Evaluación 3

Diana Laura Javier Garcia

4/11/2020

# 1. Una entidad bancaria tiene oficinas en cinco zonas (A, B, C, D y E). Se conoce el valor mensual (en miles de euros) de varias hipótecas elegidas al azar de entre las concebidas en cada una de las regiones en el último año.

1. Compruebe que se verifican los supuestos del análisis de la varianza. Pregunta: Hay diferencias en el valor medio de las hipótecas según la zona. Determine en que zonas hay diferencias.

Como se menciona en el enunciado, este problema consiste en analizar los datos de cinco grupos de hipotecas organizados según la zona, los cuales fueron obtenidos aleatoriamente. Como sabemos, el módelo del análisis de la varianza (ANOVA) permite comparar entre grupos las medias de una variable continua, por lo tanto para resolver este problema se realizó este análisis.

Para resolver este problema se inicia leyendo los datos, los cuales se encuentran en el archivo *datos*:

datos<- read.csv("datos.csv")  
names(datos)

## [1] "ï..A" "B" "C" "D" "E"

Como se puede observar, y como se mencionó anteriormente, la lectura de los datos permite mostrar las cinco muestras identificadas con la letra A hasta la E, donde el factor común es la hipóteca la cual depende de la zona.

Para obtener el tamaño de cada una de las muestras se emplearon las siguientes instrucciones:

nA<- length(datos$ï..A)

nB<- length(datos$B)

nC<- length(datos$c)

nD<- length(datos$D)

nE<- length(datos$E)

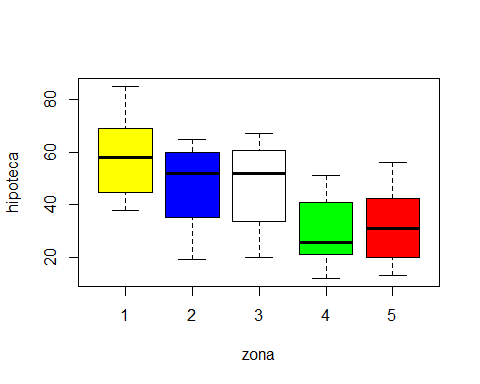
Antes de iniciar con el análisis de ANOVA, conviene tener una idea del comportamiento estadistico de los datos, para ello se estudiaron algunos parámetros estadisticos descriptivos. Así, organizando los datos en forma vertical:

datos.vert<-data.frame(zona=gl(5,12),hipoteca=c(datos$ï..A,datos$B,datos$C,datos$D,datos$E))  
datos.vert

## zona hipoteca  
## 1 1 50  
## 2 1 65  
## 3 1 68  
## 4 1 46  
## 5 1 38  
## 6 1 43  
## 7 1 70  
## 8 1 85  
## 9 1 72  
## 10 1 40  
## 11 1 57  
## 12 1 59  
## 13 2 49  
## 14 2 47  
## 15 2 30  
## 16 2 60  
## 17 2 65  
## 18 2 60  
## 19 2 19  
## 20 2 28  
## 21 2 56  
## 22 2 62  
## 23 2 55  
## 24 2 40  
## 25 3 20  
## 26 3 59  
## 27 3 67  
## 28 3 61  
## 29 3 28  
## 30 3 47  
## 31 3 29  
## 32 3 41  
## 33 3 60  
## 34 3 57  
## 35 3 61  
## 36 3 38  
## 37 4 20  
## 38 4 23  
## 39 4 38  
## 40 4 31  
## 41 4 44  
## 42 4 46  
## 43 4 27  
## 44 4 18  
## 45 4 22  
## 46 4 12  
## 47 4 24  
## 48 4 51  
## 49 5 18  
## 50 5 30  
## 51 5 22  
## 52 5 47  
## 53 5 31  
## 54 5 56  
## 55 5 15  
## 56 5 45  
## 57 5 31  
## 58 5 36  
## 59 5 40  
## 60 5 13

Así, se generó el diagrama de caja de los grupos, donde 1 se refiere a la zona A, 2 a la zona B, 3 a la C, 4 a la D y finalmente 5 a la zona E:

boxplot(hipoteca ~ zona, xlab = "zona", ylab = "hipoteca", data = datos.vert, col=c("yellow", "blue", "white", "green","red"))



Y se obtuvo el resumen estadístico de cada uno:

summary(datos)

## ï..A B C D   
## Min. :38.00 Min. :19.00 Min. :20.00 Min. :12.00   
## 1st Qu.:45.25 1st Qu.:37.50 1st Qu.:35.75 1st Qu.:21.50   
## Median :58.00 Median :52.00 Median :52.00 Median :25.50   
## Mean :57.75 Mean :47.58 Mean :47.33 Mean :29.67   
## 3rd Qu.:68.50 3rd Qu.:60.00 3rd Qu.:60.25 3rd Qu.:39.50   
## Max. :85.00 Max. :65.00 Max. :67.00 Max. :51.00   
## E   
## Min. :13.00   
## 1st Qu.:21.00   
## Median :31.00   
## Mean :32.00   
## 3rd Qu.:41.25   
## Max. :56.00

A simple vista, el diagrama de caja permite mostrar que las zonas 4 y 5, es decir D y E, tiene una media que aunque difieren entre si por tres unidades son muy parecidas. Lo mismo ocurre con las zonas B y C, que tienen una media de 47 aproximadamente, y la que difiere entre ellos es la zona A, la cual tiene la mayor media comparada con las demás zonas, y además de que tiene valores máximos y minimos muy destacables. En la siguiente subsección se realizó el análisis de ANOVA que permitió confirmar estas observaciones.

En este punto, conviene recordar las hipótesis del análisis ANOVA. En primer lugar, se parte de que la hipótesis nula es que la media de la variable estudiada es la misma en los diferentes grupos, en contra posición a la hipótesis alternativa de que al menos dos medias son diferentes. La instrucción para realizar el ANOVA es: *aov*. En este trabajo se guarda la instrucción como *modelo*. Y posteriormente , los resultados se ven con el comando *summary*, como se muestra a continuación:

modelo <- aov(hipoteca ~ zona, data = datos.vert)  
modelo

## Call:  
## aov(formula = hipoteca ~ zona, data = datos.vert)  
##   
## Terms:  
## zona Residuals  
## Sum of Squares 6672.433 11310.500  
## Deg. of Freedom 4 55  
##   
## Residual standard error: 14.34034  
## Estimated effects may be unbalanced

summary(modelo)

## Df Sum Sq Mean Sq F value Pr(>F)   
## zona 4 6672 1668.1 8.112 3.25e-05 \*\*\*  
## Residuals 55 11311 205.6   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Analizando los resultados de la tabla ANOVa tenemos las siguientes observaciones:

1. Los grados de libertad del factor son 4
2. Los grados de libertad residuales con 55
3. La suma de cuadrado de los grupos son 6672.
4. La suma de cuadrados del error son 11311.
5. Las medias correspondientes de las sumas de cuadrados de los grupos son 1668.1.
6. El valor estadístico F es 8.112.
7. El valor de p es 3.25e-05

De la observación 7, dado que p es menor que 0.05 se concluye que existe un efecto significativo de la variable zona, lo que conduce a la pregunta: ¿Cuáles son los grupos que producen estas diferencias?

Para responder a la pregunta, se empleó el método de Tukey, con la función: *TukeyHSD* y posteriormente se graficaron los intervalos de confianza.

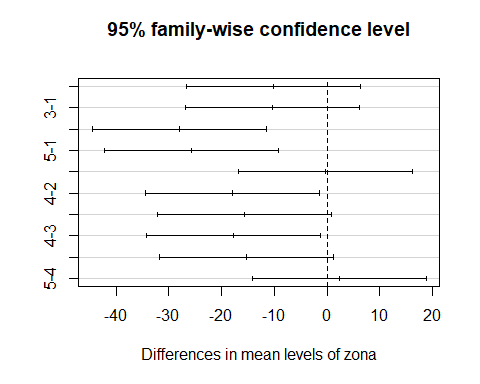
intervals = TukeyHSD(modelo)  
intervals

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = hipoteca ~ zona, data = datos.vert)  
##   
## $zona  
## diff lwr upr p adj  
## 2-1 -10.166667 -26.67805 6.3447174 0.4206561  
## 3-1 -10.416667 -26.92805 6.0947174 0.3958221  
## 4-1 -28.083333 -44.59472 -11.5719493 0.0001206  
## 5-1 -25.750000 -42.26138 -9.2386160 0.0004693  
## 3-2 -0.250000 -16.76138 16.2613840 0.9999992  
## 4-2 -17.916667 -34.42805 -1.4052826 0.0270888  
## 5-2 -15.583333 -32.09472 0.9280507 0.0731218  
## 4-3 -17.666667 -34.17805 -1.1552826 0.0303039  
## 5-3 -15.333333 -31.84472 1.1780507 0.0807170  
## 5-4 2.333333 -14.17805 18.8447174 0.9945248

print(intervals)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = hipoteca ~ zona, data = datos.vert)  
##   
## $zona  
## diff lwr upr p adj  
## 2-1 -10.166667 -26.67805 6.3447174 0.4206561  
## 3-1 -10.416667 -26.92805 6.0947174 0.3958221  
## 4-1 -28.083333 -44.59472 -11.5719493 0.0001206  
## 5-1 -25.750000 -42.26138 -9.2386160 0.0004693  
## 3-2 -0.250000 -16.76138 16.2613840 0.9999992  
## 4-2 -17.916667 -34.42805 -1.4052826 0.0270888  
## 5-2 -15.583333 -32.09472 0.9280507 0.0731218  
## 4-3 -17.666667 -34.17805 -1.1552826 0.0303039  
## 5-3 -15.333333 -31.84472 1.1780507 0.0807170  
## 5-4 2.333333 -14.17805 18.8447174 0.9945248

plot(intervals)



El método de Turkey indica que las diferencias entre medias en las que el intervalo de confianza engloba los límites inferior y superior que no contienen el valor cero, son estadísticamente significativas, además de que el valor P es menor a 0.05. Para este problema, vemos que los grupos con diferencias significativas son los siguientes:

4-3, es decir D y C

4-2, es decir D y B

5-1, es decir E y A

4-1, es decir D y A

Mientras que, los grupos que tocan en cero, es decir, que tienen diferencia entre las medias no estadisticamente significativa y con valor de p mayor a 0.05 son:

2-1, es decir B y A

3-1, es decir C y A

3-2, es decir C y B

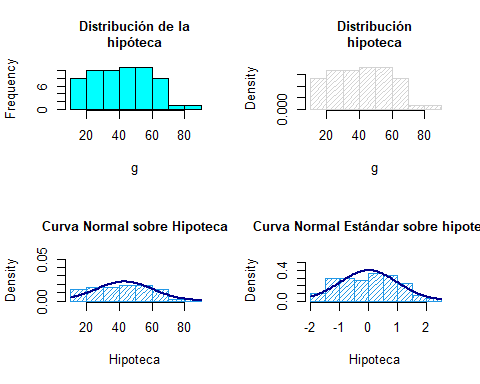
5-2, es decir E y B

5-3, es decir E y C

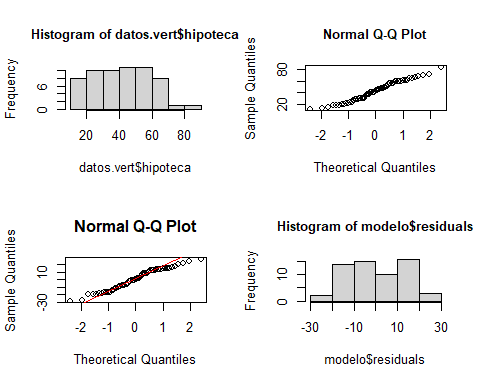
Vemos que esto coincide con la observación numéro siete del análisis descriptivo, donde el diagrama de cajas permitió notar difencias importantes entre la media del grupo A con el resto.

Para verificar la normalidad de manera grafica, se presenta a continuación el histograma con el ajuste a una curva normal. Como se puede observar, las hipotecas se ajustan normalmente.

par (mfrow=c(2,2))  
g = datos.vert$hipoteca  
m<-mean(g)  
std<-sqrt(var(g))  
  
hist(g, breaks = "Sturges", col = "cyan", main= "Distribución de la  
hipóteca", cex.main=1.0)  
hist(g, density=30, breaks = "Sturges", prob=TRUE, main= "Distribución  
hipoteca", cex.main=1.0)  
hist(g, density=30, prob=TRUE, col = 12,  
xlab="Hipoteca", ylim=c(0, 0.050),  
main="Curva Normal sobre Hipoteca", cex.main=1.0)  
curve(dnorm(x, mean=m, sd=std),  
col="darkblue", lwd=2, add=TRUE, yaxt="n")  
zg<-(g-m)/std  
hist(zg, density=30, prob=TRUE, col = 12,  
xlab="Hipoteca", ylim=c(0, 0.55),  
main="Curva Normal Estándar sobre hipoteca", cex.main=1.0)  
curve(dnorm(x, mean=0, sd=1),  
col="darkblue", lwd=2, add=TRUE, yaxt="n")



par(mfrow=c(2,2))  
hist(datos.vert$hipoteca, cex.main=1.0)  
qqnorm(datos.vert$hipoteca, cex.main=1.0)  
qqnorm(modelo$residuals)  
qqline(modelo$residuals, col = "red")  
hist(modelo$residuals, cex.main=1.0)



De la gráfica *QQ norm* de residuos, se observa que estos siguen una linea recta (color rojo), lo cual indica normalidad. Además de las gráficas, se presenta a continuación las pruebas de Shapiro Wilk y Komogorov, ambas pruebas plantean la hipótesis nula que una muestra proviene de una distribución normal, y por ende plantea una hipótesis alternativa que sostiene que la distribución no es normal. Como se puede observar, ambas pruebas tienen un p-valor mayor a 0.05, por lo tanto se concluye que los datos se distribuyen normalmente.

## Prueba de Shapiro Wilk

sw<-shapiro.test(datos.vert$hipoteca)  
print(sw)

##   
## Shapiro-Wilk normality test  
##   
## data: datos.vert$hipoteca  
## W = 0.9717, p-value = 0.1767

## Prueba de Kolmogorov

ks<-ks.test(datos.vert$hipoteca, "pnorm", mean=mean(datos.vert$hipoteca),sd=sd(datos.vert$hipoteca))

## Warning in ks.test(datos.vert$hipoteca, "pnorm", mean =  
## mean(datos.vert$hipoteca), : ties should not be present for the Kolmogorov-  
## Smirnov test

print(ks)

##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: datos.vert$hipoteca  
## D = 0.10166, p-value = 0.5647  
## alternative hypothesis: two-sided

## Prueba de igualdad de varianzas

Para verificar la igualdad de varianzas se aplicó la Prueba de Bartlett, la cual permite contrastar la igualdad de varianza en 2 o más poblaciones sin necesidad de que el tamaño de los grupos sea el mismo. Como se puede observar, el p-valor es mayor a 0.05, por lo tanto se cumple la hipotesis de que las varianzas no presentan diferencias significativas.

bartlett.test(hipoteca ~ zona, data=datos.vert)

##   
## Bartlett test of homogeneity of variances  
##   
## data: hipoteca by zona  
## Bartlett's K-squared = 0.77661, df = 4, p-value = 0.9416

Respecto al análisis estadistico se tienen las siguientes observaciones:

1. Se comprobó la hipotesis de normalidad por medio de la prueba de Shapiro Wilk y Komogorov, donde se encontraron p-valores mayores el 0.05.
2. Se comprobo la igualdad de varianzas por medio de la prueba de Bartlett, en la cual se obtuvo un p-valor mayor a 0.05.

Respecto a la hipotecas en función de la zona se encontró:

1. La zona A es la que tiene valores mayores de hipoteca.
2. La zona B y C tienen valores muy similares, menores a la zona A.
3. Las zonas que tienen menores valores de hipoteca son D y E, y son muy similares entre sí.

Finalmente se responde la pregunta:

Con base a los resultados obtenidos por el método de Turkey se encontró que los grupos con diferencias significativas son los siguientes:

4-3, es decir D y C

4-2, es decir D y B

5-1, es decir E y A

4-1, es decir D y A

Mientras que, los grupos que tienen diferencia entre las medias no estadisticamente significativa son:

2-1, es decir B y A

3-1, es decir C y A

3-2, es decir C y B

5-2, es decir E y B

5-3, es decir E y C

# 3. Una empresa dedicada a la fabricación de pañales está pensando en lazar al mercado un nuevo producto. Para ello está probando con tres materiales y tres procesos de fabricación distintos. Los datos de poder absorbente (en gramos líquido por gramo de pañal) de diez pañales, están registrados el objeto pr7.3

1. ¿El efecto del material en el poder absorbente depende del proceso o del material?
2. ¿Podría un material? ¿Y el proceso?
3. ¿La interacción del material y el proceso es significativa? Interprete el gráfico del perfil.
4. ¿Se verifican las condiciones para realizar un análisis de la varianza?

Como se indica en el enunciado, en este inciso se estudiará la absorción de pañales, los cuales han sido elaborados con tres tipos de materiales (A, B y C) y por medio de tres procesos (P1,P2 y P3), entonces el material y el procesos son los factores. Los resultados obtenidos permitiran conocer si la influencia de uno de los factores varía dependiendo de los niveles del otro factor, es decir, si la influencia del factor material sobre la absorción es distinta según el tipo de proceso empleado. Para ello se inicia leyendo los datos:

datos2<-read.csv("problema3.csv")  
datos2

## n absorcion material proceso  
## 1 1 2.09 A P1  
## 2 2 2.27 A P1  
## 3 3 2.21 A P1  
## 4 4 1.99 A P1  
## 5 5 2.51 A P1  
## 6 6 2.29 A P1  
## 7 7 2.31 A P1  
## 8 8 1.91 A P1  
## 9 9 2.15 A P1  
## 10 10 2.43 A P1  
## 11 11 2.32 A P2  
## 12 12 1.94 A P2  
## 13 13 2.10 A P2  
## 14 14 2.12 A P2  
## 15 15 2.14 A P2  
## 16 16 2.12 A P2  
## 17 17 2.30 A P2  
## 18 18 2.18 A P2  
## 19 19 2.20 A P2  
## 20 20 2.12 A P2  
## 21 21 2.04 A P3  
## 22 22 2.30 A P3  
## 23 23 1.90 A P3  
## 24 24 1.70 A P3  
## 25 25 1.92 A P3  
## 26 26 1.44 A P3  
## 27 27 1.54 A P3  
## 28 28 1.60 A P3  
## 29 29 1.86 A P3  
## 30 30 1.78 A P3  
## 31 31 1.96 B P1  
## 32 32 2.12 B P1  
## 33 33 2.46 B P1  
## 34 34 2.32 B P1  
## 35 35 2.10 B P1  
## 36 36 2.34 B P1  
## 37 37 2.28 B P1  
## 38 38 2.08 B P1  
## 39 39 2.16 B P1  
## 40 40 2.36 B P1  
## 41 41 1.69 B P2  
## 42 42 1.85 B P2  
## 43 43 1.89 B P2  
## 44 44 1.73 B P2  
## 45 45 1.69 B P2  
## 46 46 2.19 B P2  
## 47 47 1.79 B P2  
## 48 48 1.97 B P2  
## 49 49 1.99 B P2  
## 50 50 1.79 B P2  
## 51 51 2.02 B P3  
## 52 52 2.30 B P3  
## 53 53 2.30 B P3  
## 54 54 2.18 B P3  
## 55 55 2.06 B P3  
## 56 56 2.04 B P3  
## 57 57 1.92 B P3  
## 58 58 2.28 B P3  
## 59 59 2.18 B P3  
## 60 60 2.14 B P3  
## 61 61 1.80 C P1  
## 62 62 1.66 C P1  
## 63 63 1.68 C P1  
## 64 64 1.70 C P1  
## 65 65 2.00 C P1  
## 66 66 1.78 C P1  
## 67 67 1.66 C P1  
## 68 68 1.86 C P1  
## 69 69 1.98 C P1  
## 70 70 2.26 C P1  
## 71 71 2.46 C P2  
## 72 72 2.26 C P2  
## 73 73 2.22 C P2  
## 74 74 2.24 C P2  
## 75 75 2.16 C P2  
## 76 76 1.94 C P2  
## 77 77 2.38 C P2  
## 78 78 1.86 C P2  
## 79 79 2.10 C P2  
## 80 80 2.04 C P2  
## 81 81 1.62 C P3  
## 82 82 1.90 C P3  
## 83 83 1.96 C P3  
## 84 84 1.98 C P3  
## 85 85 1.64 C P3  
## 86 86 2.00 C P3  
## 87 87 1.72 C P3  
## 88 88 1.62 C P3  
## 89 89 1.84 C P3  
## 90 90 1.78 C P3

Como se puede ver en la lectura del archivo, la primera columna representa la medida de absorción, la segunda columna el material y la tercera el proceso. Para organizar los datos se escribió la siguiente instrucción:

datos2$material<-factor(datos2$material,labels = c('material1','material2','material3'))  
datos2$proceso<-factor(datos2$proceso,labels = c('proceso1','proceso2','proceso3'))  
datos2

## n absorcion material proceso  
## 1 1 2.09 material1 proceso1  
## 2 2 2.27 material1 proceso1  
## 3 3 2.21 material1 proceso1  
## 4 4 1.99 material1 proceso1  
## 5 5 2.51 material1 proceso1  
## 6 6 2.29 material1 proceso1  
## 7 7 2.31 material1 proceso1  
## 8 8 1.91 material1 proceso1  
## 9 9 2.15 material1 proceso1  
## 10 10 2.43 material1 proceso1  
## 11 11 2.32 material1 proceso2  
## 12 12 1.94 material1 proceso2  
## 13 13 2.10 material1 proceso2  
## 14 14 2.12 material1 proceso2  
## 15 15 2.14 material1 proceso2  
## 16 16 2.12 material1 proceso2  
## 17 17 2.30 material1 proceso2  
## 18 18 2.18 material1 proceso2  
## 19 19 2.20 material1 proceso2  
## 20 20 2.12 material1 proceso2  
## 21 21 2.04 material1 proceso3  
## 22 22 2.30 material1 proceso3  
## 23 23 1.90 material1 proceso3  
## 24 24 1.70 material1 proceso3  
## 25 25 1.92 material1 proceso3  
## 26 26 1.44 material1 proceso3  
## 27 27 1.54 material1 proceso3  
## 28 28 1.60 material1 proceso3  
## 29 29 1.86 material1 proceso3  
## 30 30 1.78 material1 proceso3  
## 31 31 1.96 material2 proceso1  
## 32 32 2.12 material2 proceso1  
## 33 33 2.46 material2 proceso1  
## 34 34 2.32 material2 proceso1  
## 35 35 2.10 material2 proceso1  
## 36 36 2.34 material2 proceso1  
## 37 37 2.28 material2 proceso1  
## 38 38 2.08 material2 proceso1  
## 39 39 2.16 material2 proceso1  
## 40 40 2.36 material2 proceso1  
## 41 41 1.69 material2 proceso2  
## 42 42 1.85 material2 proceso2  
## 43 43 1.89 material2 proceso2  
## 44 44 1.73 material2 proceso2  
## 45 45 1.69 material2 proceso2  
## 46 46 2.19 material2 proceso2  
## 47 47 1.79 material2 proceso2  
## 48 48 1.97 material2 proceso2  
## 49 49 1.99 material2 proceso2  
## 50 50 1.79 material2 proceso2  
## 51 51 2.02 material2 proceso3  
## 52 52 2.30 material2 proceso3  
## 53 53 2.30 material2 proceso3  
## 54 54 2.18 material2 proceso3  
## 55 55 2.06 material2 proceso3  
## 56 56 2.04 material2 proceso3  
## 57 57 1.92 material2 proceso3  
## 58 58 2.28 material2 proceso3  
## 59 59 2.18 material2 proceso3  
## 60 60 2.14 material2 proceso3  
## 61 61 1.80 material3 proceso1  
## 62 62 1.66 material3 proceso1  
## 63 63 1.68 material3 proceso1  
## 64 64 1.70 material3 proceso1  
## 65 65 2.00 material3 proceso1  
## 66 66 1.78 material3 proceso1  
## 67 67 1.66 material3 proceso1  
## 68 68 1.86 material3 proceso1  
## 69 69 1.98 material3 proceso1  
## 70 70 2.26 material3 proceso1  
## 71 71 2.46 material3 proceso2  
## 72 72 2.26 material3 proceso2  
## 73 73 2.22 material3 proceso2  
## 74 74 2.24 material3 proceso2  
## 75 75 2.16 material3 proceso2  
## 76 76 1.94 material3 proceso2  
## 77 77 2.38 material3 proceso2  
## 78 78 1.86 material3 proceso2  
## 79 79 2.10 material3 proceso2  
## 80 80 2.04 material3 proceso2  
## 81 81 1.62 material3 proceso3  
## 82 82 1.90 material3 proceso3  
## 83 83 1.96 material3 proceso3  
## 84 84 1.98 material3 proceso3  
## 85 85 1.64 material3 proceso3  
## 86 86 2.00 material3 proceso3  
## 87 87 1.72 material3 proceso3  
## 88 88 1.62 material3 proceso3  
## 89 89 1.84 material3 proceso3  
## 90 90 1.78 material3 proceso3

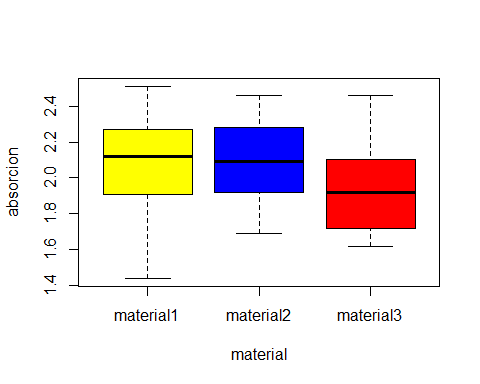
names(datos2)

## [1] "n" "absorcion" "material" "proceso"

attach(datos2)

Para tener una idea del comportamiento de los datos se presenta a continuación el diagrama de caja de la absorción en función del material:

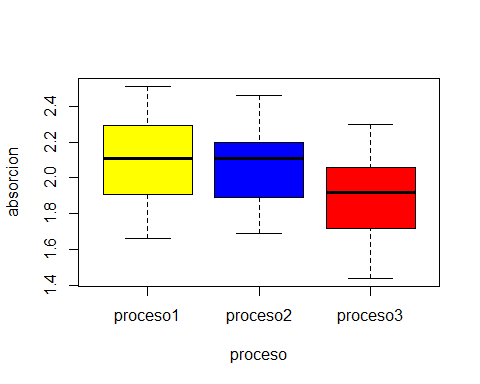
boxplot(absorcion ~ material,data = datos2, col=c("yellow", "blue", "red"))



De este diagrama presentado se observa que el material 1 y 2 proporcionan el mismo grado de absorción, a diferencia del material 3 que tiene una absorción menor. También se puede observar que el material 1 tiene una dispersión mayor, por lo tanto, si bien es cierto que el material tiene tiene la mayor absorción, tambien es cierto que presenta una gran dispersión, y esta es una observación que debe tomarse en cuenta para las conclusiones.

Por otra parte, el diagrama de caja de la absorción en función del proceso, generado a continuación, muestra que tanto el proceso 1 como el 2 dan el un valor similar de absorción, mientras que el proceso 3 brinda la menor absorción. Sin embargo, al igual que en el caso caso anterior, el proceso 1 tiene una mayor dispersión de datos.

boxplot(absorcion ~ proceso,data = datos2, col=c("yellow", "blue", "red"))



Una vez que se ha estudiaron de manera descriptiva los datos se realizó el análisis de varianza de dos factores,

El manejo de datos inicia calculando el número de replicas por cada combinación:

replicas <- tapply(absorcion, list(material,proceso), length)  
replicas

## proceso1 proceso2 proceso3  
## material1 10 10 10  
## material2 10 10 10  
## material3 10 10 10

Así, el número total de datos es:

N<-sum(replicas)  
N

## [1] 90

Además, la suma de cada una de las combinaciones se calcula como sigue:

sumacombs<-tapply(absorcion, list(material,proceso), sum)  
sumacombs

## proceso1 proceso2 proceso3  
## material1 22.16 21.54 18.08  
## material2 22.18 18.58 21.42  
## material3 18.38 21.66 18.06

Y la suma de todas las observaciones es:

sumobs<-sum(sumacombs)  
sumobs

## [1] 182.06

De esta manera, se calcula el promedio de cada observación:

promcombs<-tapply(absorcion, list(material,proceso), mean)  
promcombs

## proceso1 proceso2 proceso3  
## material1 2.216 2.154 1.808  
## material2 2.218 1.858 2.142  
## material3 1.838 2.166 1.806

Y el promedio de todas las abservaciones es:

mean(promcombs)

## [1] 2.022889

Presentando la suma de los valores para cada modalidad por separado y la suma de las observaciones según el proceso:

apply(sumacombs,1,sum)

## material1 material2 material3   
## 61.78 62.18 58.10

apply(sumacombs, 2, sum)

## proceso1 proceso2 proceso3   
## 62.72 61.78 57.56

Para obtener el ANOVA:

attach(datos2)

## The following objects are masked from datos2 (pos = 3):  
##   
## absorcion, material, n, proceso

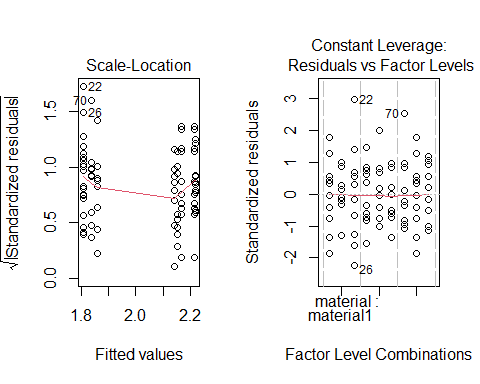
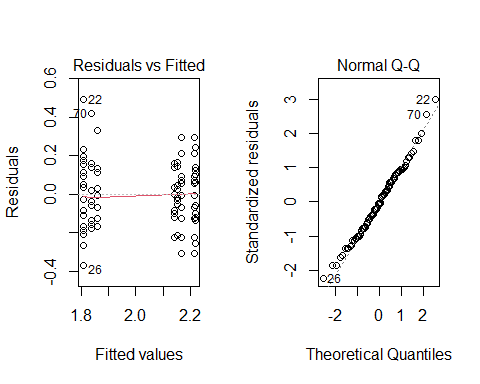
modelo2<-aov(absorcion ~ material+proceso+material\*proceso)  
summary(modelo2)

## Df Sum Sq Mean Sq F value Pr(>F)   
## material 2 0.3372 0.1686 5.561 0.005458 \*\*   
## proceso 2 0.5035 0.2518 8.303 0.000524 \*\*\*  
## material:proceso 4 1.9774 0.4943 16.304 7.79e-10 \*\*\*  
## Residuals 81 2.4560 0.0303   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

De los resultados obtenidos de la ANOVA se observa que dado que el valor de p del material y del proceso es menor a 0.05 se puede rechazar la hipótesis nula y se concluye que tanto como el material y el proceso influyen en los niveles de absorción.

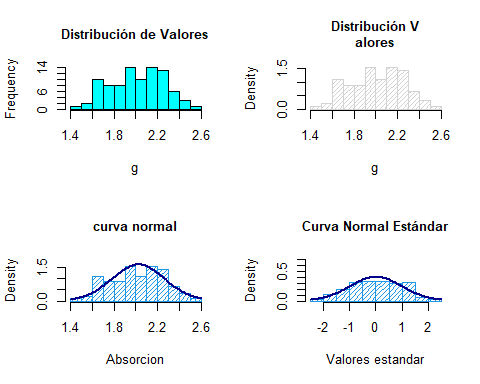
Para poder dar por válidos los resultados del ANOVA es necesario verificar que se cumplen las condiciones de un ANOVA.

par(mfrow=c(1,2))  
plot(modelo2)



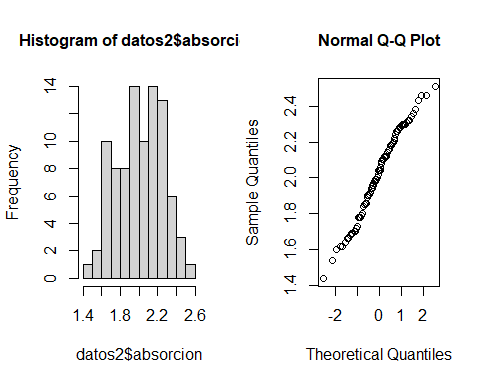
De esta figura se observa que los residuos siguen una linea recta, lo que indica normalidad. Graficando el histograma de los datos:

par (mfrow=c(2,2))  
g = datos2$absorcion  
m<-mean(g)  
std<-sqrt(var(g))  
hist(g, breaks = "Sturges", col = "cyan", main= "Distribución de Valores"  
, cex.main=1.0)  
hist(g, density=30, breaks = "Sturges", prob=TRUE, main= "Distribución V  
alores", cex.main=1.0)  
hist(g, density=30, prob=TRUE, col = 12,  
xlab="Absorcion", ylim=c(0, 1.90),  
main="curva normal", cex.main=1.0)  
curve(dnorm(x, mean=m, sd=std),  
col="darkblue", lwd=2, add=TRUE, yaxt="n")  
zg<-(g-m)/std  
hist(zg, density=30, prob=TRUE, col = 12,  
xlab="Valores estandar", ylim=c(0, 0.70),  
main="Curva Normal Estándar", cex.main=1.0)  
curve(dnorm(x, mean=0, sd=1),  
col="darkblue", lwd=2, add=TRUE, yaxt="n")



De la figura, se observa que el histograma de los datos de absorción, con clases obtenidas por el método de Sturges, siguen una distribución normal.

par(mfrow=c(1,2))  
hist(datos2$absorcion, cex.main=1.0)  
qqnorm(datos2$absorcion, cex.main=1.0)



De la figura anterior, se observa de la grafica cuantil-cualtil (q-q) sigue una linea recta y por lo tanto se cumple la condición de normalidad.

Otro modo de comprobar la normalidad de una variable es por medio del test de *Shapiro-Wilks*. Esta prueba plantea la hipótesis nula que una muestra proviene de una distribución normal, y por ende plantea una hipótesis alternativa que sostiene que la distribución no es normal.

## Prueba de Shapiro Wilk

sw2<-shapiro.test(datos2$absorcion)  
print(sw2)

##   
## Shapiro-Wilk normality test  
##   
## data: datos2$absorcion  
## W = 0.98084, p-value = 0.2067

Para verificar los resultados de la prueba de Shapiro-Wilk:

aov\_residuals2<-residuals(object = modelo2)  
shapiro.test(aov\_residuals2)

##   
## Shapiro-Wilk normality test  
##   
## data: aov\_residuals2  
## W = 0.98989, p-value = 0.7226

Como se puede observar, para un nivel de significancia del 0.05, el p-valor del 0.77 es mucho mayor que 0.05, por lo que no se rechaza la hipotesis nula, es decir, se puede concluir que la variable sigue una distribución normal.

Por otra parte, la prueba de Kolmogorov permite estudiar si una muestra procede de una población con una determinada distribución (media y desviación típica), no está limitado únicamente a la distribución normal. A continuación Se ejecuta con la función *ks.test()*, que como se puede observar, el p-valor es 0.78, que es mayor a 0.05, concluyendo así que se satisface la condición de normalidad.

ks2 <- ks.test(datos2$absorcion,"pnorm",mean=mean(datos2$absorcion),sd=sd(datos2$absorcion))

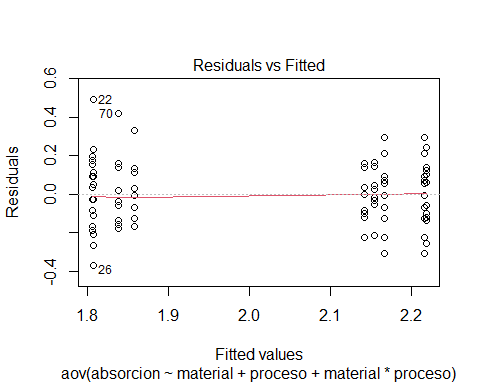
## Warning in ks.test(datos2$absorcion, "pnorm", mean = mean(datos2$absorcion), :  
## ties should not be present for the Kolmogorov-Smirnov test

print(ks2)

##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: datos2$absorcion  
## D = 0.068734, p-value = 0.7888  
## alternative hypothesis: two-sided

Finalmente, la prueba de igualdad de varianzas se analizó en primer lugar forma gráfica:

plot(modelo2,1)



Otro método que permite evalular la igualdad de varianzas es la prueba de Bartlett, esta ermite contrastar la igualdad de varianza en 2 o más poblaciones sin necesidad de que el tamaño de los grupos sea el mismo. En este caso, al evaluar esta prueba se encontró un p-valor, por un lado entre la absorción y el material, y por otro lado entre la absorción y proceso mayor a 0.05, conluyendo así que no existe una diferencia significativa entre las varianzas.

bartlett.test(absorcion ~ material)

##   
## Bartlett test of homogeneity of variances  
##   
## data: absorcion by material  
## Bartlett's K-squared = 1.1441, df = 2, p-value = 0.5644

bartlett.test(absorcion ~ proceso)

##   
## Bartlett test of homogeneity of variances  
##   
## data: absorcion by proceso  
## Bartlett's K-squared = 1.1109, df = 2, p-value = 0.5738

De las pruebas anteriores se procedió a calcular el ANOVa considerando la suma de los efectos:

modelo21<-aov(absorcion ~ material+proceso)  
summary(modelo21)

## Df Sum Sq Mean Sq F value Pr(>F)   
## material 2 0.337 0.16860 3.233 0.0444 \*  
## proceso 2 0.504 0.25176 4.827 0.0103 \*  
## Residuals 85 4.433 0.05216   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Luego, el ANOVA:

modelo22<-aov(absorcion~material+proceso+proceso\*material)  
summary(modelo22)

## Df Sum Sq Mean Sq F value Pr(>F)   
## material 2 0.3372 0.1686 5.561 0.005458 \*\*   
## proceso 2 0.5035 0.2518 8.303 0.000524 \*\*\*  
## material:proceso 4 1.9774 0.4943 16.304 7.79e-10 \*\*\*  
## Residuals 81 2.4560 0.0303   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

De la ANOVA presentada, al tener valores de p-value menores a 0.05, se observa que tanto el material como el proceso influyen en la absorción.

Para determinar los factores que influyen en la respuesta se obtiene la diferecia minima significativa (LSD)

Observación: se instalo la paquetería *agricolae*:

install.packages('agricolae', repos="http://cran.rstudio.com/")

## Installing package into 'C:/Users/diana/OneDrive/Documentos/R/win-library/4.0'  
## (as 'lib' is unspecified)

## package 'agricolae' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\diana\AppData\Local\Temp\RtmpkvvqNf\downloaded\_packages

library(agricolae)

## Warning: package 'agricolae' was built under R version 4.0.3

LSD2<-LSD.test(modelo2,"proceso")  
LSD2

## $statistics  
## MSerror Df Mean CV t.value LSD  
## 0.03032049 81 2.022889 8.607878 1.989686 0.08945552  
##   
## $parameters  
## test p.ajusted name.t ntr alpha  
## Fisher-LSD none proceso 3 0.05  
##   
## $means  
## absorcion std r LCL UCL Min Max Q25 Q50 Q75  
## proceso1 2.090667 0.2511756 30 2.027412 2.153921 1.66 2.51 1.9225 2.11 2.2875  
## proceso2 2.059333 0.2076624 30 1.996079 2.122588 1.69 2.46 1.9025 2.11 2.1975  
## proceso3 1.918667 0.2414287 30 1.855412 1.981921 1.44 2.30 1.7350 1.92 2.0550  
##   
## $comparison  
## NULL  
##   
## $groups  
## absorcion groups  
## proceso1 2.090667 a  
## proceso2 2.059333 a  
## proceso3 1.918667 b  
##   
## attr(,"class")  
## [1] "group"

Como sabemos el análisis LSD es una opción para realizar un análisis exploratorio de las diferencias en los tratamientos, se puede observar, el resultado de la LSD es 0.0894,que el es límite que permite definir diferencias entre los modos de aplicación.

Obteniendo los análisis LSD del modelo 21 y 22:

LSD21<-LSD.test(modelo21, "material")  
LSD21

## $statistics  
## MSerror Df Mean CV t.value LSD  
## 0.0521566 85 2.022889 11.28971 1.988268 0.1172422  
##   
## $parameters  
## test p.ajusted name.t ntr alpha  
## Fisher-LSD none material 3 0.05  
##   
## $means  
## absorcion std r LCL UCL Min Max Q25 Q50 Q75  
## material1 2.059333 0.2604894 30 1.976431 2.142236 1.44 2.51 1.9125 2.12 2.2550  
## material2 2.072667 0.2131040 30 1.989764 2.155569 1.69 2.46 1.9300 2.09 2.2575  
## material3 1.936667 0.2386793 30 1.853764 2.019569 1.62 2.46 1.7350 1.92 2.0850  
##   
## $comparison  
## NULL  
##   
## $groups  
## absorcion groups  
## material2 2.072667 a  
## material1 2.059333 a  
## material3 1.936667 b  
##   
## attr(,"class")  
## [1] "group"

De lo anterior, se observa que la LSD es 0.1172, que es el límite que permite observar diferencias significativas entre los materiales.

Otro modo de obtener los grupos con diferencias significativas es la prueba de Tukey:

TukeyHSD(modelo2)

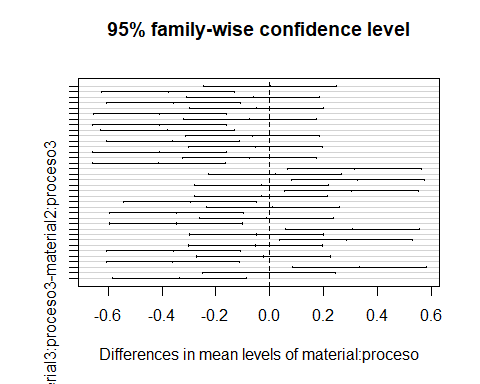
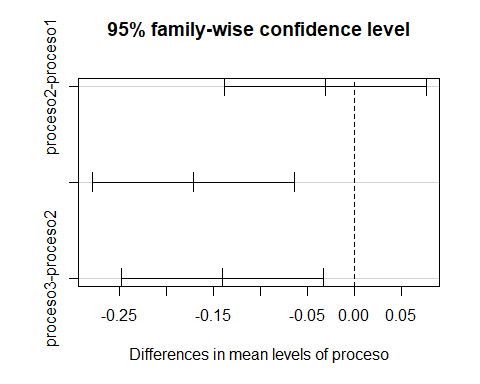
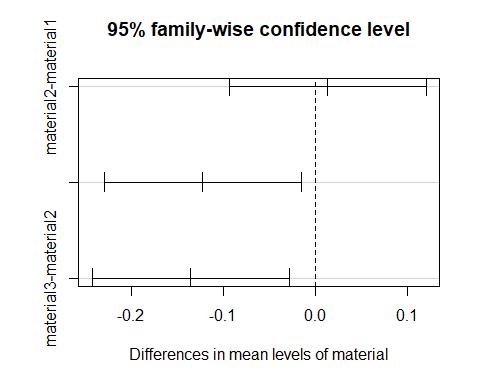
## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = absorcion ~ material + proceso + material \* proceso)  
##   
## $material  
## diff lwr upr p adj  
## material2-material1 0.01333333 -0.09400985 0.12067652 0.9527042  
## material3-material1 -0.12266667 -0.23000985 -0.01532348 0.0210775  
## material3-material2 -0.13600000 -0.24334318 -0.02865682 0.0092428  
##   
## $proceso  
## diff lwr upr p adj  
## proceso2-proceso1 -0.03133333 -0.1386765 0.07600985 0.7659619  
## proceso3-proceso1 -0.17200000 -0.2793432 -0.06465682 0.0007401  
## proceso3-proceso2 -0.14066667 -0.2480098 -0.03332348 0.0068228  
##   
## $`material:proceso`  
## diff lwr upr p adj  
## material2:proceso1-material1:proceso1 0.002 -0.2461952 0.2501952 1.0000000  
## material3:proceso1-material1:proceso1 -0.378 -0.6261952 -0.1298048 0.0001932  
## material1:proceso2-material1:proceso1 -0.062 -0.3101952 0.1861952 0.9966904  
## material2:proceso2-material1:proceso1 -0.358 -0.6061952 -0.1098048 0.0005089  
## material3:proceso2-material1:proceso1 -0.050 -0.2981952 0.1981952 0.9992847  
## material1:proceso3-material1:proceso1 -0.408 -0.6561952 -0.1598048 0.0000427  
## material2:proceso3-material1:proceso1 -0.074 -0.3221952 0.1741952 0.9891104  
## material3:proceso3-material1:proceso1 -0.410 -0.6581952 -0.1618048 0.0000386  
## material3:proceso1-material2:proceso1 -0.380 -0.6281952 -0.1318048 0.0001751  
## material1:proceso2-material2:proceso1 -0.064 -0.3121952 0.1841952 0.9958802  
## material2:proceso2-material2:proceso1 -0.360 -0.6081952 -0.1118048 0.0004626  
## material3:proceso2-material2:proceso1 -0.052 -0.3001952 0.1961952 0.9990489  
## material1:proceso3-material2:proceso1 -0.410 -0.6581952 -0.1618048 0.0000386  
## material2:proceso3-material2:proceso1 -0.076 -0.3241952 0.1721952 0.9870579  
## material3:proceso3-material2:proceso1 -0.412 -0.6601952 -0.1638048 0.0000348  
## material1:proceso2-material3:proceso1 0.316 0.0678048 0.5641952 0.0034642  
## material2:proceso2-material3:proceso1 0.020 -0.2281952 0.2681952 0.9999994  
## material3:proceso2-material3:proceso1 0.328 0.0798048 0.5761952 0.0020383  
## material1:proceso3-material3:proceso1 -0.030 -0.2781952 0.2181952 0.9999849  
## material2:proceso3-material3:proceso1 0.304 0.0558048 0.5521952 0.0057967  
## material3:proceso3-material3:proceso1 -0.032 -0.2801952 0.2161952 0.9999751  
## material2:proceso2-material1:proceso2 -0.296 -0.5441952 -0.0478048 0.0080952  
## material3:proceso2-material1:proceso2 0.012 -0.2361952 0.2601952 1.0000000  
## material1:proceso3-material1:proceso2 -0.346 -0.5941952 -0.0978048 0.0008954  
## material2:proceso3-material1:proceso2 -0.012 -0.2601952 0.2361952 1.0000000  
## material3:proceso3-material1:proceso2 -0.348 -0.5961952 -0.0998048 0.0008157  
## material3:proceso2-material2:proceso2 0.308 0.0598048 0.5561952 0.0048914  
## material1:proceso3-material2:proceso2 -0.050 -0.2981952 0.1981952 0.9992847  
## material2:proceso3-material2:proceso2 0.284 0.0358048 0.5321952 0.0131651  
## material3:proceso3-material2:proceso2 -0.052 -0.3001952 0.1961952 0.9990489  
## material1:proceso3-material3:proceso2 -0.358 -0.6061952 -0.1098048 0.0005089  
## material2:proceso3-material3:proceso2 -0.024 -0.2721952 0.2241952 0.9999973  
## material3:proceso3-material3:proceso2 -0.360 -0.6081952 -0.1118048 0.0004626  
## material2:proceso3-material1:proceso3 0.334 0.0858048 0.5821952 0.0015549  
## material3:proceso3-material1:proceso3 -0.002 -0.2501952 0.2461952 1.0000000  
## material3:proceso3-material2:proceso3 -0.336 -0.5841952 -0.0878048 0.0014196

A partir del test de Tukey se tienen las siguientes observaciones:

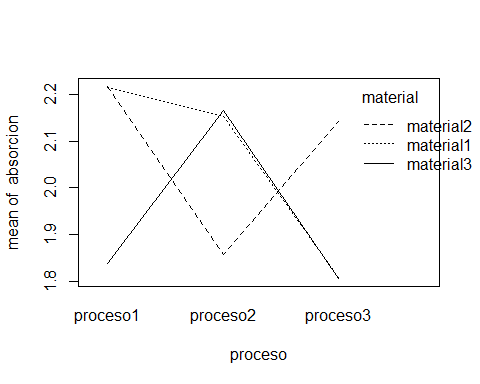
1. No se observan diferencias significativas sobre la absorción entre el material 2 y el material 1.
2. Al contrario que lo anterior, se observan diferencias significativas sobre la absorción entre usar el material 1 y 3, asi como el material 3 y 2.
3. Con respecto a los procesos, no se observan diferencias significativas sobre la absorción entre el proceso 2 y 1.
4. Sin embargo, se observan diferencias significativas entre el proceso 3 y 1, asi como el proceso 3 con el 2.

Para facilitar la comprensión entre las relaciones de las combinaciones entre materiales y procesos, se presenta el diagrama de límites. El cual se interpreta de la siguiente manera: cuando uno de los intervalos incluye el valor de cero, se infiere que la diferencia de un material, o proceso, incluye al cero, es decir, en la práctica no ocasionan diferencias entre usar uno o el otro.

plot(TukeyHSD(modelo2))



interaction.plot(proceso, material, absorcion)



A partir de los resultados presentados anteriormente se puede dar respuesta a cada una de las preguntas del enunciado:

### a) ¿El efecto del material en el poder absorbente depende del proceso o del material?

Del análisis del ANOVA, se observó que tanto el material como el proceso influyen en la capacidad de absorción. Estos resultados se pueden visualizar en el gráfico de interaación.

### b) ¿Podría *recomendar* un material? ¿Y *un* proceso?

Con base al gráfico de interacción se pueden hacer las siguientes recomendaciones respecto al material:

1. Se recomienda usar el material 1 o 2 siempre y cuando estos sean realizados por el proceso 1.
2. También puede usarse el material 3 siempre que sea realizado por el proceso 2.
3. Por ultimo, se recomienda utilizar el material 2 siempre que sea realizado por el proceso 3.

Con respecto al proceso se recomienda lo siguiente:

1. Se recomienda aplicar el proceso 1, siempre y cuando se emplee el material 1 o 2, ya que este tiene los mayores valores de absorción.
2. Después, se recomienda aplicar el proceso 2, siempre y cuando se emplee el material 1 o 3.
3. Finalmente, se recomienda aplica el proceso 3, ya que este es el que menor capacidad de absorción tiene.

### c) ¿La interacción del material y el proceso es significativa? Interprete el gráfico del perfil.

El gráfico de interacción permite hacer las siguientes interpretaciones:

1. Con respecto al material 1:
2. El máximo de poder de absorción corresponde al proceso 1 con un valor de 2.2, y disminuye en aproximadamente el en el proceso 2, siendo el proceso 3 el que menor capacidad de absorción con un valor de 1.8.
3. Con respecto al material 2:
4. La capacidad de absorción es mayor cuando se emplea el proceso 2, el cual tiene un valor de aproximadamente 2.2, se observa una absorción ligeramente menor cuando se emplea el proceso 3, sin embargo, se el proceso 2 el es que menor absorción tiene con un valor de aproximadamente 1.85.
5. Con respecto al material 3 (linea continua):
6. Se observa que el proceso 2 brinda la mayor capacidad de absorción con un valor de aproximadamente 2.2. Los proceso 1 y 2 presentan una absorción muy similar y muy baja con respecto al proceso 2.

### d) ¿Se verifican las condiciones para realizar un análisis de la varianza?

Si, se verificó la normalidad por método grafico y se observó que las mediciones siguen una distribución normal. Además, la prueba de Shapiro Wilk como de Kolmogorov permitieron corroborar lo anterior, al tener cada una un p-valor mayor al 0.05, correspondiente a la hipotesis nula de normalidad.

Así mismo, la prueba de igualdad de varianzas se estudio gráficamente por medio del gráfico de residuos y además, se aplico la prueba de Bartllet, el cual mostró un p-valor mayor al 0.05, concluyendo asi que no existe una diferencia de varianzas significativas.