Unbalanced Two Factor Analysis

The term unbalanced refers to when the numbers of observations per treatment are unequal. When data is unbalanced, still appropriate to use Anova(, type = 3) and emmeans. Bottom line: No changes are required to run the analysis with unbalanced data.

In this example, we consider 3 levels of Drug and 4 levels of Dose for a total of 12 treatment combinations.

```
library(dplyr)
library(car)
library(emmeans)
DrugTrial1 <- read.csv("C:/hess/STAT512/RNotes/ExpDesign3/ED3_Unbalanced2Factor.csv")
str(DrugTrial1)
```

```
## 'data.frame':    63 obs. of  3 variables:
##  $ Drug: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Dose: int  1 1 1 2 2 2 2 2 2 3 ...
##  $ Y   : int  45 44 43 42 44 45 43 44 45 13 ...
```

```
#Important: Need to define Drug and Dose as.factors!!!!
DrugTrial1$Drug<-as.factor(DrugTrial1$Drug)
DrugTrial1$Dose<-as.factor(DrugTrial1$Dose)
str(DrugTrial1)
```

```
## 'data.frame':    63 obs. of  3 variables:
##  $ Drug: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Dose: Factor w/ 4 levels "1","2","3","4": 1 1 1 2 2 2 2 2 2 3 ...
##  $ Y   : int  45 44 43 42 44 45 43 44 45 13 ...
```

```
SumStats <- summarize(group_by(DrugTrial1, Drug, Dose),
                n = n(),
                mean = mean(Y),
                sd   = sd(Y),
                SE   = sd/sqrt(n))
SumStats
```

```
## # A tibble: 12 x 6
## # Groups:   Drug [?]
##    Drug  Dose      n  mean    sd    SE
##    <fct> <fct> <int> <dbl> <dbl> <dbl>
##  1 1     1         3  44   1     0.577
##  2 1     2         6  43.8 1.17  0.477
##  3 1     3         4  12.8 1.26  0.629
##  4 1     4         7  23.3 1.38  0.522
##  5 2     1         6  19.3 1.37  0.558
##  6 2     2         8  38   0.926 0.327
##  7 2     3         4  45.5 1     0.5
##  8 2     4         3  26.7 0.577 0.333
##  9 3     1         3  45   0     0
## 10 3     2         7  32.3 0.756 0.286
## 11 3     3         7  37.3 1.38  0.522
## 12 3     4         5  35.6 0.894 0.40
```

```
#Change contrasts options to get meaningful Type 3 tests!
options(contrasts=c("contr.sum", "contr.poly"))
Model <- lm( Y ~ Drug*Dose, data = DrugTrial1)
```

```
Anova(Model, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: Y
##             Sum Sq Df  F value    Pr(>F)
## (Intercept)  62950  1 51723.61 < 2.2e-16 ***
## Drug           454  2   186.66 < 2.2e-16 ***
## Dose           851  3   233.09 < 2.2e-16 ***
## Drug:Dose     4830  6   661.47 < 2.2e-16 ***
## Residuals       62 51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
emout1 <- emmeans(Model, ~ Dose:Drug)
emout1
```

```
##  Dose Drug emmean    SE df lower.CL upper.CL
##  1    1      44.0 0.637 51     42.7     45.3
##  2    1      43.8 0.450 51     42.9     44.7
##  3    1      12.8 0.552 51     11.6     13.9
##  4    1      23.3 0.417 51     22.4     24.1
##  1    2      19.3 0.450 51     18.4     20.2
##  2    2      38.0 0.390 51     37.2     38.8
##  3    2      45.5 0.552 51     44.4     46.6
##  4    2      26.7 0.637 51     25.4     27.9
##  1    3      45.0 0.637 51     43.7     46.3
##  2    3      32.3 0.417 51     31.4     33.1
##  3    3      37.3 0.417 51     36.4     38.1
##  4    3      35.6 0.493 51     34.6     36.6
##
## Confidence level used: 0.95
```

Means for main effects (For Illustration)

Based on the context of the data and significant interaction, inferenence for main effects probably not of interest here. So this section of the analysis is for illustration only. Due to imbalance, the simple means and emmeans for Dose are not the same. The emmeans for Dose are calculated as if sample sizes had been equal.

```
#Simple Means for Drug
SumStats2 <- summarize(group_by(DrugTrial1, Dose),
                       n = n(),
                       mean = mean(Y))
SumStats2
```

```
## # A tibble: 4 x 3
##   Dose      n  mean
##   <fct> <int> <dbl>
## 1 1        12  31.9
## 2 2        21  37.8
## 3 3        15  32.9
## 4 4        15  28.1
```

```r
#Emmeans for Drug
emmeans(Model, ~ Dose)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
##  Dose emmean    SE df lower.CL upper.CL
##  1      36.1 0.336 51     35.4     36.8
##  2      38.0 0.242 51     37.6     38.5
##  3      31.8 0.295 51     31.3     32.4
##  4      28.5 0.302 51     27.9     29.1
##
## Results are averaged over the levels of: Drug
## Confidence level used: 0.95
```

```r
aggregate(emmean ~ Dose, data = summary(emout1), FUN = mean)
```

```
##   Dose   emmean
## 1    1 36.11111
## 2    2 38.03968
## 3    3 31.84524
## 4    4 28.51746
```