

How to capitalize on *a priori* contrasts in linear (mixed) models: A tutorial

Daniel J. Schad<sup>1</sup>, Shravan Vasishth<sup>1</sup>, Sven Hohenstein<sup>1</sup>, & Reinhold Kliegl<sup>1</sup>

<sup>1</sup> University of Potsdam, Germany

#### Author Note

Correspondence concerning this article should be addressed to Daniel J. Schad, University of Potsdam, Germany. Phone: +49-331-977-2309. Fax: +49-331-977-2087. E-mail: [danieljschad@gmail.com](mailto:danieljschad@gmail.com)

## Abstract

Factorial experiments in research on memory, language, and in other areas are often analyzed using analysis of variance (ANOVA). However, for effects with more than one numerator degrees of freedom, e.g., for experimental factors with more than two levels, the ANOVA omnibus F-test is not informative about the source of a main effect or interaction. Because researchers typically have specific hypotheses about which condition means differ from each other, a priori contrasts (i.e., comparisons planned before the sample means are known) between specific conditions or combinations of conditions are the appropriate way to represent such hypotheses in the statistical model. Many researchers have pointed out that contrasts should be “tested instead of, rather than as a supplement to, the ordinary ‘omnibus’ F test” (Hays, 1973, p. 601). In this tutorial, we explain the mathematics underlying different kinds of contrasts (i.e., treatment, sum, repeated, polynomial, custom, nested, interaction contrasts), discuss their properties, and demonstrate how they are applied in the R System for Statistical Computing (R Core Team, 2018). In this context, we explain the generalized inverse which is needed to compute the coefficients for contrasts that test hypotheses that are not covered by the default set of contrasts. A detailed understanding of contrast coding is crucial for successful and correct specification in linear models (including linear mixed models). Contrasts defined a priori yield far more useful confirmatory tests of experimental hypotheses than standard omnibus F-test. Reproducible code is available from <https://osf.io/7ukf6/>.

*Keywords:* Contrasts, null hypothesis significance testing, linear models, a priori hypotheses

How to capitalize on *a priori* contrasts in linear (mixed) models: A tutorial

## Introduction

Whenever an experimental factor comprises more than two levels, analysis of variance (ANOVA) F-statistics provide very little information about the source of an effect or interaction involving this factor. For example, let's assume an experiment with three groups of subjects. Let's also assume that an ANOVA shows that the main effect of group is significant. This of course leaves unclear which groups differ from each other and how they differ. However, scientists typically have *a priori* expectations about the pattern of means. That is, we usually have specific expectations about which groups differ from each other.

One potential strategy is to follow up on these results using t-tests. However, this approach does not consider all the data in each test and therefore loses statistical power, does not generalize well to more complex models (e.g., linear mixed-effects models), and is subject to problems of multiple comparisons. In this paper, we will show how to test specific hypotheses directly in a regression model, which gives much more control over the analysis. Specifically, we show how planned comparisons between specific conditions (groups) or clusters of conditions, are implemented as contrasts. This is a very effective way to align expectations with the statistical model. Indeed, if planned comparisons, implemented in contrasts, are defined *a priori*, and are not defined after the results are known, planned comparisons should be “tested *instead of*, rather than as a supplement to, the ordinary ‘omnibus’ F test.” (Hays, 1973, p. 601).

Every contrast consumes exactly one degree of freedom. Every degree of freedom in the ANOVA source-of-variance table can be spent to test a specific hypothesis about a difference between means or a difference between clusters of means.

Linear mixed-effects models (LMMs) (Baayen, Davidson, & Bates, 2008; Bates, Kliegl, Vasishth, & Baayen, 2015; Bates, Maechler, Bolker, & Walker, 2014; Kliegl, Masson, &

Richter, 2010; Pinheiro & Bates, 2000) are a great tool and represent an important development in statistical practice in psychology and linguistics. LMMs are often taken to replace more traditional ANOVA analyses. However, LMMs also present some challenges. One key challenge is about how to incorporate categorical effects from factors with discrete levels into LMMs. One approach to analyzing factors is to do model comparison; this is akin to the ANOVA omnibus test, and again leaves it unclear which groups differ from which others. An alternative approach is to base analyses on contrasts, which allow us to code factors as independent variables in linear regression models. The present paper explains how to understand and use contrast coding to test particular comparisons between conditions of your experiment (for a Glossary of key terms see Appendix A). Such knowledge about contrast specification is required if analysis of factors is to be based on LMMs instead of ANOVAs. Arguably, in the R System for Statistical Computing (R Core Team, 2018), an understanding of contrast specification is therefore a necessary pre-condition for the proper use of LMMs.

To model differences between categories/groups/cells/conditions, regression models (such as multiple regression, logistic regression and linear mixed models) specify a set of contrasts (i.e., which groups are compared to which baselines or groups). There are several ways to specify such contrasts mathematically, and as discussed below, which of these is more useful depends on the hypotheses about the expected pattern of means. If the analyst does not provide the specification explicitly, R will pick a default specification on its own, which may not align very well with the hypotheses that the analyst intends to test. Therefore, LMMs effectively demand that the user implement planned comparisons—perhaps almost as intended by Hays (1973). Obtaining a statistic roughly equivalent to an ordinary “omnibus” F test requires the extra effort of model comparison. This tutorial provides a practical introduction to contrast coding for factorial experimental designs that will allow scientists to express their hypotheses within the statistical model.

## Prerequisites for this tutorial

The reader is assumed to be familiar with R, and with the foundational ideas behind frequentist statistics, specifically null hypothesis significance testing. Some notational conventions: True parameters are referred to with Greek letters (e.g.,  $\mu$ ,  $\beta$ ), and estimates of these parameters have a hat ( $\hat{\cdot}$ ) on the parameter (e.g.,  $\hat{\mu}$ ,  $\hat{\beta}$ ). The mean of several parameters is written as  $\mu$  or as  $\bar{\mu}$ .

Example analyses of simulated data-sets are presented using linear models (LMs) in R. We use simulated data rather than real data-sets because this allows full control over the dependent variable.

To guide the reader, we provide a preview of the structure of the paper:

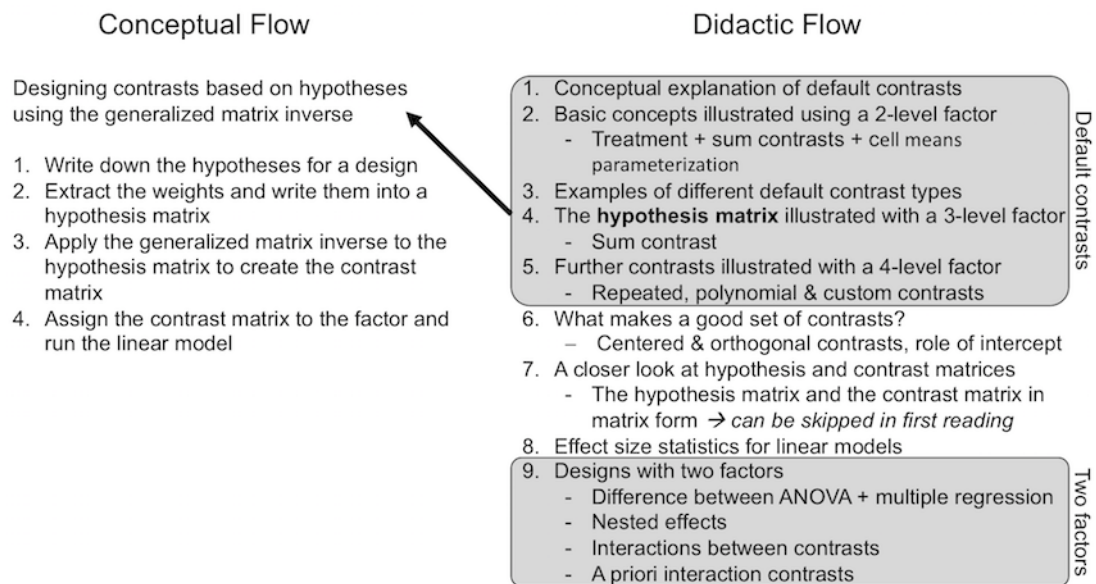
- First, basic concepts of different contrasts are explained, using a factor with two levels to explain the concepts. After a demonstration of the default contrast setting in R, `TREATMENT CONTRASTS`, a commonly used contrast for factorial designs, and `SUM CONTRASTS`, are illustrated.
- We then demonstrate how the generalized inverse (see Fieller, 2016, sec. 8.3) is used to convert what we call the hypothesis matrix to create a contrast matrix. We demonstrate this workflow of going from a hypothesis matrix to a contrast matrix for `SUM CONTRASTS` using an example data-set involving one factor with three levels. We also demonstrate the workflow for `REPEATED CONTRASTS` using an example data-set involving one factor with four levels.
- After showing how contrasts implement hypotheses as predictors in a multiple regression model, we introduce `POLYNOMIAL CONTRASTS` and `CUSTOM CONTRASTS`.
- Then we discuss what makes a good contrast. Here, we introduce two important

concepts: centering and orthogonality of contrasts, and their implications for contrast coding.

- We provide additional information about the generalized inverse, and how it is used to switch between the hypothesis matrix and the contrast matrix. The section on the matrix notation (Hypothesis matrix and contrast matrix in matrix form) can optionally be skipped on first reading.
- Then we discuss an effect size measure for contrasts.
- The next section discusses designs with two factors. We compare regression models with analysis of variance (ANOVA) in simple  $2 \times 2$  designs, and look at contrast centering, nested effects, and at a priori interaction contrasts.

Throughout the tutorial, we show how contrasts are implemented and applied in R and how they relate to hypothesis testing via the generalized inverse.

An overview of the conceptual and didactic flow is provided in Figure 1.



*Figure 1.* Conceptual and didactic flow. The left panel, conceptual flow, shows how to use the hypothesis matrix and the generalized inverse to construct custom contrast matrices for a given set of hypotheses. The right panel, didactic flow, shows an overview of the sections of this paper. The first grey box indicates the sections that introduce a set of default contrasts (treatment, sum, repeated, polynomial, and custom contrasts); the second grey box indicates sections that introduce contrasts for designs with two factors.

### Conceptual explanation of default contrasts

What are examples for different contrast specifications? One contrast in widespread use is the TREATMENT CONTRAST. As suggested by its name, the TREATMENT CONTRAST is often used in intervention studies, where one or several intervention groups receive some treatment, which are compared to a control group. For example, two treatment groups may obtain (a) psychotherapy and (b) pharmacotherapy, and they may be compared to (c) a control group of patients waiting for treatment. This implies one factor with three levels. In this setting, a TREATMENT CONTRAST for this factor makes two comparisons: it tests (1) whether the psychotherapy group is better than the control group, and (2) whether the

pharmacotherapy group is better than the control group. That is, each treatment condition is compared to the same control group or baseline condition. An example in research on memory and language may be a priming study, where two different kinds of priming conditions (e.g., phonological versus orthographic priming) are each compared to a control condition without priming.

A second contrast of widespread use is the SUM CONTRAST. This contrast compares each tested group not against a baseline / control condition, but instead to the average response across all groups. Consider an example where three different priming conditions are compared to each other, such as orthographic, phonological, and semantic priming. The question of interest may be whether two of the priming conditions (e.g., orthographic and phonological priming) elicit stronger responses than the average response across all conditions. This could be done with SUM CONTRASTS: The first contrast would compare orthographic priming with the average response, and the second contrast would compare phonological priming with the average response. SUM CONTRASTS also have an important role in factors with two levels, where they simply test the difference between those two factor levels (e.g., the difference between phonological versus orthographic priming).

A third contrast coding is REPEATED CONTRASTS. These are probably less often used in empirical studies, but are arguably the contrast of highest relevance for research in psychology and cognitive science. REPEATED CONTRASTS successively test neighboring factor levels against each other. For example, a study may manipulate the frequency of some target words into three categories of “low frequency”, “medium frequency”, and “high frequency”. What may be of interest in the study is whether low frequency words differ from medium frequency words, and whether medium frequency words differ from high frequency words. REPEATED CONTRASTS test exactly these differences between neighboring factor levels.

A fourth type of contrast is POLYNOMIAL CONTRASTS. These are useful when trends



are of interest that span across multiple factor levels. In the example with different levels of word frequency, a simple hypothesis may state that the response increases with increasing levels of word frequency. That is, one may expect that a response increases from low to medium frequency words by the same magnitude as it increases from medium to high frequency words. That is, a linear trend is expected. Here, it is possible to test a quadratic trend of word frequency; for example, when the effect is expected to be larger or smaller between medium and high frequency words compared to the effect between low and medium frequency words.

One additional option for contrast coding is provided by `HELMERT CONTRASTS`. In an example with three factor levels, for `HELMERT CONTRASTS` the first contrast codes the difference between the first two factor levels, and the second contrast codes the difference between the mean of the first two levels and the third level. (In cases of a four-level factor, the third contrast tests the difference between (i) the average of the first three levels and (ii) the fourth level.) An example for the use of `HELMERT CONTRASTS` is a priming paradigm with the three experimental conditions “valid prime”, “invalid prime 1”, and “invalid prime 2”. The first contrast may here test the difference between conditions “invalid prime 1” and “invalid prime 2”. The second contrast could then test the difference between valid versus invalid conditions. This coding would provide maximal power to test the difference between valid and invalid conditions, as it pools across both invalid conditions for the comparison.

### Basic concepts illustrated using a two-level factor

Consider the simplest case: suppose we want to compare the means of a dependent variable (DV) such as response times between two groups of subjects. R can be used to simulate data for such an example using the function `mixedDesign()` (for details regarding this function, see Appendix C). The simulations assume longer response times in condition F1 ( $\mu_1 = 0.8$  sec) than F2 ( $\mu_2 = 0.4$  sec). The data from the 10 simulated subjects are

aggregated and summary statistics are computed for the two groups.

```
library(dplyr)
# load mixedDesign function for simulating data
source("functions/mixedDesign.v0.6.3.R")
M <- matrix(c(0.8, 0.4), nrow=2, ncol=1, byrow=FALSE)
set.seed(1) # set seed of random number generator for replicability
simdat <- mixedDesign(B=2, W=NULL, n=5, M=M, SD=.20, long = TRUE)
names(simdat)[1] <- "F" # Rename B_A to F(actor)
levels(simdat$F) <- c("F1", "F2")
simdat
```

```
##      F id    DV
## 1  F1   1 0.997
## 2  F1   2 0.847
## 3  F1   3 0.712
## 4  F1   4 0.499
## 5  F1   5 0.945
## 6  F2   6 0.183
## 7  F2   7 0.195
## 8  F2   8 0.608
## 9  F2   9 0.556
## 10 F2  10 0.458
```

```
str(simdat)
```

```
## 'data.frame':   10 obs. of  3 variables:
##  $ F : Factor w/ 2 levels "F1","F2": 1 1 1 1 1 2 2 2 2 2
##  $ id: Factor w/ 10 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10
```

```
## $ DV: num 0.997 0.847 0.712 0.499 0.945 ...
```

```
table1 <- simdat %>% group_by(F) %>% # Table for main effect F
  summarize(N=n(), M=mean(DV), SD=sd(DV), SE=SD/sqrt(N) )
(GM <- mean(table1$M)) # Grand Mean
```

```
## [1] 0.6
```

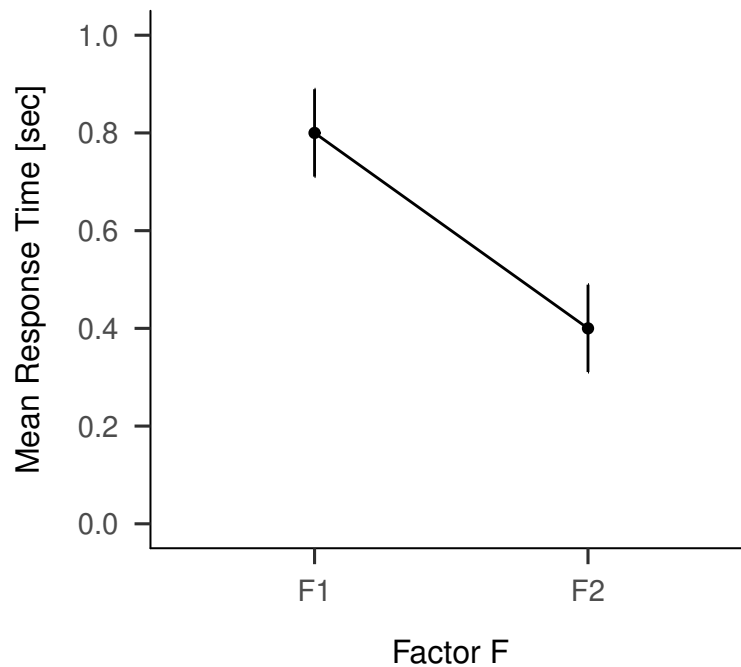


Figure 2. Means and standard errors of the simulated dependent variable (e.g., response times in seconds) in two conditions F1 and F2.

The results, displayed in Table 1 and in Figure 2, show that the assumed true condition means are exactly realized with the simulated data. The numbers are exact because the `mixedDesign()` function ensures that the data are generated so as to have the true means for each level. In real data-sets, of course, the sample means will vary from experiment to experiment.

A simple regression of DV on F yields a straightforward test of the difference between the group means. Part of the output of the `summary` function is presented below, and the

same results are also displayed in Table 2:

```
m_F <- lm(DV ~ 1 + F, simdat)
round(summary(m_F)$coef,3)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.8      0.089    8.94   0.000
## FF2             -0.4      0.126   -3.16   0.013
```

Comparing the means for each condition with the coefficients (*Estimates*) reveals that (i) the intercept (0.8) is the mean for condition F1,  $\hat{\mu}_1$ ; and (ii) the slope (FF2:  $-0.4$ ) is the difference between the true means for the two groups,  $\hat{\mu}_2 - \hat{\mu}_1$  (Bolker, 2018):

$$\begin{aligned} \text{Intercept} &= \hat{\mu}_1 = \text{estimated mean for F1} \\ \text{Slope (FF2)} &= \hat{\mu}_2 - \hat{\mu}_1 = \text{estim. mean for F2} - \text{estim. mean for F1} \end{aligned} \quad (1)$$

The new information is confidence intervals associated with the regression coefficients. The t-test suggests that response times in group F2 are lower than in group F1.

### Default contrast coding: Treatment contrasts

How does R arrive at these particular values for the intercept and slope? That is, why does the intercept assess the mean of condition F1 and how do we know the slope measures

Table 1

*Summary statistics per condition for the simulated data.*

Factor F	N data points	Estimated means	Standard deviations	Standard errors
F1	5	0.8	0.2	0.1
F2	5	0.4	0.2	0.1

Table 2

*Estimated regression model. Confidence intervals are obtained in R, e.g., using the function `confint()`.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (8)	<i>p</i>
Intercept	0.8	[0.6, 1.0]	8.94	< .001
FF2	-0.4	[-0.7, -0.1]	-3.16	.013

the difference in means between F2–F1? This result is a consequence of the default contrast coding of the factor F. R assigns TREATMENT CONTRASTS to factors and orders their levels alphabetically. The first factor level (here: F1) is coded as 0 and the second level (here: F2) is coded as 1. This is visible when inspecting the current contrast attribute of the factor using the `contrasts` command:

```
contrasts(simdat$F)
```

```
##      F2
## F1    0
## F2    1
```

Why does this contrast coding yield these particular regression coefficients? Let's take a look at the regression equation. Let  $\beta_0$  represent the intercept, and  $\beta_1$  the slope. Then, the simple regression above expresses the belief that the expected response time  $y$  is a linear function of the factor F. In a more general formulation, this is written as follows:  $y$  is a linear function of some predictor  $x$  with regression coefficients for the intercept,  $\beta_0$ , and for the factor,  $\beta_1$ :

$$y = \beta_0 + \beta_1 x \quad (2)$$

So, if  $x = 0$  (condition F1),  $y$  is  $\beta_0 + \beta_1 \cdot 0 = \beta_0$ ; and if  $x = 1$  (condition F2),  $y$  is  $\beta_0 + \beta_1 \cdot 1 = \beta_0 + \beta_1$ .

Expressing the above in terms of the estimated coefficients:

$$\begin{aligned} \text{estim. value for F1} &= \hat{\mu}_1 = \hat{\beta}_0 = \text{Intercept} \\ \text{estim. value for F2} &= \hat{\mu}_2 = \hat{\beta}_0 + \hat{\beta}_1 = \text{Intercept} + \text{Slope (FF2)} \end{aligned} \tag{3}$$

It is useful to think of such unstandardized regression coefficients as difference scores; they express the increase in the dependent variable  $y$  associated with a change in the independent variable  $x$  of 1 unit, such as going from 0 to 1 in this example. The difference between condition means is  $0.4 - 0.8 = -0.4$ , which is exactly the estimated regression coefficient  $\hat{\beta}_1$ . The sign of the slope is negative because we have chosen to subtract the larger mean F1 score from the smaller mean F2 score.

## Defining hypotheses

The analysis of the regression equation demonstrates that in the TREATMENT CONTRAST the intercept assesses the average response in the baseline condition, whereas the slope tests the difference between condition means. However, these are just verbal descriptions of what each coefficient assesses. Is it also possible to formally write down the null hypotheses that are tested by each of these two coefficients? From the perspective of formal hypothesis tests, the slope represents the main test of interest, so we consider this first. The TREATMENT CONTRAST expresses the null hypothesis that the difference in means between the two levels of the factor F is 0; formally, the null hypothesis  $H_0$  is that  $H_0 : \beta_1 = 0$ :

$$H_0 : -1 \cdot \mu_{F1} + 1 \cdot \mu_{F2} = 0 \tag{4}$$

or equivalently:

$$H_0 : \mu_{F2} - \mu_{F1} = 0 \quad (5)$$

The  $\pm 1$  weights in the null hypothesis statement directly express which means are compared by the treatment contrast.

The intercept in the TREATMENT CONTRAST expresses a null hypothesis that is usually of no interest: that the mean in condition F1 of the factor F is 0. Formally, the null hypothesis is  $H_0 : \beta_0 = 0$ :

$$H_0 : 1 \cdot \mu_{F1} + 0 \cdot \mu_{F2} = 0 \quad (6)$$

or equivalently:

$$H_0 : \mu_{F1} = 0. \quad (7)$$

The fact that the intercept term formally tests the null hypothesis that the mean of condition F1 is zero is in line with our previous derivation (see equation 1).

In R, factor levels are ordered alphabetically and by default the first level is used as the baseline in TREATMENT CONTRASTS. Obviously, this default mapping will only be correct for a given data-set if the levels' alphabetical ordering matches the desired contrast coding. When it does not, it is possible to re-order the levels. Here is one way of re-ordering the levels in R:

```
simdat$Fb <- factor(simdat$F, levels = c("F2", "F1"))
contrasts(simdat$Fb)
```

```
##      F1
## F2   0
## F1   1
```

This re-ordering did not change any data associated with the factor, only one of its attributes. With this new contrast attribute the simple regression yields the following result (see Table 3).

```
m1_mr <- lm(DV ~ 1 + Fb, simdat)
```

Table 3

*Reordering factor levels*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (8)	<i>p</i>
Intercept	0.4	[0.2, 0.6]	4.47	.002
FbF1	0.4	[0.1, 0.7]	3.16	.013

The model now tests different hypotheses. The intercept now codes the mean of condition F2, and the slope measures the difference in means between F1 minus F2. This represents an alternative coding of the treatment contrast.

## Sum contrasts

Treatment contrasts are only one of many options. It is also possible to use SUM CONTRASTS, which code one of the conditions as  $-1$  and the other as  $+1$ , effectively ‘centering’ the effects at the grand mean (GM, i.e., the mean of the two group means). Here, we rescale the contrast to values of  $-0.5$  and  $+0.5$ , which makes the estimated treatment effect the same as for treatment coding and easier to interpret.

To use this contrast in a linear regression, use the `contrasts` function (for results see



Table 4):

```
(contrasts(simdat$F) <- c(-0.5, +0.5))
```

```
## [1] -0.5  0.5
```

```
m1_mr <- lm(DV ~ 1 + F, simdat)
```

Table 4

*Estimated regression model*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (8)	<i>p</i>
Intercept	0.6	[0.5, 0.7]	9.49	< .001
F1	-0.4	[-0.7, -0.1]	-3.16	.013

Here, the slope (F1) again codes the difference of the groups associated with the first and second factor levels. It has the same value as in the TREATMENT CONTRAST. However, the intercept now represents the estimate of the average of condition means for F1 and F2, that is, the GM. This differs from the TREATMENT CONTRAST. For the scaled SUM CONTRAST:

$$\begin{aligned}
 \text{Intercept} &= (\hat{\mu}_1 + \hat{\mu}_2)/2 = \text{estimated mean of F1 and F2} \\
 \text{Slope (F1)} &= \hat{\mu}_2 - \hat{\mu}_1 = \text{estim. mean for F2} - \text{estim. mean for F1}
 \end{aligned}
 \tag{8}$$

How does R arrive at these values for the intercept and the slope? Why does the intercept assess the GM and why does the slope test the group-difference? This is the result of rescaling the SUM CONTRAST. The first factor level (F1) was coded as  $-0.5$ , and the second factor level (F2) as  $+0.5$ :

```
contrasts(simdat$F)
```

```
##      [,1]
```

```
## F1 -0.5
```

```
## F2  0.5
```

Let's again look at the regression equation to better understand what computations are performed. Again,  $\beta_0$  represents the intercept,  $\beta_1$  represents the slope, and the predictor variable  $x$  represents the factor **F**. The regression equation is written as:

$$y = \beta_0 + \beta_1 x \quad (9)$$

The group of **F1** subjects is then coded as  $-0.5$ , and the response time for the group of **F1** subjects is  $\beta_0 + \beta_1 \cdot x_1 = 0.6 + (-0.4) \cdot (-0.5) = 0.8$ . The **F2** group, to the contrary, is coded as  $+0.5$ . By implication, the mean of the **F2** group must be  $\beta_0 + \beta_1 \cdot x_1 = 0.6 + (-0.4) \cdot 0.5 = 0.4$ . Expressed in terms of the estimated coefficients:

$$\begin{aligned} \text{estim. value for F1} &= \hat{\mu}_1 = \hat{\beta}_0 - 0.5 \cdot \hat{\beta}_1 = \text{Intercept} - 0.5 \cdot \text{Slope (F1)} \\ \text{estim. value for F2} &= \hat{\mu}_2 = \hat{\beta}_0 + 0.5 \cdot \hat{\beta}_1 = \text{Intercept} + 0.5 \cdot \text{Slope (F1)} \end{aligned} \quad (10)$$

The unstandardized regression coefficient is a difference score: Taking a step of one unit on the predictor variable  $x$ , e.g., from  $-0.5$  to  $+0.5$ , reflecting a step from condition **F1** to **F2**, changes the dependent variable from  $0.8$  (for condition **F1**) to  $0.4$  (condition **F2**), reflecting a difference of  $0.4 - 0.8 = -0.4$ ; and this is again exactly the estimated regression coefficient  $\hat{\beta}_1$ . Moreover, as mentioned above, the intercept now assesses the GM of conditions **F1** and **F2**: it is exactly in the middle between condition means for **F1** and **F2**.

So far we gave verbal statements about what is tested by the intercept and the slope in

the case of the scaled SUM CONTRAST. It is possible to write these statements as formal null hypotheses that are tested by each regression coefficient: Sum contrasts express the null hypothesis that the difference in means between the two levels of factor F is 0; formally, the null hypothesis  $H_0$  is that

$$H_0 : -1 \cdot \mu_{F1} + 1 \cdot \mu_{F2} = 0 \quad (11)$$

This is the same hypothesis that was also tested by the slope in the treatment contrast. The intercept, however, now expresses a different hypothesis about the data: it expresses the null hypothesis that the average of the two conditions F1 and F2 is 0:

$$H_0 : 1/2 \cdot \mu_{F1} + 1/2 \cdot \mu_{F2} = \frac{\mu_{F1} + \mu_{F2}}{2} = 0 \quad (12)$$

In balanced data, i.e., in data-sets where there are no missing data points, the average of the two conditions F1 and F2 is the GM. In unbalanced data-sets, where there are missing values, this average is the weighted GM. To illustrate this point, consider an example with fully balanced data and two equal group sizes of 5 subjects for each group F1 and F2. Here, the GM is also the mean across all subjects. Next, consider a highly simplified unbalanced data-set, where in condition F1 two observations of the dependent variable are available with values of 2 and 3, and where in condition F2 only one observation of the dependent variables is available with a value of 4. In this data-set, the mean across all subjects is  $\frac{2+3+4}{3} = \frac{9}{3} = 3$ . However, the (weighted) GM as assessed in the intercept in a model using sum contrasts for factor F would first compute the mean for each group separately (i.e.,  $\frac{2+3}{2} = 2.5$ , and 4), and then compute the mean across conditions  $\frac{2.5+4}{2} = \frac{6.5}{2} = 3.25$ . The GM of 3.25 is different from the mean across subjects of 3.

To summarize, TREATMENT CONTRASTS and SUM CONTRASTS are two possible ways

to parameterize the difference between two groups; they test different hypotheses (there are cases, however, where the hypotheses are equivalent). `TREATMENT CONTRASTS` compare one or more means against a baseline condition, whereas `SUM CONTRASTS` allow us to determine whether we can reject the null hypothesis that a condition's mean is the same as the GM (which in the two-group case also implies a hypothesis test that the two group means are the same). One question that comes up here, is how one knows or formally derives what hypotheses a given set of contrasts tests. This question will be discussed in detail below for the general case of any arbitrary contrasts.

### Cell means parameterization

One alternative option is use the so-called cell means parameterization. In this approach, one does not estimate an intercept term, and then differences between factor levels. Instead, each degree of freedom in a design is used to simply estimate the mean of one of the factor levels. As a consequence, no comparisons between condition means are tested, but it is simply tested for each factor level whether the associated condition mean differs from zero. Cell means parameterization is specified by explicitly removing the intercept term (which is added automatically) by adding a `-1` in the regression formula:

```
m2_mr <- lm(DV ~ -1 + F, simdat)
```

Table 5

*Estimated regression model*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (8)	<i>p</i>
FF1	0.8	[0.6, 1.0]	8.94	< .001
FF2	0.4	[0.2, 0.6]	4.47	.002

Now, the regression coefficients (see the column labeled 'Estimate') estimate the mean

of the first factor level (0.8) and the mean of the second factor level (0.4). Each of these means is compared to zero in the statistical tests, and each of these means is significantly larger than zero. This cell means parameterization usually does not allow a test of the hypotheses of interest, as these hypotheses usually relate to differences between conditions rather than to whether each condition differs from zero.

### Examples of different default contrast types

Above, we introduced conceptual explanations of different default contrasts by discussing example applications. Moreover, the preceding section on basic concepts mentioned that contrasts are implemented by numerical contrast coefficients. These contrast coefficients are usually represented in matrices of coefficients. Each of the discussed default contrasts is implemented using a different contrast matrix. These default contrast matrices are available in the R System for Statistical Computing in the basic distribution of R in the stats package. Here, we provide a quick overview of the different contrast matrices specifically for the example applications discussed above.

For the TREATMENT CONTRASTS, we discussed an example where a psychotherapy group and a pharmacotherapy group are each compared to the same control or baseline group; the latter could be a group that receives no treatment. The corresponding contrast matrix is obtained using the following function call:

```
contr.treatment(3)
```

```
##    2 3
## 1 0 0
## 2 1 0
## 3 0 1
```

The number of rows specifies the number of factor levels. The three rows indicate the three levels of the factor. The first row codes the baseline or control condition (the baseline always only contains 0s as contrast coefficients), the second row codes the psychotherapy group, and the third row codes the pharmacotherapy group. The two columns reflect the two comparisons that are being tested by the contrasts: the first column tests the second group (i.e., psychotherapy) against the baseline / control group, and the second column tests the third group (i.e., pharmacotherapy) against the baseline / control group.

For the SUM CONTRAST, our example involved conditions for orthographic priming, phonological priming, and semantic priming. The orthographic and phonological priming conditions were compared to the average priming effect across all three groups. In R, there is again a standard function call for the SUM CONTRAST:

```
contr.sum(3)
```

```
##      [,1] [,2]
## 1      1     0
## 2      0     1
## 3     -1    -1
```

Again, the three rows indicate three groups, and the two columns reflect the two comparisons. The first row codes orthographic priming, the second row phonological priming, and the last row semantic priming. Now, the first column codes a contrast that compares the response in orthographic priming against the average response, and the second column codes a contrast comparing phonological priming against the average response. Why these contrasts test these hypotheses is not transparent here. We will return to this issue below.

For REPEATED CONTRASTS, our example compared response times in low frequency words vs. medium frequency words, and medium frequency words vs. high frequency words. In R, the corresponding contrast matrix is available in the MASS package (Venables &

Ripley, 2002):

```
library(MASS)
contr.sdif(3)
```

```
##      2-1    3-2
## 1 -0.667 -0.333
## 2  0.333 -0.333
## 3  0.333  0.667
```

The first row represents low frequency words, the second row medium frequency words, and the last row high frequency words. Now the first contrast (column) tests the difference between the second minus the first row, i.e., the response to medium frequency words minus response to low frequency words. The second contrast (column) tests the difference between the third minus the second row, i.e., the difference in the response to high frequency words minus the response to medium frequency words. Why the REPEATED CONTRAST tests exactly these differences is not transparent either.

Below, we will explain how these and other contrasts are generated from a careful definition of the hypotheses that one wishes to test for a given data-set. We will introduce a basic workflow for how to create one's own custom contrasts.

We discussed that for the example with three different word frequency levels, it is possible to test the hypothesis of a linear (or quadratic) trend across all levels of word frequency. Such POLYNOMIAL CONTRASTS are specified in R using the following command:

```
contr.poly(3)
```

```
##      .L      .Q
## [1,] -7.07e-01  0.408
## [2,] -7.85e-17 -0.816
```

```
## [3,] 7.07e-01 0.408
```

As in the other contrasts mentioned above, it is not clear from this contrast matrix what hypotheses are being tested. As before, the three rows represent three levels of word frequency. The first column codes a linear increase with word frequency levels. The second column codes a quadratic trend. The `contr.poly()` function tests orthogonalized trends - a concept that we will explain below.

We had also discussed HELMERT CONTRASTS above, and had given the example that one might want to compare two "invalid" priming conditions to each other, and then compare both "invalid" priming conditions to one "valid" prime condition. HELMERT CONTRASTS are specified in R using the following command:

```
contr.helmert(3)
```

```
##      [,1] [,2]
## 1     -1    -1
## 2      1    -1
## 3      0     2
```

The first row represents "invalid prime 1", the second row "invalid prime 2", and the third row the "valid prime" condition. The first column tests the difference between the two "invalid" prime conditions. The coefficients in the second column test both "invalid prime" conditions against the "valid prime" condition.

How can one make use of these contrast matrices for a specific regression analysis? As discussed above for the case of two factor levels, one needs to tell R to use one of these contrast coding schemes for a factor of interest in a linear model (LM)/regression analysis. Let's assume a data-frame called `dat` with a dependent variable `dat$DV` and a three-level factor `dat$WordFrequency` with levels low, medium, and high frequency words. One chooses



one of the above contrast matrices and ‘assigns’ this contrast to the factor. Here, we choose the REPEATED CONTRAST:

```
contrasts(dat$WordFrequency) <- contr.sdif(3)
```

This way, when running a linear model using this factor, R will automatically use the contrast matrix assigned to the factor. This is done in R with the simple call of a linear model, where `dat` is specified as the data-frame to use, where the numeric variable DV is defined as the dependent variable, and where the factor `WordFrequency` is added as predictor in the analysis:

```
lm(DV ~ 1 + WordFrequency, data=dat)
```

Assuming that there are three levels to the `WordFrequency` factor, the `lm` function will estimate three regression coefficients: one intercept and two slopes. What these regression coefficients test will depend on which contrast coding is specified. Given that we have used REPEATED CONTRASTS, the resulting regression coefficients will now test the difference between medium and low frequency words (first slope) and will test the difference between high and medium frequency words (second slope). Examples of output from regression models for concrete situations will be shown below. Moreover, it will be shown how contrast matrices are generated for whatever hypotheses one wants to test in a given data-set.

### The hypothesis matrix illustrated with a three-level factor

Consider again the example with the three low, medium, and high frequency conditions. The `mixedDesign` function can be used to simulate data from a lexical decision task with response times as dependent variable. The research question is: do response times differ as a function of the between-subject factor word frequency with three levels: low, medium, and high? We assume that lower word frequency results in longer response times. Here, we

specify word frequency as a between-subject factor. In cognitive science experiments, frequency will usually vary within subjects and between items. However, the within- or between-subjects status of an effect is independent of its contrast coding; we assume the manipulation to be between subjects for ease of exposition. The concepts presented here extend to repeated measures designs that are usually analyzed using linear mixed models.

The following R code simulates the data and computes the table of means and standard deviations for the three frequency categories:

```
M <- matrix(c(500, 450, 400), nrow=3, ncol=1, byrow=FALSE)
set.seed(1)
simdat2 <- mixedDesign(B=3, W=NULL, n=4, M=M, SD=20, long = TRUE)
names(simdat2)[1] <- "F" # Rename B_A to F(actor)/F(requency)
levels(simdat2$F) <- c("low", "medium", "high")
simdat2$DV <- round(simdat2$DV)
head(simdat2)

##           F id  DV
## 1    low  1 497
## 2    low  2 474
## 3    low  3 523
## 4    low  4 506
## 5 medium  5 422
## 6 medium  6 467

table.word <- simdat2 %>% group_by(F) %>%
  summarise(N = length(DV), M = mean(DV), SD = sd(DV), SE = sd(DV)/sqrt(N))
```

As shown in Table 6, the estimated means reflect our assumptions about the true means in the data simulation: Response times decrease with increasing word frequency. The

Table 6

*Summary statistics of the simulated lexical decision data per frequency level.*

Factor F	N data points	Estimated means	Standard deviations	Standard errors
low	4	500	20	10
medium	4	450	20	10
high	4	400	20	10

effect is significant in an ANOVA (see Table 7).

```
aovF <- aov(DV ~ 1 + F + Error(id), data=simdat2)
```

The ANOVA, however, does not tell us the source of the difference. In the following sections, we use this and an additional data-set to illustrate **SUM**, **REPEATED**, **POLYNOMIAL**, and **CUSTOM** contrasts. In practice, usually only one set of contrasts is selected when the expected pattern of means is formulated during the design of the experiment. The decision about which contrasts to use is made before the pattern of means is known.

## Sum contrasts

For didactic purposes, the next sections describe SUM CONTRASTS. Suppose that the expectation is that low-frequency words are responded to slower and medium-frequency words are responded to faster than the GM response time. Then, the research question could be: Do low-frequency words differ from the GM and do medium-frequency words differ from the GM? And if so, are they above or below the GM? We want to test the following two hypotheses:

$$H_{0_1} : \mu_1 = \frac{\mu_1 + \mu_2 + \mu_3}{3} = GM \quad (13)$$

Table 7

*ANOVA results. The effect F stands for the word frequency factor.*

Effect	<i>F</i>	<i>df</i> <sub>1</sub>	<i>df</i> <sub>2</sub>	<i>MSE</i>	<i>p</i>	$\hat{\eta}_G^2$
F	24.93	2	9	403.19	< .001	.847

and

$$H_{0_2} : \mu_2 = \frac{\mu_1 + \mu_2 + \mu_3}{3} = GM \quad (14)$$

$H_{0_1}$  can also be written as:

$$\mu_1 = \frac{\mu_1 + \mu_2 + \mu_3}{3} \quad (15)$$

$$\Leftrightarrow \mu_1 - \frac{\mu_1 + \mu_2 + \mu_3}{3} = 0 \quad (16)$$

$$\Leftrightarrow \frac{2}{3}\mu_1 - \frac{1}{3}\mu_2 - \frac{1}{3}\mu_3 = 0 \quad (17)$$

Here, the weights  $2/3, -1/3, -1/3$  are informative about how to combine the condition means to define the null hypothesis.

$H_{0_2}$  is also rewritten as:

$$\mu_2 = \frac{\mu_1 + \mu_2 + \mu_3}{3} \quad (18)$$

$$\Leftrightarrow \mu_2 - \frac{\mu_1 + \mu_2 + \mu_3}{3} = 0 \quad (19)$$

$$\Leftrightarrow -\frac{1}{3}\mu_1 + \frac{2}{3}\mu_2 - \frac{1}{3}\mu_3 = 0 \quad (20)$$

Here, the weights are  $-1/3, 2/3, -1/3$ , and they again indicate how to combine the condition means for defining the null hypothesis.

## The hypothesis matrix

The weights of the condition means are not only useful to define hypotheses. They also provide the starting step in a very powerful method which allows the researcher to generate the contrasts that are needed to test these hypotheses in a linear model. That is, what we did so far is to explain some kinds of different contrast codings that exist and what the hypotheses are that they test. That is, if a certain data-set is given and certain hypotheses exist that need to be tested in this data-set, then the procedure would be to check whether any of the contrasts that we encountered above happen to test exactly the hypotheses of interest. Sometimes it suffices to use one of these existing contrasts. However, at other times, our research hypotheses may not correspond exactly to any of the contrasts in the default set of standard contrasts provided in R. For these cases, or simply for more complex designs, it is very useful to know how contrast matrices are created. Indeed, a relatively simple procedure exists in which we write our hypotheses formally, extract the weights of the condition means from the hypotheses, and then automatically generate the correct contrast matrix that we need in order to test these hypotheses in a linear model. Using this powerful method, it is not necessary to find a match to a contrast matrix provided by the family of functions in R starting with the prefix `contr`. Instead, it is possible to simply define the hypotheses that one wants to test, and to obtain the correct contrast matrix for these in an automatic procedure. Here, for pedagogical reasons, we show some examples of how to apply this procedure in cases where the hypotheses *do* correspond to some of the existing contrasts.

Defining a custom contrast matrix involves four steps:

1. Write down the hypotheses

2. Extract the weights and write them into what we will call a *hypothesis matrix*
3. Apply the *generalized matrix inverse* to the hypothesis matrix to create the contrast matrix
4. Assign the contrast matrix to the factor and run the linear model

Let us apply this four-step procedure to our example of the SUM CONTRAST. The first step, writing down the hypotheses, is shown above. The second step involves writing down the weights that each hypothesis gives to condition means. The weights for the first null hypothesis are  $\text{wH01} = c(+2/3, -1/3, -1/3)$ , and the weights for the second null hypothesis are  $\text{wH02} = c(-1/3, +2/3, -1/3)$ .

Before writing these into a hypothesis matrix, we also define a null hypothesis for the intercept term. For the intercept, the hypothesis is that the mean across all conditions is zero:

$$H_{00} : \frac{\mu_1 + \mu_2 + \mu_3}{3} = 0 \quad (21)$$

$$H_{00} : \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 = 0 \quad (22)$$

This null hypothesis has weights of 1/3 for all condition means. The weights from all three hypotheses that were defined are now combined and written into a matrix that we refer to as the *hypothesis matrix* (Hc):

```
HcSum <- rbind(cH00=c(low= 1/3, med= 1/3, hi= 1/3),
               cH01=c(low=+2/3, med=-1/3, hi=-1/3),
               cH02=c(low=-1/3, med=+2/3, hi=-1/3))
fractions(t(HcSum))
```

```
##      cH00 cH01 cH02
```

```
## low  1/3  2/3 -1/3
## med  1/3 -1/3  2/3
## hi   1/3 -1/3 -1/3
```

Each set of weights is first entered as a row into the matrix (command `rbind()`). This has mathematical reasons that we discuss below. However, we then switch rows and columns of the matrix for easier readability using the command `t()` (this transposes the matrix, i.e., switches rows and columns).<sup>1</sup> The command `fractions()` turns the decimals into fractions to improve readability.

Now that the condition weights from the hypotheses have been written into the hypothesis matrix, the third step of the procedure is implemented: a matrix operation called the “generalized matrix inverse”<sup>2</sup> is used to obtain the contrast matrix that is needed to test these hypotheses in a linear model. In R this next step is done using the function `ginv()` from the `MASS` package. We here define a function `ginv2()` for nicer formatting of the output.<sup>3</sup>

```
ginv2 <- function(x) # define a function to make the output nicer
  fractions(provideDimnames(ginv(x), base=dimnames(x)[2:1]))
```

---

<sup>1</sup>Matrix transpose changes the arrangement of the columns and rows of a matrix, but leaves the content of the matrix unchanged. For example, for the matrix  $X$  with three rows and two columns  $X = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}$ , the transpose yields a matrix with two rows and three columns, where the rows and columns are flipped:

$$X^T = \begin{pmatrix} a & c & e \\ b & d & f \end{pmatrix}.$$

<sup>2</sup>At this point, there is no need to understand in detail what this means. We refer the interested reader to Appendix D. For a quick overview, we recommend a vignette explaining the generalized inverse in the [matlib package](#) (Friendly, Fox, & Chalmers, 2018).

<sup>3</sup>The function `fractions()` from the `MASS` package is used to make the output more easily readable, and the function `provideDimnames()` is used to keep row and column names.

Applying the generalized inverse to the hypothesis matrix results in the new matrix `XcSum`. This is the contrast matrix  $X_c$  that tests exactly those hypotheses that were specified earlier:

```
(XcSum <- ginv2(HcSum))
```

```
##      cH00 cH01 cH02
## low   1     1     0
## med   1     0     1
## hi    1    -1    -1
```

This contrast matrix corresponds exactly to the SUM CONTRASTS described above. In the case of the SUM CONTRAST, the contrast matrix looks very different from the hypothesis matrix. The contrast matrix in SUM CONTRASTS codes with +1 the condition that is to be compared to the GM. The condition that is never compared to the GM is coded as -1. Without knowing the relationship between the hypothesis matrix and the contrast matrix, the meaning of the coefficients is completely opaque.

To verify this custom-made contrast matrix, it is compared to the SUM CONTRAST matrix as generated by the R function `contr.sum()` in the `stats` package. The resulting contrast matrix is identical to the result when adding the intercept term, a column of ones, to the contrast matrix:

```
fractions(cbind(1,contr.sum(3)))
```

```
##      [,1] [,2] [,3]
## 1     1     1     0
## 2     1     0     1
## 3     1    -1    -1
```

In order to test the hypotheses, step four in our procedure involves assigning SUM



CONTRASTS to the factor **F** in our example data, and running a linear model.<sup>4</sup> This allows estimating the regression coefficients associated with each contrast. We compare these to the data in Table 6 to test whether the regression coefficients actually correspond to the differences of condition means, as intended. To define the contrast, it is necessary to remove the intercept term, as this is automatically added by the linear model function `lm()` in R.

```
contrasts(simdat2$F) <- XcSum[,2:3]
m1_mr <- lm(DV ~ 1 + F, data=simdat2)
```

Table 8

*Regression model using the sum contrast.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (9)	<i>p</i>
Intercept	450	[437, 463]	77.62	< .001
FcH01	50	[32, 69]	6.11	< .001
FcH02	0	[−18, 19]	0.01	.992

The linear model regression coefficients (see Table 8) show the GM response time of 450 ms in the intercept. Remember that the first regression coefficient **FcH01** was designed to test our first hypothesis that low frequency words are responded to slower than the GM. The regression coefficient **FcH01** (“Estimate”) of 50 exactly reflects the difference between low frequency words (500 ms) and the GM of 450 ms. The second hypothesis was that response times for medium frequency words differ from the GM. The fact that the second regression coefficient **FcH02** is exactly 0 indicates that response times for medium frequency words (450 ms) are identical with the GM of 450 ms. Although there is evidence against the null hypothesis that low-frequency words have the same reading time as the GM, there is no

<sup>4</sup>Alternative ways to specify default contrasts in R are to set contrasts globally using `options(contrasts="contr.sum")` or to set contrasts locally only for a specific analysis, by including a named list specifying contrasts for each factor in a linear model: `lm(DV ~ 1 + F, contrasts=list(F="contr.sum"))`.

evidence against the null hypothesis that medium frequency words have the same reading times as the GM.

We have now not only derived contrasts and hypothesis tests for the sum contrast, we have also used a powerful and highly general procedure that is used to generate contrasts for many kinds of different hypotheses and experimental designs.

### Further examples of contrasts illustrated with a factor with four levels

In order to understand REPEATED DIFFERENCE and POLYNOMIAL CONTRASTS, it may be instructive to consider an experiment with one between-subject factor with four levels. We simulate such a data-set using the function `mixedDesign`. The sample sizes for each level and the means and standard errors are shown in Table 9, and the means and standard errors are also shown graphically in Figure 3.

We assume that the four factor levels F1 to F4 reflect levels of word frequency, including the levels `low`, `medium-low`, `medium-high`, and `high` frequency words, and that the dependent variable reflects some response time.<sup>5</sup>

```
# Data, means, and figure
M <- matrix(c(10, 20, 10, 40), nrow=4, ncol=1, byrow=FALSE)
set.seed(1)
simdat3 <- mixedDesign(B=4, W=NULL, n=5, M=M, SD=10, long = TRUE)
names(simdat3)[1] <- "F" # Rename B_A to F(actor)
levels(simdat3$F) <- c("F1", "F2", "F3", "F4")
table3 <- simdat3 %>% group_by(F) %>%
  summarize(N=length(DV), M=mean(DV), SD=sd(DV), SE=SD/sqrt(N) )
```

---

<sup>5</sup>Qualitatively, the simulated pattern of results is actually empirically observed for word frequency effects on single fixation durations (Heister, Würzner, & Kliegl, 2012).

```
(GM <- mean(table3$M)) # Grand Mean
```

```
## [1] 20
```

Table 9

*Summary statistics for simulated data with one between-subjects factor with four levels.*

Factor F	N data points	Estimated means	Standard deviations	Standard errors
F1	5	10.0	10.0	4.5
F2	5	20.0	10.0	4.5
F3	5	10.0	10.0	4.5
F4	5	40.0	10.0	4.5

## Repeated contrasts

Arguably, the most popular contrast psychologists and psycholinguists are interested in is the comparison between neighboring levels of a factor. This type of contrast is called REPEATED CONTRAST. In our example, our research question might be whether the frequency level `low` leads to slower response times than frequency level `medium-low`, whether frequency level `medium-low` leads to slower response times than frequency level `medium-high`, and whether frequency level `medium-high` leads to slower response times than frequency level `high`.

REPEATED CONTRASTS are used to implement these comparisons. Consider first how to derive the contrast matrix for REPEATED CONTRASTS, starting out by specifying the hypotheses that are to be tested about the data. Importantly, this again applies the general strategy of how to translate (any) hypotheses about differences between groups or conditions into a set of contrasts, yielding a powerful tool of great value in many research settings. We follow the four-step procedure outlined above.

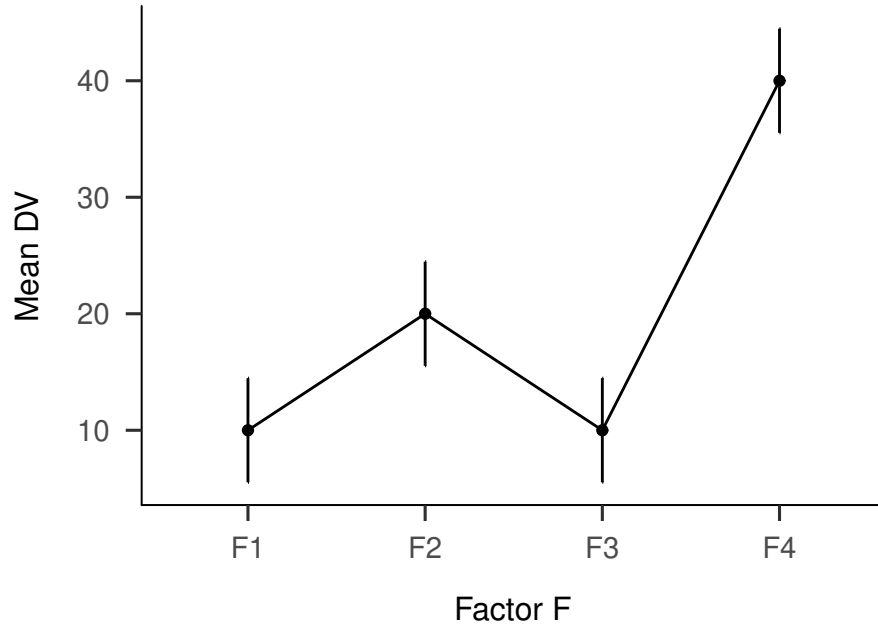


Figure 3. Means and error bars (showing standard errors) for a simulated data-set with one between-subjects factor with four levels.

The first step is to specify our hypotheses, and to write them down in a way such that their weights can be extracted easily. For a four-level factor, the three hypotheses are:

$$H_{0_{2-1}} : -1 \cdot \mu_1 + 1 \cdot \mu_2 + 0 \cdot \mu_3 + 0 \cdot \mu_4 = 0 \quad (23)$$

$$H_{0_{3-2}} : 0 \cdot \mu_1 - 1 \cdot \mu_2 + 1 \cdot \mu_3 + 0 \cdot \mu_4 = 0 \quad (24)$$

$$H_{0_{4-3}} : 0 \cdot \mu_1 + 0 \cdot \mu_2 - 1 \cdot \mu_3 + 1 \cdot \mu_4 = 0 \quad (25)$$

Here, the  $\mu_x$  are the mean response times in condition  $x$ . Each hypothesis gives weights to the different condition means. The first hypothesis ( $H_{0_{2-1}}$ ) tests the difference between condition mean for F2 ( $\mu_2$ ) minus the condition mean for F1 ( $\mu_1$ ), but ignores

condition means for F3 and F4 ( $\mu_3, \mu_4$ ).  $\mu_1$  has a weight of  $-1$ ,  $\mu_2$  has a weight of  $+1$ , and  $\mu_3$  and  $\mu_4$  have weights of 0. As the second step, the vector of weights for the first hypothesis is extracted as `c2vs1 <- c(F1=-1,F2=+1,F3=0,F4=0)`. Next, the same thing is done for the other hypotheses - the weights for all hypotheses are extracted and coded into a *hypothesis matrix* in R:

```
t(HcRE <- rbind(c2vs1=c(F1=-1,F2=+1,F3= 0,F4= 0),
               c3vs2=c(F1= 0,F2=-1,F3=+1,F4= 0),
               c4vs3=c(F1= 0,F2= 0,F3=-1,F4=+1)))
```

```
##      c2vs1 c3vs2 c4vs3
## F1      -1      0      0
## F2       1     -1      0
## F3       0      1     -1
## F4       0      0      1
```

Again, we show the transposed version of the hypothesis matrix (switching rows and columns), but now we leave out the hypothesis for the intercept (we discuss below when this can be neglected).

Next, the new contrast matrix `XcRE` is obtained. This is the contrast matrix  $X_c$  that exactly tests the hypotheses written down above:

```
(XcRE <- ginv2(HcRE))
```

```
##      c2vs1 c3vs2 c4vs3
## F1 -3/4  -1/2  -1/4
## F2  1/4  -1/2  -1/4
## F3  1/4   1/2  -1/4
## F4  1/4   1/2   3/4
```

In the case of the REPEATED CONTRAST, the contrast matrix again looks very different from the hypothesis matrix. In this case, the contrast matrix looks a lot less intuitive than the hypothesis matrix, and if one did not know the associated hypothesis matrix, it seems unclear what the contrast matrix would actually test. To verify this custom-made contrast matrix, we compare it to the REPEATED CONTRAST matrix as generated by the R function `contr.sdif()` in the MASS package (Venables & Ripley, 2002). The resulting contrast matrix is identical to our result:

```
fractions(contr.sdif(4))
```

```
##      2-1  3-2  4-3
## 1 -3/4 -1/2 -1/4
## 2  1/4 -1/2 -1/4
## 3  1/4  1/2 -1/4
## 4  1/4  1/2  3/4
```

Step four in the procedure is to apply REPEATED CONTRASTS to the factor **F** in the example data, and to run a linear model. This allows us to estimate the regression coefficients associated with each contrast. These are compared to the data in Figure 3 to test whether the regression coefficients actually correspond to the differences between successive condition means, as intended.

```
contrasts(simdat3$F) <- XcRE
m2_mr <- lm(DV ~ 1 + F, data=simdat3)
```

The results (see Table 10) show that as expected, the regression coefficients reflect exactly the differences that were of interest: the regression coefficient (“Estimate”) **Fc2vs1** has a value of 10, which exactly corresponds to the difference between condition mean for **F2** (20) minus condition mean for **F1** (10), i.e.,  $20 - 10 = 10$ . Likewise, the regression coefficient **Fc3vs2** has a value of  $-10$ , which corresponds to the difference between condition mean for

Table 10

*Repeated contrasts.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
Fc2vs1	10	[−3, 23]	1.58	.133
Fc3vs2	-10	[−23, 3]	-1.58	.133
Fc4vs3	30	[17, 43]	4.74	< .001

F3 (10) minus condition mean for F2 (20), i.e.,  $10 - 20 = -10$ . Finally, the regression coefficient **Fc4vs3** has a value of 30, which reflects the difference between condition **F4** (40) minus condition **F3** (10), i.e.,  $40 - 10 = 30$ . Thus, the regression coefficients reflect differences between successive or neighboring condition means, and test the corresponding null hypotheses.

To sum up, formally writing down the hypotheses, extracting the weights into a hypothesis matrix, and applying the generalized matrix inverse operation yields a set of contrast coefficients that provide the desired estimates. This procedure is very general: it allows us to derive the contrast matrix corresponding to any set of hypotheses that one may want to test. The four-step procedure described above allows us to construct contrast matrices that are among the standard set of contrasts in R (**REPEATED CONTRASTS** or **SUM CONTRASTS**, etc.), and also allows us to construct non-standard custom contrasts that are specifically tailored to the particular hypotheses one wants to test. The hypothesis matrix and the contrast matrix are linked by the generalized inverse; understanding this link is the key ingredient to understanding contrasts in diverse settings.

## Contrasts in linear regression analysis: The design or model matrix

We have now discussed how different contrasts are created from the hypothesis matrix. However, we have not treated in detail how exactly contrasts are used in a linear model. Here, we will see that the contrasts for a factor in a linear model are just the same thing as continuous numeric predictors (i.e., covariates) in a linear/multiple regression analysis. That is, contrasts are the way to encode discrete factor levels into numeric predictor variables to use in linear/multiple regression analysis, by encoding which differences between factor levels are tested. The contrast matrix  $X_c$  that we have looked at so far has one entry (row) for each experimental condition. For use in a linear model, however, the contrast matrix is coded into a design or model matrix  $X$ , where each individual data point has one row. The design matrix  $X$  can be extracted using the function `model.matrix()`:

```
(contrasts(simdat3$F) <- XcRE) # contrast matrix
```

```
##      c2vs1 c3vs2 c4vs3
## F1 -3/4  -1/2  -1/4
## F2  1/4  -1/2  -1/4
## F3  1/4   1/2  -1/4
## F4  1/4   1/2   3/4
```

```
(covars <- as.data.frame(model.matrix(~ 1 + F, simdat3))) # design matrix
```

```
##      (Intercept) Fc2vs1 Fc3vs2 Fc4vs3
## 1              1 -0.75  -0.5  -0.25
## 2              1 -0.75  -0.5  -0.25
## 3              1 -0.75  -0.5  -0.25
## 4              1 -0.75  -0.5  -0.25
## 5              1 -0.75  -0.5  -0.25
```



## 6	1	0.25	-0.5	-0.25
## 7	1	0.25	-0.5	-0.25
## 8	1	0.25	-0.5	-0.25
## 9	1	0.25	-0.5	-0.25
## 10	1	0.25	-0.5	-0.25
## 11	1	0.25	0.5	-0.25
## 12	1	0.25	0.5	-0.25
## 13	1	0.25	0.5	-0.25
## 14	1	0.25	0.5	-0.25
## 15	1	0.25	0.5	-0.25
## 16	1	0.25	0.5	0.75
## 17	1	0.25	0.5	0.75
## 18	1	0.25	0.5	0.75
## 19	1	0.25	0.5	0.75
## 20	1	0.25	0.5	0.75

For each of the 20 subjects, four numbers are stored in this model matrix. They represent the three values of three predictor variables used to predict response times in the task. Indeed, this matrix is exactly the design matrix  $X$  commonly used in multiple regression analysis, where each column represents one numeric predictor variable (covariate), and the first column codes the intercept term.

To further illustrate this, the covariates are extracted from this design matrix and stored separately as numeric predictor variables in the data-frame:

```
simdat3[,c("Fc2vs1", "Fc3vs2", "Fc4vs3")] <- covars[,2:4]
```

They are now used as numeric predictor variables in a multiple regression analysis:

```
m3_mr <- lm(DV ~ 1 + Fc2vs1 + Fc3vs2 + Fc4vs3, data=simdat3)
```

Table 11

*Repeated contrasts as linear regression.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
Fc2vs1	10	[−3, 23]	1.58	.133
Fc3vs2	-10	[−23, 3]	-1.58	.133
Fc4vs3	30	[17, 43]	4.74	< .001

The results show that the regression coefficients are exactly the same as in the contrast-based analysis shown in the previous section (Table 10). This demonstrates that contrasts serve to code discrete factor levels into a linear/multiple regression analysis by numerically encoding comparisons between specific condition means.

Figure 4 provides an overview of the introduced contrasts.

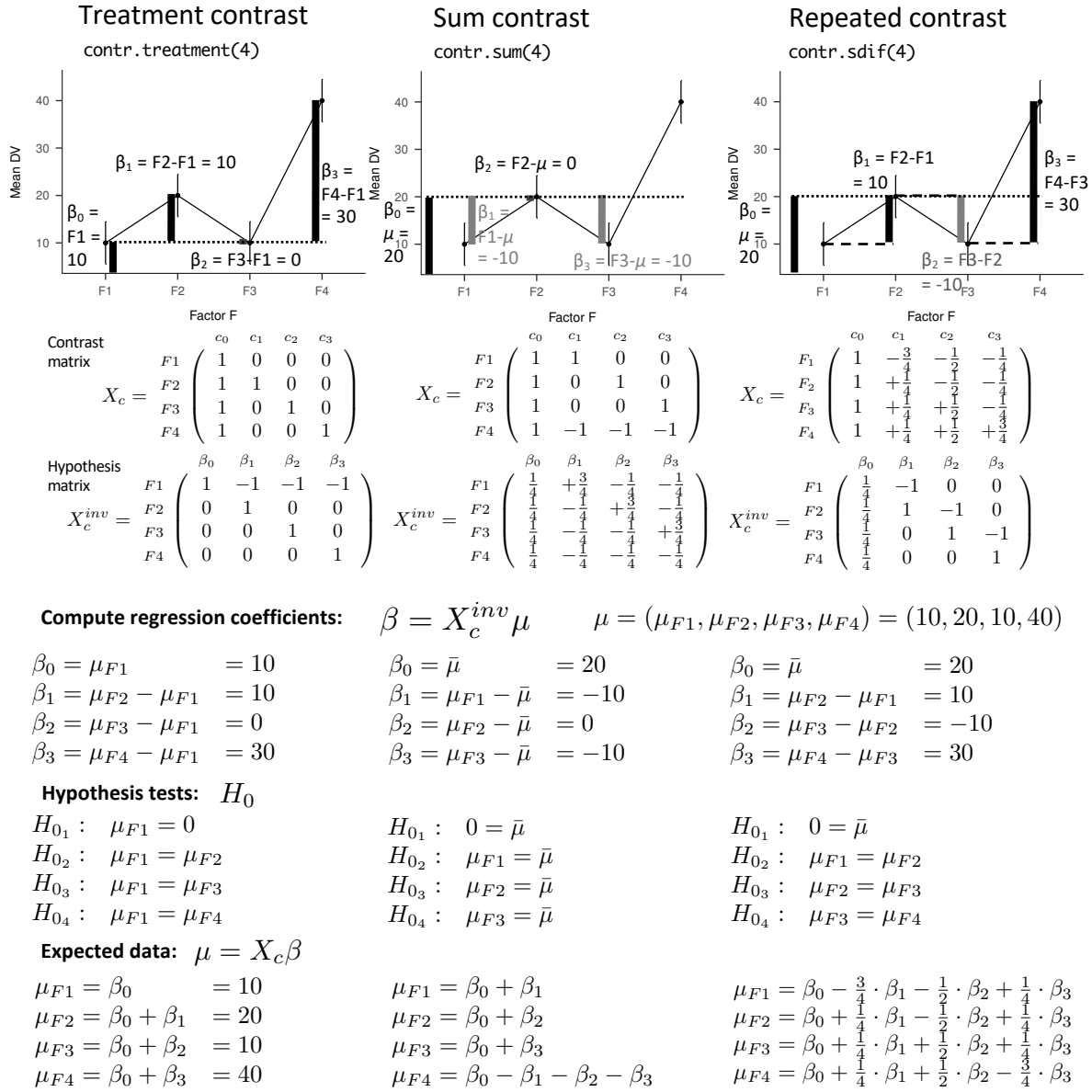


Figure 4. Overview of contrasts including treatment, sum, and repeated contrasts. From top to bottom panels, we illustrate the computation of regression coefficients, show the contrast and hypothesis matrices, formulas for computing regression coefficients, the null hypotheses tested by each coefficient, and formulas for estimated data.

## Polynomial contrasts

POLYNOMIAL CONTRASTS are another option for analyzing factors. Suppose that we expect a linear trend across conditions, where the response increases by a constant magnitude with each successive factor level. This could be the expectation when four levels of a factor reflect decreasing levels of word frequency (i.e., four factor levels: high, medium-high, medium-low, and low word frequency), where one expects the lowest response for high frequency words, and successively higher responses for lower word frequencies. The effect for each individual level of a factor may not be strong enough for detecting it in the statistical model. Specifying a linear trend in a polynomial contrast allows us to pool the whole increase into a single coefficient for the linear trend, increasing statistical power to detect the increase. Such a specification constrains the hypothesis to one interpretable degree of freedom, e.g., a linear increase across factor levels. The larger the number of factor levels, the more parsimonious are POLYNOMIAL CONTRASTS compared to contrast-based specifications as introduced in the previous sections or compared to an omnibus F-test. Going beyond a linear trend, one may also have expectations about quadratic trends. For example, one may expect an increase only among very low frequency words, but no difference between high and medium-high frequency words.

```
Xpol <- contr.poly(4)
(contrasts(simdat3$F) <- Xpol)
```

```
##           .L    .Q    .C
## [1,] -0.671  0.5 -0.224
## [2,] -0.224 -0.5  0.671
## [3,]  0.224 -0.5 -0.671
## [4,]  0.671  0.5  0.224
```

```
m1_mr.Xpol <- lm(DV ~ 1 + F, data=simdat3)
```

Table 12

*Polynomial contrasts.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
F L	18	[8, 27]	4.00	.001
F Q	10	[1, 19]	2.24	.040
F C	13	[4, 23]	3.00	.008

In this example (see Table 12), condition means increase across factor levels in a linear fashion, but the quadratic and cubic trends are also significant.

### Custom contrasts

Sometimes, a hypothesis about a pattern of means takes a form that cannot be expressed by the standard sets of contrasts available in R. For example, a theory or model may make quantitative predictions about the expected pattern of means. Alternatively, prior empirical research findings or logical reasoning may suggest a specific qualitative pattern. Such predictions could be quantitatively constrained when they come from a computational or mathematical model, but when a theory only predicts a qualitative pattern, these predictions can be represented by choosing some plausible values for the means (Baguley, 2012). For example, assume that a theory predicts for the pattern of means presented in Figure 3 that the first two means (for F1 and F2) are identical, but that means for levels F3 and F4 increase linearly. One starts approximating a contrast by giving a potential expected outcome of means, such as  $M = c(10, 10, 20, 30)$ . It is possible to turn these predicted means into a contrast by centering them (i.e., subtracting the mean of 17.5):  $M = c(-7.5,$

-7.5, 2.5, 12.5). This already works as a contrast. It is possible to further simplify this by dividing by 2.5, which yields  $M = c(-3, -3, 1, 5)$ . We will use this contrast in a regression model. Notice that if you have  $I$  conditions, you can specify  $I - 1$  contrasts. However, it is also possible to specify less contrasts than this maximum number of  $I - 1$ . The example below illustrates this point:

```
(contrasts(simdat3$F) <- cbind(c(-3, -3, 1, 5)))
```

```
##      [,1]
```

```
## [1,]  -3
```

```
## [2,]  -3
```

```
## [3,]   1
```

```
## [4,]   5
```

```
C <- model.matrix(~ 1 + F, data=simdat3)
```

```
simdat3$Fcust <- C[, "F1"]
```

```
m1_mr.Xcust <- lm(DV ~ 1 + Fcust, data=simdat3)
```

Table 13

*Custom contrasts.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (18)	<i>p</i>
Intercept	20	[14, 26]	6.97	< .001
Fcust	3	[1, 5]	3.15	.006

For cases where a qualitative pattern of means can be realized with more than one set of quantitative values, Baguley (2012) notes that often the precise numbers may not be decisive. He also suggests selecting the simplest set of integer numbers that matches the desired pattern.

### What makes a good set of contrasts?

Contrasts decompose ANOVA omnibus F tests into several component comparisons (Baguley, 2012). Orthogonal contrasts decompose the sum of squares of the F test into additive independent subcomponents, which allows for clarity in interpreting each effect. As mentioned earlier, for a factor with  $I$  levels one can make  $I - 1$  comparisons. For example, in a design with one factor with two levels, only one comparison is possible (between the two factor levels). More generally, if we have a factor with  $I_1$  and another factor with  $I_2$  levels, then the total number of conditions is  $I_1 \times I_2 = \nu$  (not  $I_1 + I_2$ !), which implies a maximum of  $\nu - 1$  contrasts.

For example, in a design with one factor with three levels, A, B, and C, in principle one could make three comparisons (A vs. B, A vs. C, B vs. C). However, after defining an intercept, only two means can be compared. Therefore, for a factor with three levels, we define two comparisons within one statistical model. F tests are nothing but combinations, or bundles of contrasts. F tests are less specific and they lack focus, but they are useful when the hypothesis in question is vague. However, a significant F test leaves unclear what effects the data actually show. Contrasts are very useful to test specific effects in the data.

One critical precondition for contrasts is that they implement different hypotheses that are not collinear, that is, that none of the contrasts can be generated from the other contrasts by linear combination. For example, the contrast  $c1 = c(1,2,3)$  can be generated from the contrast  $c2 = c(3,4,5)$  simply by computing  $c2 - 2$ . Therefore, contrasts  $c1$  and  $c2$  cannot be used simultaneously. That is, each contrast needs to encode some independent information about the data. Otherwise, the model cannot be estimated, and the `lm()` function gives an error, indicating that the design matrix is "rank deficient".

There are (at least) two criteria to decide what a good contrast is. First, *orthogonal contrasts* have advantages as they test mutually independent hypotheses about the data (see

Dobson & Barnett, 2011, sec. 6.2.5, p. 91 for a detailed explanation of orthogonality).

Second, it is crucial that contrasts are defined in a way such that they answer the research questions. This second point is crucial. One way to accomplish this, is to use the hypothesis matrix to generate contrasts, as this ensures that one uses contrasts that exactly test the hypotheses of interest in a given study.

### Centered contrasts

Contrasts are often constrained to be centered, such that the individual contrast coefficients  $c_i$  for different factor levels  $i$  sum to 0:  $\sum_{i=1}^I c_i = 0$ . This has advantages when testing interactions with other factors or covariates (we discuss interactions between factors below). All contrasts discussed here are centered except for the TREATMENT CONTRAST, in which the contrast coefficients for each contrast do not sum to zero:

```
colSums(contr.treatment(4))
```

```
## 2 3 4
```

```
## 1 1 1
```

Other contrasts, such as REPEATED CONTRASTS, are centered and the contrast coefficients for each contrast sum to 0:

```
colSums(contr.sdif(4))
```

```
## 2-1 3-2 4-3
```

```
## 0 0 0
```

The contrast coefficients mentioned above appear in the contrast matrix. By contrast, the weights in the hypothesis matrix are always centered. This is also true for the TREATMENT CONTRAST. The reason is that they code hypotheses, which always relate to comparisons between conditions or bundles of conditions. The only exception are the weights



for the intercept, which always sum to 1 in the hypothesis matrix. This is done to ensure that when applying the generalized matrix inverse, the intercept results in a constant term with values of 1 in the contrast matrix. That the intercept is coded by a column of 1s in the contrast matrix accords to convention as it provides a scaling of the intercept coefficient that is simple to interpret. An important question concerns whether (or when) the intercept needs to be considered in the generalized matrix inversion, and whether (or when) it can be ignored. This question is closely related to the concept of orthogonal contrasts, a concept we turn to below.

## Orthogonal contrasts

Two centered contrasts  $c_1$  and  $c_2$  are orthogonal to each other if the following condition applies. Here,  $i$  is the  $i$ -th cell of the vector representing the contrast.

$$\sum_{i=1}^I c_{1,i} \cdot c_{2,i} = 0 \quad (26)$$

Orthogonality can be determined easily in R by computing the correlation between two contrasts. Orthogonal contrasts have a correlation of 0. Contrasts are therefore just a special case for the general case of predictors in regression models, where two numeric predictor variables are orthogonal if they are un-correlated.

For example, coding two factors in a  $2 \times 2$  design (we return to this case in a section on ANOVA below) using SUM CONTRASTS, these sum contrasts and their interaction are orthogonal to each other:

```
(Xsum <- cbind(F1=c(1,1,-1,-1), F2=c(1,-1,1,-1), F1xF2=c(1,-1,-1,1)))
```

```
##      F1 F2 F1xF2
## [1,]  1  1      1
```

```
## [2,]  1 -1   -1
## [3,] -1  1   -1
## [4,] -1 -1    1
```

```
cor(Xsum)
```

```
##          F1 F2 F1xF2
## F1         1  0     0
## F2         0  1     0
## F1xF2      0  0     1
```

Notice that the correlations between the different contrasts (i.e., the off-diagonals) are exactly 0. SUM CONTRASTS coding one multi-level factor, however, are not orthogonal to each other:

```
cor(contr.sum(4))
```

```
##      [,1] [,2] [,3]
## [1,]  1.0  0.5  0.5
## [2,]  0.5  1.0  0.5
## [3,]  0.5  0.5  1.0
```

Here, the correlations between individual contrasts, which appear in the off-diagonals, deviate from 0, indicating non-orthogonality. The same is also true for TREATMENT and REPEATED CONTRASTS:

```
cor(contr.sdif(4))
```

```
##      2-1  3-2  4-3
## 2-1  1.000  0.577  0.333
## 3-2  0.577  1.000  0.577
## 4-3  0.333  0.577  1.000
```

```
cor(contr.treatment(4))
```

```
##           2           3           4
## 2  1.000 -0.333 -0.333
## 3 -0.333  1.000 -0.333
## 4 -0.333 -0.333  1.000
```

Orthogonality of contrasts plays a critical role when computing the generalized inverse. In the inversion operation, orthogonal contrasts are converted independently from each other. That is, the presence or absence of another orthogonal contrast does not change the resulting weights. In fact, for orthogonal contrasts, applying the generalized matrix inverse to the hypothesis matrix simply produces a scaled version of the hypothesis matrix into the contrast matrix (for mathematical details see [Appendix E](#)).

The crucial point here is the following. As long as contrasts are fully orthogonal, and as long as one does not care about the scaling of predictors, it is not necessary to use the generalized matrix inverse, and one can code the contrast matrix directly. However, when scaling is of interest, or when non-orthogonal or non-centered contrasts are involved, then the generalized inverse formulation of the hypothesis matrix is needed to specify contrasts correctly.

### The role of the intercept in non-centered contrasts

A related question concerns whether the intercept needs to be considered when computing the generalized inverse for a contrast. It turns out that considering the intercept is necessary for contrasts that are not centered. This is the case for TREATMENT CONTRASTS which are not centered; e.g., the treatment contrast for two factor levels `c1vs0` = `c(0,1)`:  $\sum_i c_i = 0 + 1 = 1$ . One can actually show that the formula to determine whether

contrasts are centered (i.e.,  $\sum_i c_i = 0$ ) is the same formula as the formula to test whether a contrast is “orthogonal to the intercept”. Remember that for the intercept, all contrast coefficients are equal to one:  $c_{1,i} = 1$  (here,  $c_{1,i}$  indicates the vector of contrast coefficients associated with the intercept). We enter these contrast coefficient values into the formula testing whether a contrast is orthogonal to the intercept (here,  $c_{2,i}$  indicates the vector of contrast coefficients associated with some contrast for which we want to test whether it is “orthogonal to the intercept”):  $\sum_i c_{1,i} \cdot c_{2,i} = \sum_i 1 \cdot c_{2,i} = \sum_i c_{2,i} = 0$ . The resulting formula is:  $\sum_i c_{2,i} = 0$ , which is exactly the formula for whether a contrast is centered. Because of this analogy, TREATMENT CONTRASTS can be viewed to be ‘not orthogonal to the intercept’. This means that the intercept needs to be considered when computing the generalized inverse for treatment contrasts. As we have discussed above, when the intercept is included in the hypothesis matrix, the weights for this intercept term should sum to one, as this yields a column of ones for the intercept term in the contrast matrix.

### A closer look at hypothesis and contrast matrices

#### Inverting the procedure: From a contrast matrix to the associated hypothesis matrix

One important point to appreciate about the generalized inverse matrix operation is that applying the inverse twice yields back the original matrix. It follows that applying the inverse operation twice to the hypothesis matrix  $H_c$  yields back the original hypothesis matrix:  $(H_c^{inv})^{inv} = H_c$ . For example, let us look at the hypothesis matrix of a REPEATED CONTRAST:

```
t(HcRE <- rbind(c2vs1=c(F1=-1,F2= 1,F3= 0),
                c3vs1=c( 0,   -1,   1)))
```

```
##      c2vs1 c3vs1
```

```
## F1    -1    0
## F2     1   -1
## F3     0    1

t(ginv2(ginv2(HcRE)))
```

```
##      c2vs1 c3vs1
## F1 -1      0
## F2  1     -1
## F3  0      1
```

It is clear that applying the generalized inverse twice to the hypothesis matrix yields back the same matrix. This also implies that taking the contrast matrix  $X_c$  (i.e.,  $X_c = H_c^{inv}$ ), and applying the generalized inverse operation, gets back the hypothesis matrix  $X_c^{inv} = H_c$ .

Why is this of interest? This means that if one has a given contrast matrix, e.g., one that is provided by standard software packages, or one that is described in a research paper, then one can apply the generalized inverse operation to obtain the hypothesis matrix. This will tell us exactly which hypotheses were tested by the given contrast matrix.

As an example, let us take a closer look at this using the TREATMENT CONTRAST. Let's start with a TREATMENT CONTRAST for a factor with three levels F1, F2, and F3. Adding a column of 1s adds the intercept (`int`):

```
(XcTr <- cbind(int=1,contr.treatment(3)))
```

```
##   int 2 3
## 1   1 0 0
## 2   1 1 0
## 3   1 0 1
```

The next step is to apply the generalized inverse operation:

```
t(ginv2(XcTr))
```

```
##   int 2  3
## 1   1 -1 -1
## 2   0  1  0
## 3   0  0  1
```

This shows the hypotheses that the TREATMENT CONTRASTS test, by extracting the weights from the hypothesis matrix. The first contrast (`int`) has weights `cH00 <- c(1, 0, 0)`. Writing this down as a formal hypothesis test yields:

$$H_{0_0} : 1 \cdot \mu_1 + 0 \cdot \mu_2 + 0 \cdot \mu_3 = 0 \quad (27)$$

That is, the first contrast tests the hypothesis  $H_{0_0} : \mu_1 = 0$  that the mean of the first factor level  $\mu_1$  is zero. As the factor level F1 was defined as the baseline condition in the treatment contrast, this means that for treatment contrasts, the intercept captures the condition mean of the baseline condition. This is the exact same result that was shown at the beginning of this paper, when first introducing treatment contrasts (see equation 6).

We also extract the weights for the other contrasts from the hypothesis matrix. The weights for the second contrast are `cH01 <- c(-1, 1, 0)`. This is written as a formal hypothesis test:

$$H_{0_1} : -1 \cdot \mu_1 + 1 \cdot \mu_2 + 0 \cdot \mu_3 = 0 \quad (28)$$

The second contrast tests the difference in condition means between the first and the second factor level, i.e., it tests the null hypothesis that the difference in condition means of

the second minus the first factor levels is zero  $H_{0_1} : \mu_2 - \mu_1 = 0$ .

We also extract the weights for the last contrast, which are `cH02 <- c(-1, 0, 1)`, and write them as a formal hypothesis test:

$$H_{0_2} : -1 \cdot \mu_1 + 0 \cdot \mu_2 + 1 \cdot \mu_3 = 0 \quad (29)$$

This contrast tests the difference between the third (F3) and the first (F1) condition means, and tests the null hypothesis that the difference is zero:  $H_{0_2} : \mu_3 - \mu_1 = 0$ . These results correspond to what we know about treatment contrasts, i.e., that treatment contrasts test the difference of each group to the baseline condition. They demonstrate that it is possible to use the generalized inverse to learn about the hypotheses that a given set of contrasts tests.

### **The importance of the intercept when transforming between the contrast matrix and the hypothesis matrix in non-centered contrasts**

The above example of the treatment contrast also demonstrates that it is vital to consider the intercept when doing the transformation between the contrast matrix and the hypothesis matrix. Let us have a look at what the hypothesis matrix looks like when the intercept is ignored in the inversion:

```
(XcTr <- contr.treatment(3))
```

```
##    2 3
## 1 0 0
## 2 1 0
## 3 0 1
```

```
t(Hc <- ginv2(XcTr))
```

```
##    2 3
## 1 0 0
## 2 1 0
## 3 0 1
```

Now, the hypothesis matrix looks very different. In fact it looks just the same as the contrast matrix. However, the hypothesis matrix does not code any reasonable hypotheses or comparisons any more: The first contrast now tests the hypothesis that the condition mean for F2 is zero,  $H_{0_1} : \mu_2 = 0$ . The second contrast now tests the hypothesis that the condition mean for F3 is zero,  $H_{0_2} : \mu_3 = 0$ . However, we know that these are the wrong hypotheses for the TREATMENT CONTRAST when the intercept is included in the model. This demonstrates that it is important to consider the intercept in the generalized inverse. As explained earlier in the section on non-/orthogonal contrasts, this is important for contrasts that are not centered. For centered contrasts, such as the SUM CONTRAST or the REPEATED CONTRAST, including or excluding the intercept does not change the results.

### The hypothesis matrix and contrast matrix in matrix form

**Matrix notation for the contrast matrix.** We have discussed above the relation of contrasts to linear/multiple regression analysis, i.e., that contrasts encode numeric predictor variables (covariates) for testing comparisons between discrete conditions in a linear/multiple regression model. The introduction of TREATMENT CONTRASTS had shown that a contrast can be used as the predictor  $x$  in the linear regression equation  $y = \beta_0 + \beta_1 \cdot x$ . To repeat: in the treatment contrast, if  $x$  is 0 for the baseline condition, the predicted data is  $\beta_0 + \beta_1 \cdot 0 = \beta_0$ , indicating the intercept  $\beta_0$  is the prediction for the mean of the baseline factor level (F1). If  $x$  is 1 (F2), then the predicted data is  $\beta_0 + \beta_1 \cdot 1 = \beta_0 + \beta_1$ .



Both of these predictions for conditions  $x = 0$  and  $x = 1$  are summarized in a single equation using matrix notation (also see equation (2) in the introduction; cf. Bolker, 2018). Here, the different possible values of  $x$  are represented in the contrast matrix  $X_c$ .

$$X_c = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad (30)$$

This matrix has one row for each condition/group of the study, i.e., here, it has 2 rows. The matrix has two columns. The second column (i.e., on the right-hand side) contains the treatment contrast with  $x_1 = 0$  and  $x_2 = 1$ . The first column of  $X_c$  contains a column of 1s, which indicate that the intercept  $\beta_0$  is added in each condition.

Multiplying<sup>6</sup> this contrast matrix  $X_c$  with the vector of regression coefficients  $\beta = (\beta_0, \beta_1)$  containing the intercept  $\beta_0$  and the effect of the factor  $x$ ,  $\beta_1$  (i.e., the slope), yields the expected response times for conditions F1 and F2,  $y_1$  and  $y_2$ , which here correspond to the condition means,  $\mu_1$  and  $\mu_2$ :

$$X_c \beta = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 1 \cdot \beta_0 + 0 \cdot \beta_1 \\ 1 \cdot \beta_0 + 1 \cdot \beta_1 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_0 + \beta_1 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \mu \quad (31)$$

---

<sup>6</sup>Matrix multiplication is defined as follows. Consider a matrix  $X$  with three rows and two columns  $X = \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ x_{3,1} & x_{3,2} \end{pmatrix}$ , and a vector with two entries  $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ . The matrix  $X$  can be multiplied with

the vector  $\beta$  as follows:  $X\beta = \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ x_{3,1} & x_{3,2} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} x_{1,1} \cdot \beta_0 + x_{1,2} \cdot \beta_1 \\ x_{2,1} \cdot \beta_0 + x_{2,2} \cdot \beta_1 \\ x_{3,1} \cdot \beta_0 + x_{3,2} \cdot \beta_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = y_{3,1}$

Multiplying an  $n \times p$  matrix with another  $p \times m$  matrix will yield an  $n \times m$  matrix. If the number of columns of the first matrix is not the same as the number of rows of the second matrix, matrix multiplication is undefined.

More compactly:

$$\mu = X_c \beta \quad (32)$$

This matrix formulation can be implemented in R. Consider again the simulated data displayed in Figure 2 with two factor levels F1 and F2, and condition means of  $\mu_1 = 0.8$  and  $\mu_2 = 0.4$ . The TREATMENT CONTRAST codes condition F1 as the baseline condition with  $x = 0$ , and condition F2 as  $x = 1$ . As shown in Table 2, the estimated regression coefficients were  $\beta_0 = 0.8$  and  $\beta_1 = -0.4$ . The contrast matrix  $X_c$  can now be constructed as follows:

```
(XcTr <- cbind(int=c(F1=1,F2=1), c2vs1=c(0,1)))
```

```
##      int c2vs1
## F1    1      0
## F2    1      1
```

The regression coefficients can be written as:

```
(beta <- c(0.8,-0.4))
```

```
## [1]  0.8 -0.4
```

```
## convert to a 2x1 vector:
```

```
beta <- matrix(beta,ncol = 1)
```

Multiplying the  $2 \times 2$  contrast matrix with the estimated regression coefficients, a  $2 \times 1$  vector, gives predictions of the condition means:

```
XcTr %*% beta # The symbol %*% indicates matrix multiplication
```

```
##      [,1]
## F1    0.8
```

```
## F2  0.4
```

As expected, the matrix multiplication yields the condition means.

**Using the generalized inverse to estimate regression coefficients.** One key question remains unanswered by this representation of the linear model: Although in our example we know the contrast matrix  $X_c$  and the condition means  $\mu$ , both the condition means  $\mu$  and the regression coefficients  $\beta$  are unknown, and need to be estimated from the data  $y$  (this is what the command `lm()` does).

Can we use the matrix notation  $\mu = X_c\beta$  (cf. equation 32) to estimate the regression coefficients  $\beta$ ? That is, can we re-formulate the equation, by writing the regression coefficients  $\beta$  on one side of the equation, such that we can compute  $\beta$ ? Intuitively, what needs to be done for this, is to “divide by  $X_c$ ”. This would yield  $1 \cdot \beta = \beta$  on the right side of the equation, and “one divided by  $X_c$ ” times  $\mu$  on the left hand side. That is, this would allow us to solve for the regression coefficients  $\beta$  and would provide a formula to compute them. Indeed, the generalized matrix inverse operation does for matrices exactly what we intuitively refer to as “divide by  $X_c$ ”, that is, it computes the inverse of a matrix such that pre-multiplying the inverse of a matrix with the matrix yields 1:  $X_c^{inv} \cdot X_c = 1$  (where 1 is the identity matrix, with 1s on the diagonal and off-diagonal 0s). For example, we take the contrast matrix of a TREATMENT CONTRAST for a factor with two levels, and we compute the related hypothesis matrix using the generalized inverse: Pre-multiplying the contrast matrix with its inverse yields the identity matrix, that is, a matrix where the diagonals are all 1 and the off-diagonals are all 0:

```
XcTr
```

```
##      int c2vs1
## F1    1      0
## F2    1      1
```

```
ginv2(XcTr)
```

```
##          F1 F2
## int      1  0
## c2vs1 -1  1
```

```
fractions( ginv2(XcTr) %*% XcTr )
```

```
##          int c2vs1
## int      1  0
## c2vs1 0  1
```

That is, multiplying equation 32 by  $X_c^{inv}$ , yields (for details see Appendix D):

$$X_c^{inv} \mu = X_c^{inv} X_c \beta \quad (33)$$

$$X_c^{inv} \mu = \beta \quad (34)$$

This shows that the generalized matrix inverse actually allows to estimate regression coefficients from the data. This is done by (matrix) multiplying the hypothesis matrix (i.e., the inverse contrast matrix) with the condition means:  $\hat{\beta} = X_c^{inv} \mu$ . Importantly, this derivation ignored residual errors in the regression equation. For a full derivation see the Appendix D.

Consider again the simple example of a TREATMENT CONTRAST for a factor with two levels.

```
XcTr
```

```
##          int c2vs1
## F1      1      0
```

```
## F2    1    1
```

Inverting the contrast matrix yields the hypothesis matrix.

```
(HcTr <- ginv2(XcTr))
```

```
##          F1 F2
## int      1  0
## c2vs1 -1  1
```

So far, we always displayed the resulting hypothesis matrix with rows and columns switched using matrix transpose (R function `t()`) to make it more easily readable. That is, each column represented one contrast, and each row represented one condition. However, in the present paragraph we discuss the matrix notation of the linear model, and we therefore show the hypothesis matrix in its original, untransposed form, where each row represents one contrast/hypothesis and each column represents one factor level. That is, the first row of `HcTr` (`int = c(1,0)`) encodes the intercept and the null hypothesis that the condition mean `F1` is zero. The second row `cH01 = c(-1,1)` encodes the null hypothesis that the condition mean of `F2` is identical to the condition mean `F1`.

When we multiply this hypothesis matrix  $H_c$  with the observed condition means  $\mu$ , this yields the regression coefficients. For illustration we use the condition means from our first example of the treatment contrast (see Figure 2), encoded in the data frame `table1` in variable `table1$M`. The resulting regression coefficients are the same values as in the `lm` command presented in Table 2:

```
mu <- table1$M
HcTr %*% mu # %*% indicates matrix multiplication.
```

```
##          [,1]
## int      0.8
```

```
## c2vs1 -0.4
```

**Estimating regression coefficients and testing hypotheses using condition means.** We explained above that the hypothesis matrix contains weights for the condition means to define the hypothesis that a given contrast tests. For example, the hypothesis matrix from a treatment contrast was used for a factor with two levels:

```
HcTr
```

```
##          F1 F2
## int      1  0
## c2vs1   -1  1
```

The first row codes the weights for the intercept. This encodes the following hypothesis:

$$H_{0_0} : 1 \cdot \mu_1 + 0 \cdot \mu_2 = \mu_1 = 0 \quad (35)$$

In fact, writing this down as a hypothesis involves a short-cut: What the hypothesis matrix actually encodes are weights for how to combine condition means to compute regression coefficients. Our null hypotheses are then tested as  $H_{0_x} : \beta_x = 0$ . For the present example, estimates for the regression coefficient for the intercept are:

$$\hat{\beta}_0 = 1 \cdot \mu_1 + 0 \cdot \mu_2 = \mu_1 \quad (36)$$

For the example data-set:

$$\hat{\beta}_0 = 1 \cdot 0.8 + 0 \cdot 0.4 = 0.8 \quad (37)$$

This is exactly the same value that the `lm()` command showed. A second step tests the null hypothesis that the regression coefficient is zero, i.e.,  $H_{0_0} : \beta_0 = 0$ .

The same analysis is done for the slope, which is coded in the second row of the hypothesis matrix. The hypothesis is expressed as:

$$H_{0_1} : -1 \cdot \mu_1 + 1 \cdot \mu_2 = \mu_2 - \mu_1 = 0 \quad (38)$$

This involves first computing the regression coefficient for the slope:

$$\hat{\beta}_0 = -1 \cdot \mu_1 + 1 \cdot \mu_2 = \mu_2 - \mu_1 \quad (39)$$

For the example data-set, this yields:

$$\hat{\beta}_1 = -1 \cdot 0.8 + 1 \cdot 0.4 = -0.4 \quad (40)$$

Again, this is the same value for the slope as given by the command `lm()`. A second step tests the hypothesis that the slope is zero:  $H_{0_1} : \beta_1 = 0$ .

To summarize, we write the formulas for both regression coefficients in a single equation. A first step writes down the hypothesis matrix:

$$H_c = X_c^{inv} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \quad (41)$$

Here, the first row contains the weights for the intercept (`int = c(1,0)`), and the second row contains the weights for the slope (`c2vs1 = c(-1,1)`). Multiplying this hypothesis matrix with the average response times in conditions F1 and F2,  $\mu_1$  and  $\mu_2$ , yields

estimates of the regression coefficients:  $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$ . They show that the intercept  $\beta_0$  is equal to the average value of F1,  $\mu_1$ , and the slope  $\beta_1$  is equal to the difference in average values between F2 minus F1,  $\mu_2 - \mu_1$ :

$$X_c^{inv} \mu = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 1 \cdot \mu_1 + 0 \cdot \mu_2 \\ -1 \cdot \mu_1 + 1 \cdot \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 - \mu_1 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \hat{\beta} \quad (42)$$

Or in short (see equation 56):

$$\hat{\beta} = X_c^{inv} \mu \quad (43)$$

These analyses show the important result that the hypothesis matrix is used to compute regression coefficients from the condition means. This important result is derived from the matrix formulation of the linear model (for details see Appendix D). This is one reason why it is important to understand the matrix formulation of the linear model.

**The design or model matrix and its generalized inverse.** Until now we have been applying the generalized inverse to the hypothesis matrix to obtain the contrast matrix  $X_c$ . The contrast matrix contains one row per condition. When performing a linear model analysis, the contrast matrix is translated into a so-called *design or model matrix*  $X$ . This matrix contains one row for every data point in the vector of data  $y$ . As a consequence, the same condition will appear more than once in the design matrix  $X$ .

The key point to note here is that the generalized inverse operation can not only be applied to the contrast matrix  $X_c$  but also to the design matrix  $X$ , which contains one row per data point. The important result is that taking the generalized inverse of this design



matrix  $X$  and multiplying it with the raw data (i.e., the dependent variable) also yields the regression coefficients:

```
contrasts(simdat$F) <- c(0,1)
data.frame(X <- model.matrix( ~ 1 + F,simdat)) # obtain design matrix X
```

```
##      X.Intercept. F1
## 1           1  0
## 2           1  0
## 3           1  0
## 4           1  0
## 5           1  0
## 6           1  1
## 7           1  1
## 8           1  1
## 9           1  1
## 10          1  1
```

```
(Xinv <- ginv2(X)) # take generalized inverse of X
```

```
##           1      2      3      4      5      6      7      8      9      10
## (Intercept) 1/5  1/5  1/5  1/5  1/5      0      0      0      0      0
## F1          -1/5 -1/5 -1/5 -1/5 -1/5  1/5  1/5  1/5  1/5  1/5
```

```
(y <- simdat$DV) # raw data
```

```
## [1] 0.997 0.847 0.712 0.499 0.945 0.183 0.195 0.608 0.556 0.458
```

```
Xinv %*% y # matrix multiplication
```

```
##           [,1]
```

```
## (Intercept)  0.8
## F1          -0.4
```

The generalized inverse automatically generates covariates that perform the averaging across the individual data points per condition. E.g., the estimate of the intercept  $\hat{\beta}_0$  is computed from the  $i$  observations of the dependent variable  $y_i$  with the formula:

$$\hat{\beta}_0 = \frac{1}{5} \cdot y_1 + \frac{1}{5} \cdot y_2 + \frac{1}{5} \cdot y_3 + \frac{1}{5} \cdot y_4 + \frac{1}{5} \cdot y_5 + 0 \cdot y_6 + 0 \cdot y_7 + 0 \cdot y_8 + 0 \cdot y_9 + 0 \cdot y_{10} = \frac{1}{5} \cdot \sum_{i=1}^5 y_i \quad (44)$$

which expresses the formula for estimating  $\mu_1$ , the mean response time in condition F1.

### The variance explained by each predictor and effect size statistics for linear contrasts

Should one include all predictors in a model? And what is the effect size of a contrast? To answer these questions, it is important to understand the concept of variance explained. This is explained next.

### Sum of squares and $r_{alerting}^2$ as measures of variability and effect size

The sum of squares is a measure of the variability in the data. The total sum of squares in a data set equals the sum of squared deviations of individual data points  $y_i$  from their mean  $\bar{y}$ :  $SS_{total} = \sum_i (y_i - \bar{y})^2$ . This can be partitioned into different components, which add up to the total sum of squares. One component is the sum of squares associated with the residuals, that is the sum of squared deviations of  $i$  individual observed data points  $y_i$  from the value predicted by a linear model  $y_{pred}$ :  $SS_{residuals} = \sum_i (y_i - y_{pred})^2$ ; another component, that we are interested in here, is the sum of squares associated with a certain

factor, which is the squared deviation of  $j$  condition means  $\mu_j$  from their mean  $\bar{\mu}$ , where each squared deviation is multiplied by the number of data points  $n_j$  that go into its computation:

$$SS_{effect} = \sum_j n_j \cdot (\mu_j - \bar{\mu})^2.$$

The sum of squares for a factor ( $SS_{effect}$ ) can be further partitioned into the contributions of  $k$  different linear contrasts:  $SS_{effect} = \sum_k SS_{contrast\ k}$ .<sup>7</sup> A measure for the effect size of a linear contrast is the proportion of variance that it explains (i.e., sum of squares *contrast*) from the total variance explained by the factor (i.e., sum of squares *effect*), that is,  $r_{alerting}^2 = SS_{contrast}/SS_{effect}$  (Baguley, 2012; Rosenthal, Rosnow, & Rubin, 2000). This is computed by entering each contrast as an individual predictor into a linear model, and by then extracting the corresponding sum of squares from the output of the `anova()` function. This is illustrated with the example of POLYNOMIAL CONTRASTS.

First, we assign POLYNOMIAL CONTRASTS to the factor F, extract the numeric predictor variables using the R function `model.matrix()`, and use these as covariates in a linear model analysis.

```
(contrasts(simdat3$F) <- contr.poly(4))
```

```
##           .L      .Q      .C
## [1,] -0.671  0.5 -0.224
## [2,] -0.224 -0.5  0.671
## [3,]  0.224 -0.5 -0.671
## [4,]  0.671  0.5  0.224
```

---

<sup>7</sup>The  $SS_{contrast}$  can be computed as follows:  $SS_{contrast} = \frac{(\sum_j c_j \mu_j)^2}{\sum_j c_j^2 / n_j}$ , where  $j$  is the number of factor levels,  $c_j$  are the contrast coefficients for each factor level,  $\mu_j$  are the condition means, and  $n_j$  is the number of data points per factor level  $j$ .

```
simdat3X <- model.matrix(~ 1 + F, data=simdat3)
simdat3[,c("cLinear", "cQuadratic", "cCubic")] <- simdat3X[,2:4]
m1_mr.Xpol2 <- lm(DV ~ 1 + cLinear + cQuadratic + cCubic, data=simdat3)
```

Next, this model is analyzed using the R function `anova()`. This yields the sum of squares (Sum Sq) explained by each of the covariates.

```
(aovModel <- anova(m1_mr.Xpol2))
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: DV
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## cLinear      1   1600    1600      16 0.0010 **
## cQuadratic    1    500     500       5 0.0399 *
## cCubic        1    900     900       9 0.0085 **
## Residuals   16   1600     100
```

```
## ---
```

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

```
# SumSq contrast
```

```
SumSq <- aovModel[1:3, "Sum Sq"]
```

```
names(SumSq) <- c("cLinear", "cQuadratic", "cCubic")
```

```
SumSq
```

```
##      cLinear cQuadratic      cCubic
```

```
##      1600         500         900
```

Summing across the three contrasts that encode the factor F allows us to compute the total effect sum of squares associated with it. Now, everything is available that is needed to

compute the  $r^2_{alerting}$  summary statistic: dividing the individual sum of squares by the total effect size of squares yields  $r^2_{alerting}$ .

```
# SumSq effect
```

```
sum(SumSq)
```

```
## [1] 3000
```

```
# r2 alerting
```

```
round(SumSq / sum(SumSq), 2)
```

```
##      cLinear cQuadratic      cCubic
##      0.53      0.17      0.30
```

The results show that the expected linear trend explains 53% of the variance in condition means of factor F. Based on the statistical test shown in the anova output, the linear trend has a significant effect on the dependent variable. However, the effect size analysis shows it does not explain the full pattern of results, as nearly half the variance associated with factor F remains unexplained by it, whereas other effects, namely non-linear trends, seem to contribute to explaining the effect of factor F. This situation is a so-called *ecumenical* outcome (Abelson & Prentice, 1997), where the a priori contrast (linear trend) is only one of several contrasts explaining the factor's effect. The fact that the a priori linear contrast is significant but has a  $r^2_{alerting}$  clearly smaller than 1, suggests that other contrasts seem to contribute to the effect of factor F. In a so-called *canonical* outcome, to the contrary,  $r^2_{alerting}$  for a contrast would approach 1, such that no other additional contrasts are needed to explain the effect of factor F.

$r^2_{alerting}$  is useful for comparing the relative importance of different contrasts for a given data-set. However, it is not a general measure of effect size, such as  $\eta^2$ , which is computed using the function call `etasq(Model)` from the package `heplots` ( $\eta^2_{partial}$ ) (Friendly, 2010).

### **Adding all contrasts associated with a factor**

The above discussion shows that the total variability associated with a factor can be composed into contributions from individual contrasts. This implies, that even if only one of the contrasts associated with a factor is of interest in an analysis, it still makes sense to include the other contrasts associated with the factor. For example, when using polynomial contrasts, even if only the linear trend is of interest in an analysis, it still makes sense to include the contrasts for the quadratic and cubic trends. This is because there is a total amount of variance associated with each factor. One can capture all of this variance by using polynomial contrasts up to  $I - 1$  degrees. That is, for a two-level factor, only a linear trend is tested. For a three-level factor, we test a linear and a quadratic trend. For a four-level factor, we additionally test a cubic trend. Specifying all these  $I - 1$  polynomial trends allows us to capture all the variance associated with the factor. In case one is not interested in the cubic trend, one can simply leave this out of the model. However, this would also mean that some of the variance associated with the factor could be left unexplained. This unexplained variance would be added to the residual variance, and would impair our ability to detect effects in the linear (or in the quadratic) trend.

This shows that if the variance associated with a factor is not explained by contrasts, then this unexplained variance will increase the residual variance, and reduce the statistical power to detect the effects of interest. It is therefore good practice to code contrasts that capture all the variance associated with a factor.

### **Examples of contrast coding in a factorial design with two factors**

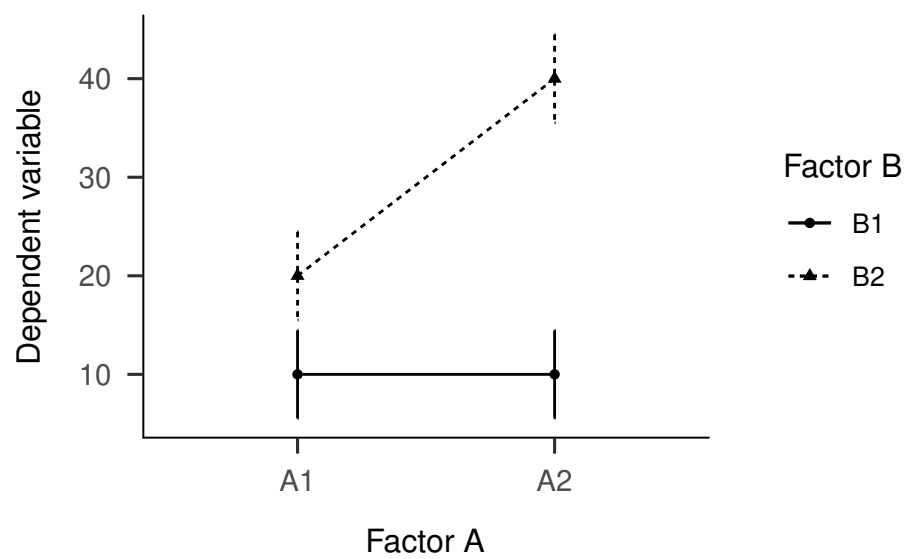
Let us assume that the exact same four means that we have simulated above actually come from an  $A(2) \times B(2)$  between-subject-factor design rather than an  $F(4)$  between-subject-factor design. We simulate the data as shown below in Table 14 and Figure

5. The means and standard deviations are exactly the same as in Figure 3.

Table 14

*Summary statistics for a two-by-two between-subjects factorial design.*

Factor A	Factor B	N data	Means	Std. dev.	Std. errors
A1	B1	5	10.0	10.0	4.5
A1	B2	5	20.0	10.0	4.5
A2	B1	5	10.0	10.0	4.5
A2	B2	5	40.0	10.0	4.5



*Figure 5.* Means and error bars (showing standard errors) for a simulated data-set with a two-by-two between-subjects factorial design.



```
# generate 2 times 2 between subjects data:
simdat4 <- mixedDesign(B=c(2,2), W=NULL, n=5, M=M, SD=10, long = TRUE)
names(simdat4)[1:2] <- c("A","B")
head(simdat4)
```

```
##      A  B id      DV
## 1 A1 B1   1 26.195
## 2 A1 B1   2  5.758
## 3 A1 B1   3 11.862
## 4 A1 B1   4  6.321
## 5 A1 B1   5 -0.136
## 6 A1 B2   6 18.380
```

```
table4 <- simdat4 %>% group_by(A, B) %>% # plot interaction
  summarize(N=length(DV), M=mean(DV), SD=sd(DV), SE=SD/sqrt(N))
GM <- mean(table4$M) # Grand Mean
```

## The difference between an ANOVA and a multiple regression

Let's compare the traditional ANOVA with a multiple regression (i.e., using contrasts as covariates) for analyzing these data.

```
# ANOVA: B_A(2) times B_B(2)
m2_aov <- aov(DV ~ A*B + Error(id), data=simdat4)

# MR: B_A(2) times B_B(2)
m2_mr <- lm(DV ~ 1 + A*B, data=simdat4)
```

The results from the two analyses, shown in Tables 15 and 16, are very different. How

Table 15

*Estimated ANOVA model.*

Effect	$F$	$df_1$	$df_2$	$MSE$	$p$	$\hat{\eta}_G^2$
A	5.00	1	16	100.00	.040	.238
B	20.00	1	16	100.00	< .001	.556
A $\times$ B	5.00	1	16	100.00	.040	.238

Table 16

*Estimated regression model.*

Predictor	$Estimate$	95% CI	$t(16)$	$p$
Intercept	10	[1, 19]	2.24	.040
AA2	0	[-13, 13]	0.00	> .999
BB2	10	[-3, 23]	1.58	.133
AA2 $\times$ BB2	20	[1, 39]	2.24	.040

do we see these are different? Notice that it is possible to compute F-values from t-values from the fact that  $F(1, df) = t(df)^2$  (Snedecor & Cochran, 1967) (where  $df$  indicates degrees of freedom). When applying this to the above multiple regression model, the F-value for factor  $A$  (i.e.,  $AA2$ ) is  $0.00^2 = 0$ . This is obviously not the same as in the ANOVA, where the F-value for factor  $A$  is 5. Likewise, in the multiple regression factor  $B$  (i.e.,  $BB2$ ) has an F-value of  $1.58^2 = 2.5$ , which also does not correspond to the F-value for factor  $B$  in the ANOVA of 20. Interestingly, however, the F-value for the interaction is identical in both models, as  $2.24^2 = 5$ .

The reason that the two results are different is that one needs SUM CONTRASTS in the linear model to get the conventional tests from an ANOVA model. (This is true for factors with two levels, but does not generalize to factors with more levels.)

```

# define sum contrasts:
contrasts(simdat4$A) <- contr.sum(2)
contrasts(simdat4$B) <- contr.sum(2)
m2_mr.sum <- lm(DV ~ 1 + A*B, data=simdat4)

# Alternative using covariates
mat_myC <- model.matrix(~ 1 + A*B, simdat4)
simdat4[, c("GM", "FA", "FB", "FAxB")] <- mat_myC
m2_mr.v2 <- lm(DV ~ 1 + FA + FB + FxAxB, data=simdat4)

```

Table 17

*Regression analysis with sum contrasts.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
A1	-5	[-10, 0]	-2.24	.040
B1	-10	[-15, -5]	-4.47	< .001
A1 × B1	5	[0, 10]	2.24	.040

When using SUM CONTRASTS, the results from the multiple regression models (see Tables 17 and 18) are identical to the results from the ANOVA (Table 15). This is visible as the F-value for factor *A* is now  $-2.24^2 = 5$ , for factor *B* it is  $-4.47^2 = 20$ , and for the interaction it is again  $2.24^2 = 5$ . All F-values are now the same as in the ANOVA model.

Next, we reproduce the  $A(2) \times B(2)$  - ANOVA with contrasts specified for the corresponding one-way  $F(4)$  ANOVA, that is by treating the  $2 \times 2 = 4$  condition means as four levels of a single factor *F*. In other words, we go back to the data frame simulated for the analysis of REPEATED CONTRASTS (see section *Further examples of contrasts illustrated with a factor with four levels*). We first define weights for condition means according to our

Table 18

*Defining sum contrasts using the `model.matrix()` function.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
FA	-5	[-10, 0]	-2.24	.040
FB	-10	[-15, -5]	-4.47	< .001
FAxB	5	[0, 10]	2.24	.040

hypotheses, invert this matrix, and use it as the contrast matrix for factor F in a LM. We define weights of  $1/4$  and  $-1/4$ . We do so because (a) we want to compare the mean of two conditions to the mean of two other conditions (e.g., factor A compares  $\frac{F1+F2}{2}$  to  $\frac{F3+F4}{2}$ ). Moreover, (b) we want to use sum contrasts, where the regression coefficients assess half the difference between means. Together (a+b), this yields weights of  $1/2 \cdot 1/2 = 1/4$ . The resulting contrast matrix contains contrast coefficients of  $+1$  or  $-1$ , showing that we successfully implemented sum contrasts. The results, presented in Table 19, are identical to the previous models.

```
t(fractions(HcInt <- rbind(A =c(F1=1/4,F2= 1/4,F3=-1/4,F4=-1/4),
                             B =c(F1=1/4,F2=-1/4,F3= 1/4,F4=-1/4),
                             AxB=c(F1=1/4,F2=-1/4,F3=-1/4,F4= 1/4))))
```

```
##      A      B      AxB
## F1  1/4  1/4  1/4
## F2  1/4 -1/4 -1/4
## F3 -1/4  1/4 -1/4
## F4 -1/4 -1/4  1/4
```

```
(XcInt <- ginv2(HcInt))

##      A  B  AxB
## F1   1   1   1
## F2   1  -1  -1
## F3  -1   1  -1
## F4  -1  -1   1

contrasts(simdat3$F) <- XcInt
m3_mr <- lm(DV ~ 1 + F, data=simdat3)
```

Table 19

*Main effects and interaction: Custom-defined sum contrasts (scaled).*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
FA	-5	[-10, 0]	-2.24	.040
FB	-10	[-15, -5]	-4.47	< .001
FAxB	5	[0, 10]	2.24	.040

This shows that it is possible to specify the contrasts not only for each factor (e.g., here in the  $2 \times 2$  design) separately. Instead, one can also pool all experimental conditions (or design cells) into one large factor (here factor F with 4 levels), and specify the contrasts for the main effects and for the interactions in the resulting one large contrast matrix simultaneously.

## Nested effects

One can specify hypotheses that do not correspond directly to main effects and interaction of the traditional ANOVA. For example, in a  $2 \times 2$  experimental design, where factor  $A$  codes word frequency (low/high) and factor  $B$  is part of speech (noun/verb), one can test the effect of word frequency within nouns and the effect of word frequency within verbs. Formally,  $A_{B1}$  versus  $A_{B2}$  are nested within levels of  $B$ . Said differently, simple effects of factor  $A$  are tested for each of the levels of factor  $B$ . In this version, we test whether there is a main effect of part of speech ( $B$ ; as in traditional ANOVA). However, instead of also testing the second main effect word frequency,  $A$ , and the interaction, we test (1) whether the two levels of word frequency,  $A$ , differ significantly for the first level of  $B$  (i.e., nouns) and (2) whether the two levels of  $A$  differ significantly for the second level of  $B$  (i.e., verbs). In other words, we test whether there are significant differences for  $A$  in each of the levels of  $B$ . Often researchers have hypotheses about these differences, and not about the interaction.

```
t(fractions(HcNes <- rbind(B =c(F1= 1/2,F2=-1/2,F3= 1/2,F4=-1/2),
                             B1xA=c(F1=-1 ,F2= 0 ,F3= 1 ,F4= 0 ),
                             B2xA=c(F1= 0 ,F2=-1 ,F3= 0 ,F4= 1 ))))
```

```
##      B      B1xA B2xA
## F1  1/2    -1     0
## F2 -1/2     0    -1
## F3  1/2     1     0
## F4 -1/2     0     1
```

```
(XcNes <- ginv2(HcNes))
```

```
##      B      B1xA B2xA
## F1  1/2 -1/2     0
```

```
## F2 -1/2    0 -1/2
## F3  1/2   1/2    0
## F4 -1/2    0  1/2

contrasts(simdat3$F) <- XcNes
m4_mr <- lm(DV ~ 1 + F, data=simdat3)
```

Table 20

*Regression model for nested effects.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
FB	-20	[-29, -11]	-4.47	< .001
FB1xA	0	[-13, 13]	0.00	> .999
FB2xA	20	[7, 33]	3.16	.006

Regression coefficients (cf. Table 20) estimate the GM, the difference for the main effect of word frequency (*A*) and the two differences (for *B*; i.e., simple main effects) within levels of word frequency (*A*).

These custom nested contrasts' columns are scaled versions of the corresponding hypothesis matrix. This is the case because the columns are orthogonal. It illustrates the advantage of orthogonal contrasts for the interpretation of regression coefficients: the underlying hypotheses being tested are already clear from the contrast matrix.

There is also a built-in R-formula specification of nested designs. The order of factors in the formula from left to right specifies a top-down order of nesting within levels, i.e., here factor *A* (word frequency) is nested within levels of the factor *B* (part of speech). This (see Table 21) yields the exact same result as our previous result based on custom nested contrasts (cf. Table 20).

```
contrasts(simdat4$A) <- c(-0.5,+0.5)
contrasts(simdat4$B) <- c(+0.5,-0.5)
m4_mr.x <- lm(DV ~ 1 + B / A, data=simdat4)
```

Table 21

*Nested effects: R-formula.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
B1	-20	[-29, -11]	-4.47	< .001
BB1 $\times$ A1	0	[-13, 13]	0.00	> .999
BB2 $\times$ A1	20	[7, 33]	3.16	.006

In cases such as these, where  $A_{B1}$  vs.  $A_{B2}$  are nested within levels of  $B$ , why is it necessary to estimate the effect of  $B$  (part of speech) at all, when the interest might only be in the effect of  $A$  (word frequency) within levels of  $B$  (part of speech)? Setting up a regression model where the main effect of  $B$  (part of speech) is removed yields:

```
# Extract contrasts as covariates from model matrix
mat_myC <- model.matrix(~ 1 + F, simdat3)
fractions(as.matrix(data.frame(mat_myC)))
```

```
##      X.Intercept.  FB    FB1xA FB2xA
## 1          1      1/2 -1/2      0
## 2          1      1/2 -1/2      0
## 3          1      1/2 -1/2      0
## 4          1      1/2 -1/2      0
## 5          1      1/2 -1/2      0
## 6          1     -1/2      0    -1/2
```



```
## 7      1      -1/2    0 -1/2
## 8      1      -1/2    0 -1/2
## 9      1      -1/2    0 -1/2
## 10     1      -1/2    0 -1/2
## 11     1       1/2  1/2    0
## 12     1       1/2  1/2    0
## 13     1       1/2  1/2    0
## 14     1       1/2  1/2    0
## 15     1       1/2  1/2    0
## 16     1      -1/2    0  1/2
## 17     1      -1/2    0  1/2
## 18     1      -1/2    0  1/2
## 19     1      -1/2    0  1/2
## 20     1      -1/2    0  1/2
```

```
# Repeat the multiple regression with covariates
simdat3[, c("GM", "B", "B1_A", "B2_A")] <- mat_myC
m2_mr.myC1 <- lm(DV ~ 1 + B + B1_A + B2_A, data=simdat3)
summary(m2_mr.myC1)$sigma # standard deviation of residual
```

```
## [1] 10
```

```
# Run the multiple regression by leaving out C1
m2_mr.myC2 <- lm(DV ~ 1 + B1_A + B2_A, data=simdat3)
summary(m2_mr.myC2)$sigma # standard deviation of residual
```

```
## [1] 14.6
```

The results (Tables 22 and 23) show that in the fully balanced data-set simulated here, the estimates for  $A_{B1}$  and  $A_{B2}$  are still the same. However, the confidence intervals and

Table 22

*Nested effects: Full model.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
B	-20	[-29, -11]	-4.47	< .001
B1 A	0	[-13, 13]	0.00	> .999
B2 A	20	[7, 33]	3.16	.006

associated p-values are now larger. One loses statistical power because the variance explained by factor  $B$  is now not taken into account in the linear regression any more. As a consequence, the unexplained variance increases the standard deviation of the residual from 10 to 14.55, and increases uncertainty about  $A_{B1}$  and  $A_{B2}$ . In unbalanced or nonorthogonal designs, leaving out the effect of  $B$  from the linear regression can lead to dramatic changes in the estimated slopes.

Of course, we can also ask the reverse question: Are the differences significant for part of speech ( $B$ ) in the levels of word frequency ( $A$ ; in addition to testing the main effect of word frequency,  $A$ ). That is, do nouns differ from verbs for low-frequency words ( $B_{A1}$ ) and do nouns differ from verbs for high-frequency words ( $B_{A2}$ )?

```
t(fractions(HcNes2 <- rbind(A =c(F1=1/2,F2= 1/2,F3=-1/2,F4=-1/2),
                             A1_B=c(F1=1 ,F2=-1 ,F3= 0 ,F4= 0 ),
                             A2_B=c(F1=0 ,F2= 0 ,F3= 1 ,F4=-1 ))))
```

```
##      A      A1_B A2_B
## F1  1/2      1     0
## F2  1/2     -1     0
## F3 -1/2      0     1
```

Table 23

*Nested effects: Without main effect of B B.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (17)	<i>p</i>
Intercept	20	[13, 27]	6.15	< .001
B1 A	0	[−19, 19]	0.00	> .999
B2 A	20	[1, 39]	2.17	.044

```
## F4 -1/2    0    -1
```

```
(XcNes2 <- ginv2(HcNes2))
```

```
##      A      A1_B A2_B
```

```
## F1  1/2  1/2    0
```

```
## F2  1/2 -1/2    0
```

```
## F3 -1/2    0  1/2
```

```
## F4 -1/2    0 -1/2
```

```
contrasts(simdat3$F) <- XcNes2
```

```
m4_mr <- lm(DV ~ 1 + F, data=simdat3)
```

Regression coefficients (cf. Table 24) estimate the GM, the difference for the main effect of word frequency ( $A$ ) and the two part of speech effects (for  $B$ ; i.e., simple main effects) within levels of word frequency ( $A$ ).

## Interactions between contrasts

We have discussed above that in a  $2 \times 2$  experimental design, the results from sum contrasts (see Table 17) are equivalent to typical ANOVA results (see Table 15). In addition, we had also run the analysis with treatment contrasts. It was clear that the results for

Table 24

*Factor B nested within Factor A.*

Predictor	<i>Estimate</i>	95% CI	<i>t</i> (16)	<i>p</i>
Intercept	20	[15, 25]	8.94	< .001
FA	-10	[−19, −1]	-2.24	.040
FA1 B	-10	[−23, 3]	-1.58	.133
FA2 B	-30	[−43, −17]	-4.74	< .001

treatment contrasts (see Table 16) did not correspond to the results from the ANOVA. However, if the results for treatment contrasts do not correspond to the typical ANOVA results, what do they then test? That is, is it still possible to meaningfully interpret the results from the treatment contrasts in a simple  $2 \times 2$  design?

This leads us to a very important principle in interpreting results from contrasts: When interactions between contrasts are included in a model, then the results of one contrast actually depend on the specification of the other contrast(s) in the analysis! This may be counter-intuitive at first. However, it is very important and essential to keep in mind when interpreting results from contrasts. How does this work in detail?

The general rule to remember is that the main effect of one contrast measures its effect at the location 0 of the other contrast(s) in the analysis. What does that mean? Let us consider the example that we use two treatment contrasts in a  $2 \times 2$  design (see results in Table 16). Let's take a look at the main effect of factor A. How can we interpret what this measures or tests? This main effect actually tests the effect of factor A at the "location" where factor B is coded as 0. Factor B is coded as a treatment contrast, that is, it codes a zero at its baseline condition, which is B1. Thus, the main effect of factor A tests the effect of A nested within the baseline condition of B. We take a look at the data presented in Figure 5, what this nested effect should be. Figure 5 shows that the effect of factor A nested

in B1 is 0. If we now compare this to the results from the linear model, it is indeed clear that the main effect of factor A (see Table 16) is exactly estimated as 0. As expected, when factor B is coded as a treatment contrast, the main effect of factor A tests the effect of A nested within the baseline level of factor B.

Next, consider the main effect of factor B. According to the same logic, this main effect tests the effect of factor B at the “location” where factor A is 0. Factor A is also coded as a treatment contrast, that is, it codes its baseline condition A1 as 0. The main effect of factor B tests the effect of B nested within the baseline condition of A. Figure 5 shows that this effect should be 10; this indeed corresponds to the main effect of B as estimated in the regression model for treatment contrasts (see Table 16, the *Estimate* for BB2). As we had seen before, the interaction term, however, does not differ between the treatment contrast and ANOVA ( $t^2 = 2.24^2 = F = 5.00$ ).

How do we know what the “location” is, where a contrast applies? For the treatment contrasts discussed here, it is possible to reason this through because all contrasts are coded as 0 or 1. However, how is it possible to derive the “location” in general? What we can do is to look at the hypotheses tested by the treatment contrasts in the presence of an interaction between them by using the generalized matrix inverse. We go back to the default treatment contrasts. Then we extract the contrast matrix from the design matrix:

```
contrasts(simdat4$A) <- contr.treatment(2)
contrasts(simdat4$B) <- contr.treatment(2)
XcTr <- simdat4 %>%
  group_by(A, B) %>%
  summarise() %>%
  model.matrix(~ 1 + A*B, .) %>%
  as.data.frame() %>% as.matrix()
rownames(XcTr) <- c("A1_B1", "A1_B2", "A2_B1", "A2_B2")
```

```
XcTr
```

```
##      (Intercept) A2 B2 A2:B2
## A1_B1           1  0  0      0
## A1_B2           1  0  1      0
## A2_B1           1  1  0      0
## A2_B2           1  1  1      1
```

This shows the treatment contrast for factors A and B, and their interaction. We now apply the generalized inverse. Remember, when we apply the generalized inverse to the contrast matrix, we obtain the corresponding hypothesis matrix (again, we use matrix transpose for better readability):

```
t(ginv2(XcTr))
```

```
##      (Intercept) A2 B2 A2:B2
## A1_B1    1      -1 -1  1
## A1_B2    0         0  1 -1
## A2_B1    0         1  0 -1
## A2_B2    0         0  0  1
```

As discussed above, the main effect of factor A tests its effect nested within the baseline level of factor B. Likewise, the main effect of factor B tests its effect nested within the baseline level of factor A.

How does this work for sum contrasts? They do not have a baseline condition that is coded as 0. In sum contrasts, however, the average of the contrast coefficients is 0. Therefore, main effects test the average effect across factor levels. This is what is typically also tested in standard ANOVA. Let's look at the example shown in Table 17: given that factor B has a sum contrast, the main effect of factor A is tested as the average across levels of factor B.

Figure 5 shows that the effect of factor A in level B1 is  $10 - 10 = 0$ , and in level B2 it is  $20 - 40 = -20$ . The average effect across both levels is  $(0 - 20)/2 = -10$ . Due to the sum contrast coding, we have to divide this by 2, yielding an expected effect of  $-10/2 = -5$ . This is exactly what the main effect of factor A measures (see Table 17, *Estimate* for A1).

Similarly, factor B tests its effect at the location 0 of factor A. Again, 0 is exactly the mean of the contrast coefficients from factor A, which is coded as a sum contrast. Therefore, factor B tests the effect of B averaged across factor levels of A. For factor level A1, factor B has an effect of  $10 - 20 = -10$ . For factor level A2, factor B has an effect of  $10 - 40 = -30$ . The average effect is  $(-10 - 30)/2 = -20$ , which again needs to be divided by 2 due to the sum contrast. This yields exactly the estimate of  $-10$  that is also reported in Table 17 (*Estimate* for B1).

Again, we look at the hypothesis matrix for the main effects and the interaction:

```
contrasts(simdat4$A) <- contr.sum(2)
contrasts(simdat4$B) <- contr.sum(2)
XcSum <- simdat4 %>%
  group_by(A, B) %>%
  summarise() %>%
  model.matrix(~ 1 + A*B, .) %>%
  as.data.frame() %>% as.matrix()
rownames(XcSum) <- c("A1_B1", "A1_B2", "A2_B1", "A2_B2")
XcSum
```

```
##      (Intercept) A1 B1 A1:B1
## A1_B1           1  1  1      1
## A1_B2           1  1 -1     -1
## A2_B1           1 -1  1     -1
```

```
## A2_B2          1 -1 -1      1
```

```
t(ginv2(XcSum))
```

```
##      (Intercept) A1    B1    A1:B1
## A1_B1  1/4          1/4  1/4  1/4
## A1_B2  1/4          1/4 -1/4 -1/4
## A2_B1  1/4         -1/4  1/4 -1/4
## A2_B2  1/4         -1/4 -1/4  1/4
```

This shows that each of the main effects now does not compute nested comparisons any more, but that they rather test their effect averaged across conditions of the other factor. The averaging involves using weights of  $1/2$ . Moreover, the regression coefficients in the sum contrast measure half the distance between conditions, leading to weights of  $1/2 \cdot 1/2 = 1/4$ .

The general rule to remember from these examples is that when interactions between contrasts are tested, what a main effect of a factor tests depends on the contrast coding of the other factors in the design! The main effect of a factor tests the effect nested within the location zero of the other contrast(s) in an analysis. If another contrast is centered, and zero is the average of this other contrasts' coefficients, then the contrast of interest tests the average effect, averaged across the levels of the other factor. Importantly, this property holds only when the interaction between two contrasts is included into a model. If the interaction is omitted and only main effects are tested, then there is no such "action at a distance".

This may be a very surprising result for interactions of contrasts. However, it is also essential to interpreting contrast coefficients involved in interactions. It is particularly relevant for the analysis of the default TREATMENT CONTRAST, where the main effects test nested effects rather than average effects.



## A priori interaction contrasts

When testing interaction effects, if at least one of the factors involved in an interaction has more than two levels, then omnibus F tests for the interaction are not informative about the source of the interaction. Contrasts are used to test hypotheses about specific differences between differences. Sometimes, researchers may have a hypothesis believed to fully explain the observed pattern of means associated with the interaction between two factors (Abelson & Prentice, 1997). In such situations, it is possible to specify an a priori interaction contrast for testing this hypothesis. Such contrasts are more focused compared to the associated omnibus F test for the interaction term. Maxwell, Delaney, and Kelley (2018), citing Abelson and Prentice (1997), write that "psychologists have historically failed to capitalize on possible advantages of testing a priori interaction contrasts even when theory dictates the relevance of the approach" (p. 346).

Abelson and Prentice (1997) demonstrate interesting example cases for such a priori interaction contrasts, which they argue are often useful in psychological studies, including what they term a "matching" pattern. We here illustrate this "matching" pattern, by simulating data from a hypothetical priming study: we assume a  $Prime(3) \times Target(3)$  between-subject-Factor design, where factor **Prime** indicates three ordinally scaled levels of different prime stimuli `c("Prime1", "Prime2", "Prime3")`, whereas factor **Target** indicates three ordinally scaled levels of different target stimuli `c("Target1", "Target2", "Target3")`. We assume a similarity structure in which **Prime1** primes **Target1**, **Prime2** primes **Target2**, and **Prime3** primes **Target3**. In addition, we assume some similarity between neighboring categories, such that, e.g., **Prime1** also weakly primes **Target2**, **Prime3** also weakly primes **Target2**, and **Prime2** also weakly primes **Target1** and **Target2**. We therefore assume the following pattern of mean response times, for which we simulate response times:

```
M9 <- matrix(c(150,175,200, 175,150,175, 200,175,150),ncol=1)
matrix(M9,nrow=3,dimnames=list(paste0("Prime",1:3),paste0("Target",1:3)))
```

```
##           Target1 Target2 Target3
## Prime1      150      175      200
## Prime2      175      150      175
## Prime3      200      175      150
```

```
set.seed(1)
simdat5 <- mixedDesign(B=c(3,3), W=NULL, n=5, M=M9, SD=50, long = TRUE)
names(simdat5)[1:2] <- c("Prime","Target")
levels(simdat5$Prime) <- paste0("Prime",1:3)
levels(simdat5$Target) <- paste0("Target",1:3)
table5 <- simdat5 %>% group_by(Prime, Target) %>% # plot interaction
  summarize(N=length(DV), M=mean(DV), SD=sd(DV), SE=SD/sqrt(N))
table5 %>% select(Prime,Target,M) %>% spread(key=Target, value=M) %>%
  data.frame()
```

```
##    Prime Target1 Target2 Target3
## 1 Prime1      150      175      200
## 2 Prime2      175      150      175
## 3 Prime3      200      175      150
```

This reflects a "matching" pattern, where relative match between prime and target is the only factor determining differences in response time between conditions. We specify this hypothesis in an a priori contrast. One way to start for formulating a contrast matrix can be to write down the expected pattern of means. Here we assume that we would a priori hypothesize that fully matching primes and targets speed up response times, and that all other combinations lead to equally slow responses:

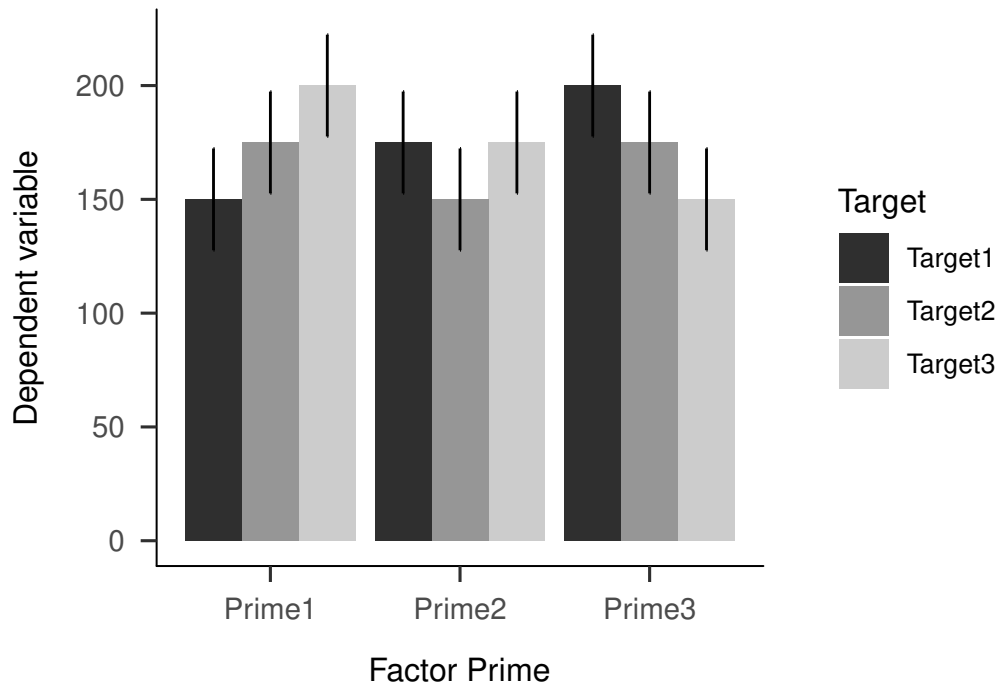


Figure 6. Means and error bars (showing standard errors) for a simulated data-set with a 3 x 3 between-subjects factorial design.

```
meansExp <- rbind(Prime1=c(Target1=150,Target2=200,Target3=200),
                  Prime2=c(      200,      150,      200),
                  Prime3=c(      200,      200,      150))
```

We transform this into a contrast matrix by subtracting the mean response time. We also scale the matrix to ease readability.

```
(XcS <- (meansExp - mean(meansExp)) * 3 / 50)
```

```
##      Target1 Target2 Target3
## Prime1     -2       1       1
## Prime2       1     -2       1
## Prime3       1       1     -2
```

This matrix is orthogonal to the main effects as all the rows and all the columns sum to zero. This is important for contrast matrices for interaction contrasts to ensure that the contrast matrix does not simply capture and test parts of the main effects. If some rows or columns from this matrix do not sum to zero, then it is important to change the interaction contrast matrix to fulfill this requirement. Based on the interaction contrast matrix, we here perform an ANOVA decomposition of the main effects, the omnibus F test for the interaction, and for the effects involved within the interaction:

```
simdat5$cMatch <- ifelse(
  as.numeric(simdat5$Prime)==as.numeric(simdat5$Target),-2,1)
mOmn <- summary(aov(DV~Prime*Target +Error(id), data=simdat5))
mCon <- summary(aov(DV~Prime*Target+cMatch+Error(id), data=simdat5))
```

The outputs from both models are shown in Table 25.

Table 25

*Interaction contrast.*

Effect	$F$	$df_1$	$df_2$	$Sum Sq$	$p$	$\hat{\eta}_G^2$
Prime	0.14	2	36	694	.871	.008
Target	0.14	2	36	694	.871	.008
Prime $\times$ Target	1.39	4	36	13889	.257	.134
Matching contrast	4.44	1	36	11111	.042	.110
Contrast residual	0.37	3	36	2778	.775	.030

The a priori matching contrast is significant, whereas the omnibus F test for the interaction is not. In addition to the matching contrast, there is also a term "contrast residual", which captures all differences in means of the  $3 \times 3$  interaction that are not captured by the matching contrast (and not by the main effects). It provides information

about how much of the interaction is explained by our a priori interaction matching contrast, and whether there is additional pattern in the data, that is not yet explained by this a priori contrast.

Formally speaking, the omnibus F test for the interaction term can be additively decomposed into the a priori matching contrast plus some residual variance associated with the interaction. That is, the 4 degrees of freedom for the omnibus interaction test are decomposed into 1 degree of freedom for the matching contrast and 3 degrees of freedom for the contrast residual:  $4\ df = 1\ df + 3\ df$ . Likewise, the total sum of squares associated with the omnibus interaction test (Sum Sq = 13889) is decomposed into the sum of squares associated with the matching contrast (Sum Sq = 11111) plus the residual contrast sum of squares (Sum Sq = 2778):  $13889 = 11111 + 2778$ . Here, the a priori matching contrast explains a large portion of the variance associated with the interaction of  $r_{alerting}^2 = 11111 / 13889 = 0.80$ , whereas the contribution of the contrast residual is small:  $r_{alerting}^2 = 2778 / 13889 = 0.20$

Interestingly, the results in Table 25 also show a significance test for the contrast residual, which here is not significant. This suggests that our a priori contrast provides a good account of the interaction, a situation that Abelson and Prentice (1997) term a *canonical* outcome. It suggests that the a priori contrast may be a good account of the true pattern of means. Sometimes, the contrast residual term is significant, clearly suggesting that systematic variance is associated with the interaction beyond what is captured in the a priori contrast. In this situation,  $r_{alerting}^2$  for the a priori contrast will often be smaller, whereas  $r_{alerting}^2$  for the contrast residuals may be larger. Abelson and Prentice (1997) term this an *ecumenical* outcome, which suggests that the a priori contrast seems to be missing some of the systematic variance associated with the interaction. For such *ecumenical* outcomes, Abelson and Prentice (1997) suggest it is important for researchers to further explore the data and to search for structure in the residuals to inform future studies and

future a priori expectations.

Even if the contrast residual term is not significant, as is the case here, it can still hide systematic structure. For example, in the present simulations, we assumed that priming did not only facilitate processing of targets from the same category, but also facilitated processing of targets from neighboring categories. We know that response times should gradually increase further towards the diagonal for off-diagonal elements of the  $3 \times 3$  design, and that additional structure is actually contained in the present data — even though the contrast residual is not significant. We could try to directly test for this structure here by formulating a contrast that captures this gradual increase in response times off from the diagonal.

This procedure of testing an a priori interaction contrast in addition to the automatically generated or "residual" interaction term is advantageous (1) because the automatically generated or "residual" interaction term consumes many degrees of freedom and (2) because it is generally difficult to interpret and not particularly scientifically meaningful itself, except for potentially pointing to residual variance in the interaction that needs to be explained.

## Summary

Contrast coding allows us to implement comparisons in a very flexible and general manner. Specific hypotheses can be formulated, coded in hypothesis matrices, and converted into contrast matrices. As Maxwell et al. (2018) put it: “[S]uch tests may preclude the need for an omnibus interaction test and may be more informative than following the more typical path of testing simple effects” (p. 347).

Barring convergence issues or the use of any kind of regularization, all sensible (non-rank-deficient) contrast coding schemes are essentially fitting the same model, since

they are all linear transformations of each other. So the utility of contrast coding is that the researcher can pick and choose how she/he wants to *interpret* the regression coefficients, in a way that's statistically well-founded (e.g., it does not require running many post-hoc tests) without changing the model (in the sense that the fit in data space does not change) or compromising model fit.

The generalized inverse provides a powerful tool to convert between contrast matrices  $X_c$  for experimental designs used in linear models as well as the associated hypothesis matrices  $H_c = X_c^{inv}$ , which define the estimation of regression coefficients and definition of hypothesis tests. Understanding these two representations provides a flexible tool for designing custom contrast matrices and for understanding the hypotheses that are being tested by a given contrast matrix.

### **From omnibus F tests in ANOVA to contrasts in regression models.**

ANOVAs can be reformulated as multiple regressions. This has the advantage that hypotheses about comparisons between specific condition means can be tested. In R, contrasts are flexibly defined for factors via contrast matrices. Functions are available to specify a set of pre-defined contrasts including TREATMENT CONTRASTS (`contr.treatment()`), SUM CONTRASTS (`contr.sum()`), REPEATED CONTRASTS (`MASS::contr.sdif()`), and POLYNOMIAL CONTRASTS (`contr.poly()`). An additional contrast that we did not cover in detail is the HELMERT CONTRAST (`contr.helmert()`). These functions generate contrast matrices for a desired number of factor levels. The generalized inverse (function `ginv()`) is used to obtain hypothesis matrices that inform about how each of the regression coefficients are estimated from condition means and about which hypothesis each of the regression coefficients tests. Going beyond pre-defined contrast functions, the hypothesis matrix is also used to flexibly define custom contrasts tailored to the specific hypotheses of interest. A priori interaction contrasts moreover allow us to formulate and test specific hypotheses about interactions in a design, and to assess their

relative importance for understanding the full interaction pattern.}

**Further readings.** There are many introductions to contrast coding. For further reading we recommend Abelson and Prentice (1997), Baguley (2012) (chapter 15), and Rosenthal et al. (2000). It is also informative to revisit the exchange between Rosnow and Rosenthal (1989), Rosnow and Rosenthal (1996), Abelson (1996), and Petty, Fabrigar, Wegener, and Priester (1996) and the earlier one between Rosnow and Rosenthal (1995) and Meyer (1991). R-specific background about contrasts can be found in section 2.3 of Chambers and Hastie (1992), section 8.3 of Fieller (2016), and section 6.2 of Venables and Ripley (2002). Aside from the default functions in base R, there are also the `CONTRAST` (Kuhn, Weston, Wing, & Forester, 2013) and `MULTCOMP` (Hothorn, Bretz, & Westfall, 2008) packages in R.

### Acknowledgements

We thank Thom Baguley, Adrian Staub, and Dave Kleinschmidt for helpful comments. Thanks to Titus von der Malsburg, Amy Goodwin Davies, and Jan Vanhove for suggestions on improvement, and to Atri Vasisht for help in designing Figure D1. This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich 1287, project number 317633480 (Limits of Variability in Language), and the Forschergruppe 1617 (grant SCHA 1971/2) at the University of Potsdam, and by the Volkswagen Foundation (grant 89 953).



## References

- Abelson, R. P. (1996). Vulnerability of contrast tests to simpler interpretations: An addendum to Rosnow and Rosenthal. *Psychological Science*, 7(4), 242–246.
- Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypothesis. *Psychological Methods*, 2(4), 315.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baguley, T. (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*. Palgrave Macmillan.
- Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. *arXiv Preprint [arXiv:1506.04967](https://arxiv.org/abs/1506.04967)*.
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bolker, B. (2018). <https://github.com/bbolker/mixedmodels-misc/blob/master/notes/contrasts.rmd>.
- Chambers, J. M., & Hastie, T. J. (1992). *Statistical models in S*. New York: Wadsworth & Brooks/Cole. Retrieved from <http://www.stats.ox.ac.uk/pub/MASS4>
- Dobson, A. J., & Barnett, A. (2011). *An introduction to generalized linear models*. CRC press.
- Fieller, N. (2016). *Basics of matrix algebra for statistics with R*. Boca Raton: CRC Press.

- Friendly, M. (2010). HE plots for repeated measures designs. *Journal of Statistical Software*, 37(4), 1–40.
- Friendly, M., Fox, J., & Chalmers, P. (2018). *Matlib: Matrix functions for teaching and learning linear algebra and multivariate statistics*. Retrieved from <https://CRAN.R-project.org/package=matlib>
- Hays, W. L. (1973). *Statistics for the social sciences* (Vol. 410). Holt, Rinehart; Winston New York.
- Heister, J., Würzner, K.-M., & Kliegl, R. (2012). Analysing large datasets of eye movements during reading. *Visual Word Recognition*, 2, 102–130.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363.
- Kliegl, R., Masson, M., & Richter, E. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, 18(5), 655–681.
- Kuhn, M., Weston, S., Wing, J., & Forester, J. (2013). The contrast package.
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2018). *Designing experiments and analyzing data: A model comparison perspective* (3rd ed.). New York: Routledge.
- Meyer, D. L. (1991). Misinterpretation of interaction effects: A reply to Rosnow and Rosenthal. *Psychological Bulletin*, 110(3), 571–573.
- Petty, R. E., Fabrigar, L. R., Wegener, D. T., & Priester, J. R. (1996). Understanding data when interactions are present or hypothesized. *Psychological Science*, 7(4), 247–252.
- Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: Basic concepts and examples. *Mixed-Effects Models in S and S-Plus*, 3–56.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and correlations in behavioral research*. New York: Cambridge University Press.

Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretations of interaction effects. *Psychological Science*, 105(1), 143–146.

Rosnow, R. L., & Rosenthal, R. (1995). “Some things you learn aren’t so”: Cohen’s paradox, Asch’s paradigm, and the interpretation of interaction. *Psychological Bulletin*, 6(1), 3–9.

Rosnow, R. L., & Rosenthal, R. (1996). Contrasts and interactions redux: Five easy pieces. *Psychological Science*, 7(4), 253–257.

Snedecor, G. W., & Cochran, W. G. (1967). *Statistical methods*. Ames, Iowa: Iowa State University Press.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S PLUS* (Fourth.). New York: Springer. Retrieved from <http://www.stats.ox.ac.uk/pub/MASS4>

Appendix A

Glossary

Table A1

*Glossary.*


---

Centered contrasts	The coefficients of a centered contrast sum to zero. If all contrasts (and covariates) in an analysis are centered, then the intercept assesses the grand mean.
Condition mean	The mean of the dependent variable for one factor level or one design cell.
Contrast coefficients	Each contrast consists of a list (vector) of contrast coefficients - there is one contrast coefficient ( $c_i$ ) for each factor level $i$ , which encodes how this factor level contributes in computing the comparison between conditions or bundles of conditions tested by the contrast.
Contrast matrix	Contains contrast coefficients either for one factor, or for all factors in the experimental design; Each condition/group or design cell is represented once. The contrast coefficients indicate the numeric predictor values that are to be used as covariates in the linear model to encode differences between factor levels.
Design / model matrix	A matrix where each data point yields one row and each column codes one predictor variable in a linear model. Specifically, the first column usually contains a row of 1s, which codes the intercept. Each of the other columns contains one contrast or one covariate.
Grand mean	Average of the means of all experimental conditions/groups
Hypothesis matrix	Each column codes one condition/group and each row codes one hypothesis. (We mostly show rows of the hypothesis matrix as columns in this paper.) Each hypothesis consists of weights of how different condition/group means are combined/compared.
Orthogonal contrasts	The coefficients of two contrasts are orthogonal to each other if they have a correlation of zero across conditions. They then represent mutually independent hypotheses about the data.

---

Table A2

*Glossary (continued).*

---

Regression coefficients	Estimates of how strongly a predictor variable or a certain contrast influences the dependent variable. They estimate the effects associated with a predictor variable or a contrast.
Weights	(in a hypothesis matrix); Each row in the hypothesis matrix encodes the hypothesis tested by one contrast. Each row (i.e., each hypothesis) is defined as a list (vector) of weights, where each factor level / condition mean has one associated weight. The weight weighs the contribution of this condition mean to defining the null hypothesis of the associated contrast. For this, the weight is multiplied with its associated condition mean.

---

## Appendix B

### Contrast names

Table B1

*Contrast names.*

R-function	contrast names	description
contr.treatment()	<b>treatment contrasts;</b> dummy contrasts	Compares each of p-1 conditions to a baseline condition.
contr.sum()	<b>sum contrasts;</b> devia- tion contrasts	Compares each of p-1 conditions to the average across conditions. In the case of only 2 factor levels, it tests half the difference between the two conditions.
contr.sum()/2	<b>scaled sum contrasts;</b> effect coding	In the case of two groups, the scaled sum contrast estimates the difference between groups.
contr.sdif()	<b>repeated contrasts;</b> sliding differ- successive difference ence contrasts; simple difference contrasts	Compares neighbouring factor levels, i.e., 2-1, 3-2, 4-3.
contr.poly()	<b>polynomial contrasts;</b> orthogonal polynomial contrasts	Tests e.g., linear or quadratic trends.
contr.helmert()	<b>Helmert contrasts;</b> dif- ference contrasts	The first contrast compares the first two conditions. The second contrast compares the average of the first two conditions to the third condition.
	<b>custom contrasts</b>	Customly defined contrasts, not from the standard set.
	<b>effect coding</b>	All centered contrasts, which give ANOVA-like test results.



## Appendix C

The R-function `mixedDesign()`

The function `mixedDesign()` can be downloaded from the OSF repository

<https://osf.io/7ukf6/>. It allows the researcher to flexibly simulate data from a wide variety of between- and within-subject designs, which are commonly analyzed with (generalized) linear models (ANOVA, ANCOVA, multiple regression analysis) or (generalized) linear mixed-effects models. It can be used to simulate data for different response distributions such as Gaussian, Binomial, Poisson, and others. It involves defining the number of between- and within-subject factors, the number of factor levels, means and standard deviations for each design cell, the number of subjects per cell, and correlations of the dependent variable between within-subject factor levels and how they vary across between-subject factor levels. The following lines of code show a very simple example for simulating a data set for an experiment with one between-subject factor with two levels.

```
(M <- matrix(c(300, 250), nrow=2, ncol=1, byrow=FALSE))

##      [,1]
## [1,]  300
## [2,]  250

set.seed(1)

simexp <- mixedDesign(B=2, W=NULL, n=5, M=M, SD=20)

str(simexp)

## 'data.frame':   10 obs. of  3 variables:
##  $ B_A: Factor w/ 2 levels "A1","A2": 1 1 1 1 1 2 2 2 2 2
##  $ id : Factor w/ 10 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10
##  $ DV : num  320 305 291 270 315 ...
```

```
simexp
```

```
##      B_A id  DV
## 1    A1   1 320
## 2    A1   2 305
## 3    A1   3 291
## 4    A1   4 270
## 5    A1   5 315
## 6    A2   6 228
## 7    A2   7 230
## 8    A2   8 271
## 9    A2   9 266
## 10   A2  10 256
```

The first and second argument of the `mixedDesign()` function specify the numbers and levels of the between- (`B =`) and within- (`W =`) subject factors. The arguments take a vector of numbers (integers). The length of the vector defines the number of factors, and each individual entry in the vector indicates the number of levels of the respective factor. For example, a  $2 \times 3 \times 4$  between-subject design with three between-subject factors, where the first factor has 2 levels, the second has 3 levels, and the third has 4 levels, would be coded as `B = c(2,3,4)`, `W = NULL`. A  $3 \times 3$  within-subject design with two within-subjects factors with each 3 levels would be coded as `B = NULL`, `W = c(3,3)`. A 2 (between)  $\times$  2 (within)  $\times$  3 (within) design would be coded `B = 2`, `W = c(2,3)`.

The third argument (`n =`) takes a single integer value indicating the number of simulated subjects for each cell of the between-subject design.

The next necessary argument (`M =`) takes as input a matrix containing the table of means for the design. The number of rows of this matrix of means is the product of the

number of levels of all between-subject factors. The number of columns of the matrix is the product of the number of levels of the within-subject factors. In the present example, the matrix **M** has just two rows each containing the mean of the dependent variable for one between-subject factor level, that is 300 and 250. Because there is no within-subject factor, the matrix has just a single column. The second data set simulated in the paper contains a between-subject factor with three levels. Accordingly, we specify three means. The `mixedDesign` function generates a data frame with one factor **F** with three levels, one factor **id** with  $3 \text{ (between-subject factor levels)} \times 4 \text{ (n)} = 12$  levels, and a dependent variable with 12 entries.

```
(M <- matrix(c(500, 450, 400), nrow=3, ncol=1, byrow=FALSE))
```

```
##      [,1]
## [1,]  500
## [2,]  450
## [3,]  400
```

```
set.seed(1)
simdat2 <- mixedDesign(B=3, W=NULL, n=4, M=M, SD=20)
names(simdat2)[1] <- "F" # Rename B_A to F(actor)/F(requency)
levels(simdat2$F) <- c("low", "medium", "high")
str(simdat2)
```

```
## 'data.frame':   12 obs. of  3 variables:
## $ F : Factor w/ 3 levels "low","medium",...: 1 1 1 1 2 2 2 2 3 3 ...
## $ id: Factor w/ 12 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ DV: num  497 474 523 506 422 ...
```

```
simdat2
```

```
##          F id  DV
## 1      low  1 497
## 2      low  2 474
## 3      low  3 523
## 4      low  4 506
## 5  medium  5 422
## 6  medium  6 467
## 7  medium  7 461
## 8  medium  8 450
## 9   high  9 414
## 10  high 10 412
## 11  high 11 402
## 12  high 12 371
```

For a 2 (between)  $\times$  2 (within) subject design, we need a  $2 \times 2$  matrix. For multiple between-subject factors, **M** has one row for each cell of the between-subject design, describing the mean for this design cell. For example, in case of a  $2 \times 3 \times 4$  design, the matrix has  $2 \cdot 3 \cdot 4 = 24$  rows. The same is true for multiple within-subject factors, where **M** has one column for each cell of the between-subject design, e.g., in a 4 (within)  $\times$  2 (within) design it has  $4 \cdot 2 = 8$  columns.

The levels of multiple between-subject factors are sorted such that the levels of the first factor vary most slowly, whereas the levels of the later factors vary more quickly. E.g.,

$$\begin{bmatrix}
 B\_A1 & B\_B1 & B\_C1 \\
 B\_A1 & B\_B1 & B\_C2 \\
 B\_A1 & B\_B2 & B\_C1 \\
 B\_A1 & B\_B2 & B\_C2 \\
 B\_A2 & B\_B1 & B\_C1 \\
 B\_A2 & B\_B1 & B\_C2 \\
 \dots & \dots & \dots
 \end{bmatrix} \tag{45}$$

The same logic also applies to the levels of the within-subject factors, which are sorted across columns.

The argument **SD** = takes either a single number specifying the standard deviation of the dependent variable for all design cells, or it takes a table of standard deviations represented as a matrix of the same dimension as the matrix of means, where each element defines the standard deviation of the dependent variable for the given design cell.

For designs with at least one within-subject factor it is necessary to define how the dependent variable correlates between pairs of within-subject factor levels across subjects. For example, when for each subject in a lexical decision task response times are measured once for content words and once for function words, yielding two levels of a within-subject factor, it is necessary to define how the two measurements correlate: that is, for subjects with a very slow response time for content words, will the response times for function words also be very slow, reflecting a high correlation? Or might it be fast as well, such that the individual's response time for content words is uninformative for its response times for function words, reflecting a correlation close to zero? When  $n_k$  within-subject design cells are present in a design, the design defines  $n_r = n_k \cdot (n_k - 1)/2$  correlations. These can be identical across levels of the between-subject factors, but they can also differ. The optional argument **R** = defines these correlations. When assuming that all correlations in a design are

identical, then it is possible to define  $\mathbf{R}$  as a single number (between -1 and +1) specifying this correlation. Under the assumption that correlations differ,  $\mathbf{R}$  takes as input a list of correlation matrices. This list contains one correlation matrix for each combination of between-subject factor levels. Each of these matrices is of dimension  $n_k \cdot n_k$ . The diagonal elements in correlation matrices need to be 1, and the upper and lower triangles must contain the exact same entries and are thus symmetrical.

## Appendix D

## Matrix notation of the linear model and the generalized matrix inverse

This section requires knowledge of some linear algebra; for a review of the important facts, see Fieller (2016) and Van de Geer (1971). Here, we briefly show the derivation for estimating regression coefficients in linear models using matrix notation. The question we address here is: given the data  $y$  and some predictor(s)  $x$ , how to estimate the parameters  $\beta$ ? The next section is adapted from lectures note by Vasishth (2018), available from <https://osf.io/ces89/>.

Consider a deterministic function  $\phi(\mathbf{x}, \beta)$  which takes as input some variable values  $x$  and some fixed values  $\beta$ . A simple example would be

$$y = \beta x \tag{46}$$

Another example with two fixed values  $\beta_0$  and  $\beta_1$  is:

$$y = \beta_0 + \beta_1 x \tag{47}$$

We can rewrite the above equation as follows.

$$\begin{aligned} y &= \beta_0 + \beta_1 x \\ &= \beta_0 \times 1 + \beta_1 x \\ &= \begin{pmatrix} 1 & x \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \\ y &= \phi(x, \beta) \end{aligned} \tag{48}$$

In a statistical model, we don't expect an equation like  $y = \phi(x, \beta)$  to fit all the points exactly. For example, we could come up with an equation that, given a person's weight, we can compute their height:

$$\text{height} = \beta_0 + \beta_1 \text{weight} \quad (49)$$

Given any single value of the weight of a person, we will probably not get a perfectly correct prediction of the height of that person. This leads us to a non-deterministic version of the above function:

$$y = \phi(x, \beta, \varepsilon) = \beta_0 + \beta_1 x + \varepsilon \quad (50)$$

Here,  $\varepsilon$  is an error random variable which is assumed to have some probability density function (specifically, the normal distribution) associated with it. It is assumed to have expectation (mean) 0, and some standard deviation (to be estimated from the data)  $\sigma$ . We can write this statement in compact form as  $\varepsilon \sim N(0, \sigma)$ .

The **general linear model** is a non-deterministic function like the one above ( $^T$  is the transformation operation on a matrix, which converts the rows of a matrix into the columns, and vice versa):

$$Y = f(x)^T \beta + \varepsilon \quad (51)$$

The matrix formulation will be written as below.  $n$  refers to the number of data points (that is,  $Y = y_1, \dots, y_n$ ), and the index  $i$  ranges from 1 to  $n$ .

$$Y = X\beta + \varepsilon \Leftrightarrow y_i = f(x_i)^T \beta + \varepsilon_i, i = 1, \dots, n \quad (52)$$



To make this concrete, suppose we have three data points, i.e.,  $n = 3$ . Then, the matrix formulation is

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \varepsilon \quad (53)$$

$$Y = X\beta + \varepsilon$$

Here,  $f(x_1)^T = (1 \ x_1)$ , and is the first row of the matrix  $X$ ,  $f(x_2)^T = (1 \ x_2)$  is the second row, and  $f(x_3)^T = (1 \ x_3)$  is the third row.

The expectation of  $Y$ ,  $E[Y]$ , is  $X\beta$ .  $\beta$  is a  $p \times 1$  matrix, and  $X$ , the **design matrix**, is an  $n \times p$  matrix.

We now provide a geometric argument for least squares estimation of the  $\beta$  parameters. When we have a deterministic model  $y = \phi(f(x)^T, \beta) = \beta_0 + \beta_1 x$ , this implies a perfect fit to all data points. This is like solving the equation  $Ax = b$  in linear algebra: we solve for  $\beta$  in  $X\beta = y$  using, e.g., Gaussian elimination (Lay, 2005).

But when we have a non-deterministic model  $y = \phi(f(x)^T, \beta, \varepsilon)$ , there is no solution. Now, the best we can do is to get  $Ax$  to be as close an approximation as possible to  $b$  in  $Ax = b$ . In other words, we try to minimize the absolute distance between  $b$  and  $Ax$ :  $|b - Ax|$ .

The goal is to estimate  $\beta$ ; we want to find a value of  $\beta$  such that the observed  $Y$  is as close to its expected value  $X\beta$  as possible. In order to be able to identify  $\beta$  from  $X\beta$ , the linear transformation  $\beta \rightarrow X\beta$  should be one-to-one, so that every possible value of  $\beta$  gives a different  $X\beta$ . This in turn requires that  $X$  be of full rank  $p$ . (Rank refers to the number of linearly independent columns or rows. The row rank and column rank of an  $m \times n$  matrix

will be the same, so we can just talk of rank of a matrix. An  $m \times n$  matrix  $X$  with  $\text{rank}(X) = \min(m, n)$  is called full rank.)

So, if  $X$  is an  $n \times p$  matrix, then it is necessary that  $n \geq p$ . There must be at least as many observations as parameters. If this is not true, then the model is said to be **over-parameterized**.

Assuming that  $X$  is of full rank, and that  $n > p$ ,  $Y$  can be considered a point in  $n$ -dimensional space and the set of candidate  $X\beta$  is a  $p$ -dimensional subspace of this space; see Figure D1. There will be one point in this subspace which is closest to  $Y$  in terms of Euclidean distance. The unique  $\beta$  that corresponds to this point is the **least squares estimator** of  $\beta$ ; we will call this estimator  $\hat{\beta}$ .

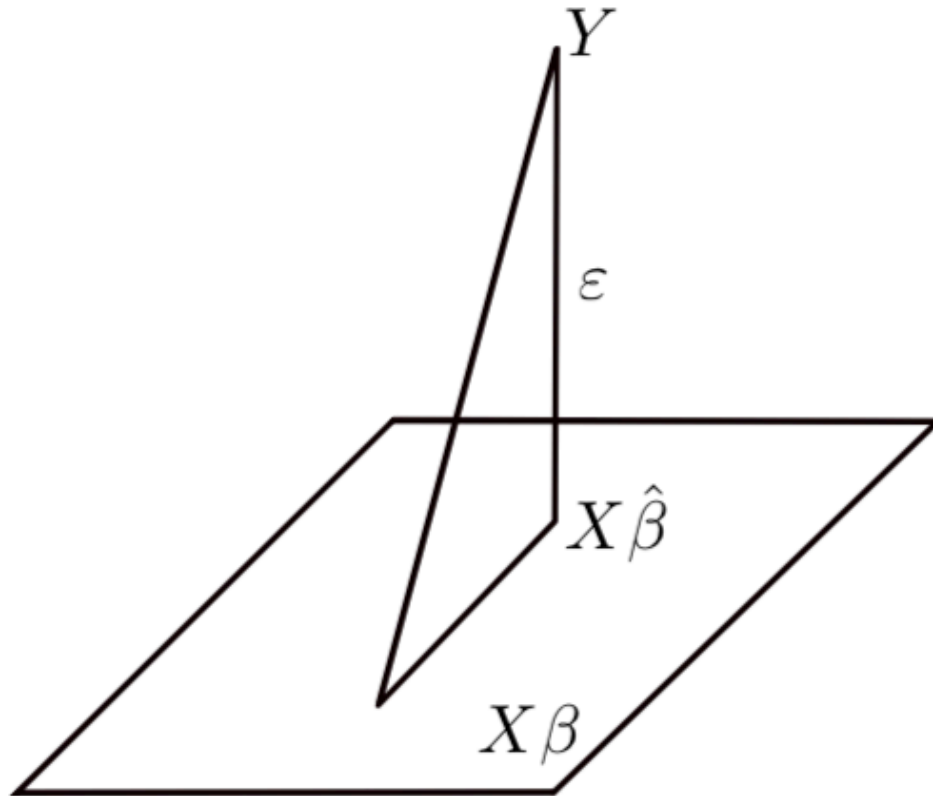


Figure D1. Geometric interpretation of least squares.

Notice that  $\varepsilon = (Y - X\hat{\beta})$  and  $X\beta$  are perpendicular to each other. Because the dot

product of two perpendicular (orthogonal) vectors is 0, we get the result:

$$(Y - X\hat{\beta})^T X\beta = 0 \Leftrightarrow (Y - X\hat{\beta})^T X = 0 \quad (54)$$

Multiplying out the terms, we proceed as follows. One result that we use here is that  $(AB)^T = B^T A^T$  (see Fieller, 2016).

$$\begin{aligned} (Y - X\hat{\beta})^T X &= 0 \\ (Y^T - \hat{\beta}^T X^T)X &= 0 \\ \Leftrightarrow Y^T X - \hat{\beta}^T X^T X &= 0 \\ \Leftrightarrow Y^T X &= \hat{\beta}^T X^T X \\ \Leftrightarrow (Y^T X)^T &= (\hat{\beta}^T X^T X)^T \\ \Leftrightarrow X^T Y &= X^T X \hat{\beta} \end{aligned} \quad (55)$$

This gives us the important result:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (56)$$

This is a key and famous result for linear models. Here, the transformation of the design matrix  $X$  into  $(X^T X)^{-1} X^T$  plays a central role for estimating regression coefficients. Indeed, this term  $(X^T X)^{-1} X^T$  is exactly the generalized matrix inverse of the design matrix  $X$ , which we can compute in R using the command `MASS::ginv()` or `matlib::Ginv()`.

Conceptually, it converts contrast matrices between two representations, where one representation defines the design matrix  $X$  used to define independent variables in the linear regression model  $y = X\beta$ , whereas the other representation, namely the generalized matrix

inverse of the design matrix  $X^+ = (X^T X)^{-1} X^T = X^{inv}$  defines weights for how observed data are combined to obtain estimated regression coefficients via  $\hat{\beta} = X^+ y$  and for what formal hypotheses are tested via the contrasts. Given the important property of the generalized matrix inverse that applying it twice yields back the original matrix, the generalized matrix inversion operation can be used to flip back and forth between these two representations of the design matrix. Fieller (2016) provides a detailed and accessible introduction to this topic.

As one important aspect, the generalized inverse of the design matrix,  $X^+$ , defines weights for each of the  $k$  factor levels of the independent variables. As estimating coefficients for all  $k$  factor levels in addition to the intercept is redundant, the design matrix  $X$  defines coefficients only for  $k - 1$  comparisons. The generalized matrix inversion can be seen as transforming between these two spaces (Venables & Ripley, 2002). A vignette explaining the generalized inverse is available for the `Ginv()` function in the [matlib package](#) (Friendly et al., 2018a).

## Appendix E

## Inverting orthogonal contrasts

When we have a given contrast matrix and we apply the generalized inverse—how exactly does the generalized inverse compute the weights in the hypothesis matrix? That is, what is the formula for computing a single hypothesis weight? We look at this question here for the restricted case that all contrasts in a given design are orthogonal. For this case, we write down the equations for computing each single weight when the generalized inverse is applied. We start out with the definition of the contrast matrix. We use a design with a single factor with three levels with two orthogonal and centered contrasts  $x_1$  and  $x_2$ .

$$X_c = \begin{pmatrix} x_{1,1} & x_{2,1} \\ x_{1,2} & x_{2,2} \\ x_{1,3} & x_{2,3} \end{pmatrix} \quad (57)$$

Under the assumption that the contrasts are orthogonal and centered, we obtain the following formulas for the weights:

$$H_c = X_c^{inv} = \begin{pmatrix} \frac{x_{1,1}}{\sum x_1^2} & \frac{x_{1,2}}{\sum x_1^2} & \frac{x_{1,3}}{\sum x_1^2} \\ \frac{x_{2,1}}{\sum x_2^2} & \frac{x_{2,2}}{\sum x_2^2} & \frac{x_{2,3}}{\sum x_2^2} \end{pmatrix} \quad (58)$$

That is, here each single weight in the hypothesis matrix is computed as  $h_{i,j} = \frac{x_{j,i}}{\sum x_j^2}$ .

## Appendix F

## Information about program version

We used R (Version 3.6.0; R Core Team, 2018) and the R-packages *bindrcpp* (Müller, 2017), *car* (Version 3.0.2; Fox & Weisberg, 2011), *dplyr* (Version 0.8.1; Wickham et al., 2017a), *forcats* (Version 0.4.0; Wickham, 2017a), *ggplot2* (Version 3.1.1; Wickham, 2009), *knitr* (Version 1.22; Xie, 2015), *MASS* (Version 7.3.51.4; Venables & Ripley, 2002), *matlib* (Friendly et al., 2018b), *papaja* (Version 0.1.0.9842; Aust & Barth, 2018), *png* (Version 0.1.7; Urbanek, 2013), *purrr* (Version 0.3.2; Henry & Wickham, 2017), *readr* (Version 1.3.1; Wickham et al., 2017b), *sjPlot* (Version 2.6.3; Lüdtke, 2018), *stringr* (Version 1.4.0; Wickham, 2018), *tibble* (Version 2.1.2; Müller & Wickham, 2018), *tidyr* (Version 0.8.3; Wickham & Henry, 2018), *tidyverse* (Version 1.2.1; Wickham, 2017b), *XML* (Lang & CRAN Team, 2017), and *xtable* (Version 1.8.4; Dahl, 2016) for all our analyses.

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>

Dahl, D. B. (2016). *Xtable: Export tables to latex or html*. Retrieved from <https://CRAN.R-project.org/package=xtable>

Fieller, N. (2016). *Basics of matrix algebra for statistics with R*. Boca Raton: CRC Press.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (Second.). Thousand Oaks CA: Sage. Retrieved from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

Friendly, M., Fox, J., & Chalmers, P. (2018a). *Matlib: Matrix functions for teaching and learning linear algebra and multivariate statistics*. Retrieved from <https://CRAN.R-project.org/package=matlib>

Friendly, M., Fox, J., & Chalmers, P. (2018b). *Matlib: Matrix functions for teaching and learning linear algebra and multivariate statistics*. Retrieved from <https://CRAN.R-project.org/package=matlib>

Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>

Lang, D. T., & CRAN Team. (2017). *XML: Tools for parsing and generating xml within r and s-plus*. Retrieved from <https://CRAN.R-project.org/package=XML>

Lay, D. C. (2005). *Linear algebra and its applications* (Third.). Addison Wesley.

Lüdecke, D. (2018). *SjPlot: Data visualization for statistics in social science*. Retrieved from <https://CRAN.R-project.org/package=sjPlot>

Müller, K. (2017). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from <https://CRAN.R-project.org/package=bindrcpp>

Müller, K., & Wickham, H. (2018). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Urbanek, S. (2013). *Png: Read and write png images*. Retrieved from <https://CRAN.R-project.org/package=png>

Van de Geer, J. P. (1971). *Introduction to multivariate analysis for the social sciences*. Freeman.

Vasishth, S. (2018). *Linear modeling: Lecture notes*. Retrieved from

<https://osf.io/ces89/>

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S PLUS* (Fourth.). New York: Springer. Retrieved from <http://www.stats.ox.ac.uk/pub/MASS4>

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <http://ggplot2.org>

Wickham, H. (2017a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://CRAN.R-project.org/package=forcats>

Wickham, H. (2017b). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>

Wickham, H. (2018). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017a). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>

Wickham, H., & Henry, L. (2018). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from <https://CRAN.R-project.org/package=tidyr>

Wickham, H., Hester, J., & Francois, R. (2017b). *Readr: Read rectangular text data*. Retrieved from <https://CRAN.R-project.org/package=readr>

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from <https://yihui.name/knitr/>