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:

Para todo : i, j, k = 1,…,t

Donde:

* es el valor o rendimiento observado en el i-ésimo variedad de Garbanzo, j-ésima nivel de riesgo, k-ésima fertilidad del suelo.
* es el efecto de la media general.
* es el efecto del i-ésima variedad de garbanzo
* es el efecto de la j-ésima nivel de riesgo
* es el efecto de la k-ésima fertilidad del suelo
* 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 2 1 C  
## 3 63 3 1 A  
## 4 97 4 1 E  
## 5 76 5 1 D  
## 6 80 1 2 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 2 3 4 5 1 2 3 4 5 ...  
## $ columna : int 1 1 1 1 1 2 2 2 2 2 ...  
## $ tratamiento: Factor w/ 5 levels "A","B","C","D",..: 2 3 1 5 4 3 1 5 4 2 ...

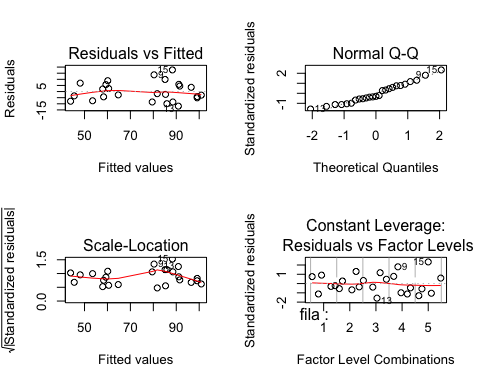
garbanzo$fila<-factor(garbanzo$fila)  
garbanzo$columna<-factor(garbanzo$columna)  
garbanzo$tratamiento<-factor(garbanzo$tratamiento)  
str(garbanzo)

## '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 2 3 4 5 1 2 3 4 5 ...  
## $ columna : Factor w/ 5 levels "1","2","3","4",..: 1 1 1 1 1 2 2 2 2 2 ...  
## $ 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 4 569.4 142.34 1.1891 0.364656   
## columna 4 193.0 48.24 0.4030 0.802966   
## 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)



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

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

library(car)

## Loading required package: carData

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

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

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 0.774835   
## C - A == 0 39.20 6.92 5.665 0.000817 \*\*\*  
## D - A == 0 33.60 6.92 4.856 0.002940 \*\*   
## E - A == 0 42.40 6.92 6.127 0.000392 \*\*\*  
## C - B == 0 31.20 6.92 4.509 0.005247 \*\*   
## D - B == 0 25.60 6.92 3.700 0.020783 \*   
## E - B == 0 34.40 6.92 4.971 0.002431 \*\*   
## D - C == 0 -5.60 6.92 -0.809 0.922791   
## E - C == 0 3.20 6.92 0.462 0.989387   
## E - D == 0 8.80 6.92 1.272 0.712151   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

Se ha encontrados diferencias altamente muy significativas entre las siguientes comparaciones de medias de rendimientos:

* Entre la media de los rendimientos obtenidos con los garbanzos de tipo C y A
* Entre la media de los rendimientos obtenidos con los garbanzos de tipo E y A

Se ha encontrados diferencias altamente significativas entre las siguientes comparaciones de medias de rendimientos:

* Entre la media de los rendimientos obtenidos con los garbanzos de tipo D y A
* Entre la media de los rendimientos obtenidos con los garbanzos de tipo C y B
* Entre la media de los rendimientos obtenidos con los garbanzos de tipo E y B

Se ha encontrados diferencias significativas entre las siguientes comparaciones de medias de rendimientos:

* Entre la media de los rendimientos obtenidos con los garbanzos de tipo D y B

Entre las otras comparaciones no se ha encontrados diferencias significativas a un nivel de significación del 10º%

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

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

Se ha ordenado de menor a mayor y se muestran los que no tienen difencias significativas :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | B | D | C | E |
| 51.2 | 59.2 | 84.8 | 90.4 | 93.6 |
| . | . | — | — | — |
| — | — |  |  |  |

Se recomienda emplear el tipo de garbanzo C y E porque no tienen diferencias significativas y se obtiene la mayor rendimiento.

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

library(agricolae)  
anva<-aov(mod)  
compd<-duncan.test(anva,"tratamiento", group=FALSE)  
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  
## difference pvalue signif. LCL UCL  
## A - B -8.0 0.2701 -23.076797 7.076797  
## A - C -39.2 0.0002 \*\*\* -55.407794 -22.992206  
## A - D -33.6 0.0005 \*\*\* -49.381079 -17.818921  
## A - E -42.4 0.0001 \*\*\* -58.890210 -25.909790  
## B - C -31.2 0.0010 \*\*\* -46.981079 -15.418921  
## B - D -25.6 0.0030 \*\* -40.676797 -10.523203  
## B - E -34.4 0.0005 \*\*\* -50.607794 -18.192206  
## C - D 5.6 0.4341 -9.476797 20.676797  
## C - E -3.2 0.6520 -18.276797 11.876797  
## D - E -8.8 0.2494 -24.581079 6.981079  
##   
## $groups  
## NULL  
##   
## attr(,"class")  
## [1] "group"

Se ha encontrados diferencias altamente muy significativas entre las siguientes comparaciones de medias de rendimientos:

* Entre la media de los rendimientos obtenidos con los garbanzos de tipo A y C
* Entre la media de los rendimientos obtenidos con los garbanzos de tipo A y D
* Entre la media de los rendimientos obtenidos con los garbanzos de tipo A y E
* Entre la media de los rendimientos obtenidos con los garbanzos de tipo B y C
* Entre la media de los rendimientos obtenidos con los garbanzos de tipo B y E

Se ha encontrados diferencias altamente significativas entre las siguientes comparaciones de medias de rendimientos:

* Entre la media de los rendimientos obtenidos con los garbanzos de tipo B y D

Entre las otras comparaciones no se ha encontrados diferencias significativas a un nivel de significación del 10º%

Se agrupa las medias de menor a mayor y se agrupa las que no tienen diferencias significativas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | B | D | C | E |
| 51.2 | 59.2 | 84.8 | 90.4 | 93.6 |
| . | . | — | — | — |
| — | — |  |  |  |

Se recomienda emplear el tipo de garbanzo C y E porque no tienen diferencias significativas y se obtiene la mayor rendimiento.

Se llega a la misma conclusión con las pruebas de Tukey y Duncan, pero se identifica que la prueba de Duncan encuentra mayores diferencias altamente muy significativa que la prueba de Tukey

#### 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 0.61606   
## C - A == 0 39.20 6.92 5.665 < 0.001 \*\*\*  
## D - A == 0 33.60 6.92 4.856 0.00136 \*\*   
## 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:

* Hay diferencia altamente muy significativa entre las variedades de garbanzo C y A (\*\*\*)
* Hay diferencia altamente muy significativa entre las variedades de garbanzo E - A (\*\*\*)
* Hay diferencia muy significativa entre las variedades de garbanzo D y A (\*\*)
* No hay diferencia entre la variedad de garbanzo B y A a un nivel de significancia del 10%

##### 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

$$ H1: 3\mu\_A + 3\mu\_B \not= 2\mu\_C + 2\mu\_D + 2\mu\_E $$

Matriz de Constrastes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Coeficientes |  |  |  |  |  |
|  | -3 | -3 | 2 | 2 | 2 |
|  | -1 | 1 | 0 | 0 | 0 |
|  | 0 | 0 | 1 | -2 | 1 |
|  | 3 | 0 | -1 | -1 | -1 |
|  | 0 | 3 | -1 | -1 | -1 |

Constrastes Ortogonales

|  |  |
| --- | --- |
| Item | $Expresi\'on$ |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

summary(anva)

## Df Sum Sq Mean Sq F value Pr(>F)   
## fila 4 569 142.3 1.189 0.364656   
## columna 4 193 48.2 0.403 0.802966   
## 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

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

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

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

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

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.