

ANOVA I II III with R_2

Sergio Camiz

2025-05-15

La tabla ANOVA que se encuentra en R no corresponde exactamente a la tabla que se ha presentado en las clases 3 y 4.

Trabajamos con los datos de Decathlon, solo considerando los resultados de las Olimpiades y modelando el puntaje con las especialidades.

```
decathlon      <- read.csv("Decathlon_cpl.csv",header=TRUE,row.names = 1)
events         <- decathlon[14:41,1:10]
attach(events)
score          <- decathlon[14:41,12]
n              <- dim(events)[1]
SSst           <- sum(score^2) - n*(mean(score))^2
```

Vemos lo que sale, considerando tres eventos de decathlon para modelar al puntaje. Los comandos estándar para el anova son `anova()` y `aov()`. El primero es más completo, el segundo resultando casi idéntico solo saliendo su `summary()`.

```
lm1             <- lm(score ~ X100m + High.jump + Shot.put, data=events)
summary(lm1)
```

```
##
## Call:
## lm(formula = score ~ X100m + High.jump + Shot.put, data = events)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -495.37 -128.44   15.34  127.87  421.54
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13281.94    2555.96   5.196 2.53e-05 ***
## X100m        -854.45     187.95  -4.546 0.000132 ***
## High.jump    1344.48     567.48   2.369 0.026214 *
## Shot.put      98.41      61.01   1.613 0.119823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 208.3 on 24 degrees of freedom
## Multiple R-squared:  0.7222, Adjusted R-squared:  0.6874
## F-statistic: 20.79 on 3 and 24 DF,  p-value: 7.354e-07
```

```
anova(lm1)
```

```
## Analysis of Variance Table
##
```

```
## Response: score
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X100m      1 1909787 1909787 44.0155 7.317e-07 ***
## High.jump  1  683884  683884 15.7617 0.0005682 ***
## Shot.put   1  112885  112885  2.6017 0.1198227
## Residuals 24 1041334    43389
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov(lm1)
```

```
## Call:
## aov(formula = lm1)
##
## Terms:
##           X100m High.jump Shot.put Residuals
## Sum of Squares 1909787.3  683884.4 112884.9 1041334.3
## Deg. of Freedom      1          1          1          24
##
## Residual standard error: 208.3001
## Estimated effects may be unbalanced
```

```
summary(aov(lm1))
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X100m      1 1909787 1909787 44.016 7.32e-07 ***
## High.jump  1  683884  683884 15.762 0.000568 ***
## Shot.put   1  112885  112885  2.602 0.119823
## Residuals 24 1041334    43389
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Estaremos viendo de comprender su contenido más adelante. De hecho, para conseguir la tabla ANOVA, se necesita una elaboración específica basada sobre los resultados de `lm()`. La forma de la tabla es la siguiente:

#	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Cuadrados promedios</i>	<i>F</i>	<i>Prob.</i>
# Regression	<i>p-1 (sin alfa)</i>	<i>SSr=eta'eta-n ym²</i>	<i>MSr</i>	<i>MSr/MSe</i>	<i>p</i>
# Error	<i>n-p</i>	<i>SSE= SSt-SSr</i>	<i>MSe</i>		
# Total	<i>p-1 (centrado)</i>	<i>SSt=y'y-n ym²</i>			

La función `ANOVA()` proporciona la tabla, solo ingresando la lista de resultados que sale de `lm()`.

```
ANOVA <- function(m) {
  # computations
  y <- m$model$score # criterion
  eta <- m$fitted.values # fitted
  res <- m$residuals # residuals
  df1 <- length(m$coefficients)-1 # number of regressors (without 1)
  df2 <- m$df.residual # residual degrees of freedom
  dft <- df1 + df2 # total degrees of freedom
  n <- dft + 1 # number of units
  sst <- t(y) %*% y - n * mean(y)^2 # total centred sum of squares
  ssr <- t(eta) %*% eta - n * mean(y)^2 # regression sum of squares
  sse <- t(res) %*% res # error sum of squares
  msr <- ssr / df1 # mean regression sum of squares
  mse <- sse / df2 # mean error sum of squares
  f <- msr / mse # ratio msr/mse = f
}
```

```

prf          <- pf(f,df1,df2,lower.tail = FALSE) # probability of  $f > F(df1,df2)$ 
# creates anova table
anova        <- matrix(NA,nrow=3,ncol=5,dimnames=list(c("Regression","Error",
               "Total"), c("Deg.freedom","Sum of squares","Mean square",
               "F","Probability")))
anova[,1]     <- c(df1,df2,dft)
anova[,2]     <- round(c(ssr,sse,sst),8)
anova[1:2,3]  <- round(c(msr,mse),8)
anova[1,4:5]  <- round(c(f,prf),8)
out           <- print(anova,na.print = "")
}

```

Aplicada a nuestro modelo, se resulta:

```
ANOVA(lm1)
```

	Deg.freedom	Sum of squares	Mean square	F	Probability
Regression	3	2706557	902185.55	20.79299	7.4e-07
Error	24	1041334	43388.93		
Total	27	3747891			

Esta corresponde exactamente a la forma general de ANOVA. Por otro lado, las funciones que calculan el ANOVA proporcionadas para R se refieren a los varios regresores individualmente. De hecho, comparando la de arriba con la siguiente

```
anova(lm1)
```

```

## Analysis of Variance Table
##
## Response: score
##      Df Sum Sq Mean Sq F value    Pr(>F)
## X100m   1 1909787 1909787 44.0155 7.317e-07 ***
## High.jump 1  683884   683884 15.7617 0.0005682 ***
## Shot.put  1  112885   112885  2.6017 0.1198227
## Residuals 24 1041334    43389
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

se observa que la suma de cuadrados debido a la regresión está desagregada en los tres regresores, mientras el intercepta no se encuentra, así que se considera la tabla centrada.

Por otro lado, si se cambia el orden de los regresores, se encuentra otra tabla:

```

lm2          <- lm(score ~ High.jump + Shot.put + X100m, data=events)
summary(lm2)

```

```

##
## Call:
## lm(formula = score ~ High.jump + Shot.put + X100m, data = events)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -495.37 -128.44   15.34  127.87  421.54
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13281.94    2555.96   5.196 2.53e-05 ***
## High.jump    1344.48     567.48   2.369 0.026214 *

```

```
## Shot.put      98.41      61.01    1.613 0.119823
## X100m        -854.45     187.95   -4.546 0.000132 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 208.3 on 24 degrees of freedom
## Multiple R-squared:  0.7222, Adjusted R-squared:  0.6874
## F-statistic: 20.79 on 3 and 24 DF,  p-value: 7.354e-07
anova(lm2)
```

```
## Analysis of Variance Table
##
## Response: score
##          Df Sum Sq Mean Sq F value    Pr(>F)
## High.jump  1 1473411 1473411 33.9582 5.217e-06 ***
## Shot.put   1  336370   336370  7.7524 0.0102971 *
## X100m      1  896776   896776 20.6683 0.0001317 ***
## Residuals 24 1041334    43389
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Come se ve, mientras el modelo es el mismo, la tabla ANOVA es diferente. ¿Que se ha ocurrido?

Para entender esto, hay que considerar tres tipos de ANOVA: I, II y III, que tratan la ANOVA considerando los regresores separadamente. Véase, por ejemplo, <https://mcfromnz.wordpress.com/2011/03/02/anova-type-iiii-ss-explained/>.

Las sumas de cuadrados del ANOVA de tipo I corresponden a la suma de cuadrados de cada fuente después de las previas: $SS(A)$, $SS(B|A)$, $SS(AB|A,B)$. Entonces si se cambia el orden estas cambian también: solo no cambia el residuo y el total.

Las sumas de cuadrados del ANOVA de tipo II corresponden a la suma de cuadrados de cada fuente tirando las interacciones: $SS(A|AB)$, $SS(B|AB)$. Entonces cambiando el orden, no cambian.

Las sumas de cuadrados del ANOVA de tipo III corresponden a la suma de cuadrados de cada fuente tirando todas las demás y las interacciones: $SS(A|B,AB)$, $SS(B|A,AB)$.

Vamos distinguiendo los tres tipos con funciones `anova1`, `anova2`, y `anova3`, sino que se corresponden respectivamente a `anova()`, `Anova()` (de la librería `car`) y `Anova(type=3)`.

En primero lugar, vamos ajuntando a las salidas de `anova` la fila final que les falta. Por esto definimos tres funciones, correspondientes a las ANOVA de tipo I, II y III. Efectivamente las nuevas funciones llaman `anova` y `Anova` de la librería `car`, incluyendo por la III el parámetro `type=3`.

```
anova1 <- function(lm){
  a <- anova(lm)
  rbind(a,"Total"=c(n,sum(a[2]),NA,NA,NA))
}
anova2 <- function(lm){
  a <- car::Anova(lm)
  rbind(a,"Total"=c(sum(a[1]),n,NA,NA))
}
anova3 <- function(lm){
  a <- car::Anova(lm,type=3)
  rbind(a,"Total"=c(sum(a[1]),n,NA,NA))
}
```

Estamos trabajando con los datos de Decathlon. Nos limitamos al puntaje y a las diez especialidades

correspondientes a los 28 atletas de las olimpiades. Para tener regresores ortogonales, le extraemos dos componentes principales: la Dim_1 y la Dim_4.

```
dec.pca      <- PCA(events,graph=FALSE)
v1           <- dec.pca$ind$coord[,1]
v2           <- dec.pca$ind$coord[,4]
```

Estimamos el primero modelo: dos regresores ortogonales y interacción: después de la estimación corremos tres análisis de varianza diferentes.

```
lm3 = lm(score~v1*v2)
summary(lm3)
```

```
##
## Call:
## lm(formula = score ~ v1 * v2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -129.401  -25.979   -3.361   32.201   80.160
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8051.5357     9.3161  864.259 < 2e-16 ***
## v1           188.9624     4.9711   38.013 < 2e-16 ***
## v2           75.9005    11.1395    6.814 4.78e-07 ***
## v1:v2        -0.1623     4.6214   -0.035  0.972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 49.3 on 24 degrees of freedom
## Multiple R-squared:  0.9844, Adjusted R-squared:  0.9825
## F-statistic: 506.1 on 3 and 24 DF,  p-value: < 2.2e-16
```

```
cat("\n ANOVA table\n")
```

```
##
## ANOVA table
```

```
ANOVA(lm3)
```

```
##              Deg.freedom Sum of squares Mean square      F Probability
## Regression           3    3689568.07 1229856.023 506.0885          0
## Error                24     58322.89   2430.121
## Total                27    3747890.96
```

```
cat("\n ANOVA table type I\n")
```

```
##
## ANOVA table type I
```

```
anova1(lm3)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: score
```

```
##              Df  Sum Sq Mean Sq  F value    Pr(>F)
## v1             1 3544541 3544541 1458.5864 < 2.2e-16 ***
## v2             1  145024  145024   59.6777 5.827e-08 ***
```

```

## v1:v2      1      3      3      0.0012      0.9723
## Residuals 24    58323    2430
## Total      28 3747891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type II\n")

##
## ANOVA table type II
anova2(lm3)

## Anova Table (Type II tests)
##
## Response: score
##           Sum Sq Df    F value    Pr(>F)
## v1          3544541  1 1458.5864 < 2.2e-16 ***
## v2          145024  1   59.6777 5.827e-08 ***
## v1:v2           3  1    0.0012   0.9723
## Residuals    58323 24
## Total        3747891 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type III\n")

##
## ANOVA table type III
anova3(lm3)

## Anova Table (Type III tests)
##
## Response: score
##           Sum Sq Df    F value    Pr(>F)
## (Intercept) 1815162366  1 7.4694e+05 < 2.2e-16 ***
## v1           3511417  1 1.4450e+03 < 2.2e-16 ***
## v2           112821  1 4.6426e+01 4.778e-07 ***
## v1:v2           3  1 1.2000e-03   0.9723
## Residuals      58323 24
## Total        1818844929 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n Estimated standard deviation of residuals = ",sqrt(anova1(lm3)[4,3]))

##
## Estimated standard deviation of residuals = 49.29625

cat("\n F of the model (MSr/MSe) = ",
    (sum(anova1(lm3)[1:3,2])/sum(anova1(lm3)[1:3,1]))/(anova1(lm3)[4,2]/anova1(lm3)[4,1]))

##
## F of the model (MSr/MSe) = 506.0885

cat("\n Difference SSt - Total SS AnovaIII = ",SSt-anova3(lm3)[6,1],"\n")

##

```

```
## Difference SSt - Total SS AnovaIII = -1815097038
```

En la salida del modelo lineal se encuentra el Residual Standard Error que es el desvío estándar de los residuos (raíz cuadrada de MSe).

Igualmente, el F que resulta corresponde en el anova (anova1) a la razón entre el promedio de los tres cuadrados promedios de los tres regresores (intercepta, v1, v2) (o sea $MSr = SSr/df$) y MSe.

Las probabilidades a lado de los beta están asociadas al t correspondiente y se refiere a la variación correspondiente a beta(i) con otros regresores fijos. Son las mismas a lado del F en el anova3 (Anova con type=3).

Como en este caso los regresores están ortogonales, no hay variación en la suma de cuadrados en el ANOVA I debido a un intercambio. Igualmente por esta razón las sumas de cuadrados en ANOVA II son las mismas.

En anova1 y anova2 el total corresponde a la suma de cuadrados centrados SStc del y. En anova3 se debería corresponder a la suma de cuadrados, ya que hay una suma de cuadrados del promedio arriba. Pero hay una diferencia entre dichos totales que no explico: aparentemente, hay un error en el cálculo.

Aquí vemos como, sin interacción no cambia el orden en anova I intercambiando v1 y v2 si los regresores son ortogonales:

```
lm4 = lm(score~v1+v2)
summary(lm4)

##
## Call:
## lm(formula = score ~ v1 + v2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -129.76  -25.77   -3.38   32.30   80.27
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8051.536     9.128  882.058 < 2e-16 ***
## v1           188.979     4.848   38.978 < 2e-16 ***
## v2           75.715     9.603    7.884 3.06e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48.3 on 25 degrees of freedom
## Multiple R-squared:  0.9844, Adjusted R-squared:  0.9832
## F-statistic: 790.7 on 2 and 25 DF,  p-value: < 2.2e-16

cat("\n ANOVA table\n")

##
## ANOVA table

ANOVA(lm4)

##              Deg.freedom Sum of squares Mean square      F Probability
## Regression              2    3689565.07 1844782.536 790.7219          0
## Error                   25     58325.89   2333.036
## Total                   27    3747890.96

cat("\n ANOVA table type I\n")

##
```

```
## ANOVA table type I
anova1(lm4)

## Analysis of Variance Table
##
## Response: score
##           Df Sum Sq Mean Sq F value    Pr(>F)
## v1          1 3544541 3544541 1519.283 < 2.2e-16 ***
## v2          1 145024 145024 62.161 3.062e-08 ***
## Residuals 25 58326 2333
## Total      28 3747891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type II\n")

##
## ANOVA table type II
anova2(lm4)

## Anova Table (Type II tests)
##
## Response: score
##           Sum Sq Df F value    Pr(>F)
## v1          3544541 1 1519.283 < 2.2e-16 ***
## v2          145024 1 62.161 3.062e-08 ***
## Residuals    58326 25
## Total        3747891 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type III\n")

##
## ANOVA table type III
anova3(lm4)

## Anova Table (Type III tests)
##
## Response: score
##           Sum Sq Df F value    Pr(>F)
## (Intercept) 1815162366 1 778025.968 < 2.2e-16 ***
## v1          3544541 1 1519.283 < 2.2e-16 ***
## v2          145024 1 62.161 3.062e-08 ***
## Residuals    58326 25
## Total        1818910257 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n Difference SSt - Total SS AnovaIII = ", SSt-anova3(lm4)[5,1], "\n")

##
## Difference SSt - Total SS AnovaIII = -1815162366

lm4a = lm(score~v2+v1)
summary(lm4a)
```



```
##
## Call:
## lm(formula = score ~ v2 + v1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -129.76  -25.77   -3.38   32.30   80.27
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8051.536      9.128  882.058 < 2e-16 ***
## v2           75.715      9.603   7.884 3.06e-08 ***
## v1          188.979      4.848  38.978 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48.3 on 25 degrees of freedom
## Multiple R-squared:  0.9844, Adjusted R-squared:  0.9832
## F-statistic: 790.7 on 2 and 25 DF,  p-value: < 2.2e-16

cat("\n ANOVA table\n")

##
## ANOVA table

ANOVA(lm4a)

##              Deg.freedom Sum of squares Mean square      F Probability
## Regression              2    3689565.07 1844782.536 790.7219          0
## Error                   25     58325.89   2333.036
## Total                    27    3747890.96

cat("\n ANOVA table type I\n")

##
## ANOVA table type I

anova1(lm4a)

## Analysis of Variance Table
##
## Response: score
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## v2              1 145024  145024    62.161 3.062e-08 ***
## v1              1 3544541 3544541 1519.283 < 2.2e-16 ***
## Residuals      25   58326    2333
## Total          28 3747891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type II\n")

##
## ANOVA table type II

anova2(lm4a)

## Anova Table (Type II tests)
##
```

```
## Response: score
##           Sum Sq Df  F value    Pr(>F)
## v2          145024  1    62.161 3.062e-08 ***
## v1          3544541  1 1519.283 < 2.2e-16 ***
## Residuals    58326 25
## Total        3747891 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\n ANOVA table type III\n")
```

```
##
## ANOVA table type III
```

```
anova3(lm4a)
```

```
## Anova Table (Type III tests)
```

```
##
```

```
## Response: score
```

```
##           Sum Sq Df    F value    Pr(>F)
## (Intercept) 1815162366  1 778025.968 < 2.2e-16 ***
## v2          145024  1    62.161 3.062e-08 ***
## v1          3544541  1 1519.283 < 2.2e-16 ***
## Residuals    58326 25
## Total        1818910257 28
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\n Difference SSt - Total SS AnovaIII  = ",SSt-anova3(lm4a)[5,1],"\n")
```

```
##
```

```
## Difference SSt - Total SS AnovaIII  = -1815162366
```

En este caso también las sumas de cuadrados son las mismas en anova3, incluso que no hay diferencia entre suma de cuadrados total con y'y. Entonces, la diferencia encontrada depende de la presencia de la interacción.

Al contrario si cambia, si los regresores no son ortogonales. Vemos aquí una regresión del puntaje con dos especialidades:

```
lm5 = lm(score~X100m+Discus)
summary(lm5)
```

```
##
```

```
## Call:
```

```
## lm(formula = score ~ X100m + Discus)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -603.70 -108.65    3.95  163.94  598.93
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17351.00    2423.79   7.159 1.67e-07 ***
## X100m        -1016.39    201.51  -5.044 3.33e-05 ***
## Discus         40.45     14.11   2.867 0.00829 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 235.2 on 25 degrees of freedom
## Multiple R-squared:  0.6309, Adjusted R-squared:  0.6014
## F-statistic: 21.37 on 2 and 25 DF,  p-value: 3.88e-06

cat("\n ANOVA table\n")

##
## ANOVA table

ANOVA(lm5)

##           Deg.freedom Sum of squares Mean square      F Probability
## Regression           2      2364655  1182327.27  21.36886    3.88e-06
## Error                25      1383236    55329.46
## Total                 27      3747891

cat("\n ANOVA table type I\n")

##
## ANOVA table type I

anova1(lm5)

## Analysis of Variance Table
##
## Response: score
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## X100m       1 1909787 1909787  34.5166 3.955e-06 ***
## Discus      1  454867  454867   8.2211 0.008287 **
## Residuals   25 1383236    55329
## Total       28 3747891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type II\n")

##
## ANOVA table type II

anova2(lm5)

## Anova Table (Type II tests)
##
## Response: score
##           Sum Sq Df F value    Pr(>F)
## X100m      1407668  1 25.4416 3.327e-05 ***
## Discus     454867  1  8.2211 0.008287 **
## Residuals 1383236 25
## Total     3245772 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type III\n")

##
## ANOVA table type III

anova3(lm5)

## Anova Table (Type III tests)
```

```
##
## Response: score
##           Sum Sq Df F value    Pr(>F)
## (Intercept) 2835412  1 51.2460 1.675e-07 ***
## X100m       1407668  1 25.4416 3.327e-05 ***
## Discus      454867  1  8.2211 0.008287 **
## Residuals   1383236 25
## Total       6081184 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lm5a = lm(score~Discus+X100m)
summary(lm5a)

##
## Call:
## lm(formula = score ~ Discus + X100m)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -603.70 -108.65   3.95  163.94  598.93
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17351.00    2423.79   7.159 1.67e-07 ***
## Discus       40.45       14.11   2.867 0.00829 **
## X100m      -1016.39     201.51  -5.044 3.33e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 235.2 on 25 degrees of freedom
## Multiple R-squared:  0.6309, Adjusted R-squared:  0.6014
## F-statistic: 21.37 on 2 and 25 DF,  p-value: 3.88e-06

cat("\n ANOVA table\n")

##
## ANOVA table
ANOVA(lm5a)

##           Deg.freedom Sum of squares Mean square      F Probability
## Regression           2      2364655  1182327.27 21.36886    3.88e-06
## Error                25      1383236    55329.46
## Total                27      3747891

cat("\n ANOVA table type I\n")

##
## ANOVA table type I
anova1(lm5a)

## Analysis of Variance Table
##
## Response: score
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Discus     1  956987  956987  17.296 0.0003292 ***
```

```
## X100m      1 1407668 1407668 25.442 3.327e-05 ***
## Residuals 25 1383236 55329
## Total      28 3747891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\n ANOVA table type II\n")
```

```
##
## ANOVA table type II
```

```
anova2(lm5a)
```

```
## Anova Table (Type II tests)
##
## Response: score
##           Sum Sq Df F value    Pr(>F)
## Discus      454867  1  8.2211 0.008287 **
## X100m      1407668  1 25.4416 3.327e-05 ***
## Residuals 1383236 25
## Total       3245772 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\n ANOVA table type III\n")
```

```
##
## ANOVA table type III
```

```
anova3(lm5a)
```

```
## Anova Table (Type III tests)
##
## Response: score
##           Sum Sq Df F value    Pr(>F)
## (Intercept) 2835412  1 51.2460 1.675e-07 ***
## Discus      454867  1  8.2211 0.008287 **
## X100m      1407668  1 25.4416 3.327e-05 ***
## Residuals   1383236 25
## Total       6081184 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\n Difference SSt - Total SS AnovaIII = ", SSt-anova3(lm5a)[5,1], "\n")
```

```
##
## Difference SSt - Total SS AnovaIII = -2333293
```

Em este caso si resulta una diferencia en los dos modelos así como en el anova de tipo I. Asimismo las sumas totales varian. Lo mismo se ocurre, si incluimos a las interacciones.

```
lm6 = lm(score~X100m*Discus)
summary(lm6)
```

```
##
## Call:
## lm(formula = score ~ X100m * Discus)
##
## Residuals:
```

```
##      Min      1Q Median      3Q      Max
## -585.9 -127.1   -2.9  153.3  621.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11632.59   25460.91  -0.457   0.652
## X100m         1662.48    2351.26   0.707   0.486
## Discus         680.32     559.75   1.215   0.236
## X100m:Discus  -59.17      51.75  -1.143   0.264
##
## Residual standard error: 233.8 on 24 degrees of freedom
## Multiple R-squared:  0.65, Adjusted R-squared:  0.6062
## F-statistic: 14.86 on 3 and 24 DF,  p-value: 1.12e-05
```

```
cat("\n ANOVA table\n")
```

```
##
## ANOVA table
```

```
ANOVA(lm6)
```

```
##              Deg.freedom Sum of squares Mean square      F Probability
## Regression           3      2436122    812040.65 14.85702    1.12e-05
## Error                24      1311769     54657.04
## Total                27      3747891
```

```
cat("\n ANOVA table type I\n")
```

```
##
## ANOVA table type I
```

```
anova1(lm6)
```

```
## Analysis of Variance Table
##
## Response: score
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## X100m           1 1909787 1909787 34.9413 4.241e-06 ***
## Discus           1  454867  454867  8.3222 0.008145 **
## X100m:Discus     1   71467   71467  1.3076 0.264115
## Residuals       24 1311769   54657
## Total          28 3747891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sum(anova1(lm6)[1:3,2])
```

```
## [1] 2436122
```

```
cat("\n ANOVA table type II\n")
```

```
##
## ANOVA table type II
```

```
anova2(lm6)
```

```
## Anova Table (Type II tests)
##
## Response: score
##              Sum Sq Df F value    Pr(>F)
```

```

## X100m          1407668  1 25.7546 3.439e-05 ***
## Discus         454867  1  8.3222 0.008145 **
## X100m:Discus    71467  1  1.3076 0.264115
## Residuals      1311769 24
## Total          3245772 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

sum(anova2(lm6)[1:3,1])

## [1] 1934003

sum(anova2(lm6)[1:3,1])+2*anova2(lm6)[3,1]

## [1] 2076937

cat("\n ANOVA table type III\n")

##
## ANOVA table type III

anova3(lm6)

## Anova Table (Type III tests)
##
## Response: score
##              Sum Sq Df F value Pr(>F)
## (Intercept)   11409  1  0.2087 0.6519
## X100m          27325  1  0.4999 0.4863
## Discus         80739  1  1.4772 0.2360
## X100m:Discus   71467  1  1.3076 0.2641
## Residuals     1311769 24
## Total         1502709 28

sum(anova3(lm6)[1:3,1])

## [1] 119472.6

cat("\n Difference SSt - Total SS AnovaIII = ",SSt-anova3(lm6)[5,1],"\n")

##
## Difference SSt - Total SS AnovaIII = 2436122

lm6a = lm(score~Discus*X100m)
summary(lm6a)

##
## Call:
## lm(formula = score ~ Discus * X100m)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -585.9  -127.1    -2.9   153.3   621.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11632.59   25460.91  -0.457   0.652
## Discus        680.32     559.75   1.215   0.236
## X100m        1662.48    2351.26   0.707   0.486

```

```

## Discus:X100m      -59.17      51.75  -1.143    0.264
##
## Residual standard error: 233.8 on 24 degrees of freedom
## Multiple R-squared:  0.65, Adjusted R-squared:  0.6062
## F-statistic: 14.86 on 3 and 24 DF,  p-value: 1.12e-05

cat("\n ANOVA table\n")

##
## ANOVA table

ANOVA(lm6a)

##           Deg.freedom Sum of squares Mean square      F Probability
## Regression           3      2436122    812040.65 14.85702    1.12e-05
## Error                24      1311769     54657.04
## Total                27      3747891

cat("\n ANOVA table type I\n")

##
## ANOVA table type I

anova1(lm6a)

## Analysis of Variance Table
##
## Response: score
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Discus      1  956987  956987 17.5089 0.0003304 ***
## X100m       1 1407668 1407668 25.7546 3.439e-05 ***
## Discus:X100m 1   71467   71467  1.3076 0.2641150
## Residuals   24 1311769   54657
## Total      28 3747891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type II\n")

##
## ANOVA table type II

anova2(lm6a)

## Anova Table (Type II tests)
##
## Response: score
##           Sum Sq Df F value    Pr(>F)
## Discus      454867  1  8.3222  0.008145 **
## X100m      1407668  1 25.7546 3.439e-05 ***
## Discus:X100m  71467  1  1.3076  0.264115
## Residuals   1311769 24
## Total      3245772 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n ANOVA table type III\n")

##

```



```
## ANOVA table type III
```

```
anova3(lm6a)
```

```
## Anova Table (Type III tests)
```

```
##
```

```
## Response: score
```

```
##           Sum Sq Df F value Pr(>F)
## (Intercept)    11409  1  0.2087 0.6519
## Discus         80739  1  1.4772 0.2360
## X100m          27325  1  0.4999 0.4863
## Discus:X100m   71467  1  1.3076 0.2641
## Residuals     1311769 24
## Total         1502709 28
```

```
cat("\n Difference SSt - Total SS AnovaIII = ", SSt-anova3(lm6a)[5,1], "\n")
```

```
##
```

```
## Difference SSt - Total SS AnovaIII = 2436122
```

En conclusión, el empleo de las funciones de ANOVA propuestas puede ser complicado, incluso que la ANOVA de tipo III, como está propuesta, es diferente de las demás, ya que incluye el intercepta en todos los cálculos. Como el ANOVA de tipo I proporciona los regresores en el orden de ingreso, tirando los efectos de los previos, lo que aparece mejor sería de correr en primero lugar el ANOVA de tipo II, luego correr el modelo ordenando los regresores según la F decreciente. Por consecuencia, el ANOVA de tipo I estará compartiendo la suma de cuadrados por orden decreciente de importancia. Claro que la ANOVA de tipo II ya indica más correctamente cuales regresores se pueden tirar en cuanto la probabilidad de contribuir al modelo de manera significativa se encuentra.