How to Obtain Appropriate Type III SS in R

Cornell Statistical Consulting Unit

Be cautious about the anova() command in R. After estimating a linear model, researchers can either look at the beta coefficients, which are contained in the summary() of the linear model, or look at fixed-effects tests using the anova() command.The anova() command performs sequential tests (type I sum of squares), meaning that the order of terms in the table matters.

Many researchers prefer to calculate the p-value of each term controlling for all other terms in the model (type III sum of squares). For balanced designs, the type I and the type III tests will be identical, but this is not true for unbalanced designs. For unbalanced designs, type III tests can be obtained in R using the Anova() command in the car package. However, users of R must be aware that the default settings for the type III anova table will not produce the same results as SAS and JMP in the case of unbalanced designs. Only after changing the contrast for unordered factor variables from contr.treatment to contr.sum will the results of the type III anova table match other statistical programs.

For an example, we will take the mtcars dataset. Two predictor variables of interest are vs and am, which we need to specify as categorical variables for the analysis.

```
library(car)

## Loading required package: carData

data(mtcars)
mtcars$vs<-as.factor(mtcars$vs)
mtcars$am<-as.factor(mtcars$am)
```

These are the default contast settings in R.

```
options('contrasts')

## $contrasts
##         unordered           ordered
## "contr.treatment"      "contr.poly"
```

Next we estimate a linear model.

```
lm1<-lm(mpg~ vs+am+vs:am, data=mtcars)
```

The type I sums of squares can be obtained from the anova() command.

```
anova(lm1)

## Analysis of Variance Table
##
```

```
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## vs         1 496.53  496.53 41.1963 5.981e-07 ***
## am         1 276.03  276.03 22.9021 4.984e-05 ***
## vs:am      1  16.01   16.01  1.3283    0.2589
## Residuals 28 337.48   12.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Many R users use the Anova() command from the car package to obtain type III sums of squares. However, because of the default contrasts, this will not match other statistical software programs.

```
Anova(lm1, type=3)

## Anova Table (Type III tests)
##
## Response: mpg
##              Sum Sq Df  F value    Pr(>F)
## (Intercept) 2718.03  1 225.5116 6.344e-15 ***
## vs           143.28  1  11.8878  0.001805 **
## am            88.36  1   7.3311  0.011420 *
## vs:am         16.01  1   1.3283  0.258855
## Residuals    337.48 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In order to get appropriate p-values and have consistent results across statistical programs, you need to change the contrast of unordered factors to contr.sum, which uses effect coding instead of dummy coding.

```
options(contrasts=c("contr.sum", "contr.poly"))
```

The following shows the difference between the two contrasts.

```
contr.treatment(c("x1", "x2"))

##    x2
## x1  0
## x2  1

contr.sum(c("x1", "x2"))

##    [,1]
## x1    1
## x2   -1
```

Now we re-estimate the model.

```
lm2<-lm(mpg~ vs+am+vs:am, data=mtcars)
```

We can see that the type I sums of squares have not changed.

```
anova(lm2)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## vs         1 496.53  496.53 41.1963 5.981e-07 ***
## am         1 276.03  276.03 22.9021 4.984e-05 ***
## vs:am      1  16.01   16.01  1.3283    0.2589
## Residuals 28 337.48   12.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

However, the sums of squares and results have changed in the type III sums of squares table.

```
Anova(lm2, type=3)

## Anova Table (Type III tests)
##
## Response: mpg
##              Sum Sq Df   F value    Pr(>F)
## (Intercept) 13144.3  1 1090.5690 < 2.2e-16 ***
## vs            382.5  1   31.7337 4.931e-06 ***
## am            283.7  1   23.5400 4.159e-05 ***
## vs:am          16.0  1    1.3283    0.2589
## Residuals     337.5 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
