodels)
con = c(0, -2, 1, 0, 0)
fit.contrast(mod.dca, "tratamiento", con)

##              Estimate Std. Error t value Pr(>|t|)
## tratamiento c=( 0 -2 1 0 0 ) 21.66667  2.767777 7.828183 1.123894e-06
```

Como el pvalor < 0.05 , rechazamos la hipótesis nula. Por lo que el promedio de floración al aplicar el fertilizante YaraTera KRISTA K es superior al fertilizante YaraMila COMPLEX en más de dos días.

- g) **Al aplicar los fertilizantes T_2 y T_3 (en forma conjunta), ¿el tiempo medio en hacer efecto la floración es menor que al aplicar los fertilizantes T_4 y T_5 (en forma conjunta)?**

Formalizamos la expresión

$$\mu_2 + \mu_3 < \mu_4 + \mu_5$$

$$\mu_2 + \mu_3 - \mu_4 - \mu_5 < 0$$

Nos planteamos las hipótesis $H_0: \mu_2 + \mu_3 - \mu_4 - \mu_5 \geq 0$ $H_a: \mu_2 + \mu_3 - \mu_4 - \mu_5 < 0$

Y por medio de los contrastes ortogonales realizamos la comparación

```
con = c(0, 0.5, 0.5, -0.5, -0.5)
fit.contrast(mod.dca, "tratamiento", con)

##              Estimate Std. Error t value
## tratamiento c=( 0 0.5 0.5 -0.5 -0.5 ) -12.625  1.212779 -10.40997
##              Pr(>|t|)
## tratamiento c=( 0 0.5 0.5 -0.5 -0.5 ) 2.937746e-08
```

Como el pvalor < 0.05 , rechazamos la hipótesis nula. Por lo que al aplicar los fertilizantes YaraMila COMPLEX y YaraTera KRISTA K en forma conjunta el tiempo medio en hacer efecto la floración es menor que al aplicar los fertilizantes (N20, P20, K20) y (N18, P15, K10) en forma conjunta).