Factorial ANOVA in R

Notation:

\[ DV \sim IV_1 \times IV_2 \] is the same as \[ DV \sim IV_1 + IV_2 + IV_1 : IV_2 \]

"." means 'interaction between'

1. Examine IVs and DV.
   > attach(ToothGrowth)
   a. Scale? Number of levels?
   b. Are IVs in the right format for R?
     b.i. E.g. IV – dose, 3-levels, 0.5, 1, 2 – make sure it's not treating the factor as numerical data:
     > str(ToothGrowth)
     'data.frame': 60 obs. of 3 variables:
     $ len: num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
     $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
     $ dose: num 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
     b.ii. Change it: (following long command should be entered in two parts)
     > ToothGrowth$dose = factor(ToothGrowth$dose, levels=c(0.5,1.0,2.0),
     + labels=c("low","med","high"))
     > str(ToothGrowth)
     'data.frame': 60 obs. of 3 variables:
     $ len: num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
     $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
     $ dose: Factor w/ 3 levels "low","med","high": 1 1 1 1 1 1 1 1 1 1 ...
     b.iii. Examine some of the data frame to make sure
     Look at every 5th observation between 1 and 60
     > ToothGrowth[seq(1,60,5),]
     len supp dose
     1 4.2 VC low
     6 10.0 VC low
     11 16.5 VC med
     16 17.3 VC med
     21 23.6 VC high
     26 32.5 VC high
     31 15.2 OJ low
     36 10.0 OJ low
     41 19.7 OJ med
     46 25.2 OJ med
     51 25.5 OJ high
     56 30.9 OJ high
     c. Check cell sizes are equal (ish) using replications()
     > replications(len ~ supp * dose, data=ToothGrowth)
     supp dose supp:dose
     30 20 10
     > replications(len ~ supp * dose, data=ToothGrowth[1:58,])
     $supp
     supp
     OJ VC
     28 30
     $dose
dose
     low med high
     20 20 18
     $supp:dose
d. Graphical representation – visualise the data

```r
> boxplot(len ~ supp * dose, data=ToothGrowth,
+ ylab="Tooth Length", main="Boxplots of Tooth Growth Data")
```

![Boxplots of Tooth Growth Data](image)

- There appear to be differences between the means.
- Variances don’t seem to be very similar, which could be an issue.
  - See Bartlett’s test for Homogeneity of Variance below.

Main effect of Vitamin Type

![Tooth Growth Data by Vitamin Type](image)

- Means don’t look significantly different, but similar spread in each group
• Similar spread within each group; means of low/ medium and low/high seem to be different. But there is overlap between med and high.

Look at the interaction.

**How to make an interaction plot in R**

- There seems to be no difference between supp at high dose
- There seems to be a main effect of dose – higher dose results in higher tooth length
- There doesn’t seem to be much of a main effect of supp – there is little difference between the 2 groups overall.
e. Numerical summary of the data?

Means of simple main effects – i.e. each level of dose at each level of vitamin type (this is essentially a numerical summary of the interaction plot):

```r
> with(ToothGrowth, tapply(len, list(supp, dose), mean))
  low  med  high
OJ 13.23 22.70 26.06
VC  7.98 16.77 26.14
```

Variance:

```r
> with(ToothGrowth, tapply(len, list(supp, dose), var))
  low       med      high
OJ 19.889 15.295556  7.049333
VC  7.544  6.326778 23.018222
```

OR – perform the ANOVA, save the output into a model output and ask for this data:

```r
> aov.out = aov(len ~ supp * dose, data=ToothGrowth)

“We want to look at length as a function of supplement and dose with all possible interactions between the factors”

> model.tables(aov.out, type="means", se=T)

“I want the means and standard errors of the data”

<table>
<thead>
<tr>
<th>Tables of means</th>
<th>Grand mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18.81333</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>supp</th>
<th>supp</th>
</tr>
</thead>
<tbody>
<tr>
<td>OJ</td>
<td>VC</td>
</tr>
<tr>
<td>20.663</td>
<td>16.963</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dose</th>
<th>dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>med</td>
</tr>
<tr>
<td>10.605</td>
<td>19.735</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>supp:dose</th>
<th>dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>supp low</td>
<td>med</td>
</tr>
<tr>
<td>OJ 13.23</td>
<td>22.70</td>
</tr>
<tr>
<td>VC 7.98</td>
<td>16.77</td>
</tr>
</tbody>
</table>

Standard errors for differences of means

<table>
<thead>
<tr>
<th>supp</th>
<th>dose</th>
<th>supp:dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9376</td>
<td>1.1484</td>
<td>1.6240</td>
</tr>
</tbody>
</table>

| replic. | 30 | 20 | 10 |
```

```r
> bartlett.test(len ~ supp * dose, data=ToothGrowth)

Bartlett test of homogeneity of variances

data:  len by supp by dose
Bartlett's K-squared = 1.4217, df = 1, p-value = 0.2331
```
• Non-significant, therefore there is no difference between the variances – despite appearances to the contrary in the plot above – we can assume the variances to be homogenous.

2. **Run a factorial ANOVA**
   • Although we’ve already done this to get descriptives, previously, we do:

   ```r
   > aov.out = aov(len ~ supp * dose, data=ToothGrowth)
   ```

   *NB: For more factors, list all the factors after the tilde separated by asterisks. This gives a model with all possible main effects and interactions. To leave out interactions, separate the factor names with plus signs rather than asterisks.*

   • Turn off stars showing significance values if you want?

   ```r
   > options(show.signif.stars=F)
   > summary(aov.out)
   ```

   ![Main effect of way in which supplement was administered, associated degrees of freedom and p-value (significant)](image)

   There’s a main effect of supplement ($F(1,54)=15.572; p<.01$) and dose ($F(2,54)=92.0; p<.01$) (i.e. a difference between at least 2 of the means), as well as an interaction between them($F(2,54)=4.107; p<.05$).

   Where do these differences lie – i.e. between which group means?

3. **Run post hoc tests – Tukey HSD, Bonferroni …**

   ```r
   > TukeyHSD(aov.out)
   ```
• All significant simple main effects, except highlighted ones.
• Significant main effect of dose and way supplement was administered

conf.level= changes the confidence level

"which=" option specifies which comparisons we want

E.g. TukeyHSD(aov.out, which=c("dose"), conf.level=.99) compares main effect of dose at a .99 probability level.

Graphical display – to plot CIs from Tukey:

> plot(TukeyHSD(aov.out))
Multiple t-tests with Bonferroni Adjustments

Pairwise comparisons using t tests with pooled SD
data:  len and dose

low     med
med  2.0e-08 -
high 4.4e-16 4.3e-05

P value adjustment method: bonferroni

Contrasts

- Analysis of treatment contrasts assumes a balanced design, homogeneity of variance, and additive effects (the effect of a treatment is to add a constant amount to each subject's score, plus or minus a bit of random error).

> options("contrasts")

4. Plot Assumptions
   a. Look at regression for interpretation
   > plot(aov.out)