

Anova special cases with R: unbalanced and nested Anova

stats package - No install required

Y : numeric continuous variable
A, B, C, ... : factor (categorical) variables

Goal: Analyse the variation of between group means of a variable Y, where groups have different sizes and are defined by the different levels of the factors considered.

Handling unbalance

When?

Unequal number of obs for all possible combinations of factors levels

Model building

If you use ANOVA type I SS, results depend on the factors order:

```
fm <- aov(Y ~ A + B, data = data);summary(fm)
fmInv <- aov(Y ~ B + A, data = data);summary(fmInv)
```

```
> summary(fm)
      Df Sum Sq Mean Sq F value    Pr(>F)
A      1  92.62   92.62  37.431 5.59e-06 ***
B      2   42.64   21.32   8.615  0.002 **
```

```
> summary(fmInv)
      Df Sum Sq Mean Sq F value    Pr(>F)
B      2  53.09   26.55  12.24 0.000383 ***
A      1  76.80   76.80  35.42 9.95e-06 ***
```

Better use ANOVA type II SS:

```
drop1(object = fm, test = "F")
drop1(object = fmInv, test = "F")
```

```
<none>      Df Sum of Sq    RSS    AIC F value    Pr(>F)
A           1   76.802 117.996 43.609  35.424 9.949e-06 ***
B           2   47.950  89.144 35.159  11.058 0.0006532 ***
```

```
<none>      Df Sum of Sq    RSS    AIC F value    Pr(>F)
B           2   47.950  89.144 35.159  11.058 0.0006532 ***
A           1   76.802 117.996 43.609  35.424 9.949e-06 ***
```

Including info at different levels: nested Anova

When?

Each group is divided into two or more subgroups:



Data structure

Explore the data structure:

```
data$A_B = factor(data$A:data$B)
xtabs(formula = Y ~ A + A_B,
      data = data)
```

	A_B
A	1:D 1:N 2:D 2:N 3:D 3:N
1	0.631 0.634 0.000 0.000 0.000 0.000
2	0.000 0.000 0.603 0.605 0.000 0.000
3	0.000 0.000 0.000 0.000 0.623 0.618
4	0.000 0.000 0.000 0.000 0.000 0.000
5	0.000 0.000 0.000 0.000 0.000 0.000

Model building: including different levels

```
fm1 = aov(formula = Y ~ A + A/B, data = data)
summary(fm1)
```

```
> summary(fm1)
      Df Sum Sq Mean Sq F value    Pr(>F)
A       4 0.0003033 7.583e-05  8.766 3.52e-05 ***
A:B     5 0.0000186 3.720e-06  0.430  0.825
```