

Transcript of Mick Crawley's R course 2010

Imperial College London, Silwood Park

Emanuel G Heitlinger

Disclaimer: The following document is a private transcript of Mick Crawley's R-course. I am a participant in this course and my writeup has in no way been approved by Mick Crawley (from whom the ideas behind the code and teaching concepts are) or any of his staff.

Day 5: ANOVA 2

Two way factorial Anova

If a experiment is replicated within each of its experimental plots we can use factorial Anova.

```
> fa <- read.table("factorial.txt", header = TRUE)
```

To calculate the usual SSA and SSB (sums of squares for the two factors) we need our function from non-factorial Anova again. Remember:

$$SSA = \frac{\sum C^2}{n} - \frac{(\sum y)^2}{kn}$$

```
> CF <- function(x) sum(x)^2/length(x)
> SSA.aov <- function(x, fac) {
+   sumCs <- sum(as.vector((tapply(x, fac, sum)^2)/tapply(x,
+   fac, length)))
+   sumCs - CF(x)
+ }
> ssa <- SSA.aov(fa$growth, fa$diet)
> ssb <- SSA.aov(fa$growth, fa$coat)
```

Factorial Anova involves additionally looking at the **interaction sum of squares (SSAB)**.

$$SSAB = \frac{\sum Q^2}{n} - SSA - SSB - CF$$

Where CF is as usual the correction factor

$$CF = \frac{(\sum y)^2}{n}$$

and Q/n is the total of each factor combination over the numbers in this combination. We can calculate this using the combined factor levels of two factors.

```
> Q.aov <- function(x, fac) {
+   sum(as.vector((tapply(x, fac, sum)^2)/tapply(x, fac, length)))
+ }
> ssab <- Q.aov(fa$growth, fa$diet:fa$coat) - ssa - ssb - CF(fa$growth)
> sst <- sum(fa$growth^2) - sum(fa$growth)^2/length(fa$growth)
> c(DIET = ssa, COAT = ssb, INTERACTION = ssab, residuals.error = sst -
+   ssa - ssb - ssab)
```

DIET	COAT	INTERACTION	residuals.error
2.6600000	2.6133333	0.6866667	2.1700000

I leave the proper table out now, as it should be clear how to go from the sums of squares with the degrees of freedoms to the probabilities via variances (mean squares) and F statistics.

```
> fa.mod <- aov(fa$growth ~ fa$diet * fa$coat)
> summary(fa.mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fa\$diet	2	2.66000	1.33000	3.6774	0.09069 .
fa\$coat	1	2.61333	2.61333	7.2258	0.03614 *
fa\$diet:fa\$coat	2	0.68667	0.34333	0.9493	0.43833
Residuals	6	2.17000	0.36167		

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

This is a first example of model simplification. As the interaction-term of the two factors seems to be non-significant we try a model without it:

```
> fa.mod2 <- update(fa.mod, ~. - fa$diet:fa$coat)
> summary(fa.mod2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fa\$diet	2	2.6600	1.33000	3.7246	0.07190 .
fa\$coat	1	2.6133	2.61333	7.3186	0.02685 *
Residuals	8	2.8567	0.35708		

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

We have to use an Anova on the two models to compare whether the first explains the data significantly better.

```
> anova(fa.mod, fa.mod2)
```

Analysis of Variance Table

```
Model 1: fa$growth ~ fa$diet * fa$coat
Model 2: fa$growth ~ fa$diet + fa$coat
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      6 2.1700
2      8 2.8567 -2  -0.68667 0.9493 0.4383
```

The simpler model is not significantly different from the complicated one and we have to think about further simplification from the second model.

```
> fa.mod3 <- update(fa.mod2, ~. - fa$diet)
> summary(fa.mod3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fa\$coat	1	2.6133	2.61333	4.7372	0.05457 .
Residuals	10	5.5167	0.55167		

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

```
> anova(fa.mod2, fa.mod3)
```

Analysis of Variance Table

```
Model 1: fa$growth ~ fa$diet + fa$coat
Model 2: fa$growth ~ fa$coat
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
```

```

1      8 2.8567
2     10 5.5167 -2      -2.66 3.7246 0.0719 .

```

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

The model 3 is not significantly worse than 2 and we are left with no significance, but there was some effect of coat when diet was included. We need to take a closer look:

```
> tapply(fa$growth, fa$diet, mean)
```

```

  A    B    C
7.70 7.95 8.80

```

Diet A and B seem to produce similar responses and can therefore be included in the model as a single factor level. Try modelling with this new two level factor.

```

> new.fac <- as.factor(ifelse(fa$diet == "C", "C", "AB"))
> fa.mod4 <- update(fa.mod3, ~. + new.fac)
> summary(fa.mod4)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fa\$coat	1	2.6133	2.6133	7.8882	0.02042 *
new.fac	1	2.5350	2.5350	7.6518	0.02189 *
Residuals	9	2.9817	0.3313		

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

```
> anova(fa.mod4, fa.mod3)
```

Analysis of Variance Table

Model 1: fa\$growth ~ fa\$coat + new.fac

Model 2: fa\$growth ~ fa\$coat

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	9	2.9817				
2	10	5.5167	-1	-2.535	7.6518	0.02189 *

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

The new model is significantly better than the old one, because reducing the factor levels to 2 saved us a degree of freedom. In fact this is a first example of contrasts, we will meet them in another practical in more detail.

Is this enough, or do we also have to include the interaction term of diet and the new factor?

```
> fa.mod5 <- update(fa.mod4, ~. + fa$coat:new.fac)
> summary(fa.mod5)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fa\$coat	1	2.61333	2.61333	7.7432	0.02383 *
new.fac	1	2.53500	2.53500	7.5111	0.02542 *
fa\$coat:new.fac	1	0.28167	0.28167	0.8346	0.38766
Residuals	8	2.70000	0.33750		

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

```
> anova(fa.mod4, fa.mod5)
```

Analysis of Variance Table

Model 1: fa\$growth ~ fa\$coat + new.fac

Model 2: fa\$growth ~ fa\$coat + new.fac + fa\$coat:new.fac

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	9	2.9817				
2	8	2.7000	1	0.28167	0.8346	0.3877

No, the interaction is not significant and model 5 is not better than 4. Model 4 (fa.mod4) is our minimal adequate model.

Three way factorial Anova

We load a dataset, that has three factors and repeated measurements for factor combinations.

```
> dada <- read.table("daphnia.txt", header = TRUE)
> attach(dada)
```

The simple sums of squares and the interaction sums of squares are calculated exactly as above. The new quantity to calculate is the **three way interaction sum of squares(SSABC)**

$$SSABC = \frac{\sum T^2}{n} - SSA - SSB - SSC - SSAB - SSAC - AABC - CF$$

We can calculate all these quantities using functions we already have:

```
> sst <- sum(Growth.rate^2) - sum(Growth.rate)^2/length(Growth.rate)
> ssa <- SSA.aov(Growth.rate, Water)
> ssb <- SSA.aov(Growth.rate, Detergent)
> ssc <- SSA.aov(Growth.rate, Daphnia)
> ssab <- Q.aov(Growth.rate, Water:Detergent) - ssa - ssb - CF(Growth.rate)
> ssac <- Q.aov(Growth.rate, Water:Daphnia) - ssa - ssc - CF(Growth.rate)
> ssbc <- Q.aov(Growth.rate, Detergent:Daphnia) - ssb - ssc - CF(Growth.rate)
> ssabc <- Q.aov(Growth.rate, Water:Detergent:Daphnia) - ssa -
+   ssb - ssc - ssab - ssac - ssbc - CF(Growth.rate)
> sse <- sst - ssa - ssb - ssc - ssab - ssac - ssbc - ssabc
> round(c(SST = sst, SSA = ssa, SSB = ssb, SSC = ssc, SSAB = ssab,
+   SSAC = ssac, SSBC = ssbc, SSABC = ssabc, SSE = sse), 3)
```

SST	SSA	SSB	SSC	SSAB	SSAC	SSBC	SSABC	SSE
117.157	1.985	2.212	39.178	0.175	13.732	20.601	5.848	33.428

Compare this sloppy presentation of the different sums of squares to these from the model:

```
> fac <- aov(Growth.rate ~ Water * Detergent * Daphnia)
> summary(fac)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Water	1	1.985	1.9851	2.8504	0.0978380 .
Detergent	3	2.212	0.7372	1.0586	0.3754783
Daphnia	2	39.178	19.5889	28.1283	8.228e-09 ***
Water:Detergent	3	0.175	0.0583	0.0837	0.9686075
Water:Daphnia	2	13.732	6.8660	9.8591	0.0002587 ***
Detergent:Daphnia	6	20.601	3.4334	4.9302	0.0005323 ***
Water:Detergent:Daphnia	6	5.848	0.9746	1.3995	0.2343235
Residuals	48	33.428	0.6964		

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

The three way interaction can be removed and one of the two way interactions:

```
> fac2 <- update(fac, ~. - Water:Detergent:Daphnia)
> summary(fac2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Water	1	1.985	1.9851	2.7293	0.1043277
Detergent	3	2.212	0.7372	1.0136	0.3939083
Daphnia	2	39.178	19.5889	26.9329	7.705e-09 ***
Water:Detergent	3	0.175	0.0583	0.0801	0.9705247
Water:Daphnia	2	13.732	6.8660	9.4401	0.0003049 ***
Detergent:Daphnia	6	20.601	3.4334	4.7206	0.0006253 ***
Residuals	54	39.275	0.7273		

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

```
> anova(fac, fac2)
```

Analysis of Variance Table

Model 1: Growth.rate ~ Water * Detergent * Daphnia

Model 2: Growth.rate ~ Water + Detergent + Daphnia + Water:Detergent +
Water:Daphnia + Detergent:Daphnia

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	48	33.428				
2	54	39.275	-6	-5.8476	1.3995	0.2343

```
> fac3 <- update(fac2, ~. - Water:Detergent)
```

```
> summary(fac3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Water	1	1.985	1.9851	2.8681	0.0958083 .
Detergent	3	2.212	0.7372	1.0651	0.3711288
Daphnia	2	39.178	19.5889	28.3031	2.907e-09 ***
Water:Daphnia	2	13.732	6.8660	9.9204	0.0002009 ***
Detergent:Daphnia	6	20.601	3.4334	4.9608	0.0003798 ***
Residuals	57	39.450	0.6921		

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

```
> anova(fac2, fac3)
```

Analysis of Variance Table

Model 1: Growth.rate ~ Water + Detergent + Daphnia + Water:Detergent +
Water:Daphnia + Detergent:Daphnia

Model 2: Growth.rate ~ Water + Detergent + Daphnia + Water:Daphnia + Detergent:Daphnia

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	54	39.275				
2	57	39.450	-3	-0.17486	0.0801	0.9705

To get a instant view on the means we can use

```
> model.tables(fac3)
```

Tables of effects

Water

Water

	Tyne	Wear
	-0.16604	0.16604

Detergent

Detergent

	BrandA	BrandB	BrandC	BrandD
	0.03293	0.15814	0.10261	-0.29367

Daphnia

Daphnia

	Clone1	Clone2	Clone3
	-1.0120	0.7252	0.2868

Water:Daphnia

Daphnia

Water	Clone1	Clone2	Clone3
Tyne	0.1941	-0.6048	0.4107
Wear	-0.1941	0.6048	-0.4107

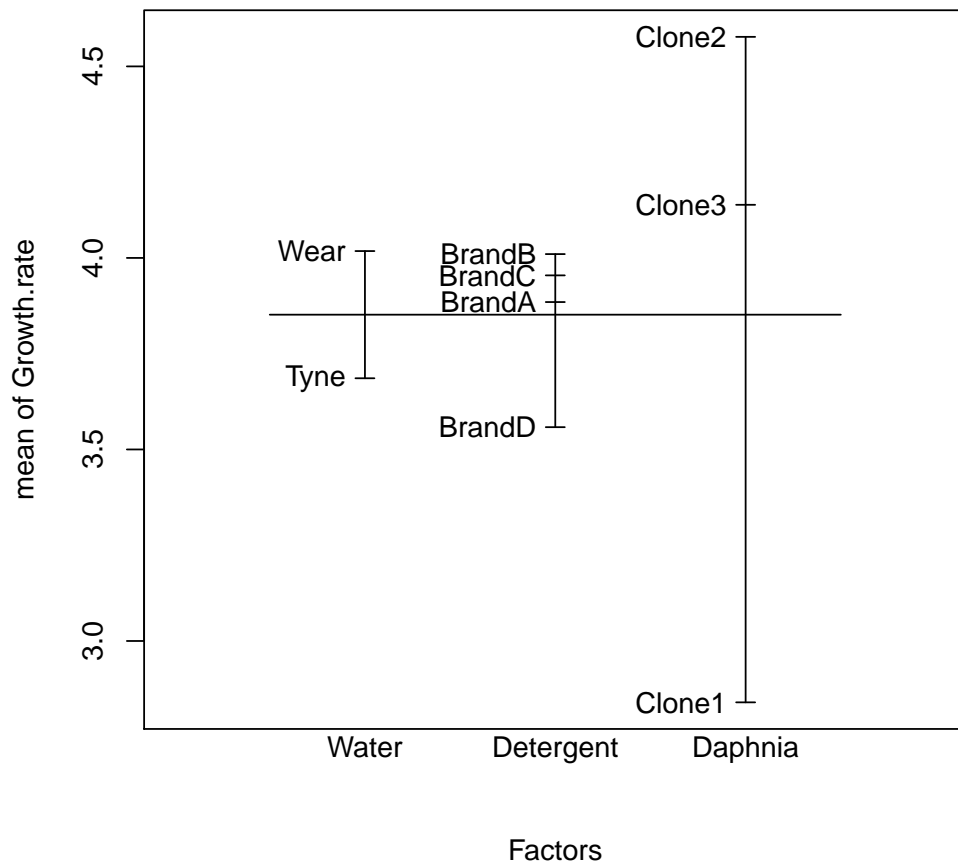
Detergent:Daphnia

Daphnia

Detergent	Clone1	Clone2	Clone3
BrandA	-0.1406	-0.6910	0.8316
BrandB	-0.0689	-0.3323	0.4012
BrandC	0.1289	0.0931	-0.2219
BrandD	0.0806	0.9303	-1.0109

To illustrate we can use plot.design and interaction.plot

```
> plot.design(dada)
```

```
> interaction.plot(Detergent, Daphnia, Growth.rate)
```

