Análisis y Diseño de Experimentos

Segundo Trabajo Encargado

Se realizó un experimento para observar el rendimiento en kilogramos por parcela de 5 variedades de garbanzo (A, B, C, D, E) en el cual se tuvo que utilizar el diseño Cuadrado Latino. Las filas fueron definidas como niveles de riego y las columnas como fertilidad del suelo. Los datos se presentan a continuación

Niveles	Fertilidad del suelo				
de riego	1	2	3	4	5
1	B = 65	C = 80	A = 55	E = 83	D = 80
2	C = 95	A = 60	E = 94	D = 95	B = 62
3	A = 63	E = 98	D = 79	B = 69	C = 100
4	E = 97	D =94	B = 46	C = 71	A = 42
5	D = 76	B = 54	C = 106	A = 36	E = 96

a. Presente el modelo aditivo lineal e interprete cada uno de sus componentes en términos del enunciado.

El modelo aditivo lineal para un Diseño Cuadrado Latino es el siguiente:

$$Y_{(i)jk} = \mu + \tau_{(i)} + \beta_j + \gamma_k + \epsilon_{(i)jk}$$

Para todo : i, j, k = 1,. . . ,t

Donde:

- $Y_{(i)jk}$ es el valor o rendimiento observado en el i-ésimo variedad de Garbanzo, j-ésima nivel de riesgo, k-ésima fertilidad del suelo.
- μ es el efecto de la media general.
- $\tau_{(i)}$ es el efecto del i-ésima variedad de garbanzo
- β_j es el efecto de la j-ésima nivel de riesgo
- γ_k es el efecto de la k-ésima fertilidad del suelo
- $\epsilon_{(i)jk}$ es el efecto del error experimental en el i-ésimo variedad de Garbanzo, j-ésima nivel de riesgo, k-ésima fertilidad del suelo.
- t es el número de tratamientos que es igual al número de niveles de riego y al número de fertilidades de suelo.

b. Realice el diagnóstico del modelo utilizando gráfico y pruebas de hipótesis

```
library(agricolae)
garbanzo <- read.table("garbanzo.txt", header = TRUE)
head(garbanzo)
```

```
##   rendimiento fila columna tratamiento
## 1          65    1       1           B
## 2          95    1       2           C
## 3          63    1       3           A
## 4          97    1       4           E
## 5          76    1       5           D
## 6          80    2       1           C
```

```
str(garbanzo)
```

```
## 'data.frame':    25 obs. of  4 variables:
##  $ rendimiento: int  65 95 63 97 76 80 60 98 94 54 ...
```

```
##  $ fila       : int  1 1 1 1 1 2 2 2 2 2 ...
##  $ columna    : int  1 2 3 4 5 1 2 3 4 5 ...
##  $ tratamiento: Factor w/ 5 levels "A","B","C","D",..: 2 3 1 5 4 3 1 5 4 2 ...
```

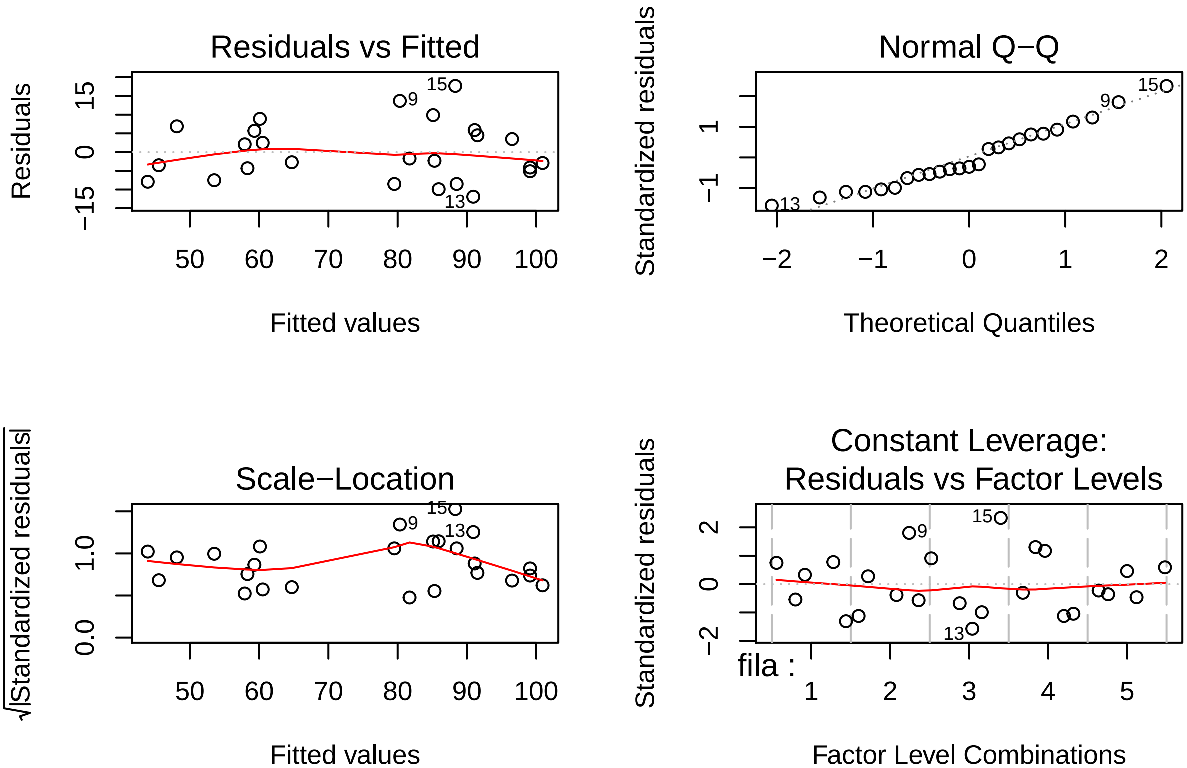
```
garbanzo$fila<-factor(garbanzo$fila)
garbanzo$columna<-factor(garbanzo$columna)
garbanzo$tratamiento<-factor(garbanzo$tratamiento)
str(garbanzo)
```

```
## 'data.frame':    25 obs. of  4 variables:
##  $ rendimiento: int  65 95 63 97 76 80 60 98 94 54 ...
##  $ fila       : Factor w/ 5 levels "1","2","3","4",..: 1 1 1 1 1 2 2 2 2 2 ...
##  $ columna    : Factor w/ 5 levels "1","2","3","4",..: 1 2 3 4 5 1 2 3 4 5 ...
##  $ tratamiento: Factor w/ 5 levels "A","B","C","D",..: 2 3 1 5 4 3 1 5 4 2 ...
```

```
mod<-lm(rendimiento~.,data=garbanzo)
anva<-anova(mod)
anva
```

```
## Analysis of Variance Table
##
## Response: rendimiento
##             Df Sum Sq Mean Sq F value   Pr(>F)
## fila         4  193.0   48.24  0.4030 0.802966
## columna      4  569.4  142.34  1.1891 0.364656
## tratamiento  4 7458.6 1864.64 15.5767 0.000107 ***
## Residuals   12 1436.5  119.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(mod)
```




```r
ri<-rstandard(mod)
shapiro.test(ri)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ri
## W = 0.95981, p-value = 0.4108
```

```r
library(car)
```

```
## Loading required package: carData
```

```r
ncvTest(mod)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.4684703, Df = 1, p = 0.49369
```

d. Realice la prueba de Tukey. De sus conclusiones y recomendaciones en término de las variables en estudio

```r
library(multcomp)
```

```
## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'TH.data'
```

```
## The following object is masked from 'package:MASS':
##
##     geyser
```

```r
amod<-aov(rendimiento~.,data=garbanzo)
comptrat<-glht(amod,linfct=mcp(tratamiento="Tukey"))
summary(comptrat)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = rendimiento ~ ., data = garbanzo)
##
## Linear Hypotheses:
##            Estimate Std. Error t value Pr(>|t|)
## B - A == 0     8.00       6.92   1.156 0.774801
## C - A == 0    39.20       6.92   5.665 0.000851 ***
## D - A == 0    33.60       6.92   4.856 0.002939 **
## E - A == 0    42.40       6.92   6.127 0.000383 ***
## C - B == 0    31.20       6.92   4.509 0.005225 **
## D - B == 0    25.60       6.92   3.700 0.020780 *
## E - B == 0    34.40       6.92   4.971 0.002434 **
## D - C == 0    -5.60       6.92  -0.809 0.922779
## E - C == 0     3.20       6.92   0.462 0.989395
## E - D == 0     8.80       6.92   1.272 0.712185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

e.Realice la prueba de Duncan. Compare con los resultados de la prueba de Tukey

```r
library(agricolae)
anva<-aov(mod)
compd<-duncan.test(anva,"tratamiento")
compd
```

```
## $statistics
##    MSerror Df  Mean      CV
##   119.7067 12 75.84 14.4265
##
## $parameters
##     test      name.t ntr alpha
##   Duncan tratamiento   5  0.05
##
## $duncan
##      Table CriticalRange
## 2 3.081307      15.07680
## 3 3.225244      15.78108
## 4 3.312453      16.20779
## 5 3.370172      16.49021
##
## $means
##   rendimiento       std r Min Max Q25 Q50 Q75
```

```
## A        51.2 11.691878 5  36  63  42  55  60
## B        59.2  9.203260 5  46  69  54  62  65
## C        90.4 14.501724 5  71 106  80  95 100
## D        84.8  8.983318 5  76  95  79  80  94
## E        93.6  6.107373 5  83  98  94  96  97
##
## $comparison
## NULL
##
## $groups
##   rendimiento groups
## E        93.6      a
## C        90.4      a
## D        84.8      a
## B        59.2      b
## A        51.2      b
##
## attr(,"class")
## [1] "group"
```

f.Asumiendo que la variedad A es el testigo, realice la prueba de Dunnett

```
compdunett<-glht(anva,linfct=mcp(tratamiento="Dunnett"))
summary(compdunett)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: aov(formula = mod)
##
## Linear Hypotheses:
##            Estimate Std. Error t value Pr(>|t|)
## B - A == 0     8.00       6.92   1.156  0.61601
## C - A == 0    39.20       6.92   5.665  < 0.001 ***
## D - A == 0    33.60       6.92   4.856  0.00144 **
## E - A == 0    42.40       6.92   6.127  < 0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Según la prueba de Duncan se ha encontrado:

- No hay diferencia entre la variedad de garbanzo B y A
- Hay diferencia altamente muy significativa entre las variedades de garbanzo C y A (***)
- Hay diferencia muy significativa entre las variedades de garbanzo D y A (**)
- Hay diferencia altamente muy significativa entre las variedades de garbanzo E - A (***)

g. Se desea comparar la media de los rendimientos obtenidos con la variedades A y B versus la media de los rendimientos obtenidos con las variedades C, D y E. Obtenga los contrastes ortogonales y utilice la prueba de F para probar el contraste dado

$$HO : 3\mu_A + 3\mu_B = 2\mu_C + 2\mu_D + 2\mu_E$$

$$H1 : 3\mu_A + 3\mu_B \neq 2\mu_C + 2\mu_D + 2\mu_E$$

Matriz de Constrastes

Coeficientes	T_A	T_B	T_C	T_D	T_E
C_{1i}	-3	-3	2	2	2
C_{2i}	-1	1	0	0	0
C_{3i}	0	0	1	-2	1
C_{4i}	3	0	-1	-1	-1
C_{5i}	0	3	-1	-1	-1

Constrastes Ortogonales

Item	*Expresión*
Q_1	$-3Y_A - 3Y_B + 2Y_C + 2Y_D + 2Y_E$
Q_2	$-Y_A + Y_B$
Q_3	$Y_C - 2Y_D + Y_E$
Q_4	$3Y_A - Y_C - Y_D - Y_E$
Q_5	$3Y_B - Y_C - Y_D - Y_E$

```r
summary(anva)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## fila         4    193    48.2   0.403 0.802966
## columna      4    569   142.3   1.189 0.364656
## tratamiento  4   7459  1864.6  15.577 0.000107 ***
## Residuals   12   1436   119.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
yp <- sort(tapply(garbanzo$rendimiento,garbanzo$tratamiento,mean))
yp
```

```
##    A    B    D    C    E
## 51.2 59.2 84.8 90.4 93.6
```

```r
c1 <- c(-3,-3,2,2,2)
tc <- (t(c1)%*%yp)/sqrt((119.7/4)*sum(c1^2))
tc
```

```
##          [,1]
## [1,] 6.888616
```

```r
pvalue <- 2*(1-pt(tc,12))
pvalue
```

```
##              [,1]
## [1,] 1.677831e-05
```

El p.valor es menor a 0.01, por lo tanto hay evidencia estadística para decir que se encontró una diferencia altamente significa, por lo tanto se acepta la hipótesis alterna, es decir que la media de los rendimientos obtenidos con la variedades de garbanzos A y B es diferente a la media de los rendimientos obtenidos con las variedades de garbanzos C, D y E.