11

AN INTRODUCTION TO RESTRIKTOR

Evaluating informative hypotheses for linear models

Leonard Vanbrabant

DEPARTMENT OF DATA ANALYSIS, GHENT UNIVERSITY, GHENT, BELGIUM

Yves Rosseel

DEPARTMENT OF DATA ANALYSIS, GHENT UNIVERSITY, GHENT, BELGIUM

Introduction

In this chapter we introduce the R package `restriktor` that enables easy application of evaluating informative hypotheses. In many psychological fields, researchers have specific expectations about the relation between the means of different groups or between (standardized) regression coefficients. For example, in experimental psychology, it is often tested whether the mean reaction time *increases* or *decreases* for different treatment groups (see, for example, Kofler et al., 2013). In clinical trials, it is often tested whether a particular treatment is *better* or *worse* than other treatments (see, for example, Roberts, Roberts, Jones, & Bisson, 2015). In observational studies, researchers often have clear ideas about whether the direction of the effects is *positive* or *negative* (see, for example, Richardson, Abraham, & Bond, 2012), indicated by symbols like “<” and “>”. Testing such specific expectations directly is known under various names, such as *one-sided testing*, *order-constrained hypothesis testing*, *constrained statistical inference*, and *informative hypothesis testing*. For the remainder of this chapter, we will refer to this kind of analysis as informative hypothesis testing (IHT; Hoijtink, 2012).

Many applied researchers are already familiar with IHT in the context of the classical one-sided *t*-test, where one mean is restricted to be greater or smaller than a fixed value (e.g., $\mu_1 > 0$) or another mean (e.g., $\mu_1 > \mu_2$). The method of constraining parameters readily extends to the AN(C)OVA and multiple regression (e.g., linear, logistic, Poisson) setting where more than one constraint can be imposed on the (adjusted) means or regression coefficients (Silvapulle & Sen, 2005). IHT has

several benefits compared to classical null-hypothesis significance testing (e.g., $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against H_{alt} : not all four means are equal). First, testing specific expectations directly does not require multiple significance tests (Hojtink, 2012; Klugkist, Van Wesel, & Bullens, 2011; Van de Schoot et al., 2011). In this way, we avoid an inflated Type I error rate or a decrease in power that results from corrections of the significance level α . Second, to avoid multiple testing issues with ordered means, an ANOVA is often combined with contrasts to directly test the specific pattern. However, contrast tests are not the same as informative hypothesis tests (Baayen, Klugkist, & Mechsner, 2012). Third, incorporating order constraints in the analysis will result in substantially more power (e.g., Bartholomew, 1961a, 1961b; Kuiper & Hoijtink, 2010; Perlman, 1969; Robertson, Wright, & Dykstra, 1988; Vanbrabant, Van de Schoot, & Rosseel, 2015; Van de Schoot & Strohmeier, 2011). Vanbrabant et al. (2015) showed that using ordered means and multiple one-sided regression coefficients yields adequate power with 50% of the sample size required by ANOVA and regression (respectively).

Evaluating an informative hypothesis requires two hypothesis tests, which are in the statistical literature often called *hypothesis test Type A* and *hypothesis test Type B*. Under the null hypothesis test of hypothesis test Type A, only the parameters (e.g., means or regression coefficients) that are involved in the order-constrained hypothesis are constrained to be equal (e.g., $H_{A0} : \mu_1 = \mu_2 = \mu_3 = \mu_4$) and it is tested against the order-constrained hypothesis (e.g., $H_{A1} : \mu_1 < \mu_2 < \mu_3 < \mu_4$). For hypothesis test Type B, the null hypothesis states that all restrictions hold in the population (e.g., $H_{B0} : \mu_1 < \mu_2 < \mu_3 < \mu_4$) and it is tested against the hypothesis where no constraints are imposed on the parameters (e.g., H_{B1} : at least one restriction is violated), although some equality constraints (if present) may be preserved under the alternative unconstrained hypothesis. Rejecting the null hypothesis would mean that at least one order constraint is violated. To find evidence in favor of an order-constrained hypothesis, a combination of hypothesis test Type B and hypothesis test Type A (in this order) is used. The rationale is that if hypothesis test Type B is not significant, we do not reject the null hypothesis that all restrictions hold in the population. However, hypothesis test Type B cannot make a distinction between inequality and equality constraints. Therefore, if hypothesis test Type B is not significant, the next step is to evaluate hypothesis test Type A. If we reject H_{A0} we can conclude that at least one inequality constraint is strictly true. Then, if we combine the evidence of hypothesis test Type B and hypothesis Type A, we can say that we have found indirect evidence in favor of (or against) the order-constrained hypothesis.

In the remainder of this chapter, we demonstrate for four examples how to evaluate informative hypotheses using *restriktor*. For each example, we show (1) how to set up the constraint syntax, (2) how to test the informative hypothesis, and (3) how to interpret the results. In the first example, we impose order constraints on the means of a one-way ANOVA model. In the second example, we impose order constraints on the means of an ANOVA model, where we test whether the effect size is at least small according to guidelines for Cohen's d . In the

third example, we impose order constraints on the standardized regression coefficients of a linear model. In the fourth example, we impose order constraints on newly defined parameters; that is, on three covariate-conditional effects of gender on the outcome variable. To ensure the reproducibility of chapter results, the data sets for each of the examples are available in the `restriktor` package. More information about how to import your own data into R can be found online at www.restriktor.ugent.be/tutorial/importdata.html. Before we continue with the examples, we first explain how to get started. The annotated R code described below can also be found on the Open Science Framework (osf.io/am7pr/).

Getting started

Installing `restriktor`

To install `restriktor`, open R, and type:

```
install.packages("restriktor")
```

If the `restriktor` package is installed, the package needs to be loaded into R. This can be done by typing:

```
library(restriktor)
```

If the package is loaded, the following startup message should be displayed (note that the version number 0.2–15 changes in future releases):

```
## This is restriktor 0.2-15
## restriktor is BETA software! Please report any bugs.
```

A more detailed description about how to get started with `restriktor` can be found online at restriktor.org/gettingstarted.html.

The constraint syntax

The easiest way in `restriktor` to construct the constraint syntax for factors is to use the factor-level names (e.g., A, B, C), preceded by the factor name (e.g., Group). For covariates, we can refer simply by their name. Order constraints are defined via inequality constraints (<, or >) or by equality constraints (==). The constraint syntax is enclosed within single quotes. For example, for a simple order-constrained hypothesis with three means (i.e., $H: \mu_1 < \mu_2 < \mu_3$), the constraint syntax might look as follows:

```
myConstraints <- ' GroupA < GroupB
                  GroupB < GroupC '
```

More information about the constraint syntax can be found online at restriktor.org/tutorial/syntax.html.

Testing the informative hypothesis

In `restriktor`, the `iht()` function is used for IHT. The minimal requirements for this function are a constraint syntax and a fitted unconstrained model. In an unconstrained model no (in)equality constraints are imposed on the means or regression coefficients. Currently, `iht()` can deal with unconstrained models of class `lm` (standard linear model/ANOVA), `mlm` (multivariate linear model), `rlm` (robust linear model) and `glm` (generalized linear model). By default, the function uses the F-bar test statistic (Kudô, 1963; Wolak, 1987). The F-bar statistic is an adapted version of the classical F statistic and can deal with order constraints. More information about all available options can be found online at restriktor.org/tutorial/contest.html.

Estimation of the restricted estimates and inference

Instead of testing the informative hypothesis, the (restricted) regression coefficients/means might be of interest. In this case, the `restriktor()` function can be used. The first argument to `restriktor()` is the fitted unconstrained linear model. The second argument is the constraint syntax. The output shows the restricted estimates and the corresponding standard errors, *t*-test statistics, two-sided *p*-values, and the multiple R^2 . The output also provides information about the type of computed standard errors. By default, conventional standard errors are computed but heteroskedastic robust standard errors are also available. Again, more information about all available options can be found online at restriktor.org/tutorial/restriktor.html.

Example 1: Ordered-constrained means of a one-way ANOVA model

In this example, we use the “anger management” data set. These data denote a person’s decrease in aggression level between week 1 (intake) and week 8 (end of training) for four different treatment groups of anger management training, namely (1) no training, (2) physical training, (3) behavioral therapy, and (4) a combination of physical exercise and behavioral therapy. The purpose of the study was to test the assumption that the exercises would be associated with a reduction in the mean aggression levels. In particular, the hypothesis of interest was $H_1 : \mu_{\text{No}} < \{\mu_{\text{Physical}} = \mu_{\text{Behavioral}}\} < \mu_{\text{Both}}$. This hypothesis states that the decrease in aggression levels is smallest for the “no training” group, larger for the “physical training” and “behavioral therapy” group, with no preference for either method, and largest in the “combination of physical exercise and behavioral therapy” group (Hoijsink, 2012, pp. 5–6).

In practice, hypothesis H_1 is usually evaluated with an ANOVA, where the null hypothesis $H_0 : \mu_{\text{No}} = \mu_{\text{Physical}} = \mu_{\text{Behavioral}} = \mu_{\text{Both}}$ is tested against the

unconstrained hypothesis H_{unc} : not all four means are equal. The results from the global F -test revealed that the four means are not equal ($F_{(4,36)} = 18.62$, $p < .001$). At this point, we do not know anything about the ordering of the means. Therefore, the next step would be to use pairwise comparisons with corrections for multiple testing (e.g., Bonferroni, Tukey, and FDR). The results with FDR (False Discovery Rate) adjusted p -values showed three significant ($p \leq .05$) mean differences (MD), namely between the “Behavioral-No” exercises ($MD = 3.3$, $p = .001$), the “Behavioral-Physical” exercises ($MD = 2.3$, $p = .018$) and the “Both-Physical” exercises ($MD = 3.3$, $p = .001$). A graphical representation of the means is shown in Figure 11.1. Based on the results of the global F test and the pairwise comparisons, it would not be an easy task to derive an unequivocal conclusion about hypothesis H_1 .

In what follows, we show all steps and the **restriktor** syntax to evaluate the informative hypothesis H_1 directly.

Step 1: Set up the constraint syntax

In R, categorical predictors are represented by “factors”. For example, the “Group” variable has four factor levels: “No”, “Physical”, “Behavioral”, and “Both”. In addition, the factor levels are presented in alphabetical order and it may therefore be convenient to re-order the levels. This can be done in R by typing:

FIGURE 11.1 Means plot: reduction of aggression levels after eight weeks of anger management training

```
AngerManagement$Group <- factor(AngerManagement$Group,
                                levels = c("No", "Physical",
                                             "Behavioral",
                                             "Both"))
```

Next, the constraint syntax for hypothesis H_1 might look as follows:

```
myConstraints1 <- ' GroupNo          < GroupPhysical
                  GroupPhysical == GroupBehavioral
                  GroupBehavioral < GroupBoth '
```

Step 2: Test the informative hypothesis

Since an ANOVA model is a special case of the multiple regression model, we can use the linear model for our ANOVA example. Then, we can fit the unconstrained linear model as follows:

```
fit_ANOVA <- lm(Anger ~ -1 + Group, data = AngerManagement)
```

The tilde $\sim$ is the regression operator. On the left-hand side of the operator we have the response variable `Anger` and on the right-hand side we have the factor `Group`. We removed the intercept (-1) from the model so that the estimates reflect the group means. Next, we can test the informative hypothesis using the `iht()` function. This is done as follows:

```
iht(fit_ANOVA, myConstraints1)
```

The first argument to `iht()` is the fitted unconstrained linear model. The second argument is the constraint syntax. By default, the function prints an overview of all available hypothesis tests. The results are shown below. Some parts are removed due to its length.

Restriktor: restricted hypothesis tests (36 residual degrees of freedom):

Multiple R-squared reduced from 0.674 to 0.608

Constraint matrix:

	GroupNo	GroupPhysical	GroupBehavioral	GroupBoth	op	rhs	active
1:	0	1	-1	0	==	0	yes
2:	-1	1	0	0	>=	0	no
3:	0	0	-1	1	>=	0	no

Overview of all available hypothesis tests:

```
Global test: H0: all parameters are restricted to be equal (==)
              vs. HA: at least one inequality restriction is strictly true (>)
Test statistic: 25.4061, p-value: <0.0001
```

```
Type A test: H0: all restrictions are equalities (==)
              vs. HA: at least one inequality restriction is strictly true (>)
Test statistic: 25.4061, p-value: <0.0001
```

```
Type B test: H0: all restrictions hold in the population
              vs. HA: at least one restriction is violated
Test statistic: 7.2687, p-value: 0.04518
```

At the top of the output the constraint matrix is shown. This matrix is constructed internally based on the text-based constraint syntax but could have been constructed manually. The constraint matrix is comparable to the contrast matrix but treated differently in the constraint framework. The “active” column indicates if a constraint is violated or not. If no constraints are active, this would mean that all constraints are in line with the data. In the remainder, an overview of the available hypothesis tests is given. Information about how to obtain a more detailed output for each hypothesis test can be found in the help file or online at restriktor.org/tutorial/contest.html.

Step 3: Interpret the results

To evaluate the informative hypothesis H_1 , we first conduct hypothesis test Type B. Not rejecting this hypothesis test would mean that the order constraints are in line with the data. The results from hypothesis test Type B, however, show that hypothesis H_1 is rejected in favor of the best-fitting (i.e., unconstrained) hypothesis ($\bar{F}_{(0,1,2;36)}^B = 7.27, p = .045$)¹. In other words, the constraints are not supported by the data and we conclude that the informative hypothesis H_1 does not hold.

Estimation of the restricted estimates and inference

Instead of testing the informative hypothesis H_1 , the restricted means might be of interest. The restricted means can be computed as follows:

```
restr_ANOVA <- restriktor(fit_ANOVA, constraints = myConstraints1)
```

By default, the `print()` function prints a brief overview of the restricted means:

```
print(restr_ANOVA)
```

Call:

```
conLM.lm(object = fit_ANOVA, constraints = myConstraints1)
```

```
restriktor (0.1-80.711): restricted linear model:
```

Coefficients:

GroupNo	GroupPhysical	GroupBehavioral	GroupBoth
-0.20	1.95	1.95	4.10

We can clearly see that the `GroupPhysical` and the `GroupBehavioral` means are constrained to be equal. If desired, a more extensive output can be requested using the `summary()` function:

```
summary(restr_ANOVA)

Call:
conLM.lm(object = fit_ANOVA, constraints = myConstraints1)

Restriktor: restricted linear model:

Residuals:
    Min       1Q   Median       3Q      Max
-3.100 -1.275  -0.025  1.200  5.050

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
GroupNo      -0.20000    0.65233   -0.3066  0.7609210
GroupPhysical  1.95000    0.46127    4.2275  0.0001544 ***
GroupBehavioral 1.95000    0.46127    4.2275  0.0001544 ***
GroupBoth     4.10000    0.65233    6.2851  2.895e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.0629 on 36 degrees of freedom
Standard errors: standard
Multiple R-squared reduced from 0.674 to 0.608

Generalized Order-Restricted Information Criterion:
  Loglik Penalty   Goric
-84.1621  2.8918 174.1079
```

The output shows the restricted group means and the corresponding (standard) standard errors, *t*-test statistics and two-sided *p*-values. The multiple $R^2 = .674$ refers to the unconstrained model and the $R^2 = .608$ refers to the order-constrained model. The reduction in R^2 provides additional evidence that at least one order constraint is violated. Both R^2 s are equal only if all constraints are in line with the data. The last part of the output provides information for model selection using the generalized order-restricted information criterion (GORIC), which is a modification of the Akaike information criterion. More information and an example can be found online at restriktor.org/tutorial/example6.html.

Example 2: Ordered-constrained means with effect sizes

The *p*-value is not a measure for the size of an effect (Nickerson, 2000). Therefore, in an AN(C)OVA the question should be whether the differences between the group means are relevant. To answer this question, the popular effect-size measure Cohen's *d* (Cohen, 1988) can be used, and is given by: $d = (\mu_{\max} - \mu_{\min}) / \sigma_{\text{pooled}}$, where $\mu_{\max}$ is the largest and $\mu_{\min}$ is the smallest of the *m* means, and σ_{pooled} is the

pooled standard deviation within the populations. According to Cohen, values of 0.2, 0.5, and 0.8 indicate a small, medium, and large effect, respectively.

In this example, we use the Zelazo, Zelazo, and Kolb (1972) data set, which is available in **restriktor**. The data consist of ages in months at which a child starts to walk for four treatment groups. For simplicity we only consider three treatment groups. The excluded group is the “Control” group. The first treatment group (“Active”) received a special walking exercise for 12 minutes per day beginning at the age of one week old and lasting seven weeks. The second group (“Passive”) received daily exercises but not the special walking exercises. The third group (“No”) were checked weekly for progress (the other two groups got daily exercises) but they did not receive any special exercises. The purpose of the study was to test the claim that the walking exercises are associated with a reduction in the mean age at which children start to walk.

If we ignore the effect sizes, the informative hypothesis can be formulated as: $H_2 : \mu_{\text{Active}} < \mu_{\text{Passive}} < \mu_{\text{No}}$. The results from hypothesis test Type B ($\bar{F}_{(0,1,2;14)}^B = 0$, $p = 0$) and hypothesis test Type A ($\bar{F}_{(0,1,2;14)}^A = 5.978$, $p = .028$) provide evidence in favor of the informative hypothesis. However, for practical relevance of the treatments, the mean differences between the groups should at least indicate a small effect. To answer this question, we reformulate hypothesis H_2 such that the effect sizes are included. The pooled within group standard deviation equals 1.516:

$$H_2^d = \frac{(\mu_{\text{Passive}} - \mu_{\text{Active}})}{1.516} > 0.2$$

$$\frac{(\mu_{\text{No}} - \mu_{\text{Passive}})}{1.516} > 0.2.$$

This hypothesis states that we expect at least $0.2 * 1.516$ standard deviations between the means, which indicates a small effect size. Next, we show how to evaluate this informative hypothesis.

Step 1: Set up the constraint syntax

Again, we use the factor-level names preceded by the factor name to construct the constraint syntax. The effect sizes can be easily computed within the constraint syntax using the arithmetic operator `/`:

```
myConstraints2 <- ' (GroupPassive - GroupActive) / 1.516 > 0.2
                  (GroupNo - GroupPassive) / 1.516 > 0.2 '
```

Step 2: Test the informative hypothesis

Since we excluded the “Control” group, we need to take a subset of the original data. The `subset()` function in R is an easy way to select observations. This is done in R by typing:

```
subData <- subset(ZelazoKolb1972, subset = Group != "Control")
```

Then, the unconstrained linear model can be fit as follows:

```
fit_ANOVAd <- lm(Age ~ -1 + Group, data = subData)
```

Next, we test the informative hypothesis using the fitted unconstrained model `fit_ANOVAd` and the constraint syntax `myConstraints2`:

```
iht(fit_ANOVAd, constraints = myConstraints2)
```

Step 3: Interpret the results

The results from hypothesis test Type B ($\bar{F}_{(0,1,2;14)}^B = 0$, $p = 1$) and hypothesis test Type A ($\bar{F}_{(0,1,2;14)}^A = 3.19$, $p = .089$) show that if we include a small effect size in the informative hypothesis, the initial significant results become irrelevant. This clearly demonstrates the importance of including effect sizes in the hypothesis.

Example 3: Order-constrained (standardized) linear regression coefficients

In this example, we show how order constraints can be imposed on the standardized regression coefficients, denoted by β^Z , of a linear model. We use the “exam” data set, which is available in `restriktor`. The model relates students’ “exam scores” (“Scores”) to the “averaged point score” (“APS”), the amount of “study hours” (“Hours”), and “anxiety score” (“Anxiety”). It is hypothesized that “APS” is the strongest predictor, followed by “study hours” and “anxiety scores”, respectively. In symbols, this informative hypothesis can be written as $H_3 : \beta_{APS}^Z > \beta_{Hours}^Z > \beta_{Anxiety}^Z$. Since the hypothesis is in terms of which predictor is stronger, we should be aware that each predictor has its own scale. To avoid spurious conclusions, the predictor variables should be standardized first². This can be done in R by typing:

```
Exam$Hours_Z <- (Exam$Hours - mean(Exam$Hours)) / sd(Exam$Hours)
Exam$Anxiety_Z <- (Exam$Anxiety - mean(Exam$Anxiety)) / sd(Exam$Anxiety)
Exam$APS_Z <- (Exam$APS - mean(Exam$APS)) / sd(Exam$APS)
```

Step 1: Set up the constraint syntax

Then, the constraint syntax corresponding H_3 might look as follows:

```
myConstraints3 <- ' APS_Z > Hours_Z
                  Hours_Z > Anxiety_Z '
```

Step 2: Test the informative hypothesis

Next, we fit the unconstrained linear model. The response variable is “Scores” and the predictor variables are the three centered covariates:

```
fit_exam <- lm(Scores ~ APS_Z + Hours_Z + Anxiety_Z,
data = Exam)
```

The informative hypothesis H_3 can be evaluated using the unconstrained model `fit_exam` and the constraint syntax `myConstraints3`:

```
iht(fit_exam, constraints = myConstraints3)
```

Step 3: Interpret the results

The results from hypothesis test Type B show that the order-constrained hypothesis is not rejected in favor of the unconstrained hypothesis ($\bar{F}_{(0,1,2;16)}^B = 0, p = 1$). The results from hypothesis test Type A show that the null hypothesis is rejected in favor of the order-constrained hypothesis ($\bar{F}_{(0,1,2;16)}^A = 12.38, p = .003$). Thus, we have found strong evidence in favor of the informative hypothesis H_3 .

Example 4: Testing order constraints on newly defined parameters

Here, we show how order constraints can be imposed between newly defined parameters, e.g., simple slopes. The original data are based on two cohort studies of children from 0 to 4 and 8 to 18 years old with burns, and their parents (e.g., Bakker, Van der Heijden, Van Son, & Van Loey, 2013; Egberts et al., 2016). Since the original data are not publicly accessible, we simulated data based on the original model parameters. This simulated data set is available in `restriktor`. For illustrative reasons we focus only on the data provided by the mother. For the current illustration we included five predictor variables in the data set: a child’s gender (0 = boys, 1 = girls), age, the estimated percentage of the total body surface area affected by second or third degree burns (“TBSA”), and parental guilt and anger feelings in relation to the burn event. The model relates post-traumatic stress symptoms (PTSS) to the five predictor variables and can be written as a linear function:

$$\begin{aligned} \text{PTSS}_i \sim & \beta_{\text{intercept}} + \beta_1 \text{gender}_i + \beta_2 \text{age}_i + \beta_3 \text{guilt}_i + \beta_4 \text{anger}_i + \beta_5 \text{TBSA}_i \\ & + \beta_6 (\text{gender}_i * \text{guilt}_i) + \beta_7 (\text{gender}_i * \text{anger}_i) + \beta_8 (\text{gender}_i * \text{TBSA}_i) + \varepsilon_i, \end{aligned}$$

where $\beta_{\text{intercept}}$ is the intercept, β_1 to β_5 are the regression coefficients for the main effects, and β_6 to β_8 are the regression coefficients for the interaction effects.

We hypothesized that the mean difference in PTSS between mothers of girls and mothers of boys would increase for simultaneously higher levels of guilt, anger, and TBSA. To test this informative hypothesis, we selected three different settings for guilt, anger, and TBSA, namely small, medium, and large. For illustrative reasons, for the small level we chose the values 0, 0, 1 for guilt, anger, and TBSA respectively. For the medium level we chose the variable means, which are 2.02, 2.06, and 8.35, respectively, and for the large level we chose 4, 4, and 20, respectively. Then, the resulting three effects (small, medium, large) can be calculated respectively as follows:

$$\text{smallEffect} = \beta_1 + \beta_6 0 + \beta_7 0 + \beta_8 1$$

$$\text{mediumEffect} = \beta_1 + \beta_6 2.02 + \beta_7 2.06 + \beta_8 8.35$$

$$\text{largeEffect} = \beta_1 + \beta_6 4 + \beta_7 4 + \beta_8 20.$$

Note that each effect reflects a mean difference between boys and girls. Then, the informative hypothesis can be expressed as:

$$H_4 : \text{smallEffect} < \text{mediumEffect} < \text{largeEffect}.$$

Step 1: Set up the constraints syntax

A convenient feature of the `restriktor` constraint syntax is the option to define new parameters, which take on values that are an arbitrary function of the original model parameters. This can be done using the `=` operator. In this way, we can compute the desired effects and impose order constraints among these effects. Then, the constraint syntax might look as follows:

```
myConstraints4 <- 'smallEffect := gender + 0*gender.guilt +
                  0*gender.anger +
                  1*gender.TBSA

                  mediumEffect := gender + 2.02*gender.guilt +
                  2.06*gender.anger +
                  8.35*gender.TBSA

                  largeEffect := gender + 4*gender.guilt +
                  4*gender.anger +
                  20*gender.TBSA

                  smallEffect < mediumEffect
                  mediumEffect < largeEffect'
```

It is important to note that variable/factor names of the interaction effects in objects of class `lm`, `rlm`, `glm`, and `mlm` contain a semi-colon (`:`) between the variable

names (e.g., `gender:guilt`). To use these parameters in the constraint syntax, the semi-colon must be replaced by a dot (.) (e.g., `gender.guilt`).

Step 2: Test the informative hypothesis

Based on outlier diagnostics³ we identified 13 outliers (approximately 4.7% of the data). Therefore, we use robust methods. The unconstrained robust linear model using MM estimation (Yohai, 1987) can be fitted as follows:

```
library(MASS)
fit_rburns <- rlm(PSS ~ gender*guilt + gender*anger +
                  gender*TBSA + age,
                  data = Burns, method = "MM")
```

On the right-hand side of the regression operator (`~`) we included the three interaction effects using the `*` operator. The main effects are automatically included. Note that the interaction operator `*` is not an arithmetic operator as used in the constraint syntax. Then, the informative hypothesis can be evaluated as follows:

```
iht(fit_rburns, constraints = myConstraints4)
```

Step 3: Interpret the results

The results from hypothesis test Type B ($\bar{F}_{MM(0,1,2;269)}^B = 0, p = 1$) show that the order-constrained hypothesis is not rejected in favor of the unconstrained hypothesis. The results from hypothesis test Type A show that the null hypothesis is rejected in favor of the order-constrained hypothesis ($\bar{F}_{MM(0,1,2;269)}^A = 5.35, p = .044$). Hence, we can conclude that the data provide enough evidence that the gender effect increases for higher levels of guilt, anger, and TBSA.

The non-robust results from hypothesis test Type A would have led to a different conclusion, namely that the null hypothesis would not have been rejected in favor of the order-constrained hypothesis ($\bar{F}_{(0,1,2;269)}^A = 3.65, p = .107$). This clearly demonstrates that ignoring outliers may result in misleading conclusions.

Conclusion

IHT has been shown to have major benefits compared to classical null-hypothesis testing. Unfortunately, applied researchers have been unable to use these methods because user-friendly freeware and a clear tutorial were not available. Therefore, in this chapter we introduced the user-friendly R package **restriktor** for evaluating (robust) informative hypotheses. The procedure was illustrated using four examples. For each example, we showed how to set

up the constraint syntax, how to evaluate the informative hypothesis and how to interpret the results. All results were obtained by the default settings of the software package `restriktor`. If desired, they can readily be adjusted.

We only discussed frequentist methods for evaluating informative hypotheses. Of course, examples 1–4 could have been evaluated in the Bayesian framework; see Chapter 12 (Zondervan-Zwijnenburg & Rijshouwer; see also Berger & Mortera, 1999; Gu, Mulder, Deković, & Hoijtink, 2014; Hoijtink, 2012; Klugkist, Laudy, & Hoijtink, 2005; Mulder, Hoijtink, & Klugkist, 2010) but we believe that the frequentist methods are a welcome addition to the applied user's toolbox and may help convince applied users unfamiliar with Bayesian statistics to include order constraints in their hypothesis. In addition, robust IHT as discussed in this chapter does not seem to exist in the Bayesian framework (yet).

It must be noted that the `restriktor` package is not finished yet, but it is already very useful for most users. The package is actively maintained, and new options are being added. We advise the reader to monitor the `restriktor` website (restriktor.org) for updates.

Notes

- 1 The null distribution is a mixture of F distributions mixed over the degrees of freedom. Therefore, in this example, the p -value $Pr(\bar{F} \geq \bar{F}_{obs})$ approximately equals $w_0 Pr(F_{0,36} \geq \bar{F}_{obs}) + w_1 Pr(F_{1,36} \geq \bar{F}_{obs}/1) + w_2 Pr(F_{2,36} \geq \bar{F}_{obs}/2)$, where $Pr(F_{0,36} \geq \bar{F}_{obs})$ equals 0 by definition. Hence the notation $\bar{F}_{(0,1,2;36)}$. w is the level probability, the probability that the order-constrained maximum likelihood estimates have j levels (under the null-hypothesis), where m = the number of inactive order constraints; and the w_m sum to 1.
- 2 Standardized regression coefficients can be obtained by standardizing all the predictor variables before including them in the model. For example: $Z(APS_i) = (APS_i - \text{mean}(APS))/sd(APS)$, where sd is the standard deviation.
- 3 The outliers were identified with robust Mahalanobis distances larger than the 99.5% quantile of a χ^2_8 distribution.

References

- Baayen, C., Klugkist, I., & Mechsner, F. (2012). A test of order-constrained hypotheses for circular data with applications to human movement science. *Journal of Motor Behavior*, 44(5), 351–363.
- Bakker, A., Van der Heijden, P. G. M., Van Son, M. J. M., & Van Loey, N. E. E. (2013). Course of traumatic stress reactions in couples after a burn event to their young child. *Health Psychology*, 32(10), 1076–1083.
- Bartholomew, D. J. (1961a). Ordered tests in the analysis of variance. *Biometrika*, 48(3/4), 325–332.
- Bartholomew, D. J. (1961b). A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(2), 239–281.
- Berger, J. O., & Mortera, J. (1999). Default Bayes factors for non-nested hypothesis testing. *Journal of the American Statistical Association*, 94(446), 542–554.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Egberts, M. R., Van de Schoot, R., Boekelaar, A., Hendrickx, H., Geenen, R., & Van Loey, N. E. E. (2016). Child and adolescent internalizing and externalizing problems 12 months postburn: The potential role of preburn functioning, parental posttraumatic stress, and informant bias. *European Child & Adolescent Psychiatry*, 25(7), 791–803.
- Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, 19(4), 511–527.
- Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Boca Raton, FL: Taylor & Francis.
- Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, 10(4), 477–493. doi:10.1037/1082-989X.10.4.477.
- Klugkist, I., Van Wesel, F., & Bullens, J. (2011). Do we know what we test and do we test what we want to know? *International Journal of Behavioral Development*, 35(6), 550–560. doi:10.1177/0165025411425873.
- Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review*, 33(6), 795–811.
- Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, 50(3/4), 403–418.
- Kuiper, R. M., & Hoijtink, H. (2010). Comparisons of means using exploratory and confirmatory approaches. *Psychological Methods*, 15(1), 69–86.
- Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, 140(4), 887–906. doi:10.1016/j.jspi.2009.09.022.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5(2), 241–301.
- Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *Annals of Mathematical Statistics*, 40(2), 549–567.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387.
- Roberts, N. P., Roberts, P. A., Jones, N., & Bisson, J. I. (2015). Psychological interventions for post-traumatic stress disorder and comorbid substance use disorder: A systematic review and meta-analysis. *Clinical Psychology Review*, 38, 25–38.
- Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York, NY: John Wiley.
- Silvapulle, M. J., & Sen, P. K. (2005). *Constrained statistical inference: Order, inequality, and shape constraints*. Hoboken, NJ: John Wiley & Sons.
- Wolak, F. (1987). An exact test for multiple inequality and equality constraints in the linear regression model. *J. Am. Stat. Assoc.*, 82, 782–793.
- Van de Schoot, R., Hoijtink, H., Mulder, J., Van Aken, M. A. G., Orobio de Castro, B., Meeus, W., & Romeijn, J. W. (2011). Evaluating expectations about negative emotional states of aggressive boys using Bayesian model selection. *Developmental Psychology*, 47(1), 203–212.
- Van de Schoot, R., & Strohmeier, D. (2011). Testing informative hypotheses in SEM increases power: An illustration contrasting classical hypothesis testing with a parametric bootstrap approach. *International Journal of Behavioral Development*, 35(2), 180–190.

- Vanbrabant, L., Van de Schoot, R., & Rosseel, Y. (2015). Constrained statistical inference: Sample-size tables for ANOVA and regression. *Frontiers in Psychology*, 5, 1–8.
- Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*, 15(2), 642–656.
- Zelazo, P. R., Zelazo, N. A., & Kolb, S. (1972). “Walking” in the newborn. *Science*, 176 (4032), 314–315.