

Supporting information for Analysis of statistical interactions in factorial experiments

Andreas Kitsche, Frank Schaarschmidt

January 19, 2014

Abstract

This manuscript provides some supporting information to the article of Kitsche and Schaarschmidt (2014). Additionally, it presents the application of the `iacontrast()` function in the `statint` package to generate a labelled product type interaction contrast matrix for a detailed analysis of statistical interactions in the two-way layout. Furthermore, the analysis of statistical interactions of some selected data sets from the agronomy and crop sciences are demonstrated in case studies.

Contents

1	Usage of <code>iacontrast()</code> function	2
2	Bush beans data set	3
3	Lettuce data set	5
4	Water deficiency data set	9
5	Oats Field Trial	11

1 Usage of `iacontrast()` function

The function `iacontrast()` generates a labelled product type interaction contrast matrix for the analysis of statistical interactions in two-way layouts. The user has to commit the two factor variables `fa` and `fb` and the types of the two one-way contrasts `typea` and `typeb`. Alternatively, the user can apply a-priori defined contrast one-way contrast matrices `cma` and `cmb`.

The output is a list with two elements: i) `fab` a new factor generated via all combinations of the levels of the two factors, ii) `cmab` labelled interaction contrast matrix

In the following the interaction contrast matrix is generated using the Dunnett type contrast (comparisons to a control) for the first factor and the Tukey type contrast (all pairwise comparisons) for the second factor.

```
# generating factor variables
fa <- factor(rep(rep(LETTERS[1:4], rep(3, 4)), each = 3))
fb <- factor(rep(rep(c(1, 2, 3), rep(3, 3)), times = 4))
# library(devtools)
# install_github(username='AKitsche', repo='statint')
library(statint)
iacontrast(fa = fa, fb = fb, typea = "Dunnett", typeb = "Tukey")
```

	A:1	A:2	A:3	B:1	B:2	B:3	C:1	C:2	C:3	D:1	D:2	D:3
((B - A):2) - ((B - A):1)	1	-1	0	-1	1	0	0	0	0	0	0	0
((B - A):3) - ((B - A):1)	1	0	-1	-1	0	1	0	0	0	0	0	0
((B - A):3) - ((B - A):2)	0	1	-1	0	-1	1	0	0	0	0	0	0
((C - A):2) - ((C - A):1)	1	-1	0	0	0	0	-1	1	0	0	0	0
((C - A):3) - ((C - A):1)	1	0	-1	0	0	0	-1	0	1	0	0	0
((C - A):3) - ((C - A):2)	0	1	-1	0	0	0	0	-1	1	0	0	0
((D - A):2) - ((D - A):1)	1	-1	0	0	0	0	0	0	0	-1	1	0
((D - A):3) - ((D - A):1)	1	0	-1	0	0	0	0	0	0	-1	0	1
((D - A):3) - ((D - A):2)	0	1	-1	0	0	0	0	0	0	0	-1	1

We next demonstrate the usage of prespecified user defined contrasts.

```
#generating user defined one-way contrast matrices
ContrastsA <- matrix(c(1, 1, -1, -1,
                      1, -1, 0, 0,
                      0, 0, 1, -1), nrow=3, byrow=TRUE)
ContrastsB <- matrix(c(-1, 0.5, 0.5,
                      0, 1, -1), nrow=2, byrow=TRUE)
iacontrast(fa=fa, fb=fb, cma=ContrastsA, cmb=ContrastsB)
```

	A:1	A:2	A:3	B:1	B:2	B:3	C:1	C:2	C:3	D:1	D:2	D:3
((A,B - C,D):2,3) - ((A,B - C,D):1)	-1.0	0.5	0.5	-1.0	0.5	0.5	1.0	-0.5	-0.5	1.0	-0.5	-0.5
((A,B - C,D):2) - ((A,B - C,D):3)	0.0	1.0	-1.0	0.0	1.0	-1.0	0.0	-1.0	1.0	0.0	-1.0	1.0
((A - B):2,3) - ((A - B):1)	-1.0	0.5	0.5	1.0	-0.5	-0.5	0.0	0.0	0.0	0.0	0.0	0.0
((A - B):2) - ((A - B):3)	0.0	1.0	-1.0	0.0	-1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
((C - D):2,3) - ((C - D):1)	0.0	0.0	0.0	0.0	0.0	0.0	-1.0	0.5	0.5	1.0	-0.5	-0.5
((C - D):2) - ((C - D):3)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	-1.0	0.0	-1.0	1.0

2 Bush beans data set

This example was published in Petersen (1985, p. 154). The data used within the `statint` package are simulated according to those in Petersen (1985).

This experiment was conducted to investigate the effect of row spacing on the yield of different varieties of bush beans. Due to the different growth habits of the considered varieties it was assumed that the spacing effect differs between the varieties. The selected four varieties differ such that "New Era" and "Big Green" form low, bushy plants and the two varieties "Little Gem" and "Red Lake" form erect plants with few branches. The chosen row spacings were of 20, 40 and 60 cm between rows. A randomized complete block design with four blocks and 12 plots per block was used. The yield of dried beans in kilograms per plot was determined after harvest time.

We first analyse the data using a two-way ANOVA while including the block effect as fixed effect.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety	3	1332.56	444.19	34.22	0.0000
Spacing	2	72.67	36.33	2.80	0.0754
Block	3	341.90	113.97	8.78	0.0002
Variety:Spacing	6	871.00	145.17	11.18	0.0000
Residuals	33	428.35	12.98		

As mentioned in Kitsche and Schaarschmidt (2014) that an alternative ANOVA is also meaningful for this example, where the hierarchical structure between the growth type and the varieties within the growth type is taken into account. The resulting three-factorial ANOVA.

Fore a detailed analysis of the variety by spacing interaction we generate the appropriate one-way contrasts according to Table 6 and 7 in Kitsche and Schaarschmidt (2014). Afterwards, the corresponding product type interaction contrast matrix (see Table 8 in Kitsche and Schaarschmidt (2014)) is generated using the `iacontrast()` function.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	341.90	113.97	8.78	0.0002
Type	1	105.02	105.02	8.09	0.0076
Spacing	2	72.67	36.33	2.80	0.0754
Type:Spacing	2	748.17	374.08	28.82	0.0000
Type:Variety	2	1227.54	613.77	47.28	0.0000
Type:Spacing:Variety	4	122.83	30.71	2.37	0.0730
Residuals	33	428.35	12.98		

```

VarMat    <- matrix(c(0.5, 0.5, -0.5, -0.5,
                      1,  -1,  0,  0,
                      0,  0,  1,  -1), nrow=3, byrow=TRUE)
SpaceMat  <- matrix(c(1, -1, 0,
                      1, 0, -1,
                      0, 1, -1), nrow=3, byrow=TRUE)
InteractionMat <- iacontrast(fa=beans$Variety, fb=beans$Spacing,
                             cma=VarMat, cmb=SpaceMat)

```

The hypothesis formulated in terms of the generated interaction contrasts are further tested using the `glht()` function in the package `multcomp` (Hothorn et al., 2008). The corresponding multiplicity adjusted p-values and simultaneous confidence intervals are available using the functions `summary()` and `confint()`. This will generate Table 9 in Kitsche and Schaarschmidt (2014).

```

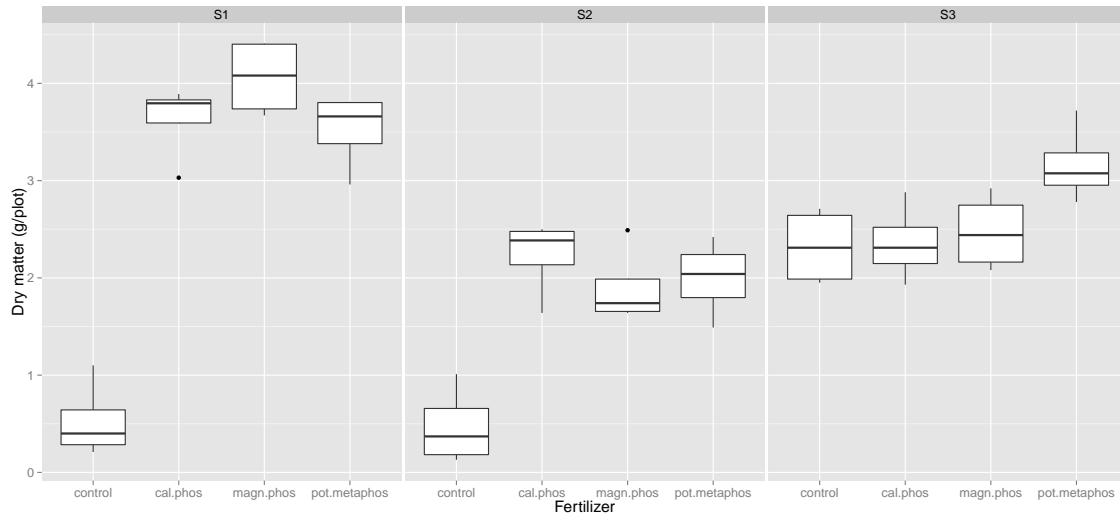
#generating a new factor for the cell means
beans$VarSpace <- InteractionMat$fab
#fitting a cell means model with the new factor
CMM <- lm(Yield ~ VarSpace + Block -1, data = beans)
library(multcomp)
MultTest <- glht(model=CMM,
                 linfct = mcp(VarSpace=InteractionMat$cmab))
summary(MultTest) #calculating adjusted p-values
confint(MultTest) #calculating simultaneous confidence intervals

```

3 Lettuce data set

The data set `lettuce` presents a simulated data set, inspired by the example presented in Bradu and Gabriel (1974). The example is an experiment conducted to analyse the effects of soil type and phosphate fertilizers on lettuce crops. The primary response variable was dry matter measured in grams per plot. Three different soil types, namely S1, S2 and S3, and four different levels of phosphate fertilization (including an untreated control) were investigated in a balanced, completely cross-classified treatment structure, laid out as completely randomized design with four replications per treatment combination.

```
data(lettuce)
str(lettuce)
#reorder Fertilizer levels
lettuce$Fertilizer <- factor(lettuce$Fertilizer,
                             levels=c("control", "cal.phos", "magn.phos", "pot.metaphos"))
```



The fertilizer factor consists of an untreated control group and three different phosphorus fertilizers. Objective for this factor is to estimate the increase in yield that results applying each of the fertilizers compared to the untreated control. The second factor, soil type, contains no further substructure, and interest is in comparing all three soil types among each other. The product type interaction contrasts allow to interpret to what extend the difference in yield between the three phosphate fertilizers compared to the control varies between the three soil types.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Fertilizer	3	26.41	8.80	55.00	0.0000
Soil	2	14.08	7.04	43.98	0.0000
Fertilizer:Soil	6	14.65	2.44	15.25	0.0000
Residuals	36	5.76	0.16		

The interaction contrasts compare the differences of the fertilizer effects between the soil types, where the fertilizer effects of interest are restricted to the differences of each fertilizer to the control group. To test the corresponding hypothesis formulated by the interaction contrasts the function `glht()` in the package `multcomp` (Hothorn et al., 2008) is used. The corresponding multiplicity adjusted p-values and simultaneous confidence intervals are available using the functions `summary()` and `confint()`.

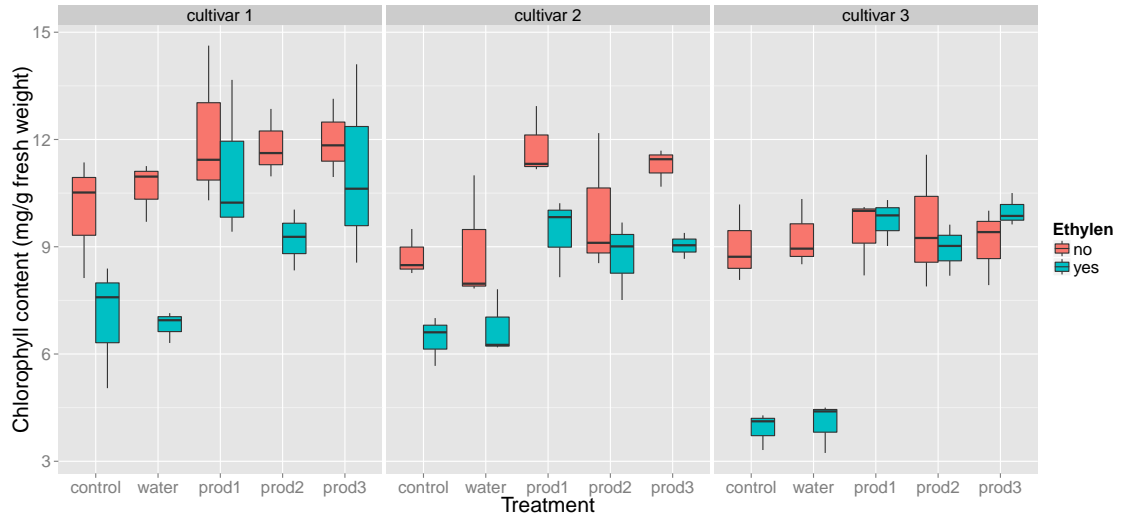
```
#define appropriate user defined contrast matrices
SoilMatrix <- matrix(c(1, -1, 0,
                      1,  0,-1,
                      0,  1,-1), nrow=3, byrow=TRUE)
FertMatrix <- matrix(c(1, -1, 0, 0,
                      1,  0,-1, 0,
                      1,  0, 0,-1), nrow=3, byrow=TRUE)
InteractionContrasts <- iacontrast(fa=lettuce$Fertilizer,
                                  fb=lettuce$Soil,
                                  cma=FertMatrix,
                                  cmb=SoilMatrix)

lettuce$FertSoil <- InteractionContrasts$fab
#fitting a cell means model
CMM <- lm(Weight ~ FertSoil -1, data = lettuce)
#Multiple Comparisons
MultTest <- glht(model=CMM,
                 linfct = mcp(FertSoil=InteractionContrasts$cmab))
summary(MultTest) #calculating adjusted p-values
confint(MultTest) #calculating simultaneous confidence intervals
```

	Estimate	Std.Error	t value	Pr(> t)
((control - cal.phos):S1) - ((control - cal.phos):S2)	-1.34	0.40	-3.36	0.01
((control - cal.phos):S1) - ((control - cal.phos):S3)	-3.06	0.40	-7.65	0.00
((control - cal.phos):S2) - ((control - cal.phos):S3)	-1.72	0.40	-4.30	0.00
((control - magn.phos):S1) - ((control - magn.phos):S2)	-2.10	0.40	-5.25	0.00
((control - magn.phos):S1) - ((control - magn.phos):S3)	-3.38	0.40	-8.45	0.00
((control - magn.phos):S2) - ((control - magn.phos):S3)	-1.28	0.40	-3.21	0.02
((control - pot.metaphos):S1) - ((control - pot.metaphos):S2)	-1.47	0.40	-3.67	0.01
((control - pot.metaphos):S1) - ((control - pot.metaphos):S3)	-2.15	0.40	-5.38	0.00
((control - pot.metaphos):S2) - ((control - pot.metaphos):S3)	-0.69	0.40	-1.71	0.45

Ethylene data set This data set results from an experiment to analyse ethylene receptor blockers for improvement of postharvest characteristics of Pelargonium zonale hybrids. A detailed description of the experiment can be found in Kitsche and Schaarschmidt (2014).

```
data(ethylen)
str(ethylen)
ethylen$Cult <- factor(ethylen$Cult,
                      levels=c(1,2,3),
                      labels=c("cultivar 1","cultivar 2","cultivar 3"))
ethylen$Treatment <- factor(ethylen$Treatment,
                           levels=c("control","water","prod1","prod2","prod3"))
```



The following one-way contrast matrices are used to analyse the ethylene by treatment interaction.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Cult	2	53.22	26.61	16.00	0.0000
Eth	1	98.85	98.85	59.43	0.0000
Treatment	4	156.05	39.01	23.46	0.0000
Cult:Eth	2	0.37	0.19	0.11	0.8938
Cult:Treatment	8	3.05	0.38	0.23	0.9840
Eth:Treatment	4	35.65	8.91	5.36	0.0009
Cult:Eth:Treatment	8	28.73	3.59	2.16	0.0436
Residuals	60	99.80	1.66		

```
#definition of appropriate contrasts
#1. contrast for the different ethylen treatments
ContrMat_Ethylen <- matrix(c(1, -1),
                             nrow=1, byrow=FALSE)

#2. contrasts for Treatments
ContrMat_Treatment <- matrix(c(1/2, 1/2, -1/3, -1/3, -1/3,
                                0, 0, 1, -1, 0,
                                0, 0, 1, 0, -1,
                                0, 0, 0, 1, -1),
                              nrow=4, byrow=TRUE)
```

Since there is also a significant cultivar by treatment by ethylene interaction, the interaction contrast matrix has to be defined manually by using the Kronecker product.

```
ContrMat_Cult <- matrix(c(1/3, 1/3, 1/3),
                        nrow=1)

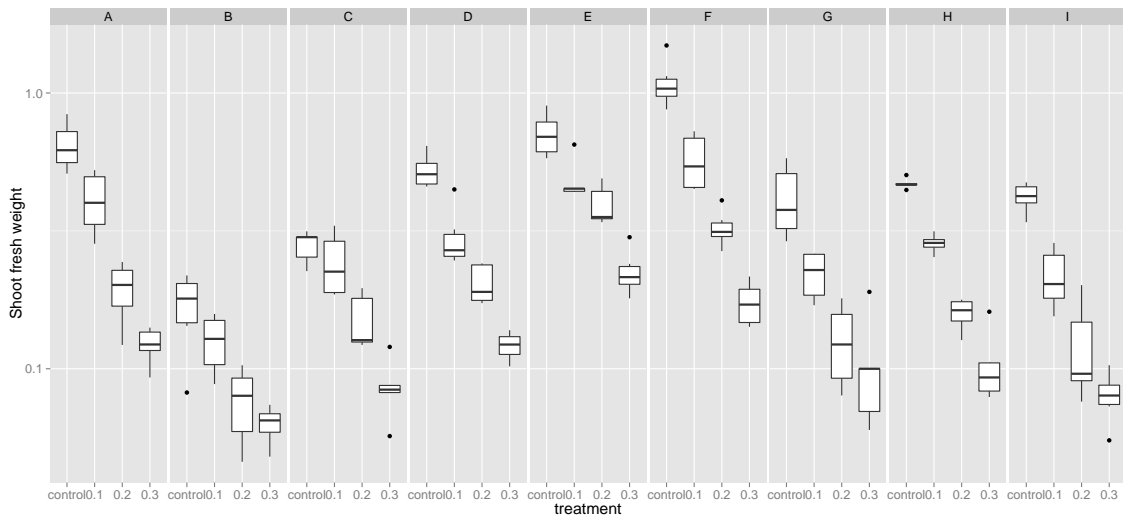
#interaction contrast matrix
InteractionContrasts <- kronecker(ContrMat_Cult,
                                   kronecker(ContrMat_Treatment, ContrMat_Ethylen))

#fitting the appropriate cell means model
ethylen$CultTreatEth <- with(ethylen, Cult:Treatment:Eth)
CMM <- lm(Chloro8 ~ CultTreatEth-1, data=ethylen)
library(multcomp)
MultTest <- glht(CMM, linfct=mcp(CultTreatEth=InteractionContrasts))
summary(MultTest)
confint(MultTest)
```


4 Water deficiency data set

This data set was part of an experiment conducted at the Woody Plant and Propagation Physiology Section in the Institute of Horticultural Production Systems at the Leibniz Universität Hannover in 2013. The goal of the experiment was to investigate a potential differing response among nine varieties of starch potato (*Solanum tuberosum*) on water deficiency stress. The following nine varieties of starch potato were selected for cultivation: A, B, C, D, E, F, G, H and I. Each variety was cultivated under four different water conditions: standard control medium, 0.1M, 0.2M and 0.3M sorbitol medium. An increasing sorbitol concentration was used to induce an increasing water deficiency stress on the plants. For each treatment by variety combination 6 replicates were investigated. The trial was planned in a completely cross-classified treatment structure, in a completely randomized design. A total number of 13 objects were discarded from the analysis because of infection diseases. For each experimental unit the fresh weight of the shoot was measured. To achieve approximately equal variances over the factor levels of the response variable a log transformation was performed.

```
library(statint)
data(potato)
str(potato)
potato$treatment <- factor(potato$treatment,
                           levels=c("control", "0.1", "0.2", "0.3"))
```



The researcher was interested in selecting those varieties with a different response on water deficiency in contrast to the general water deficiency response. Additionally, interest was in determining on which level of water deficiency stress those potential deviations occur.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	3	57.02	19.01	356.98	0.0000
genotype	8	40.96	5.12	96.16	0.0000
treatment:genotype	24	2.93	0.12	2.30	0.0012
Residuals	167	8.89	0.05		

```

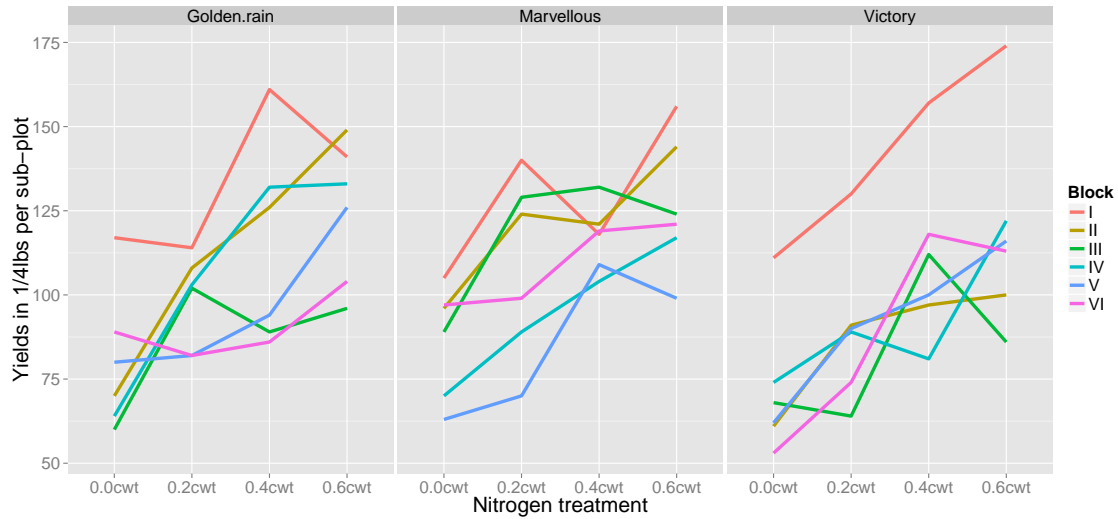
InteractionContrasts <- iacontrast(fa=potato$treatment,
                                   fb=potato$genotype,
                                   typea="Dunnett", typeb="GrandMean")
potato$GenTreat <- InteractionContrasts$fab
CellMeansModel <- lm(log(shoot) ~ GenTreat-1, data=potato)
#conducting user defined multiple comparisons
MultTest <- glht(model=CellMeansModel,
                 linfct = mcp(GenTreat=InteractionContrasts$cmab))
summary(MultTest)#calculating adjusted p-values
ConfInt <- confint(MultTest)#calculating simultaneous confidence intervals

```

	Estimate	Std.Error	t value	Pr(> t)
((0.1 - control):A) - ((0.1 - control):B,C,D,E,F,G,H,I)	0.00	0.13	0.03	1.00
((0.1 - control):B) - ((0.1 - control):A,C,D,E,F,G,H,I)	0.21	0.13	1.69	0.87
((0.1 - control):C) - ((0.1 - control):A,B,D,E,F,G,H,I)	0.32	0.14	2.36	0.36
((0.1 - control):D) - ((0.1 - control):A,B,C,E,F,G,H,I)	-0.09	0.13	-0.75	1.00
((0.1 - control):E) - ((0.1 - control):A,B,C,D,F,G,H,I)	0.09	0.13	0.72	1.00
((0.1 - control):F) - ((0.1 - control):A,B,C,D,E,G,H,I)	-0.18	0.13	-1.41	0.97
((0.1 - control):G) - ((0.1 - control):A,B,C,D,E,F,H,I)	-0.13	0.13	-1.05	1.00
((0.1 - control):H) - ((0.1 - control):A,B,C,D,E,F,G,I)	-0.02	0.14	-0.18	1.00
((0.1 - control):I) - ((0.1 - control):A,B,C,D,E,F,G,H)	-0.21	0.13	-1.64	0.90
((0.2 - control):A) - ((0.2 - control):B,C,D,E,F,G,H,I)	-0.22	0.13	-1.75	0.83
((0.2 - control):B) - ((0.2 - control):A,C,D,E,F,G,H,I)	0.22	0.13	1.73	0.85
((0.2 - control):C) - ((0.2 - control):A,B,D,E,F,G,H,I)	0.36	0.14	2.66	0.19
((0.2 - control):D) - ((0.2 - control):A,B,C,E,F,G,H,I)	0.05	0.13	0.38	1.00
((0.2 - control):E) - ((0.2 - control):A,B,C,D,F,G,H,I)	0.40	0.13	3.21	0.04
((0.2 - control):F) - ((0.2 - control):A,B,C,D,E,G,H,I)	-0.21	0.13	-1.66	0.89
((0.2 - control):G) - ((0.2 - control):A,B,C,D,E,F,H,I)	-0.20	0.13	-1.60	0.92
((0.2 - control):H) - ((0.2 - control):A,B,C,D,E,F,G,I)	-0.10	0.14	-0.70	1.00
((0.2 - control):I) - ((0.2 - control):A,B,C,D,E,F,G,H)	-0.31	0.13	-2.43	0.31
((0.3 - control):A) - ((0.3 - control):B,C,D,E,F,G,H,I)	-0.23	0.13	-1.82	0.79
((0.3 - control):B) - ((0.3 - control):A,C,D,E,F,G,H,I)	0.49	0.14	3.50	0.02
((0.3 - control):C) - ((0.3 - control):A,B,D,E,F,G,H,I)	0.24	0.14	1.72	0.85
((0.3 - control):D) - ((0.3 - control):A,B,C,E,F,G,H,I)	-0.03	0.13	-0.22	1.00
((0.3 - control):E) - ((0.3 - control):A,B,C,D,F,G,H,I)	0.28	0.13	2.22	0.47
((0.3 - control):F) - ((0.3 - control):A,B,C,D,E,G,H,I)	-0.40	0.13	-3.18	0.04
((0.3 - control):G) - ((0.3 - control):A,B,C,D,E,F,H,I)	-0.00	0.13	-0.00	1.00
((0.3 - control):H) - ((0.3 - control):A,B,C,D,E,F,G,I)	-0.11	0.14	-0.78	1.00
((0.3 - control):I) - ((0.3 - control):A,B,C,D,E,F,G,H)	-0.24	0.13	-1.88	0.75

5 Oats Field Trial

In this section the analysis of statistical interaction in a split-plot field trial is demonstrated. The underlying methodology is roughly described in Kitchie and Schaarschmidt (2014). Therefore, the `oats` dataset from the R add-on package `MASS` is used Venables and Ripley (2002). In this trial the yield of three different varieties of oats was investigated under four levels of manurial treatment. For a detailed description we refer to the `MASS` package documentation of the data set.



```
library(nlme)
fit <- lme(Y ~ N + V + N:V, data = oats, random = ~1 | B/V)
anova(fit)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	45.00	245.14	0.00
N	3	45.00	37.69	0.00
V	2	10.00	1.49	0.27
N:V	6	45.00	0.30	0.93

Although there is no evidence for an interaction in this trial, we further assume that interest is in the analysis of the nitrogen by variety interaction for illustrative purposes.

```
InteractionContrasts <- iacontrast(fa = oats$N, fb = oats$V, typea = "Tukey",
  typeb = "Tukey", abbrevnames = list(minlength = 4))
oats$NV <- InteractionContrasts$fab
CMM <- lme(Y ~ NV, data = oats, random = ~1 | B/V) #fitting the cell means model
anova(CMM)
# In principle, there may be more appropriate denominator degrees of freedom
# here. For consistency with ANOVA we select 43
MultTest <- glht(CMM, mcp(NV = InteractionContrasts$cmab), df = 43)
```

```
summary(MultTest)
confint(MultTest)
```

References

- Bradu, D. and Gabriel, K. R. (1974). Simultaneous Statistical Inference on Interactions in Two-Way Analysis of Variance. *Journal of the American Statistical Association*, 69(346):428–436.
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363.
- Kitsche, A. and Schaarschmidt, F. (2014). Analysis of statistical interactions in factorial experiments. *Journal of Agronomy and Crop Science*.
- Petersen, R. G. (1985). *Design and analysis of experiments*, volume 66 of *Statistics*. Dekker, New York NY u.a.
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.