

**JOURNAL OF ANIMAL ECOLOGY**

**SUPPLEMENTARY MATERIAL FOR:**

**Analysis of Variance with Unbalanced Data: An Update for Ecology & Evolution**

**Andy Hector, Stefanie von Felten and Bernhard Schmid**

**CONTENTS**

S1. Unbalanced Datasets

S2. A Brief History of ANOVA Sums of Squares

S3. The Intercept and the Model Matrix in ANOVA

S4. SAS Type I-IV Sums of Squares

S5. Using Indicator Dummy Variables to Adjust for Interactions

S6. Attempting to Restore Balance Using Adjusted Sums of Squares

S7. Missing and double-counted sums of squares

S8. Testing Main Effects Versus Interactions

S9. Marginality: Special Cases

S10. Marginality: Tukey's Example

## **S1. Unbalanced Datasets**

Fully balanced datasets have the same number of replicates for each treatment combination. A design can be unbalanced but orthogonal when the replicates are spread across treatments in the same proportions (Grafen and Hails, 2002). It is a good idea to check if explanatory variables are correlated by regressing one on another. If the regression term is zero then the two explanatory variables are orthogonal. Factors must be first converted into sets of binary indicator (or 'dummy') variables and then the signs of the correlations can be explored.

## **S2. A Brief History of ANOVA Sums of Squares**

Following Fisher's (1925) invention of analysis of variance using sequential sums of squares, Yates (1933, 1934) developed two alternative types of sums of squares specifically for unbalanced data. He used the following terminology (with some of the modern equivalents given above following in parentheses):

- 'Proportionate class numbers' (sequential - or SAS type I - sums of squares applied to balanced but non-orthogonal datasets with proportionate replication of the different treatment combinations as described in S1).
- 'Fitting constants' (adjusted sums of squares with higher-level terms omitted; SAS type II).
- 'Weighted squares of means' (adjusted sums of squares with higher-level terms included; SAS Type III).

Herr (1986) reviews the development of unbalanced ANOVA up until the arrival of modern statistical computer software saying that, "*we were ready to rediscover Yates. This rediscovery process has not been without controversy, so it appears that what has been done since 1965 has not resolved many questions. It seems that...in*

*general we took various methods and, treating them as more or less equal, programmed our statistical packages to use whatever method we had learned. Then we championed this method as **the method.**"* [original emphasis]

Different computer software packages then developed different approaches to unbalanced ANOVA (for a recent survey see Langsrud 2003). Some packages (e.g. S-Plus) use sequential sums of squares by default while others (e.g. GLIM, GenStat, and the base distribution of R) *only* provide sequential sums of squares. Other statistical software packages have devised their own suite of types of sums of squares, the most well-known of which is the SAS system of type I, II, III *etc.* sums of squares (SAS Institute Inc., 1985). This paper does not review the alternative forms of adjusted sums of squares available in different software packages (Langsrud, 2003). Many packages specify higher-terms included adjusted sums of squares (SAS type III) as the default setting (e.g. JMP; Minitab, SPSS, STATISTICA, Systat), either alone or in combination with sequential sums of squares (e.g. SAS). This widespread specification of higher-terms included adjusted sums of squares as the default setting in many statistical packages has led to their widespread application (either knowingly or unknowingly).

### **S3. The Intercept and the Model Matrix in ANOVA**

Statistical software packages use different formulations of the ANOVA model regarding the intercept. There are two main approaches. One option is to take the grand mean as the intercept in the ANOVA model (mixed-effect model ANOVA software often does this). However, because we are often not interested in the grand mean many statistical packages take one of the treatment level means as intercept instead (the arbitrary rules for selecting which treatment level vary between software

packages). We found it easier to write about the models in general using the formulation where the intercept is the grand mean. However, the intercept can be thought of simply as a baseline from which to calculate the sums of squares and the issues discussed are not dependent on whether this is the grand mean or one of the factor levels. Our model (1) in effects notation:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

could be rewritten in the more concise matrix notation as follows:

$$y = X\beta + \varepsilon$$

Where X is the 'model' or 'design matrix' which is made up of a series of binary indicator (or 'dummy') variables that indicate which treatment each data point is subject to and  $\beta$  is a vector of the estimated coefficients (intercepts or slopes). To give a simpler example, for the main effects model *without interaction*, T + S, the model matrix (ignoring missing values) is:

	(Intercept)	treatmentremoval	initial.sizesmall
1	1	0	1
2	1	0	1
3	1	0	0
4	1	0	0
5	1	0	0
6	1	0	0
7	1	1	1
8	1	1	1
9	1	1	1
10	1	1	0
11	1	1	0

With the system of so-called 'treatment' contrasts that we use here (see Venables and Ripley, 2002) the complete column of 1's headed '(Intercept)' indicates that the mean size of large study organisms in the control treatment provides the baseline for the calculation of all fitted values. The middle column indicates which

fitted values need to take into account the coefficient for the removal treatment and the right-hand column indicates which fitted values need to take into account the coefficient for small study organisms. Each fitted value is then calculated by taking the value of the intercept and adding to it the value of the coefficients times their respective binary dummy variables. Where the binary indicator variable contains a value of 1 the coefficient is included in the calculation and where the binary indicator variable contains the value 0 the coefficient is not included in the calculation.

#### **S4. SAS Type I-IV Sums of Squares**

The SAS categorisation of sums of squares that uses roman numerals is less informative than the term adjusted (or marginal, partial, conditional or unweighted) which give a clue as to how these sums of squares are calculated. Moreover, the different SAS types are not mutually exclusive. For a sequential analysis of a 2-way ANOVA *without interactions* (e.g. S+T) the SAS type I sum of squares for the main effect in last place (T) will also be its type II sum of squares (because in both cases the sum of squares for the second main effect has been adjusted for the preceding main effect of S). SAS type III sums of squares only apply to analyses of datasets without missing cells and where one or more interactions are included in the model (if all interactions are dropped from the model SAS type III sums of squares essentially collapse to type II).

SAS type IV sums of squares are intended for situations with missing cells and use the so-called cell-means model approach (Searle, 1987, SAS Institute Inc., 1985). However, due to the complexity of datasets with missing cells, SAS type IV sums of squares may not automatically produce the most appropriate and informative tests of the questions of interest and are therefore not widely recommended or used. Instead, it

is generally recommended that the analysts create their own contrast variables and use a multiple regression-type approach to test the hypotheses of interest. This can also be done using judicious application of the cell-means approach where the full multi-way ANOVA is treated as if it were a one-way ANOVA using a single explanatory variable that has as many levels as there are combinations of the original design variables and contrast variables are then created to test the questions of interest (Neter et al., 1996). Once again, the models should fit the objectives of the analysis.

### **S5. Using Indicator Dummy Variables to Adjust for Interactions**

When some software packages (e.g. GenStat, GLIM and the base distribution of R) encounter an interaction in a linear model formula when one of the main effects has not been specified they simply add the sum of squares for the 'missing' main effect to that for the interaction term and report the sum of these two values.

To obtain adjusted sums of squares using software that only uses sequential sums of squares a new indicator or 'dummy' variable (let's call it TS) must be created to specify the interaction as if it were a third explanatory variable that is unrelated to the main effects. The indicator variable that specifies the interaction independent of the main effects can then be placed anywhere in the model sequence and models 17 - 19 in the main text can be fitted to produce the sequential ANOVAs shown in Table 4a-c. The final term from each sequential model (ignoring the residual error, which is the same in all cases) is taken to form the composite table of adjusted sums of squares (Table 4d).

Above we have described a method for forming a model matrix using binary indicator variables that is sometimes known as the system of 'effects' or 'treatment contrasts'. Higher-order terms included sums of squares cannot be calculated with

treatment contrasts and are instead often calculated using another form of dummy variable coding that use -1 and +1 instead, the so-called 'sum-to-zero' or 'sigma constraints'. For the example analysis, using the sum-to-zero contrasts produces the following model matrix (where 'intercept' denotes the 'grand mean'):

```
(Intercept) treatment initial.size treatment:initial.size
1           1           -1           -1
1           1           -1           -1
1           1            1            1
1           1            1            1
1           1            1            1
1           1            1            1
1           -1           -1            1
1           -1           -1            1
1           -1           -1            1
1           -1            1           -1
1           -1            1           -1
```

However, following the marginality relations described above, some software will not allow a main effect to be fitted after an interaction in which it is involved even with sum-to-zero or sigma constraints coding. If the analyst wishes to bypass the marginality constraints and adjust main effects for interactions a trick is then required to mimic the process that is performed (behind the scene) by the software that does produce adjusted sums of squares. To produce adjusted sums of squares for the example analysis using the base distribution of R we created a dummy variable to code for the interaction independent of the main effects (the base distribution for R calculates only sequential sums of squares, although some additional packages can generate higher-terms included sums of squares if desired (Venables, 2000 , Fox, 2002)). For a simple two-way factorial design where both factors are simply present or absent, this dummy variable for the interaction can be made by giving the dummy variable one code when both main effects are present or both absent, and a different

code when one main effect is present but the other is absent. In our two-way factorial, initial size refers to two treatment groups (small, large). To get the dummy variable for the interaction, we arbitrarily coded treatment as control = 1 or removal = 2 and initial size as large = 1 and 2 = small, and the interaction dummy variable as 1 if treatment and initial size differed in the code (i.e., 1 and 2 or 2 and 1) and as 2 if they had the same code (see supplementary text file of R code) to add the new dummy variable, `ts`, to the dataset:

	height	initial.size	treatment	size	treat	ts
1	50	small	control	2	1	2
2	57	small	control	2	1	2
3	91	large	control	1	1	1
4	94	large	control	1	1	1
5	102	large	control	1	1	1
6	110	large	control	1	1	1
7	57	small	removal	2	2	1
8	71	small	removal	2	2	1
9	85	small	removal	2	2	1
10	105	large	removal	1	2	2
11	120	large	removal	1	2	2

### **S6. Attempting to Restore Balance Using Adjusted Sums of Squares**

Higher-terms included adjusted sums of squares can be thought of as testing variables in unbalanced datasets as if those datasets were actually balanced and orthogonal. For example, we can compare the analysis of the example data using the higher-terms included adjusted sums of squares (Table 4d) with one where missing values are replaced with the cell means to produce a balanced ( $N = 4 \times 4 = 16$ ) dataset. The example dataset has two categorical variables, each with two levels, and therefore four cell means (the average of the values in each treatment combination) as shown in Table 1. To form a balanced dataset we would replace the missing values with the appropriate cell mean. Performing an ANOVA on this modified dataset produces F

values for the main effect of initial size (71.24), removal treatment (8.84) and interaction (0.17) that are in the same relative proportions as those from the analysis with higher-terms included adjusted sums of squares (Table 4d) but which are larger (by a factor of 1.58) because of the artificially inflated sample size (from 11 to 16 values). Thus, higher-terms included adjusted sums of squares attempt to restore balance without replacing missing observations with average values.

### **S7. Missing and Double-Counted Sums of Squares**

Because ANOVA tables of adjusted sums of squares do not come from a single decomposition of the total sum of squares but are instead composite tables where the sums of squares for different explanatory variables come from different sequential analyses they are not additive. This can be easily seen from the fact that adding up the adjusted sums of squares in a given table produces a value that does not match the total of the additive sequential sums of squares for the same data. Aitkin (1977) reproduces four alternative sequential and adjusted sums of squares for a dataset analysed by Francis (1973) - which we reproduce in our Table S2. Comparison of the sums of squares in Table S2 gives an example (in the final column) of how using adjusted sums of squares can lead to a substantial reduction in the total. Aitkin argues that only the comparison of alternative sequential analyses reveals the appropriate models whereas the use of adjusted sums of squares alone can obscure important effects. This occurs due to the collinearity produced by imbalance.

### **S8. Testing Main Effects Versus Interactions**

Many statisticians have argued that testing a null hypothesis of no main effect in the presence of a significant interaction is unrealistic because the interaction implies main

effects. The exception would be an 'interaction only' situation where both main effects must be present to cause an effect and otherwise cancel out to produce zero main effects. While many consider this situation too unlikely to occur (e.g. Nelder 1994), Stewart-Oaten (1995) suggests some possible hypothetical biological situations where such special cases could exist. To this list we would add the possible special cases of genotype-by-environment interactions (e.g. Lewontin, 1974, Lewontin, 2006) and home-away effects in reciprocal transplant experiments (Antonovics and Primack, 1982, Turkington and Harper, 1979, Joshi et al., 2001). A main effect could be tested to see if it was significantly smaller than the interaction (i.e. tested versus the interaction as error term) by taking the reciprocal of the resulting F-value as the test statistic with the corresponding degrees of freedom.

One situation where the main effects are still of some interest despite a significant interaction is when the sums of squares for the main effects are considerably larger than that for the interaction. In this case, Mead, Curnow & Hasted (2003) suggest presenting estimates for the main effects alongside the estimates for the interaction model. Nelder & Lane (1995) go further and argue that when the main effects are large relative to the interaction this can be examined by testing the main effects against the interaction (not the residual error). The hypothesis being tested is not whether the main effects are null (as when testing against the residual) but whether the main effects show excess variation over and above that displayed by the interaction.

Another topic relating to ANOVA that has seen marked development during the last decade is the use of modern mixed-effects models (that is, those performed using maximum likelihood or REML). Mixed-effects models are so called because they contain a mixture of fixed and random effects. It is difficult to give a consistent

definition of fixed and random effects (Green and Tukey, 1960, Gelman and Hill, 2007) but a good example of fixed effects are experimental treatments where we are specifically interested in the treatment-level means. An example of a random effect is a blocking variables where we may, or may not, be interested in predicted values for the particular blocks but where we are usually more interested in the estimated level of overall variability (the variance components). The discussion so far has followed a classical fixed-effects approach to ANOVA where both factors are fixed treatments and where the presence of appreciable interaction usually means there is little interest in the main effects. The situation is different for mixed-effects models that contain interactions between fixed and random effects (the interaction of a fixed and random effect is random). In this case, it is usual to look at the main effect of the fixed variable even if the interaction term is significant. The main effect shows us the average effect of the fixed effect variable across the levels of the random effect and the interaction quantifies how variable the effect is (the variance component for the random interaction term).

### **S9. Marginality: Special Cases**

Nelder (1994) gives the following example (his Figure 2). Consider a linear model of the analysis of covariance type which involves one continuous explanatory variable ( $X$ ), one factor ( $A$ ) and their interaction ( $A.X$ ). Normally, the main effects would be considered marginal to the interaction. In decreasing order of complexity we would usually test five models with: (i) differences in slopes and intercepts ( $A + X + A.X$ ); (ii) a common slope but different intercepts ( $A + X$ ); (iii) a common slope and intercept ( $X$ ); (iv) or, no slope but differences in intercept ( $A$ ); and (v) no effect of either factor. Normally it would make no sense to try and remove the main effect for

the factor A in the presence of a significant A.X interaction as this would force a common intercept for the different slopes for the two groups at the zero value of the continuous explanatory variable and why should this occur? One special case could be a slope-ratio assay (Nelder 1994) or similar situation. For example, imagine a drug trial over time where both control and treatment group have zero level of the drug at the start of the trial. In a case like this it could be argued that both groups should have a common intercept at time zero even in the presence of different slopes for the response over the duration of the trial.

### **S10. Marginality: Tukey's Example**

Imagine a confectionery manufacturer that sells two flavours of a particular type of sweet but wants to specialise on only the most popular flavour. In a simple scenario a random sample of people are tested to see which flavour they prefer. The popularity of the two flavours is then analysed and the more popular flavour selected for future production. However, imagine the sex of the tasters is also recorded and when included in the analysis reveals a significant flavour-by-sex interaction (males preferring one flavour and females the other). Tukey argues that in a case like this it would make sense, despite the significant interaction of flavour with sex, to go ahead and test the null hypothesis of no main effect for a difference in popularity between flavours by testing flavour against the residual. In response, Nelder & Lane (1995) agreed that there would be many situations like this where marginal means for main effects would want to be *estimated* despite a significant interaction, but he saw this more as a question of prediction than of hypothesis testing and model selection. A similar example can also be found in Frane & Jennrich (1977).

### **S11. Glossary**

**Covariate:** Here, we use this term to mean a variable of secondary interest that is included in an analysis to account for potentially influential effects not controlled by the experimental design (specifically: to account for differences in initial size at the beginning of the example). Elsewhere, the term is sometimes more generally used to refer to the continuous explanatory variable in a classical analysis of covariance design, or even more generally as a synonym for any explanatory variable (whether categorical or continuous).

**Factor:** A categorical explanatory variable (e.g. control vs. neighbour removal treatment).

**Factor level:** The example analysis contains two factors each with two levels (given in parentheses): size (large or small) and removal of neighbours (removed or not removed). In total the design has four factor level combinations (or 'cells' in a tabular representation of the design).

**Interaction:** ANOVA and linear models assume that the effects of variables are additive. That is, if a first explanatory variable causes a change  $\Delta A$  in the response and a second explanatory variable causes a change  $\Delta B$ , then the expected change when both variables are acting is simply the sum of the two:  $\Delta A + \Delta B$ . If the observed change is significantly greater or less than expected, then there is an interaction. Positive or synergistic interactions occur when the response is more than additive and negative or inhibitory interactions occur when the response is less than expected.

**REML:** An extension of maximum likelihood (restricted/reduced/residual maximum likelihood) developed specifically for use with the mixed-effects models discussed in supplement S8.

**Term:** We use this as a general word to refer to both continuous and categorical explanatory variables plus any interactions in a statistical model formula (e.g. "terms in the model", often called "source of variation" in ANOVA tables).

**Variable:** We use this to refer to responses (outcomes) and to categorical and continuous explanatory variables.

### Supplementary References

Antonovics, J. & Primack, R. B. (1982) Experimental Ecological Genetics in *Plantago*

.6. The Demography of Seedling Transplants of *P. lanceolata*. *Journal of Ecology*, **70**, 55-75.

Fisher, R. A. (1925) *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh.

Fox, J. (2002) *An R and S-Plus companion to applied regression*, Sage Publications, Thousand Oaks.

Francis, I. (1973) Comparison of several analyses of variance programs. *Journal of the American Statistical Society*, **68**, 860-865.

Frane, J. & Jennrich, R. (1977) A reformulation of linear models - Discussion. *Journal of Royal Statistical Society. Series A*, **140**, 73-74.

Grafen, A. & Hails, R. (2002) *Modern Statistics for the Life Sciences*, Oxford University Press.

Joshi, J., Schmid, B., Caldeira, M. C., Dimitrakopoulos, P. G., Good, J., Harris, R., Hector, A., Huss-Danell, K., Jumpponen, A., Minns, A., Mulder, C. P. H., Pereira, J. S., Prinz, A., Scherer-Lorenzen, M., Siamantziouras, A. S. D., Terry, A. C., Troumbis, A. Y. & Lawton, J. H. (2001) Local adaptation enhances performance of common plant species. *Ecology Letters*, **4**, 536-544.

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

- Langsrud, Y. (2003) ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics And Computing*, **13**, 163-167.
- Lewontin, R. C. (1974) The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, **26**, 400-411.
- Lewontin, R. C. (2006) The analysis of variance and the analysis of causes (Reprinted from *Am J Hum Genet*, vol 26, pg 400-411, 1974). *International Journal of Epidemiology*, **35**, 520-525.
- Mead, R., Curnow, R. N. & Hasted, A. M. (2003) *Statistical Methods in Agriculture and Experimental Biology*, Chapman & Hall/CRC, London.
- Nelder, J. & Lane, P. (1995) The Computer Analysis Of Factorial Experiments: In Memoriam - Frank Yates. *The American Statistician*, **49**, 382-385.
- Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W. (1996) *Applied Linear Statistical Models*, Irwin, Chicago.
- SAS Institute Inc. (1985) SAS Users Guide: Statistics. SAS Institute Inc., Cary, NC, USA.
- Searle, S. R. (1987) *Linear Models for Unbalanced Data*, John Wiley & Sons, New York.
- Turkington, R. & Harper, J. L. (1979) Growth, Distribution and Neighbor Relationships of *Trifolium-Repens* in a Permanent Pasture .4. Fine-Scale Biotic Differentiation. *Journal of Ecology*, **67**, 245-254.
- Venables, W. N. (2000 ) Exegeses on Linear Models (paper presented to the S-Plus User's Conference, Washington D.C., 8-9th October 1998). Washington DC.
- Venables, W. N. & Ripley, B. D. (2002) *Modern applied statistics with S*, Springer-Verlag, Berlin.

**HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL**

Yates, F. (1933) The principles of orthogonality and confounding in replicated experiments. *The Journal of Agricultural Sciences*, **23**, 108-145.

Yates, F. (1934) The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Society*, **29**, 51-66.

**HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL**

<b>Source</b>	<b>DF</b>	<b>SS (SAS type I)</b>	<b>F</b>	<b>P</b>
Treatment	1	35.3	0.33	0.583
Size	1	4846.0	45.37	0.0003
T x S	2	11.4	0.11	0.753
Error	6	747.7		

<b>Source</b>	<b>DF</b>	<b>SS (SAS type II)</b>	<b>F</b>	<b>P</b>
Treatment	1	590.2	5.52	0.051
Size	1	4846.0	45.37	0.0003
T x S	2	11.4	0.11	0.753
Error	6	747.7		

<b>Source</b>	<b>DF</b>	<b>SS (SAS type III)</b>	<b>F</b>	<b>P</b>
Treatment	1	597.2	5.59	0.050
Size	1	4807.9	45.01	0.0003
T x S	2	11.4	0.11	0.753
Error	6	747.7		

**Table S1:** ANOVA tables reproduced from Shaw & Mitchell-Olds (1993). Note that all three models use only one of the two possible sequences of terms in the model formula.

Source	DF	SS	SS	SS	SS
		sequential S+R+S.R	sequential R+S+R.S	adjusted higher-terms omitted	adjusted higher-terms included
Mean	1	12983	12983	12983	7306
Sex	1	44	29	29	11
Religion	4	54	69	54	58
Interaction	4	10	10	10	10
Residual	1300	2989	2989	2989	2989
Total	1410	16080	16080	16065	10374

**Table S2** Sequential and adjusted sums of squares for the analysis of an unbalanced 2 x 5 (sex x religion) classification by Francis (1973) following the re-assessment by Aitkin (1977). Francis recommended the higher-terms included adjusted sums of squares but note that the main effect for sex is much reduced compared to the other analyses and the total of the sums of squares is much less than for the sequential models. Aitkin suggests that comparison of the two alternative sequential sums of squares gives a more complete understanding of the data. For simplicity, the grand mean has been omitted as has a fifth set of incorrect sums of squares from the original analysis.

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

**Appendix:** The following code for performing the analysis using the base distribution of R is supplied as a supplementary text file. Comments follow the hash symbol, #.

The input code is preceded by the > symbol with the R output interspersed.

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 > ### R code for the example analysis in:
2 > ### Analysis of Variance with Unbalanced Data: An Update for Ecology & Evolution
3 > ### by Andy Hector, Stefanie von Felten and Bernhard Schmid
4 >
5 > rm( list = ls( all = TRUE ) ) # clear workspace
6 >
7 > ## TABLE 1: HYPOTHETICAL EXAMPLE DATA
8 > # reproduced from Shaw & Mitchell-Olds (1993) Ecology 74(6), pp. 1638-1645
9 >
10 > # the data:
11 >
12 > height <- c( 50, 57, 91, 94, 102, 110, 57, 71, 85, 105, 120 )
13 > initial.size <- c( rep( "small", 2 ), rep( "large", 4 ), rep( "small", 3 ), rep( "large", 2 ) )
14 > treatment <- c( rep( "control", 6 ), rep( "removal", 5 ) )
15 > Shaw.MitchellOlds <- data.frame( height, initial.size, treatment )
16 > Shaw.MitchellOlds
17
18   height initial.size treatment
19 1      50      small  control
20 2      57      small  control
21 3      91      large  control
22 4      94      large  control
23 5     102      large  control
24 6     110      large  control
25 7      57      small  removal
26 8      71      small  removal
27 9      85      small  removal
28 10     105      large  removal
29 11     120      large  removal
30 >
31 > # cell means:
32 >
33 > tapply( height, list(initial.size, treatment), mean )
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1      control removal
2 large   99.25   112.5
3 small   53.50    71.0
4 >
5 > ## TABLE 2: SEQUENTIAL SUMS OF SQUARES
6 >
7 > # TABLE 2a: sequential analysis fitting treatment first:
8 > # reproduces published Shaw and Mitchell-Olds ANOVA table 2, top
9 >
10 > anova( txs.model <- lm( height ~ treatment * initial.size , data = Shaw.MitchellOlds ) )
11
12 Analysis of Variance Table
13
14 Response: height
15
16      Df Sum Sq Mean Sq F value    Pr(>F)
17 treatment      1    35.3     35.3  0.3309 0.5831478
18 initial.size    1 4846.0  4846.0 45.3658 0.0002686 ***
19 treatment:initial.size 1   11.4     11.4  0.1068 0.7533784
20 Residuals      7   747.8    106.8
21 ---
22 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
23 >
24 > # TABLE 2b: sequential analysis fitting size first:
25 >
26 > anova( sxt.model <- lm( height ~ initial.size * treatment, data = Shaw.MitchellOlds ) )
27
28 Analysis of Variance Table
29
30 Response: height
31
32      Df Sum Sq Mean Sq F value    Pr(>F)
33 initial.size    1 4291.2  4291.2 40.1718 0.0003896 ***
34 treatment      1   590.2    590.2  5.5249 0.0510495 .
35 initial.size:treatment 1   11.4     11.4  0.1068 0.7533784
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 Residuals          7  747.7  106.8
2 ---
3 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
4 >
5 > # total sums of squares:
6 >
7 > anova( sst <- lm( height ~ 1 , data = Shaw.MitchellOlds ) )
8
9 Analysis of Variance Table
10
11 Response: height
12      Df Sum Sq Mean Sq F value Pr(>F)
13 Residuals 10 5640.5    564.1
14 >
15 > # 5640.5
16 >
17 > # correlation of explanatory variables
18 > #main effects model, no interaction
19 >
20 > ts.model <- lm( height ~ treatment + initial.size , data = Shaw.MitchellOlds )
21 > summary(ts.model,cor=T)
22
23 Call:
24 lm(formula = height ~ treatment + initial.size, data = Shaw.MitchellOlds)
25
26 Residuals:
27      Min       1Q   Median       3Q      Max
28 -13.1053  -6.2105   0.8947   4.7895  14.8947
29
30 Coefficients:
31              Estimate Std. Error t value Pr(>|t|)
32 (Intercept)       98.58         4.47  22.055 1.89e-08 ***
33 treatmentremoval  15.26         6.12   2.494  0.0373 *
```

**HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL**

```

1  initial.sizesmall  -43.74      6.12  -7.146  9.75e-05  ***
2  ---
3  Signif. codes:  0  '***' 0.001  '**' 0.01  '*' 0.05  '.' 0.1  ' ' 1
4
5  Residual standard error: 9.741 on 8 degrees of freedom
6  Multiple R-squared:  0.8654,    Adjusted R-squared:  0.8318
7  F-statistic: 25.72 on 2 and 8 DF,  p-value: 0.0003281
8
9  Correlation of Coefficients:
10         (Intercept) treatmentremoval
11  treatmentremoval  -0.46
12  initial.sizesmall -0.46      -0.27
13
14  >
15  > ### TABLE 3: ADJUSTED SUMS OF SQUARES WITH HIGHER-LEVEL TERMS OMITTED
16  >
17  > # TABLE 3a: interaction (order of factors not important+)
18  >
19  > anova( txs.model<-lm( height ~ treatment * initial.size, data = Shaw.MitchellOlds ) )
20
21  Analysis of Variance Table
22
23  Response: height
24
25         Df Sum Sq Mean Sq F value    Pr(>F)
26  treatment      1   35.3    35.3  0.3309 0.5831478
27  initial.size    1 4846.0  4846.0 45.3658 0.0002686 ***
28  treatment:initial.size 1   11.4    11.4  0.1068 0.7533784
29  Residuals      7   747.8   106.8
30  ---
31  Signif. codes:  0  '***' 0.001  '**' 0.01  '*' 0.05  '.' 0.1  ' ' 1
32  >
33  > # TABLE 3b: main effect of treatment
34  >

```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 > anova( st.model<-lm( height ~ initial.size + treatment, data = Shaw.MitchellOlds ) )
2
3 Analysis of Variance Table
4
5 Response: height
6           Df Sum Sq Mean Sq F value    Pr(>F)
7 initial.size  1 4291.2   4291.2 45.2208 0.0001489 ***
8 treatment    1   590.2    590.2   6.2193 0.0372980 *
9 Residuals    8   759.2     94.9
10 ---
11 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
12 >
13 > # TABLE 3c: main effect of size
14 >
15 > anova( ts.model<-lm( height ~ treatment + initial.size, data = Shaw.MitchellOlds ) )
16
17 Analysis of Variance Table
18
19 Response: height
20           Df Sum Sq Mean Sq F value    Pr(>F)
21 treatment    1    35.3     35.3   0.3725   0.5586
22 initial.size  1 4846.0   4846.0 51.0676 9.746e-05 ***
23 Residuals    8   759.2     94.9
24 ---
25 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
26 >
27 > # TABLE 3d: composite table of adjusted sums of squares from the three sequential models
28 >
29 > adjusted.total <- 11.4 + 590.2 + 4846.0 + 747.8 ; adjusted.total
30 [1] 6195.4
31 >
32 > # [1] 6195.4 (Greater than SST from sequential model)
33 >
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 >
2 > ### TABLE 4: ADJUSTED SUMS OF SQUARES WITH HIGHER-LEVEL TERMS INCLUDED
3 >
4 > # type III tests by fitting the interaction as a third variable
5 > # code interaction as separate variable
6 >
7 > Shaw.MitchellOlds$size <- c( rep( 2, 2 ), rep( 1, 4 ), rep( 2, 3 ), rep( 1, 2 ) ) # numeric
8 > Shaw.MitchellOlds$treat <- c( rep( 1, 6 ), rep( 2, 5 ) ) # numeric
9 > Shaw.MitchellOlds$ts <- c( rep( 2, 2 ), rep( 1, 7 ), rep( 2, 2 ) )
10 > Shaw.MitchellOlds
11
12     height initial.size treatment size treat ts
13 1      50      small  control     2     1  2
14 2      57      small  control     2     1  2
15 3      91      large  control     1     1  1
16 4      94      large  control     1     1  1
17 5     102      large  control     1     1  1
18 6     110      large  control     1     1  1
19 7      57      small  removal     2     2  1
20 8      71      small  removal     2     2  1
21 9      85      small  removal     2     2  1
22 10     105      large  removal     1     2  2
23 11     120      large  removal     1     2  2
24 >
25 > # TABLE 4a: for interaction
26 >
27 > anova( tsi.model <- lm( height ~ treatment + initial.size + ts, data = Shaw.MitchellOlds ) )
28
29 Analysis of Variance Table
30
31 Response: height
32           Df Sum Sq Mean Sq F value    Pr(>F)
33 treatment  1   35.3    35.3  0.3309 0.5831478
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1  initial.size  1 4846.0  4846.0 45.3658 0.0002686 ***
2  ts           1   11.4   11.4  0.1068 0.7533784
3  Residuals    7  747.8   106.8
4  ---
5  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
6  >
7  > # equivalent to:
8  >
9  > anova( lm( height ~ treatment * initial.size, data = Shaw.MitchellOlds ) )
10
11 Analysis of Variance Table
12
13 Response: height
14              Df Sum Sq Mean Sq F value    Pr(>F)
15 treatment      1   35.3    35.3  0.3309 0.5831478
16 initial.size    1 4846.0  4846.0 45.3658 0.0002686 ***
17 treatment:initial.size 1   11.4    11.4  0.1068 0.7533784
18 Residuals      7  747.8   106.8
19 ---
20 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
21 >
22 > # TABLE 4b: main effect of treatment
23 >
24 > anova( ist.model <- lm( height ~ ts + initial.size + treatment , data = Shaw.MitchellOlds ) )
25
26 Analysis of Variance Table
27
28 Response: height
29              Df Sum Sq Mean Sq F value    Pr(>F)
30 ts           1   43.7    43.7  0.4090 0.5428396
31 initial.size  1 4251.9  4251.9 39.8039 0.0004007 ***
32 treatment    1  597.2   597.2  5.5906 0.0500130 .
33 Residuals    7  747.7   106.8
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 ---
2 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3 >
4 > # TABLE 4c: main effect of size
5 >
6 > anova( its.model <- lm( height ~ ts + treatment + initial.size , data = Shaw.MitchellOlds ) )
7
8 Analysis of Variance Table
9
10 Response: height
11           Df Sum Sq Mean Sq F value    Pr(>F)
12 ts         1   43.7    43.7  0.4090 0.5428396
13 treatment  1   41.2    41.2  0.3854 0.5543760
14 initial.size 1 4807.9  4807.9 45.0091 0.0002752 ***
15 Residuals   7   747.7   106.8
16 ---
17 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
18 >
19 > # TABLE 4d: composite table of adjusted sums of squares from the three sequential models
20 >
21 > ( adj.total <- 747.8 + 597.2 + 4807.9 + 11.4 )
22 [1] 6164.3
23 >
24 > # [1] 6164.3 (Greater than SST from sequential model)
25 >
26 > # correlation of explanatory variables
27 >
28 > txs.model <- lm( height ~ treatment*initial.size , data = Shaw.MitchellOlds )
29 > summary(txs.model,cor=T)
30
31 Call:
32 lm(formula = height ~ treatment * initial.size, data = Shaw.MitchellOlds)
33
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 Residuals:
2      Min       1Q   Median       3Q      Max
3 -1.400e+01 -6.375e+00  1.110e-16  5.500e+00  1.400e+01
4
5 Coefficients:
6             Estimate Std. Error t value Pr(>|t|)
7 (Intercept)      99.250      5.168  19.206 2.58e-07 ***
8 treatmentremoval    13.250      8.951   1.480  0.18233
9 initial.sizesmall  -45.750      8.951  -5.111  0.00138 **
10 treatmentremoval:initial.sizesmall  4.250     13.005   0.327  0.75338
11 ---
12 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
13
14 Residual standard error: 10.34 on 7 degrees of freedom
15 Multiple R-squared:  0.8674,    Adjusted R-squared:  0.8106
16 F-statistic: 15.27 on 3 and 7 DF,  p-value: 0.00187
17
18 Correlation of Coefficients:
19             (Intercept) treatmentremoval initial.sizesmall
20 treatmentremoval      -0.58
21 initial.sizesmall      -0.58      0.33
22 treatmentremoval:initial.sizesmall  0.40      -0.69      -0.69
23
24 >
25 >
26 > ### THE EXAMPLE DATASET REVISITED
27 > # Our favoured model: sequential model fitting initial size first
28 > # With and without non-significant interaction
29 >
30 > anova( sxt.model <- lm( height ~ initial.size * treatment, data = Shaw.MitchellOlds ) )
31
32 Analysis of Variance Table
33
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 Response: height
2           Df Sum Sq Mean Sq F value    Pr(>F)
3 initial.size      1 4291.2   4291.2 40.1718 0.0003896 ***
4 treatment         1  590.2    590.2  5.5249 0.0510495 .
5 initial.size:treatment 1   11.4     11.4  0.1068 0.7533784
6 Residuals         7   747.7    106.8
7 ---
8 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
9 >
10 > anova( s_t.model <- lm( height ~ initial.size + treatment, data = Shaw.MitchellOlds ) )
11
12 Analysis of Variance Table
13
14 Response: height
15           Df Sum Sq Mean Sq F value    Pr(>F)
16 initial.size      1 4291.2   4291.2 45.2208 0.0001489 ***
17 treatment         1  590.2    590.2  6.2193 0.0372980 *
18 Residuals         8   759.2     94.9
19 ---
20 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
21 >
22 >
23 > # initial size as a covariate
24 > # to adjust for differences in initial size before assessing the effect of the treatment
25 > # This analysis suggests:
26 > # no interaction between initial size and treatment
27 > # That initial size accounts for most of the differences in final size;
28 > # There is a treatment effect that is close to significant (at the conventional 5 % level).
29 > # The alternative sequential model fitting treatment first...
30 > # ...from Shaw and Mitchell-Olds, Ecology (1993)
31 >
32 > anova( txs.model <- lm( height ~ treatment*initial.size, data = Shaw.MitchellOlds) )
33
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 Analysis of Variance Table
2
3 Response: height
4           Df Sum Sq Mean Sq F value    Pr(>F)
5 treatment    1   35.3    35.3  0.3309 0.5831478
6 initial.size  1 4846.0  4846.0 45.3658 0.0002686 ***
7 treatment:initial.size  1   11.4    11.4  0.1068 0.7533784
8 Residuals    7   747.8   106.8
9 ---
10 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
11 >
12 > # ...does not adjust for differences in initial size and suggests no effect of treatment.
13 > # Overall, in this case - where there is no interaction -
14 > # the type III and II sum of squares produce a similar result to our a priori type I sequence.
15 >
16 >
17 > ### Additional material
18 > # model matrix with treatment contrasts
19 > # treatment contrasts are default in R, but here we show how to set them for clarity
20 > # the first contrast argument gives the function to be used with unordered factors,
21 > # the second the function to be used with ordered factors.
22 >
23 > options( contrasts = c( "contr.treatment", "contr.poly" ) )
24 > anova( ts.model<-lm( height ~ treatment + initial.size, data = Shaw.MitchellOlds ) )
25
26 Analysis of Variance Table
27
28 Response: height
29           Df Sum Sq Mean Sq F value    Pr(>F)
30 treatment    1   35.3    35.3  0.3725  0.5586
31 initial.size  1 4846.0  4846.0 51.0676 9.746e-05 ***
32 Residuals    8   759.2    94.9
33 ---
```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2
3  > # main effect size
4
5  > model.matrix(ts.model)
6
7  (Intercept) treatmentremoval initial.sizesmall
8  1           1           0           1
9  2           1           0           1
10 3           1           0           0
11 4           1           0           0
12 5           1           0           0
13 6           1           0           0
14 7           1           1           1
15 8           1           1           1
16 9           1           1           1
17 10          1           1           0
18 11          1           1           0
19 attr(,"assign")
20 [1] 0 1 2
21 attr(,"contrasts")
22 attr(,"contrasts")$treatment
23 [1] "contr.treatment"
24
25 attr(,"contrasts")$initial.size
26 [1] "contr.treatment"
27
28 > #main effects model, no interaction
29
30 > txs.model <- lm( height ~ treatment*initial.size , data = Shaw.MitchellOlds )
31 > model.matrix(txs.model)
32
33 (Intercept) treatmentremoval initial.sizesmall treatmentremoval:initial.sizesmall
```

HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```

1      1      1      0      1      0
2      2      1      0      1      0
3      3      1      0      0      0
4      4      1      0      0      0
5      5      1      0      0      0
6      6      1      0      0      0
7      7      1      1      1      1
8      8      1      1      1      1
9      9      1      1      1      1
10     10     1      1      0      0
11     11     1      1      0      0
12     attr(,"assign")
13     [1] 0 1 2 3
14     attr(,"contrasts")
15     attr(,"contrasts")$treatment
16     [1] "contr.treatment"
17
18     attr(,"contrasts")$initial.size
19     [1] "contr.treatment"
20
21     > #model with interaction
22     >
23     > # model matrix with sum-to-zero contrasts
24     > options( contrasts = c( "contr.sum", "contr.poly" ) )
25     > # sum-to-zero contrasts
26     > txs.model <- lm( height ~ treatment*initial.size , data = Shaw.MitchellOlds )
27     > model.matrix(txs.model)
28
29     (Intercept) treatment1 initial.size1 treatment1:initial.size1
30     1      1      1      -1      -1
31     2      1      1      -1      -1
32     3      1      1      1      1
33     4      1      1      1      1

```

## HECTOR, VON FELTEN & SCHMID SUPPLEMENTARY MATERIAL

```
1 5 1 1 1 1
2 6 1 1 1 1
3 7 1 -1 -1 1
4 8 1 -1 -1 1
5 9 1 -1 -1 1
6 10 1 -1 1 -1
7 11 1 -1 1 -1
8 attr(,"assign")
9 [1] 0 1 2 3
10 attr(,"contrasts")
11 attr(,"contrasts")$treatment
12 [1] "contr.sum"
13
14 attr(,"contrasts")$initial.size
15 [1] "contr.sum"
16
17
```