



Tutorial 9.6a - Factorial ANOVA

14 Jan 2014

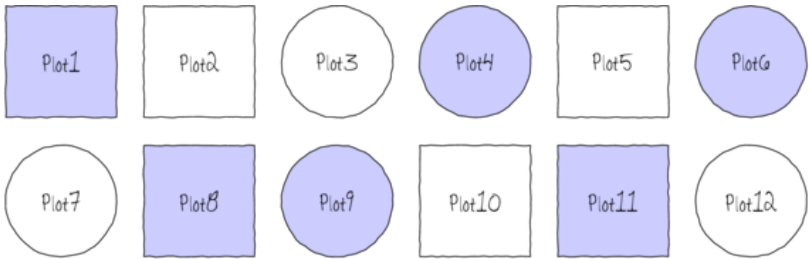
If you are completely ontop of the conceptual issues pertaining to factorial ANOVA, and just need to use this tutorial in order to learn about factorial ANOVA in R, you are invited to skip down to [the section on Factorial ANOVA in R](#).

Section 1: Overview

Factorial designs are an extension of single factor ANOVA designs in which additional factors are added such that each level of one factor is applied to all levels of the other factor(s) and these combinations are replicated.

For example, we might design an experiment in which the effects of temperature (high vs low) and fertilizer (added vs not added) on the growth rate of seedlings are investigated by growing seedlings under the different temperature and fertilizer combinations.

The following diagram depicts a very simple factorial design in which there are two factors (Shape and Color) each with two levels (Shape: square and circle; Color: blue and white) and each combination has 3 replicates.



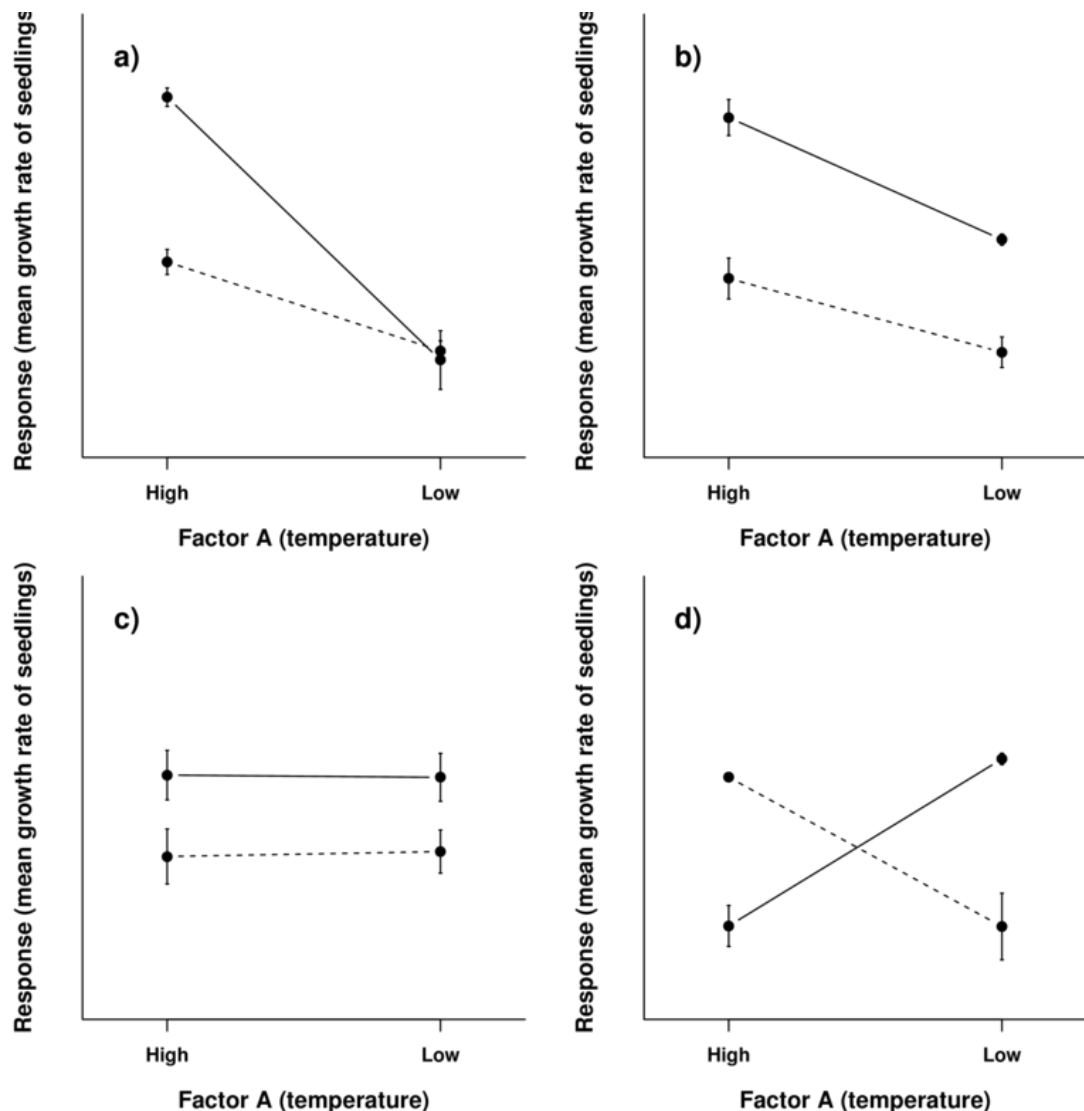
In addition to investigating the impacts of the main factors, factorial designs allow us to investigate whether the effects of one factor are consistent across levels of another factor. For example, is the effect of temperature on growth rate the same for both fertilized and unfertilized seedlings and similarly, does the impact of fertilizer treatment depend on the temperature under which the seedlings are grown?

Arguably, these interactions give more sophisticated insights into the dynamics of the system we are investigating. Hence, we could add additional main effects, such as soil pH, amount of water etc along with all the two way (temp:fert, temp:pH, temp:water, etc), three-way (temp:fert:pH, temp:pH:water), four-way etc interactions in order to explore how these various factors interact with one another to effect the response.

However, the more interactions, the more complex the model becomes to specify, compute and interpret - not to mention the rate at which the number of required observations increases.

To appreciate the interpretation of interactions, consider the following figure that depicts fictitious two factor (temperature and fertilizer) designs.

- Contents
 - Overview
 - Linear model
 - Null hypotheses
 - Analysis of Variance
 - Assumptions
 - Planned and unplanned comparisons
 - Unbalanced designs
 - ANOVA in R
 - Unbalanced designs



It is clear from the top-right figure that whether or not there is an observed effect of adding fertilizer or not depends on whether we are focused on seedlings growth under high or low temperatures. Fertilizer is only important for seedlings grown under high temperatures. In this case, it is not possible to simply state that there is an effect of fertilizer as it depends on the level of temperature. Similarly, the magnitude of the effect of temperature depends on whether fertilizer has been added or not.

Such interactions are represented by plots in which lines either intersect or converge. The top-right and bottom-left figures both depict parallel lines which are indicative of no interaction. That is, the effects of temperature are similar for both fertilizer added and controls and vice versa. Whilst the former displays an effect of both fertilizer and temperature, the latter only fertilizer is important.

Finally, the bottom-right figure represents a strong interaction that would mask the main effects of temperature and fertilizer (since the nature of the effect of temperature is very different for the different fertilizer treatments and visa verse).

Factorial designs can consist:

- entirely of crossed [fixed factors](#) (**Model I ANOVA** - most common) in which conclusions are restricted to the specific combinations of levels selected for the experiment,
- entirely of crossed random factors (**Model II ANOVA**) or
- a mixture of crossed fixed and random factors (**Model III ANOVA**).

The latter are useful for investigating the generality of a main treatment effect (fixed) over broad spatial, temporal or biological levels of organization. That is whether the observed effects of temperature and/or fertilizer (for example) are observed across the entire genera or country.

Section 2: Linear model

As with single factor ANOVA, the linear model could be constructed as either effects or cell means parameterization, although only effects parameterization will be considered here (as cell means parameterization is of little use in frequentist factorial ANOVA).

The linear models for two and three factor design are:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

where μ is the overall mean, α is the effect of Factor A, β is the effect of Factor B, γ is the effect of Factor C and ε is the random unexplained or residual component.

Note that although the linear models for Model I, Model II and Model III designs are identical, the interpretation of terms (and thus null hypothesis) differ.

Recall from the tutorial on single factor ANOVA, that categorical variables in linear models are actually re-parameterized dummy codes - and thus the α term above, actually represents one or more dummy codes. Thus, if we actually had two levels

of Factor A (A1 and A2) and three levels of Factor B (B1, B2, B3), then the fully parameterized linear model would be:

$$y = \beta_{A1B1} + \beta_{A2B1-A1B1} + \beta_{A1B2-A1B1} + \beta_{A1B3-A1B1} + \beta_{A2B2-A1B2-A2B1-A1B1} + \beta_{A2B3-A1B3-A2B1-A1B1}$$

Thus, such a model would have six parameters to estimate (in addition to the variance).

Section 3: Null hypotheses

There are separate null hypothesis associated with each of the main effects and the interaction terms.

3.1: Model 1 - fixed effects

3.1.1: Factor A

$H_0(A): \mu_1 = \mu_2 = \dots = \mu_i = \mu$ (the population group means are all equal).

The mean of population 1 is equal to that of population 2 and so on, and thus all population means are equal to an overall mean. If the effect of the i^{th} group is the difference between the i^{th} group mean and the overall mean ($\alpha_i = \mu_i - \mu$) then the H_0 can alternatively be written as:

$H_0(A): \alpha_1 = \alpha_2 = \dots = \alpha_i = 0$ (the effect of each group equals zero).

If one or more of the α_i are different from zero (the response mean for this treatment differs from the overall response mean), the null hypothesis is rejected indicating that the treatment has been found to affect the response variable.

Note, as with multiple regression models, these "effects" represent *partial effects*. In the above, the effect of Factor A is actually the effect of Factor A **at the first level of the Factor(s)**.

3.1.2: Factor B

$H_0(B): \mu_1 = \mu_2 = \dots = \mu_i = \mu$ (the population group means are all equal - at the first level of Factor A).

Equivalent interpretation to Factor A above.

3.1.3: A: Interaction

$H_0(AB): \mu_{ij} = \mu_i + \mu_j - \mu$ (the population group means are all equal).

For any given combination of factor levels, the population group mean will be equal to the difference between the overall population mean and the simple additive effects of the individual factor group means. That is, the effects of the main treatment factors are purely additive and independent of one another. This is equivalent to $H_0(AB): \alpha\beta_{ij} = 0$, no interaction between Factor A and Factor B.

3.2: Model 2 - random effects

3.2.1: Factor A

$H_0(A): \sigma_a^2 = 0$ (population variance equals zero)

There is no added variance due to all possible levels of A.

3.2.2: Factor B

$H_0(B): \sigma_a^2 = 0$ (population variance equals zero)

There is no added variance due to all possible levels of B.

3.2.3: A:B Interaction

$H_0(AB): \sigma_{a\beta}^2 = 0$ (population variance equals zero)

There is no added variance due to all possible interactions between all possible levels of A and B.

3.3: Model 3 - mixed effects

3.3.1: Fixed factor - e.g. A

$H_0(A): \mu_1 = \mu_2 = \dots = \mu_i = \mu$ (the population group means are all equal)

The mean of population 1 (pooled over all levels of the random factor) is equal to that of population 2 and so on, and thus all population means are equal to an overall mean pooling over all possible levels of the random factor. If the effect of the i^{th} group is the difference between the i^{th} group mean and the overall mean ($\alpha_i = \mu_i - \mu$) then the H_0 can alternatively be written as:

$H_0(A): \alpha_1 = \alpha_2 = \dots = \alpha_i = 0$ (no effect of any level of this factor pooled over all possible levels of the random factor)

3.3.2: Random factor - e.g. B

$H_0(B): \sigma_a^2 = 0$ (population variance equals zero)

There is no added variance due to all possible levels of B.

3.3.3: A:B Interaction

The interaction of a fixed and random factor is always considered a random factor.

$H_0(AB): \sigma_{a\beta}^2 = 0$ (population variance equals zero)

There is no added variance due to all possible interactions between all possible levels of A and B.

Section 4: Analysis of Variance

When fixed factorial designs are balanced, the total variance in the response variable can be sequentially partitioned into what is explained by each of the model terms (factors and their interactions) and what is left unexplained. For each of the specific null hypotheses, the overall unexplained variability is used as the denominator in F -ratio calculations (see tables below), and when a null hypothesis is true, an F -ratio should follow an F distribution with an expected value less than 1.

Random factors are added to provide greater generality of conclusions. That is, to enable us to make conclusions about the effect of one factor (such as whether or not fertilizer is added) over all possible levels (not just those sampled) of a random factor (such as all possible locations, seasons, varieties etc). In order to expand our conclusions beyond the specific levels used in the design, the hypothesis tests (and thus *F*-ratios) must reflect this extra generality by being more conservative.

The appropriate *F*-ratios for fixed, random and mixed factorial designs are presented in tables below. Generally, once the terms (factors and interactions) have been ordered into a hierarchy (single factors at the top, highest level interactions at the bottom and terms of same order given equivalent positions in the hierarchy), the denominator for any term is selected as the next appropriate random term (an interaction that includes the term to be tested) encountered lower in the hierarchy. Interaction terms that contain one or more random factors are considered themselves to be random terms, as is the overall residual term (as all observations are assumed to be random representations of the entire population(s)).

Note, when designs include a mixture of fixed and random crossed effects, exact denominator degrees of freedoms for certain *F*-ratios are undefined and traditional approaches adopt rather inexact estimated approximate or "Quasi" *F*-ratios.

Pooling of non-significant *F*-ratio denominator terms, in which lower random terms are added to the denominator (provided $\alpha > 0.25$), may also be useful.

For random factors within mixed models, selecting *F*-ratio denominators that are appropriate for the intended hypothesis tests is a particularly complex and controversial issue. Traditionally, there are two alternative approaches and whilst the statistical resumes of each are complicated, essentially they differ in whether or not the interaction term is constrained for the test of the random factor.

The constrained or restricted method (**Model I**), stipulates that for the calculation of a random factor *F*-ratio (which investigates the added variance added due to the random factor), the overall effect of the interaction is treated as zero. Consequently, the random factor is tested against the residual term (see tables below).

The unconstrained or unrestrained method (**Model II**) however, does not set the interaction effect to zero and therefore the interaction term is used as the random factor *F*-ratio denominator (see tables below). This method assumes that the interaction terms for each level of the random factor are completely independent (correlations between the fixed factor must be consistent across all levels of the random factor).

Some statisticians maintain that the independence of the interaction term is difficult to assess for biological data and therefore, the restricted approach is more appropriate. However, others have suggested that the restricted method is only appropriate for balanced designs.

4.1: Quasi *F*-ratios

An additional complication for three or more factor models that contain two or more random factors, is that there may not be a single appropriate interaction term to use as the denominator for many of the main effects *F*-ratios. For example, if Factors A and B are random and C is fixed, then there are two random interaction terms of equivalent level under Factor C ($A' \times C$ and $B' \times C$).

As a result, the value of the of the Mean Squares expected when the null hypothesis is true cannot be easily defined. The solutions for dealing with such situations (quasi *F*-ratios involve adding (and subtracting) terms together to create approximate estimates of *F*-ratio denominators. Alternatively, for random factors, variance components with confidence intervals can be used.)

These solutions are sufficiently unsatisfying as to lead many biostatisticians to recommend that factorial designs with two or more random factors should avoided if possible. Arguably however, [linear mixed effects models](#) offer more appropriate solutions to the above issues as they are more robust for unbalanced designs, accommodate covariates and provide a more comprehensive treatment and overview of all the underlying data structures.

Factor	d.f	MS	A,B fixed	A,B random	A fixed, B random	
					Restricted	Unrestricted
A	$a - 1$	MS_A	MS_A/MS_{Resid}	$MS_A/MS_{A:B}$	$MS_A/MS_{A:B}$	$MS_A/MS_{A:B}$
B	$b - 1$	MS_B	MS_B/MS_{Resid}	$MS_B/MS_{A:B}$	MS_B/MS_{Resid}	$MS_B/MS_{A:B}$
A:B	$(b - 1)(a - 1)$	$MS_{A:B}$	$MS_{A:B}/MS_{Resid}$	$MS_{A:B}/MS_{A:B}$	$MS_{A:B}/MS_{Resid}$	$MS_{A:B}/MS_{Resid}$
Residual (= $N'(B : A)$)	$(n - 1)ba$	MS_{Resid}				
R syntax						
Type I SS (Balanced)			<pre>> summary(lm(y ~ A * B, data)) > anova(lm(y ~ A * B, data))</pre>			
Type II SS (Unbalanced)			<pre>> summary(lm(y ~ A * B, data)) > library(car) > Anova(lm(y ~ A * B, data), type = "II")</pre>			
Type III SS (Unbalanced)			<pre>> summary(lm(y ~ A * B, data)) > library(car) > contrasts(data\$A) <- contr.sum > contrasts(data\$B) <- contr.sum > Anova(lm(y ~ A * B, data), type = "III")</pre>			
Variance components			<pre>> library(lme4) > summary(lmer(y ~ 1 + (1 A) + (1 B) + (1 A:B), + data))</pre>			

4.2: Interactions and main effects tests

Note that for fixed factor models, when null hypotheses of interactions are rejected, the null hypothesis of the individual constituent factors are unlikely to represent the true nature of the effects and thus are of little value.

The nature of such interactions are further explored by fitting simpler linear models (containing at least one less factor) separately for each of the levels of the other removed factor(s). Such **Main effects tests** are based on a subset of the data, and therefore estimates of the overall residual (unexplained) variability are unlikely to be as precise as the estimates based on the global model. Consequently, *F*-ratios involving MS_{Resid} should use the estimate of MS_{Resid} from the global model rather than that based on the smaller, theoretically less precise subset of data.

For random and mixed models, since the objective is to generalize the effect of one factor *over and above* any interactions with other factors, the main factor effects can be interpreted even in the presence of significant interactions. Nevertheless, it should be noted that when a significant interaction is present in a mixed model, the power of the main fixed effects will be reduced

(since the amount of variability explained by the interaction term will be relatively high, and this term is used as the denominator for the F -ratio calculation).

Section 5: Assumptions

Hypothesis tests assume that the residuals are:

- normally distributed. Boxplots using the appropriate scale of replication (reflecting the appropriate residuals/ F -ratio denominator (see table above) should be used to explore normality. Scale transformations are often useful.
- equally varied. Boxplots and plots of means against variance (using the appropriate scale of replication) should be used to explore the spread of values. Residual plots should reveal no patterns. Scale transformations are often useful.
- independent of one another.

Section 6: Planned and unplanned comparisons

As with single factor analysis of variance, [planned and unplanned multiple comparisons](#) (such as Tukey's test) can be incorporated into or follow the linear model respectively so as to further investigate any patterns or trends within the main factors and/or the interactions. As with single factor analysis of variance, the contrasts must be defined prior to fitting the linear model, and no more than $p - 1$ (where p is the number of levels of the factor) contrasts can be defined for a factor.

Section 7: Unbalanced designs

A factorial design can be thought of as a table made up of rows (representing the levels of one factor), columns (levels of another factor) and cells (the individual combinations of the set of factors), see the series of tables below. The top two tables depict a balanced two factor (3x3) design in which each cell (combination of factor levels) has three replicate observation.

Whilst the middle left table does not have equal sample sizes in each cell, the sample sizes are in proportion and as such, does not present the issues discussed below for unbalanced designs.

Tables middle-right and bottom-left are considered unbalanced as is the bottom-right (which is a special case in which the entire cell is missing).

Cell means structure				Balanced design (3 replicates)			
	High	Medium	Low		High	Medium	Low
Full shade	μ_{FH}	μ_{FM}	μ_{FL}	Full shade	XXX	XXX	XXX
Partial shade	μ_{PH}	μ_{PM}	μ_{PL}	Partial shade	XXX	XXX	XXX
Control	μ_{CH}	μ_{CM}	μ_{CL}	Control	XXX	XXX	XXX

Proportionally balanced design (2-3 replicates)				Unbalanced design (2-3 replicates)			
	High	Medium	Low		High	Medium	Low
Full shade	XXX	XXX	XXX	Full shade	XX	XXX	XXX
Partial shade	XX	XX	XX	Partial shade	XXX	XXX	XXX
Control	XXX	XXX	XXX	Control	XXX	XXX	XXX

Unbalanced design (1-3 replicates)				Missing cells design (3 replicates)			
	High	Medium	Low		High	Medium	Low
Full shade	XX	XXX	XXX	Full shade	XXX	XXX	
Partial shade	XXX	X	XX	Partial shade	XXX	XXX	XXX
Control	XXX	XXX	XXX	Control	XXX	XXX	XXX

7.1: Missing observations

In addition to impacting on normality and homogeneity of variance, unequal sample sizes in factorial designs have major implications for the partitioning of the total sums of squares into each of the model components.

For balanced designs, the total sums of squares (SS_{Total}) is equal to the additive sums of squares of each of the components (including the residual). For example, in a two factor balanced design, $SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Resid}$. This can be represented diagrammatically by a Venn Diagram in which each of the SS for the term components butt against one another and are surrounded by the SS_{Resid} (see the top-right figure below).

However, in unbalanced designs, the sums of squares will be non-orthogonal and the sum of the individual components does not add up to the total sums of squares. Diagrammatically, the SS of the terms intersect or are separated (see~Figure~??b and ??g respectively).

In regular **sequential sums of squares (Type I SS)**, the sum of the individual sums of squares must be equal to the total sums of squares, the sums of squares of the last factor to be estimated will be calculated as the difference between the total sums of squares and what has already been accounted for by other components.

Consequently, the order in which factors are specified in the model (and thus estimated) will alter their sums of squares and therefore their F -ratios (see~Figure~??c-d).

To overcome this problem, traditionally there are two other alternative methods of calculating sums of squares.

- **Type II (hierarchical) SS** estimate the sums of squares of each term as the improvement it contributes upon addition of that term to a model of greater complexity and lower in the hierarchy (recall that the hierarchical structure descends from the simplest model down to the fully populated model). The SS for the interaction as well as the first factor to be estimated are the same as for Type I SS . Type II SS estimate the contribution of a factor over and above the contributions of other factors of equal or lower complexity but not above the contributions of the interaction terms or terms nested within the factor. However, these sums of squares are weighted by the sample sizes of each level and therefore are biased towards the trends produced by the groups (levels) that have higher sample sizes. As a result of the weightings, Type II SS actually test hypotheses about really quite complex combinations of factor levels. Rather than test a hypothesis that $\mu_{High} = \mu_{Medium} = \mu_{Low}$, Type II SS might be testing that $4 \times \mu_{High} = 1 \times \mu_{Medium} = 0.25 \times \mu_{Low}$.
- **Type III (marginal or orthogonal) SS** estimate the sums of squares of each term as the improvement based on a comparison of models with and without the term and are unweighted by sample sizes. Type III SS essentially measure just the unique contribution of each factor over and above the contributions of the other factors and interactions. For

unbalanced designs, Type III SS essentially test equivalent hypotheses to balanced Type I SS and are therefore arguably more appropriate for unbalanced factorial designs than Type II SS. Importantly, Type III SS are only interpretable if they are based on orthogonal contrasts (such as sum or helmert contrasts and not treatment contrasts).

The choice between Type II and III SS clearly depends on the nature of the question. For example, if we had measured the growth rate of seedlings subjected to two factors (temperature and fertilizer), Type II SS could address whether there was an effect of temperature across the level of fertilizer treatment, whereas Type III SS could assess whether there was an effect of temperature within each level of the fertilizer treatment.

7.2: Missing cells

Section 8: ANOVA in R

8.1: Scenario and Data

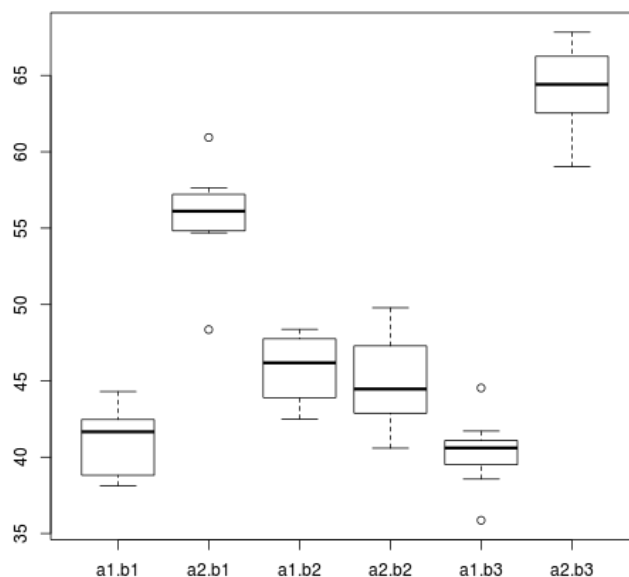
Imagine we have designed an experiment in which we had measured the response (y) under a combination of two different potential influences (Factor A: levels $a1$ and $a2$; and Factor B: levels $b1$, $b2$ and $b3$), each combination replicated 10 times ($n = 10$). As this section is mainly about the generation of artificial data (and not specifically about what to do with the data), understanding the actual details are optional and can be safely skipped. Consequently, I have folded (toggled) this section away.

[Details of data generation](#)

8.2: Exploratory data analysis

8.2.1: Normality and Homogeneity of variance

```
> boxplot(y ~ A * B, data)
```



Conclusions:

- there is no evidence that the response variable is consistently non-normal across all populations - each boxplot is approximately symmetrical
- there is no evidence that variance (as estimated by the height of the boxplots) differs between the five populations. . More importantly, there is no evidence of a relationship between mean and variance - the height of boxplots does not increase with increasing position along the y-axis. Hence there is no evidence of non-homogeneity

Obvious violations could be addressed either by:

- transform the scale of the response variables (to address normality etc). Note transformations should be applied to the entire response variable (not just those populations that are skewed).

8.3: Model fitting or statistical analysis

In R, simple ANOVA models (being regression models) are fit using the `lm()` function. The most important (=commonly used) parameters/arguments for simple regression are:

- formula:** the linear model relating the response variable to the linear predictor
- data:** the data frame containing the data

```
> data.lm <- lm(y ~ A * B, data)
```

8.4: Model evaluation

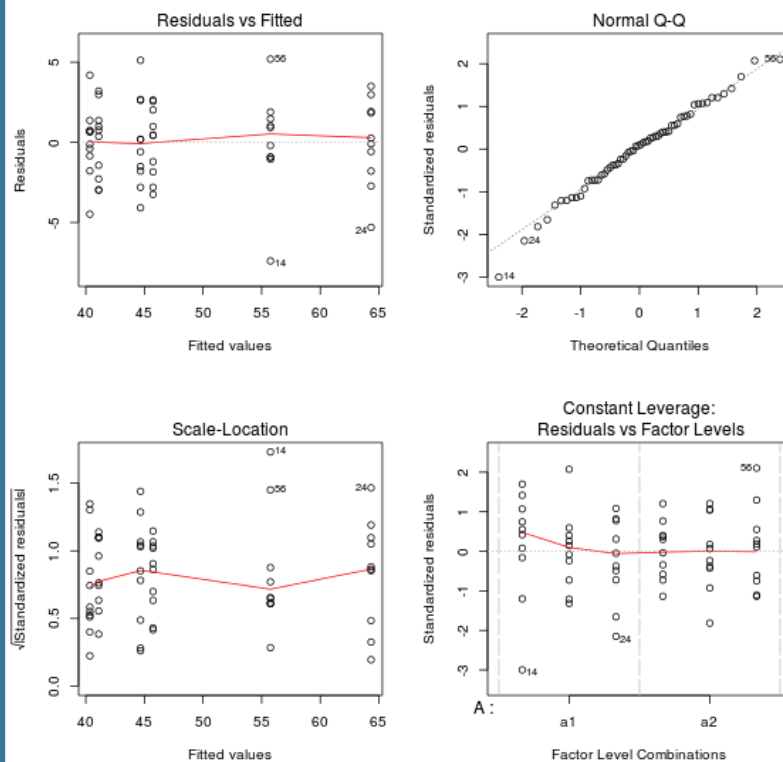
Prior to exploring the model parameters, it is prudent to confirm that the model did indeed fit the assumptions and was an appropriate fit to the data.

Whilst ANOVA models are multiple regression models, some of the regression usual diagnostics are not relevant. For example, an observation cannot logically be an outlier in x-space - it must be in one of the categories. In this case, each observation comes from one of five treatment groups. No observation can be outside of these categories and thus there are no opportunities for outliers in x-space. Consequently, Leverage (and thus Cook's D) diagnostics have no real meaning in ANOVA.

8.4.1: Residuals

As the residuals are the differences between the observed and predicted values along a vertical plane, they provide a measure of how much of an outlier each point is in y-space (on y-axis). Outliers are identified by relatively large residual values. Residuals can also be standardized and studentized, the latter of which can be compared across different models and follow a t distribution enabling the probability of obtaining a given residual can be determined. The patterns of residuals against predicted y values (residual plot) are also useful diagnostic tools for investigating linearity and homogeneity of variance assumptions.

```
> par(mfrow = c(2, 2))
> plot(data.lm)
```



Conclusions:

- there are no obvious patterns in the residuals (top-left plot). Specifically, there is no obvious 'wedge-shape' suggesting that we still have no evidence of non-homogeneity of variance
- Whilst the data do appear to deviate slightly from normality according to the Q-Q normal plot, it is only a mild deviation and given the robustness of linear models to mild non-normality, the Q-Q normal plot is really only of interest when there is also a wedge in the residuals (as it could suggest that non-homogeneity of variance could be due to non-normality and that perhaps normalizing could address both issues).
- The lower-left figure is a re-representation of the residual plot (yet with standardized residuals - hence making it more useful when the model has applied alternate variance structures) and therefore don't provide any additional insights if the default variance structure is applied.
- The lower-right figure represents the residuals for each combination of levels of the two factors. In examining this figure, we are looking to see whether the spread of the residuals is similar for each factor combination - which it is.

8.4.2: Extractor functions

There are a number of **extractor functions** (functions that extract or derive specific information from a model) available including:

Extractor	Description
<code>residuals()</code>	Extracts the residuals from the model
<code>fitted()</code>	Extracts the predicted (expected) response values (on the link scale) at the observed levels of the linear predictor
<code>predict()</code>	Extracts the predicted (expected) response values (on either the link, response or terms (linear predictor) scale)
<code>coef()</code>	Extracts the model coefficients
<code>confint()</code>	Calculate confidence intervals for the model coefficients
<code>summary()</code>	Summarizes the important output and characteristics of the model
<code>anova()</code>	Computes an analysis of variance (variance partitioning) from the model
<code>plot()</code>	Generates a series of diagnostic plots from the model
<code>effect()</code>	effects package - estimates the marginal (partial) effects of a factor (useful for plotting)
<code>avPlot()</code>	car package - generates partial regression plots

8.5: Exploring the model parameters, test hypotheses

If there was any evidence that the assumptions had been violated, then we would need to reconsider the model and start the process again. In this case, there is no evidence that the test will be unreliable so we can proceed to explore the test statistics. There are two broad hypothesis testing perspectives we could take:

- We could explore the individual effects. By default, the `lm()` function dummies codes the categorical variable for treatment contrasts (where each of the groups are compared to the first group). These parameter estimates are best viewed in conjunction with an interaction plot.

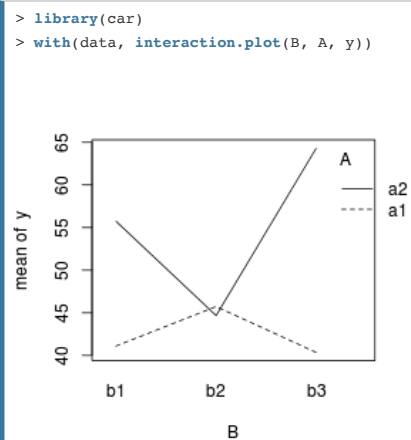
```
> summary(data.lm)

Call:
lm(formula = y ~ A * B, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-7.394 -1.575  0.228  1.558  5.191

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  41.099     0.822   50.01 < 2e-16 ***
Aa2          14.651     1.162   12.61 < 2e-16 ***
Bb2           4.639     1.162    3.99  0.0002 ***
Bb3          -0.752     1.162   -0.65  0.5202
Aa2:Bb2      -15.718     1.644   -9.56  3.2e-13 ***
Aa2:Bb3       9.335     1.644    5.68  5.5e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.6 on 54 degrees of freedom
Multiple R-squared:  0.925,    Adjusted R-squared:  0.918
F-statistic: 132 on 5 and 54 DF,  p-value: <2e-16
```



Conclusions:

- The row in the Coefficients table labeled (Intercept) represents the first combination of treatment groups. So the Estimate for this row represents the estimated population mean for a1b1. Hence, the mean of population a1b1 (Factor A level 1 and Factor B level 1) is estimated to be 41.099 with a standard error (precision) of 0.822. Note that the rest of the columns (t value and Pr(>|t|)) represent the outcome of hypothesis tests. For the first row, this hypothesis test has very little meaning as it is testing the null hypothesis that the mean of population A is zero.
- Each of the subsequent rows represent the estimated effects of each of the other population combinations from population a1b1. For example, the Estimate for the row labeled Aa2 represents the estimated **difference** in means between a2b1 and the reference group (a1b1). Hence, the population mean of a2b1 is estimated to be 14.651 units higher than the mean of a1b1 and this 'effect' is statistically significant (at the 0.05 level).

This can be seen visually in the interaction plot, where the mean of a1b1 is approximately 41 and the mean of a2b1 is approximately 55.5 (i.e. approx. 14.5 units higher).

- Similarly, a1b2 and a1b3 represent the differences in their respective means (approx 4.639 and -0.752) from the mean on a1b1.
- So far, those estimates indicate that the "effect" of Factor B is that the response increases slightly between b1 and b2 before returning down again by b3. The "effect" of Factor A suggests that a2 is approximately 15 units greater than a1. However, these are **partial effects**. It is clear, that the patterns described between b1, b2 and b3 are only appropriate for a1. The patterns between b1, b2 and b3 are very different for b2.
- The last two parameters (rows of the table), indicate the deviations of the patterns in additional levels of each factor are from the partial effects suggested by the main terms (A and B). The item Aa2:Bb2 is the deviation of mean a2b2 from what would be expected based on the deviations of a2b1 and a1b2 from a1b1. For example, we know that a1b2 is 4.639 units higher than a1b1 and we know that a2b1 is 14.651 units higher than a1b1. So if the model was purely additive (no interaction), we would expect the mean of a2b2 would be approximately 60.389. And Aa2:Bb2 would equal 0.

The value of Aa2:Bb2 of -15.718 indicates that a2b2 is considerably less than 0. So interaction terms, the effects of b2-b1 are considerably stronger (larger magnitude) at a2 than they are at a1.

- The R^2 value is 0.92 suggesting that approximately 92% of the total variance in y can be explained by the multiplicative model. The R^2 value is thus a measure of the strength of the relationship.

Alternatively, we could calculate the 95% confidence intervals for the parameter estimates.

```
> confint(data.lm)

            2.5 %    97.5 %
(Intercept) 39.451  42.746
Aa2         12.321  16.982
Bb2          2.308   6.969
Bb3         -3.082   1.578
Aa2:Bb2     -19.014 -12.423
Aa2:Bb3       6.040  12.630
```

- In addition (or alternatively), we could explore the collective null hypothesis (that all the population means are equal) by partitioning the variance into explained and unexplained components (in an ANOVA table). Since the design was balanced, we can use regular Type I SS.

```
> anova(data.lm)

Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)
A       1  2353    2353   348.4 < 2e-16 ***
B       2   511     255    37.8 5.4e-11 ***
A:B     2  1603     802   118.7 < 2e-16 ***
Residuals 54    365       7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusions:

- When an interaction term is present, it is important that these term(s) be looked at first (starting with the highest level interaction term). In this case, there is a single interaction term (A:B). There is clear inferential evidence of an interaction between Factor A and B, suggesting (as we have already seen) that the effect on Factor A on the response is not consistent across all levels of Factor B and vice versa.

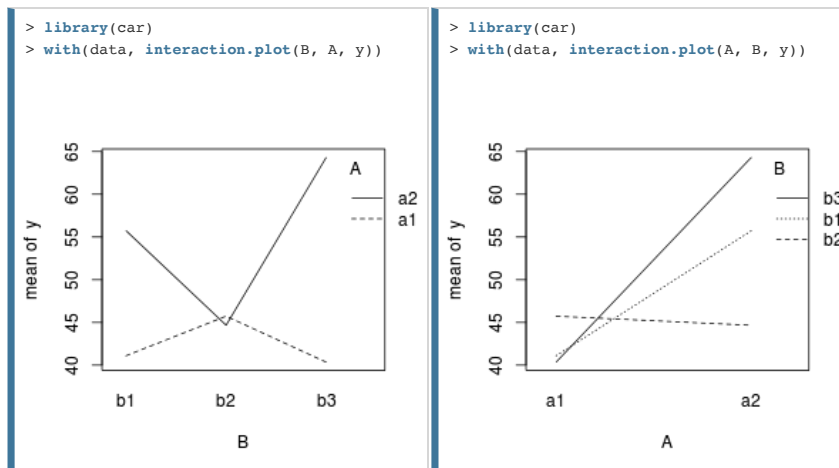
It should be fairly evident that making any conclusions about the main effects (in the presence of a significant interaction), is misleading and well short of the full complexity of the situation.

- If the interaction term was non-significant, then we could proceed to draw conclusions about the main effects of Factor A and B as we would then effectively have an additive model (no interaction). On that note, some argue that if the interaction term has a p -value > 0.25 , then the term can be removed from the model. The interaction term uses up (in this case) 2 degrees of freedom (that would otherwise be in the residuals). If the interaction term is not explaining much (and thus absorbing unexplained variability from the residuals, whilst incurring a cost of degrees of freedom), then the tests of main effects will be more powerful if the interaction is removed.

8.6: Dealing with interactions

In the current working example, we have identified that there is a significant interaction between Factor A and Factor B. Our exploration of the regression coefficients (from the `summary()` function, indicated that the pattern between b_1 , b_2 and b_3 might differ between a_1 and a_2 .

Similarly, if we consider the coefficients from the perspective of Factor A (and again an interaction plot might help), we can see that the patterns between a_1 and a_2 are similar for b_1 and b_3 , yet very different for b_2 .



At this point, we can then split the two-factor model up into a series of single-factor ANOVA models, either:

- examining the effects of Factor B separately for each level of Factor A (two single-factor ANOVA'S) or
- examining the effects of Factor A separately for each level of Factor B (three single-factor ANOVA'S)

When doing so, it is often considered appropriate to use the global residual term as the residual term. Hence we run the single-factor ANOVA's (main effects tests) and then modify the residual term as well as the F -ratio and P -value to reflect this.

There are three ways of performing such **simple main effects tests**:

- Perform simple one-way ANOVA's on subsets of data and manually adjust the F -ratios and P values
- Refit the model including a new modified interaction term designed to capture the other levels of one of the factors. This option can then generate the new regression coefficients as well as the ANOVA table.
- Apply a specific contrast. This approach only generates the new regression coefficients (and associated inference tests)

Manual adjustments

I will only demonstrate exploring the effects of Factor A at b_1 and b_2 as an example.

```
> ## Examine the effects of Factor A at level b1 of Factor B
> summary(lm(y ~ A, data = subset(data, B == "b1")))
Call:
lm(formula = y ~ A, data = subset(data, B == "b1"))

Residuals:
    Min       1Q   Median       3Q      Max
-7.394 -1.153  0.562  1.390  5.191

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  41.099     0.881    46.6   <2e-16 ***
Aa2          14.651     1.246    11.8   7e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.79 on 18 degrees of freedom
Multiple R-squared:  0.885,    Adjusted R-squared:  0.878
F-statistic: 138 on 1 and 18 DF,  p-value: 6.98e-10

> anova(lm(y ~ A, data = subset(data, B == "b1")))
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value Pr(>F)
A         1   1073    1073    138 7e-10 ***
Residuals 18    140         8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> ## Use global MSresid instead
> (MS_A <- anova(lm(y ~ A, data = subset(data, B == "b1")))$"Mean Sq"[1 ])
[1] 1073

> (DF_A <- anova(lm(y ~ A, data = subset(data, B == "b1")))$Df[1 ])
[1] 1
```

```

> (MS_resid <- anova(data.lm)$"Mean Sq"[4 ])
[1] 6.754

> (DF_resid <- anova(data.lm)$Df[4 ])
[1] 54

> (F_ratio <- MS_A/MS_resid)
[1] 158.9

> (P_value <- 1 - pf(F_ratio, DF_A, DF_resid))
[1] 0

>
>
> ## Examine the effects of Factor A at level b2 of Factor B
> summary(lm(y ~ A, data = subset(data, B == "b2")))
Call:
lm(formula = y ~ A, data = subset(data, B == "b2"))

Residuals:
    Min       1Q   Median       3Q      Max
-4.083 -1.805  0.181  2.155  5.115

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  45.737      0.792   57.78  <2e-16 ***
Aa2          -1.067      1.120   -0.95   0.35
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.5 on 18 degrees of freedom
Multiple R-squared:  0.048,    Adjusted R-squared:  -0.00486
F-statistic: 0.908 on 1 and 18 DF,  p-value: 0.353

> anova(lm(y ~ A, data = subset(data, B == "b2")))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value Pr(>F)
A           1    5.7    5.69    0.91  0.35
Residuals 18  112.8    6.27
---
> ## Use global MSresid instead
> (MS_A <- anova(lm(y ~ A, data = subset(data, B == "b2")))$"Mean Sq"[1 ])
[1] 5.69

> (DF_A <- anova(lm(y ~ A, data = subset(data, B == "b2")))$Df[1 ])
[1] 1

> (MS_resid <- anova(data.lm)$"Mean Sq"[4 ])
[1] 6.754

> (DF_resid <- anova(data.lm)$Df[4 ])
[1] 54

> (F_ratio <- MS_A/MS_resid)
[1] 0.8425

> (P_value <- 1 - pf(F_ratio, DF_A, DF_resid))
[1] 0.3628

```

There is a clear evidence of an effect of Factor A at level b1, yet no evidence of an effect of Factor A at b2.

Refit the model with modified interaction term

```

> data <- within(data, {
+   A2 = ifelse(B == "b1", "a1", as.character(A))
+ })
> data <- within(data, {
+   AB = interaction(A2, B)
+ })
> anova(update(data.lm, ~AB + .))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value Pr(>F)
AB           4   3393     848    126 <2e-16 ***
A            1   1073     1073    159 <2e-16 ***
Residuals 54     365         7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>
> data <- within(data, {
+   A2 = ifelse(B == "b2", "a1", as.character(A))
+ })
> data <- within(data, {
+   AB = interaction(A2, B)
+ })
> anova(update(data.lm, ~AB + .))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value Pr(>F)

```

```

AB          4    4461    1115    165.13 <2e-16 ***
A           1         6         6      0.84    0.36
Residuals 54     365         7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Same conclusion as above.

Apply specific contrasts

This process will perform three contrasts - the effect of Factor A (a2-a1) at each level of Factor B.

```

> library(contrast)
> contrast(data.lm, a = list(A = "a2", B = levels(data$B)), b = list(A = "a1", B = levels(data$B)))
lm model parameter contrast

Contrast S.E. Lower Upper t df Pr(>|t|)
14.651 1.162 12.321 16.982 12.61 54 0.0000
-1.067 1.162 -3.397 1.263 -0.92 54 0.3628
23.987 1.162 21.657 26.317 20.64 54 0.0000

```

Factor A level a2 is significantly higher than a1 at levels b1 and b3, yet not different at level b2.

Section 9: Unbalanced designs

As mentioned earlier, unbalanced designs result in under (or over) estimates of main effects via sequential sums of Squares. To illustrate this, we will remove a couple of observations from our fabricated data set so as to yield an unbalanced design.

```

> data1 <- data[c(-3, -6, -12), ]
> with(data1, table(A, B))
      B
A     b1 b2 b3
a1    10  9 10
a2    10 10  8

> ## If the following function returns a list (implying that the number of replicates is different for
> ## different combinations of factors), then the design is NOT balanced
> replications(~A * B, data1)
$A
A
a1 a2
29 28

$B
B
b1 b2 b3
20 19 18

$`A:B`
      B
A     b1 b2 b3
a1    10  9 10
a2    10 10  8

```

I will now fit the linear model (and summarize as an ANOVA table) twice - once with Factor A first and the second time with Factor B first.

```

> anova(lm(y ~ A * B, data1))
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value Pr(>F)
A       1   2001     2001  294.7 < 2e-16 ***
B       2    408      204   30.1 2.4e-09 ***
A:B     2   1527      763  112.4 < 2e-16 ***
Residuals 51    346         7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(lm(y ~ B * A, data1))
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value Pr(>F)
B       2    297      149   21.9 1.4e-07 ***
A       1   2112     2112  311.0 < 2e-16 ***
B:A     2   1527      763  112.4 < 2e-16 ***
Residuals 51    346         7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Normally, the order in which terms are included is irrelevant. However, when the design is not balanced (as is the case here), the first term (last term estimated via sequential sums of squares) will be calculated as whatever is left after all other terms have been estimated. Notice that the SS etc for the terms A and B differ between the two fits. In both cases, the first term is underestimated.

Whilst in this case, it has no impact on the inferential outcomes, in other cases it can do. **Note, it is only the ANOVA table that is affected by this, the estimated coefficients will be unaffected.**

If the ANOVA table is required, then the appropriate way of modelling these data is:

- Start off by defining strictly orthogonal contrasts (thus not treatment contrasts). Either sum, Helmert, polynomial or even user-defined contrasts are appropriate options.

```
> contrasts(data1$A) <- contr.sum
> contrasts(data1$B) <- contr.sum
```

- Fit the linear model (which will incorporate the contrasts).

```
> data.lml <- lm(y ~ A * B, data = data1)
```

- Generate the ANOVA table with Type III (marginal) sums of squares. For illustration purposes, I will generate ANOVA tables based on models fit with A first and B first to show that they come out the same.

```
> library(car)
> Anova(data.lml, type = "III")
Anova Table (Type III tests)

Response: y
          Sum Sq Df F value    Pr(>F)
(Intercept) 134308  1 19772.9 < 2e-16 ***
A             2176  1   320.4 < 2e-16 ***
B             445  2    32.7 7.1e-10 ***
A:B          1527  2   112.4 < 2e-16 ***
Residuals    346 51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> Anova(lm(y ~ B * A, data = data1), type = "III")
Anova Table (Type III tests)

Response: y
          Sum Sq Df F value    Pr(>F)
(Intercept) 134308  1 19772.9 < 2e-16 ***
B             445  2    32.7 7.1e-10 ***
A             2176  1   320.4 < 2e-16 ***
B:A          1527  2   112.4 < 2e-16 ***
Residuals    346 51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>
> ## OR with Type II (hierarchical) sums of squares
> Anova(data.lml, type = "II")
Anova Table (Type II tests)

Response: y
          Sum Sq Df F value    Pr(>F)
A             2112  1   311.0 < 2e-16 ***
B             408  2    30.1 2.4e-09 ***
A:B          1527  2   112.4 < 2e-16 ***
Residuals    346 51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> Anova(lm(y ~ B * A, data = data1), type = "II")
Anova Table (Type II tests)

Response: y
          Sum Sq Df F value    Pr(>F)
B             408  2    30.1 2.4e-09 ***
A             2112  1   311.0 < 2e-16 ***
B:A          1527  2   112.4 < 2e-16 ***
Residuals    346 51
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```