**Segundo Trabajo Encargado de Análisis y Diseño de Experimentos**

1. 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
2. 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 |

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

**Respuesta:**

El modelo aditivo lineal para el diseño de cuadrado latino es:

Donde:

: Es el valor o rendimiento observado en el i-ésimo tratamiento, j-ésima fila, k-ésima columna. En ese sentido, para este caso sería el rendimiento observado por parcela de una variedad de garbanzo (la variedad de garbanzo es el tratamiento), dado un nivel de riego y fertilidad de suelo. Esto queda registrado en cada celda del cuadrado latino.

: Es el efecto de la media general, para este caso sería el efecto común dado a cada parcela, pues el rendimiento se observa por parcela, siendo la parcela la unidad experimental.

: Es el efecto del i-ésimo tratamiento, en este caso el efecto de la variedad de garbanzo “i”, para este ejercicio existen 5 variedades de garbanzo. Por lo tanto, existen 5 tratamientos.

: Es el efecto del j-ésimo bloque fila, en este caso el efecto atribuido a cada uno de los cinco niveles de riego que se plantean en este ejercicio.

: Es el efecto del k-ésimo bloque columna, en este caso el efecto atribuido a cada uno de los cinco niveles de fertilidad de suelo.

: Es el efecto del error experimental en el i-ésimo tratamiento, j-ésimo bloque fila, k-ésimo bloque columna. Es decir, para este caso, el efecto del error experimental para la variedad de garbanzo “i”, dado un nivel de riego “j” y un nivel de fertilidad de suelo “k”. Donde i, j, k van del 1 al 5.

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

**Respuesta**

Los supuestos del modelo son:

1. Aditividad: Los efectos del modelo son aditivos.
2. Linealidad: Las relaciones entre los efectos del modelo son lineales.
3. Normalidad: Los errores del modelo tiene una distribución Normal con media cero y variancia *σ*2.
4. Independencia. Los resultados obtenidos en el experimento son independientes entre sí.
5. Homogeneidad de Variancias: Las diferentes poblaciones generadas por la aplicación de los diferentes tratamientos tienen variancias iguales (*σ*2).
6. No existe interacción entre los bloques por filas y columnas y los tratamientos.

Para realizer el diagnóstico del modelo y las pruebas de hipótesis elaboramos el siguiente código en RStudio:

garbanzo<-read.delim("clipboard")

names(garbanzo)

#[1] "rendimiento" "fila" "columna" "tratamiento"

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

# 7 60 2 2 A

# 8 98 3 2 E

# 9 94 4 2 D

# 10 54 5 2 B

# 11 55 1 3 A

# 12 94 2 3 E

# 13 79 3 3 D

# 14 46 4 3 B

# 15 106 5 3 C

# 16 83 1 4 E

# 17 95 2 4 D

# 18 69 3 4 B

# 19 71 4 4 C

# 20 36 5 4 A

# 21 80 1 5 D

# 22 62 2 5 B

# 23 100 3 5 C

# 24 42 4 5 A

# 25 96 5 5 E

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 ...

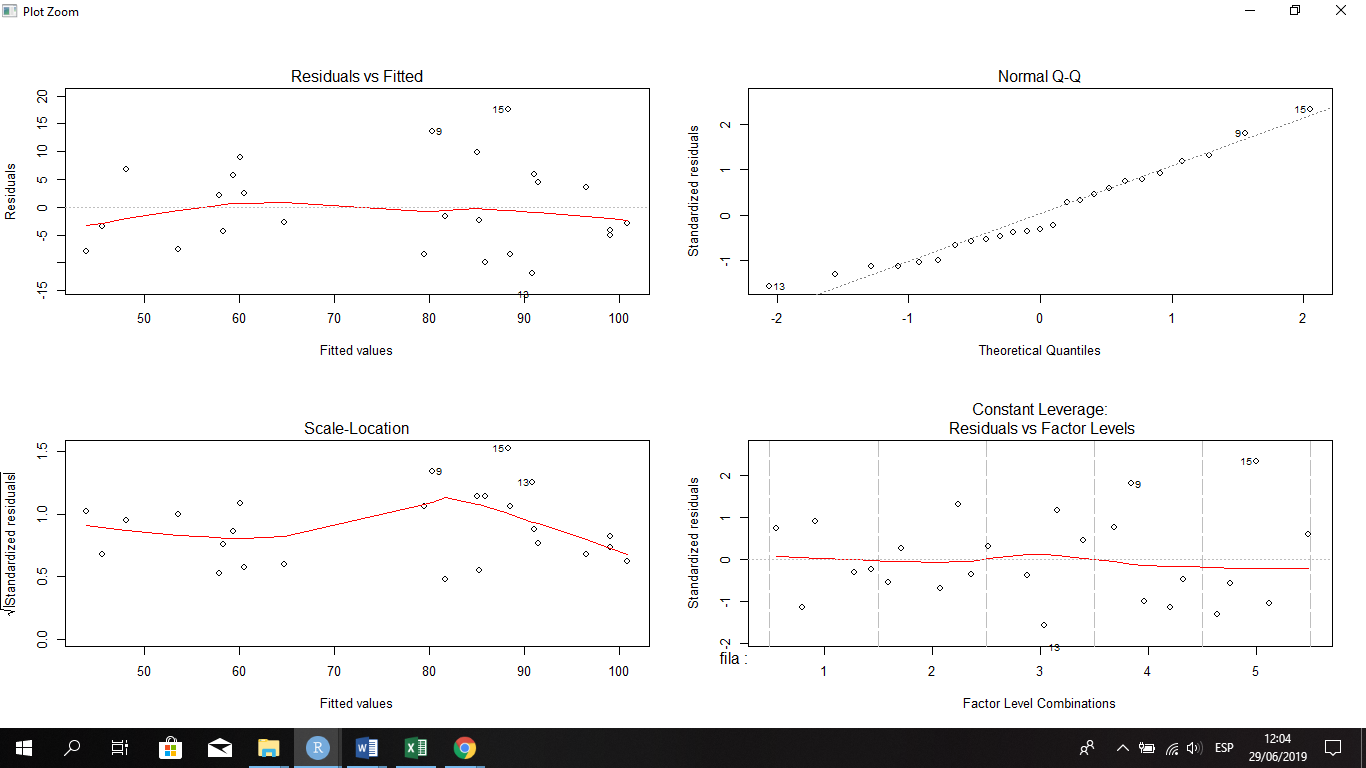
# $ tratamiento: Factor w/ 5 levels "A","B","C","D",..: 2 3 1 5 4 3 1 5 4 2 ...

attach(garbanzo)

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

par(mfrow=c(2,2))

plot(mod)



Análisis de los cuatros gráficos

**En el primer gráfico** se puede observar que la regresión no paramétrica del rendimiento predicho sobre los residuales (lowes) crece en el tramo de 50 a 60, luego decrece ligeramente hasta que el rendimiento predicho toma el valor de 80, luego crece ligeramente y en el tramo de 90 a 100 decrece, los valores de los residuales de la regresión no paramétrica van de aproximadamente de -4 a 1, la mayor parte de la regresión no paramétrica coincide con la recta horizontal que pasa por el origen, esto indica que cumple con el supuesto de linealidad del modelo.

**En el segundo gráfico** se observa que los cuantiles de los residuales estandarizados casi se sobreponen a la recta que contiene los cuantiles de la distribución normal estándar, pero al inicio se observan valores alejados a la recta punteada, esto estaría indicando que posiblemente cumpla con el supuesto de distribución normal de los errores, para verificarlo se realizará la prueba de normalidad de los errores.

> ri<-rstandard(mod)

> shapiro.test(ri)

Shapiro-Wilk normality test

data: ri

W = 0.95981, p-value = 0.4108

El pvalor de 0.4108 indica que no se rechaza la hipótesis nula de normalidad de los residuos estandarizados. Por lo tanto, se cumple el supuesto que los errores del modelo siguen una distribución normal.

**En el tercer gráfico** se observa la regresión no paramétrica de los valores predichos del rendimiento sobre la raíz cuadrada de los valores absolutos de los residuales estandarizados, se puede observar que conforme aumenta los valores predichos, la variabilidad de los residuales disminuye al inicio hasta 0.8, luego crece hasta el valor de 1.2 y finalmente decae de forma constante. Esto afectaría el cumplimiento del supuesto de homogeneidad de variancia, entonces se tendrá que verificar mediante prueba de hipótesis.

**Prueba de Score**

**Hipótesis:**

Ho : La varianza del error es constante

H1 : La varianza del error no es constante

library(car)

ncvTest(mod)

# Non-constant Variance Score Test

# Variance formula: ~ fitted.values

# Chisquare = 0.4684703, Df = 1, p = 0.49369

Al ser una muestra pequeña se trabajará con un nivel de significancia de 0.1. Se observa el pvalor de 0.49369 está por encima de dicho valor por lo que existe suficiente evidencia estadística para no rechaza la hipótesis planteada. Por lo tanto, se concluye que la varianza del error es constante, cumpliéndose así con el supuesto de homogeneidad de varianzas.

**En el cuarto gráfico** se observan los residuales estandarizados. Siendo los límites ±2, se aprecia que la observación 15 es un valor extremo. Se recomienda realizar las pruebas respectivas para saber que está ocurriendo.

1. ***Presente el cuadro ANVA y pruebe la hipótesis respectiva.***

***Se puede afirmar que la media del rendimiento obtenido con la variedad D supera a la media del rendimiento obtenido con la variedad A en 20 kilogramos por parcela. Realice la prueba de hipótesis correspondiente.***

**Respuesta**

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

H0: Los cinco tipos de garbanzo tienen el mismo efecto de rendimiento por parcela.

H1: Con al menos uno de tipos de garbanzo se obtiene un efecto diferente en rendimiento por parcela.

Estadístico de Prueba:

~ 

Regla de Decisión:

La hipótesis nula se rechaza con un nivel de significación *α* si el *Fc*resulta mayor que el valor de tabla .

Fc = 1864.64 / 119.71 = 15.58

F de tabla: F (1-0.1,4,12) = F (0.9, 4, 12) = 2.48

El estadístico de prueba es *Fc* = 15.58. Los valores de la tabla para un nivel de significación del 10% es 2.48. Dado que el estadístico de prueba resulta mayor que el valor de F de tabla, la prueba resultó muy significativa, se rechaza H0 a un nivel de significación del 10%.

Por otra parte, el mismo cuadro que genera el R nos muestra que el F calculado del tratamiento tienen una valor de 15.5767 con Pvalor de 0.000107, altamente muy significativa, por lo que se rechaza la hipótesis planteada.

En conclusión, existe suficiente evidencia estadística para aceptar que con al menos uno de los garbanzos se obtiene un efecto diferente. Es decir, rendimientos diferentes en la parcela.

**Se puede afirmar que la media del rendimiento obtenido con la variedad D supera a la media del rendimiento obtenido con la variedad A en 20 kilogramos por parcela.**

**Respuesta**

ti<-ypi-yp

ti

# A B C D E

# -24.64 -16.64 14.56 8.96 17.76

El t calculado será:

= 1.96536717

El valor del t de tabla para un nivel de significación de 10% es t(0.90,12) = 1.356. Como el valor calculado es mayor al valor de tabla: **rechazamos la hipótesis planteada**.

En conclusión, no existe suficiente evidencia estadística para aceptar que la media del rendimiento obtenido con la variedad D supera a la media del rendimiento obtenido con la variedad A en 20 Kilogramos por parcela.

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

**Respuesta**

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 0.774826   
## C - A == 0 39.20 6.92 5.665 0.000818 \*\*\*  
## D - A == 0 33.60 6.92 4.856 0.002932 \*\*   
## E - A == 0 42.40 6.92 6.127 0.000382 \*\*\*  
## C - B == 0 31.20 6.92 4.509 0.005293 \*\*   
## D - B == 0 25.60 6.92 3.700 0.020834 \*   
## E - B == 0 34.40 6.92 4.971 0.002448 \*\*   
## D - C == 0 -5.60 6.92 -0.809 0.922782   
## E - C == 0 3.20 6.92 0.462 0.989393   
## E - D == 0 8.80 6.92 1.272 0.712196   
## ---  
## 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 10%

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 el mayor rendimiento.

1. ***Realice la prueba de Duncan. Compare con los resultados de la prueba de Tukey***

**Respuesta**

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 el 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

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

**Respuesta**

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 0.61608   
## C - A == 0 39.20 6.92 5.665 < 0.001 \*\*\*  
## D - A == 0 33.60 6.92 4.856 0.00154 \*\*   
## 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)

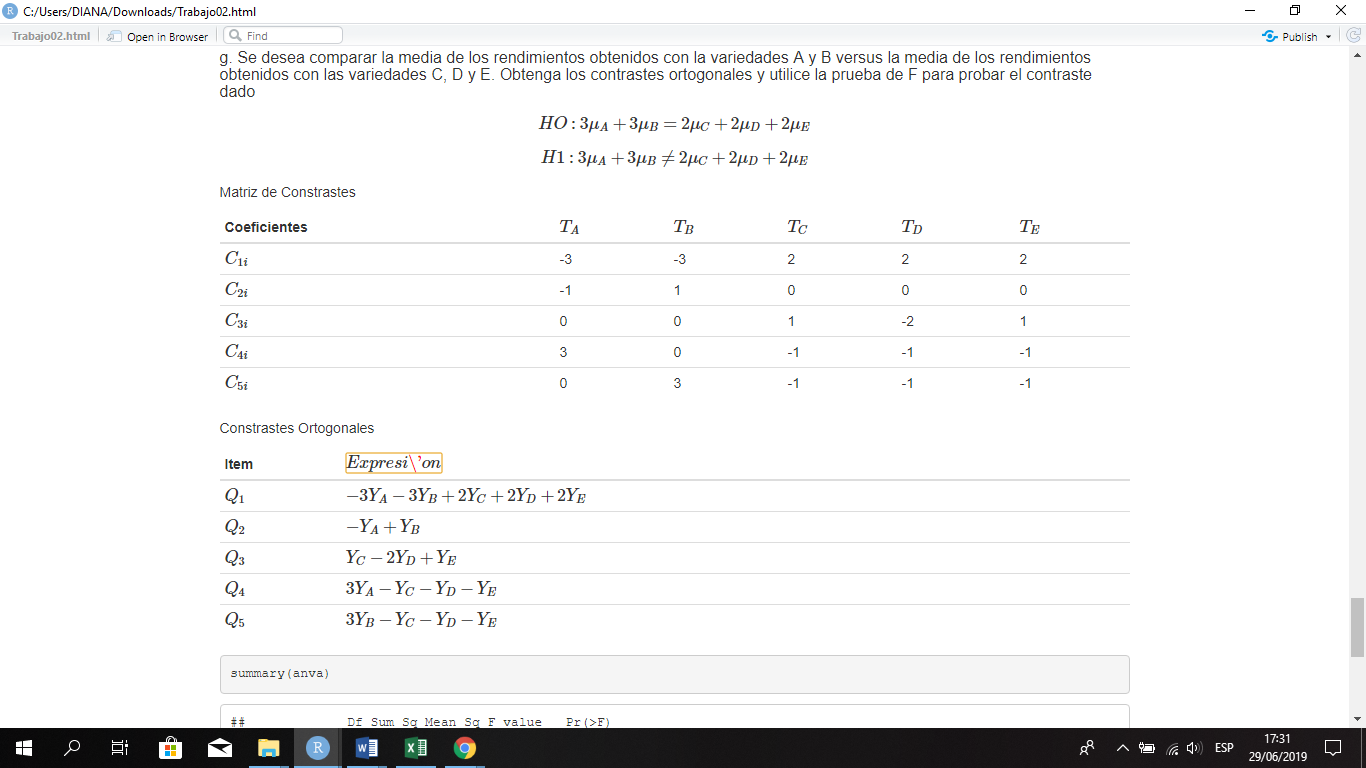
Según la prueba de Duncan se ha encontrado:

* 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%

1. ***Se desea comparar la media de los rendimientos obtenidos con las 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.***

**Observación**: Para prueba de hipótesis concluya en término de pvalue de acuerdo a la escala dado en clase.

**Respuesta**



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.