- # ANOVA DE DOS FACTORES CON INTERACCIÓN.

Lo primero cargar los paquetes correspondientes

library(nortest)

library(car)

library(RcmdrMisc)

library(DescTools)

library(PMCMR)

library(phia)

Es parecido al anova de un factor. La diferencia es que debido a que existen dos factores y por lo tanto se puede considerar que hay o que no interacción.

La función que vamos a utilizar, aov, por defecto hace la suma de cuadrados tipo I. Es un procedimiento jerarquico va incluyendo los factores uno a uno, y los factores van explicando lo que queda del residual de haber metido el factor anterior. El problema es que cuando no es balanceado el diseño, dependiendo como metes los factores cambian las sumas de cuadrados de cada factor.

La forma general

ResAnova2 <- aov(nombrearchivo$nombrevariable ~ nombrearchivo$nombrefactor1 * nombrearchivo$nombrefactor2)

En primer lugar vamos a definir los factores:

datos2$tiposuelo=factor(datos2$tiposuelo, labels=c("Acido", "Alcalino"))

datos2$Abono=factor(datos2$Abono, labels=c("A","B","C"))

CUIDADO: En la plantilla aparece la variable cuantitativa como datos2$abundancia pero en el archivo hay un error en su denominación y es abundancia. Por lo tanto, hay que poner lo del archivo de datos.

```
> anova2<-aov(datos2$abundacia~datos2$tiposuelo*datos2$Abono)
> summary(anova2)
```

```
                              Df Sum Sq Mean Sq F value  Pr(>F)
datos2$tiposuelo               1     18   18.00   2.038 0.17894
datos2$Abono                   2     48   24.00   2.717 0.10634
datos2$tiposuelo:datos2$Abono  2    144   72.00   8.151 0.00581 **
Residuals                     12    106    8.83
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Nos aparece las diferentes sumas de cuadrados y se detecta que la interacción es significativa.

En anova se puede hacer con otra función que es mas general y que permite el cálculo de las diferentes sumas de cuadrados. La función el lm. Esta función si no le dices nada calcula la suma de cuadrados tipo II, es similar a la suma de cuadrados tipo I, pero la diferencia es que la interacción la calcula como las diferencias de las sumas de cuadrados residual del modelo sin interacción y el modelo completo con interacción. Si fuera el diseño balanceado (con el mismo tamaño muestral para la combinación de niveles de los factores) entonces todas las sumas de cuadrados deberían coincidir. Sin embargo, con esta función no ocurre, aunque si coincide lo importante que es la suma de cuadrados residual y de interacción. Pero no que procedimiento utiliza para estimar los coeficientes del modelo.

Para realizar esa suma de cuadrados se debe escribir:

```
> anova3<-lm(datos2$abundacia~datos2$tiposuelo*datos2$Abono)
> Anova(anova3, type= "III")
```

```
Response: datos2$abundacia
                              Sum Sq Df F value  Pr(>F)
(Intercept)                       48  1  5.4340 0.03800 *
datos2$tiposuelo                  54  1  6.1132 0.02936 *
datos2$Abono                      24  2  1.3585 0.29388
datos2$tiposuelo:datos2$Abono    144  2  8.1509 0.00581 **
Residuals                        106 12
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Como se ve la suma de cuadrados de los factores principales no coincide.

Si hiciéramos la función sin especificar el tipo de suma de cuadrados sería:

```
> anova3<-lm(datos2$abundacia~datos2$tiposuelo*datos2$Abono)
> Anova(anova3)
```

```
Anova Table (Type II tests)

Response: datos2$abundacia
                              Sum Sq Df F value  Pr(>F)
datos2$tiposuelo                  18  1  2.0377 0.17894
datos2$Abono                      48  2  2.7170 0.10634
datos2$tiposuelo:datos2$Abono    144  2  8.1509 0.00581 **
Residuals                        106 12
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

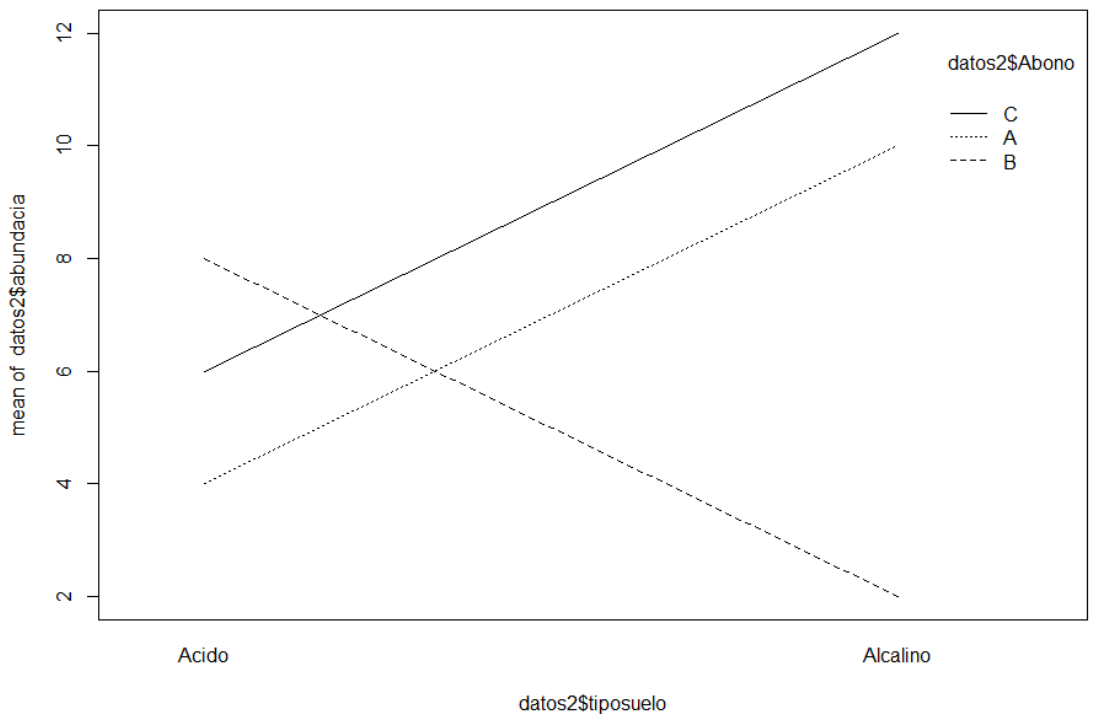
Como se puede ver la suma de cuadrados por defecto es la tipo II y ahora se coincide con la de aov.

Un vez realizada la anova, pasamos a interpretar la interacción de manera gráfica.

interaction.plot(datos2$tiposuelo, datos2$Abono, datos2$abundacia, ylim=c(?,?), col=c("red", "blue", "green"), ylab= "nombre de la variable",xlab= "nombredeX",trace.label= "nombredeY")

Aquí se ha puesto una forma mas general para que se muestre que hay mas argumentos para modificar el gráfico, escala de y, colores para las líneas, nombres de ejes etc. Si no ponemos nada hace el gráfico por defecto.

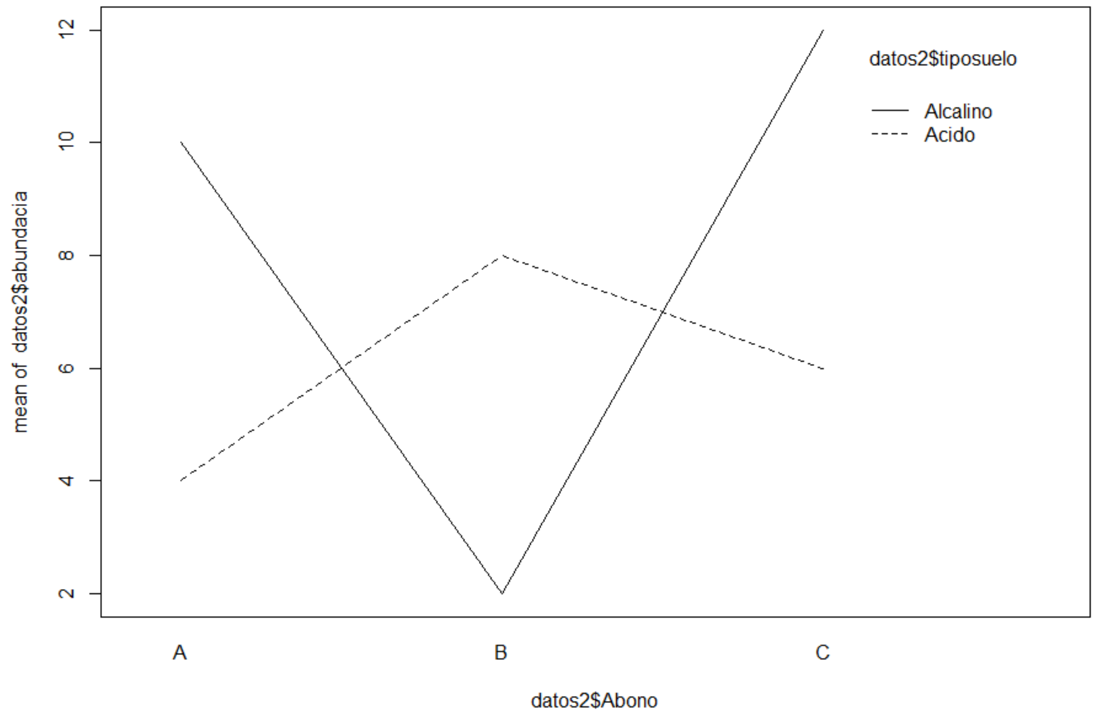
```
> interaction.plot(datos2$tiposuelo, datos2$Abono, datos2$abundacia)
```



Se puede ver como hay cruce de información y por lo tanto las diferencias entre los niveles del abono en el tipo de suelo ácido son diferentes a los del tipo alcalino. Si no hubiera interacción las diferencias entre los niveles de abono tendría que ser muy parecido independientemente del nivel de tipo de suelo (líneas paralelas).

Podemos hacer el otro gráfico, y analizar el comportamiento de los tipos de suelo para cada abono.

```
> interaction.plot(datos2$Abono, datos2$tiposuelo, datos2$abundacia)
```



Para deshacer la interacción se puede hacer el análisis para cada nivel del factor por separado, pero teniendo en cuenta el error utilizado en el análisis de dos variables con interacción. Por lo tanto, hay cargar el paquete phia (que ya se ha cargado). No se podría hacer lo que vamos hacer sin ese paquete.

El paquete phia utiliza el archivo de resultados del anova realizado con la función lm. En nuetro caso es anova 3.

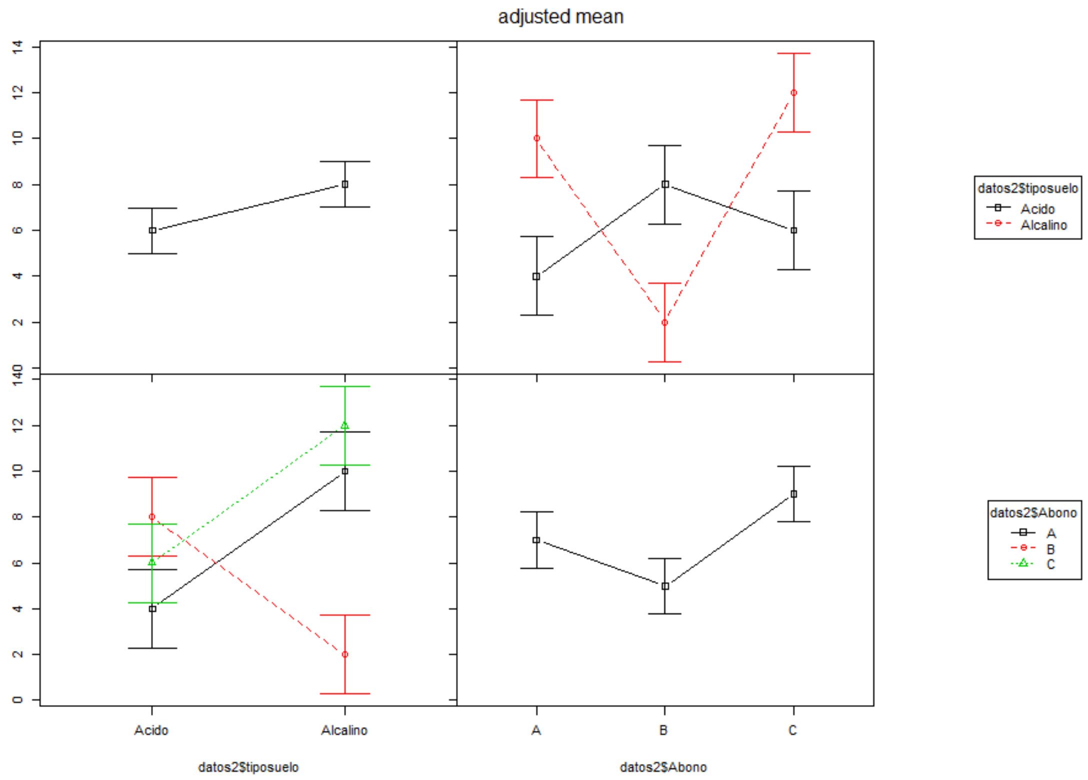
Lo primero que vamos a calcular son la medias para las diferentes combinaciones de nivel. Se van a guarda en un archivo de resultados que hemos llamado meansanova3, pero podríamos llamarle de otra manera.

```
> meansanova3<-interactionMeans(anova3)
> meansanova3
```

```
  datos2$tiposuelo datos2$Abono adjusted mean std. error
1            Acido            A             4   1.715938
2         Alcalino            A            10   1.715938
3            Acido            B             8   1.715938
4         Alcalino            B             2   1.715938
5            Acido            C             6   1.715938
6         Alcalino            C            12   1.715938
>
```

Uno de los resultados interesantes de este paquete es el grafico de interacción que hace:

```
> plot(meansanova3)
```



Pinta los gráficos de los factores principales y los de interacción intercambiando los factores.

Si ahora se quiere hacer los contrastes tras el anova considerando el error del anova con interacción se debe hacer las siguientes funciones (cuidado, los nombres de los factores entre comillas)

```
testInteractions(anova3,fixed="datos2$tiposuelo",across="datos2$Abono"
)
```

```
F Test:
P-value adjustment method: holm
          datos2$Abono1 datos2$Abono2 Df Sum of Sq      F   Pr(>F)
  Acido              -2             2  2        24 1.3585 0.293883
Alcalino             -2           -10  2       168 9.5094 0.006704 **
Residuals                             12       106
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Para el nivel de cada tipo de suelo contrasta las diferencias de las medias abono A menos el abono C contra la diferencia del abono B frente al abono C. Y se ve que esas diferencias son distintas en el tipo de suelo acido con respecto al Alcalino. Esta función hace los contraste se toma un nivel de referencias y se hacen todas la diferencias con respecto a este. El ajuste del p-valor que hace por defecto es el ajuste de Holm-Bonferroni. Podríamos cambiarlo sin mas que añadir un argumento nuevo.

```
> testInteractions(anova3,fixed="datos2$tiposuelo",across="datos2$Abon
o", p.adjust.method = "bonferroni")
```

```
F Test:
P-value adjustment method: holm
           datos2$Abono1 datos2$Abono2 Df Sum of Sq      F   Pr(>F)
   Acido              -2             2  2        24 1.3585 0.293883
Alcalino              -2           -10  2       168 9.5094 0.006704 **
Residuals                              12       106
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Podemos hacer las diferencias intercambiando los factores:

```
F Test:
P-value adjustment method: holm
          Value Df Sum of Sq      F  Pr(>F)
A            -6  1        54 6.1132 0.08808 .
B             6  1        54 6.1132 0.08808 .
C            -6  1        54 6.1132 0.08808 .
Residuals       12       106
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Se observa como las diferencias de los dos niveles de suelo son diferentes en el tipo de abono B con respecto al resto.

Aunque realmente el contraste que sería el mas adecuado con relación a la interacción y que SPSS no hace es

Si las diferencias de los niveles de tipo de suelo, son distintas en relación a los niveles del otro factor.

```
> testInteractions(anova3)
```

```
F Test:
P-value adjustment method: holm
                       Value Df Sum of Sq      F  Pr(>F)
Acido-Alcalino : A-B     -12  1       108 12.226 0.01323 *
Acido-Alcalino : A-C       0  1         0  0.000 1.00000
Acido-Alcalino : B-C      12  1       108 12.226 0.01323 *
Residuals                    12       106
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Se observa que la diferencia de abundancia entre acido y alcalino es diferente en el abono B con respecto a los otros dos niveles, pero no respecto de A y C.

Si no hubiera interacción estas diferencias entre tipo de suelo tendrían que ser muy parecidas en los tres niveles de abono.