Is it worth the effort? Penalized ordinal regression versus common social science methods for survey data.

Ludwig-Maximilians-Universität München
Institut für Statistik
Masterthesis

Author: Ruben Hartmann
Supervisor: Prof. Dr. Thomas Augustin
Munich, January 13, 2022

Abstract

Having ordinal variables as both response and predictor in a regression analysis is a frequent phenomenon in social sciences. Still, proper treatment via ordinal regression is rarely realised. A common method used instead is linear regression. This thesis evaluates whether familiarising oneself with the lesser known method is worth the effort or whether the usual approaches suffice. As is shown the performance of the penalized proportional odds model is better than that of the linear regression. The selected ordinal regression, a penalized proportional odds model, is motivated by an applied sociological question of whether authoritarian attitudes in young people are linked to experienced deprivation. Findings prove that this is not the case. A recently developed extension of the `ordPens` package allows for the implementation of the proportional odds model with first- and second-order difference generalised ridge penalties in `mgcv::gam()`. An evaluation based on simulated data concludes that the new implementation is recommended for first-order difference penalties and that the confidence intervals are reliable.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The proper treatment of ordered categorical data is a continuous issue in statistical model building (Atkinson, 1988; Jöreskog and Moustaki, 2001; Wang et al., 2014) while at the same time this type of data is one of the most common in social sciences (Jöreskog and Moustaki, 2001). However, in one of the most well-known handbooks on social research in Germany, methods for metric and categorical variables are presented, but ordinal data methods are not even mentioned (Blasius and Baur, 2019). In the same way, a rough overview over the articles concerning regression in the renowned german *Kölner Zeitschrift für Soziologie und Sozialpsychologie* showed several multinomial logistic regressions and not one on ordinal regression in the last five years[1]. One reason for this lack of attention might be the rather restricted number of available implementations, as an overview over those in the statistical open-source software R (R Core Team, 2021) will show. Another reason might be the extra effort for scientists to approach a new method with less examples in the sociological literature, or even the lack of knowledge about more adequate procedures. No matter what the reason may be, it often leads to inadequate modelling, for example ordinal variables are being treated as metric (Blasius and Baur, 2019; Tutz and Gertheiss, 2014). A very common method is regression analysis, where linear regression is conducted instead of ordinal regression.

This thesis contributes to closing this gap in the social sciences. In order to do so, the sociological question is pursued whether young people's attitudes towards authoritarian statements can be explained by the discrimination they experienced. Survey data is provided for answering it (chapter 2). The correct regression models for ordinal responses are

[1]Method: Social Sciences Citation Index (SSCI), search term: ((SO=(kolner zeitschrift fur soziologie und sozialpsychologie)) AND ALL=(regression) plus individual check for the method description in the abstract.

discussed (chapter 3) and supplemented with the adequate penalization terms for ordinal covariates (chapter 4). The penalized ordinal regression is carried out in the statistical open-source software R (R Core Team, 2021). As will be outlined, the required model has only recently been implemented via the combination of two R packages and has not been applied to the exact same regression family yet (chapter 4). Therefore, a simulation study is conducted to investigate the properties of the model (chapter 5). Now, the initially posed sociological question is answered, exemplifying how the model can be applied in practice (chapter 6). Also, it is examined whether the results of the new model are a substantial improvement compared to the common alternative (chapter 7). Findings are wrapped up in the last chapter (chapter 8).

Chapter 2

Methodological Motivation

2.1 Sociological Motivation

Studies on authoritarianism regularly demonstrate the correlation between authoritarianism and derogatory attitudes towards other groups (e.g. Group-Focused Enmity, GFE) (Brähler and Decker, 2018; Heitmeyer, 2012; Cribbs and Austin, 2011; Asbrock et al., 2010; Heitmeyer, 2002; Rippl et al., 2000b; Adorno et al., 1950). There exist several terms and closely related definitions for authoritarian attitudes, one being the authoritarian personality (Adorno et al., 1950). During their socialisation, authoritarian personalities did not develop mechanisms to deal with new and unknown situations independently ("crisis situations", Oesterreich (2005), p. 284). They can be defined by four characteristics (Oesterreich (2005) following Adorno et al. (1950)). First, they avoid new and unfamiliar situations. Second, they are characterised by rigid behavioural patterns in which tried and tested strategies can be applied. Thus, they try to avoid the risk of change. A third characteristic is their submission to authority and conformity to established values and the fourth is hostility towards others. By which influences and at what age an authoritarian personality is formed, and whether it is a fundamental trait or rather an orientation, is a continued debate (for an overview see Rippl et al. (2000a)).

Jürgen Mansel and Viktoria Spaiser for example observe that young people's own experiences of discrimination play a significant role in the devaluation of other groups (Mansel and Spaiser, 2013, p. 254). This raises the question of whether there is also a connection between their personal experiences of discrimination and authoritarian attitudes. One possibility to investigate this subject is fitting a regression model on authoritarian attitudes and own experiences of discrimination. This path will be followed here. In doing so, those

socio-demographic characteristics are included into the analysis that have proven to be relevant in connection with GFE and authoritarianism in previous studies. These are the level of education (Brähler and Decker (2018), p. 89, Rippl and Seipel (2018), p. 251, Zick et al. (2016), Mansel and Spaiser (2013), p. 254), regional affiliation to East or West Germany (Brähler and Decker (2018), p. 122, Baier et al. (2010), p. 323), gender (Brähler and Decker, 2018) and migration background (Mansel and Spaiser, 2013).

The survey AID:A 2019 (Aufwachsen in Deutschland: Alltagswelten) (Kuger et al., 2020) comprises questions which can be employed for this undertaking and will be delineated in the following. The survey investigates the everyday life of young people. Children and young people up to 33 years of age ("target persons") are interviewed as well as their household and/or parents.

2.2 Data Description

The original dataset as provided by the data repository of the Deutsches Jugendinstitut (Deutsches Jugendinstitut, 2022) contains more than 1400 variables. Those relevant for the analysis are now presented[1]. The socio-demographic variables are *age* (in years), *gender* ("female", "male", "none of the above"), *residency* (the federal state the person lives in) and *migration background*, which is operationalised by several questions about the migration background in first (own), second (parents) and third (grandparents) generation. The educational background is operationalised by the years spent on education (variable name *bija*). The survey comprises a scale on authoritarianism which consists of six items:

> *To what extent do you agree with the following statements? Please answer with values between 1 "strongly agree" and 6 "strongly disagree".*
>
> 1. *I admire people who have the ability to dominate others. (aut_beherrschen)*
> 2. *I always try to please my parents. (aut_eltern)*
> 3. *I try to always be on the side of the strongest. (aut_staerkere)*
> 4. *New and unusual situations make me uncomfortable. (aut_neues)*
> 5. *I try to always do things in the usual way. (aut_gewohnt)*
> 6. *I avoid people who are different from me. (aut_andere)*

[1]The survey is in German. Translations into English were made by the author.

It maps the sub-dimensions "authoritarian subordination" (items 1-3) and "conventionalism" (items 4-6)(Weigelt, 2020). Each item is taken as an individual response variable. To operationalise the measurement of experienced discrimination, the survey offers a question block on experienced deprivation, illuminating six aspects:

> *It can happen that one is disadvantaged in life. There can be different reasons for this. For each reason I give you, please tell me how often you have been disadvantaged in your life (always or almost always - very often - often - sometimes - rarely - never - denied - don't know)*
>
> 1. *Because of your gender (ben_gender)*
> 2. *Because of the social and financial situation in your family (ben_sozfin)*
> 3. *Because you or your family are not from Germany (ben_migration)*
> 4. *Because of your weight (ben_gew)*
> 5. *Because of a disability or physical impairment (ben_beh)*
> 6. *Because of your religion (ben_rel)*

The question asks solely about perceived disadvantage. The differentiation between perceived disadvantage and real discrimination is not made at this point, since at the moment when those affected define the situation as a real experience of disadvantage, the consequences are real (Thomas theorem, see Thomas and Thomas (1928)) Feelings of disadvantage could therefore have the same effects on members of the society's majority as on members of marginalised groups (cf. Mansel and Spaiser (2013), p. 21). The questions do not cover all aspects of discrimination but as the focus of this analysis is on the statistical matter, these questions will suffice for a first exploration of the field. The mentioned variables are extracted and prepared for later analysis.

2.3 Data Preparation

First, the survey is filtered by target person to exclude parents and siblings. The original data comprises target persons until the age of 33. The survey was generally conducted with targets until 32 years of age and some bivariate inspections show outlier-like behaviour for 33-years-old people. Therefore people aged 33 are excluded. Also people younger than 16 are excluded, because younger targets were not asked the question on authoritarian

attitudes. *Gender* is a variable taking the three values "female", "male" and "none of the above". As the latter got selected by only three surveyed, it is excluded. The third deprivation-item ("Because you or your family is not from Germany") is a filtered question only asked to people identified as immigrants who make up an only small part of the survey population. Later modelling is based on complete rows, so this question would have strongly reduced the data set. The item is therefore excluded. In order to measure the influence of residency as described above, a new dummy variable *ostd_ohne_b* is created based on the federal state. It takes the value "1" for eastern federal states and the value "0" for western federal states and Berlin. As there are several variables measuring the migration background in first (own), second (parents) and third (grandparents) generation, a new dummy variable *migration* is created taking the value "1" if any migration background exists and "0" if none is present. As a last step, the response categories of the question on experienced deprivation are ordered reversely to how they are in the original data. They are recoded in increasing frequency ("Never", "Rarely", "Sometimes", "Often", "Very often", "Always or almost always") for easier interpretation. The readily prepared dataset is depicted in table 2.1 and stored for the penalized ordinal regression in chapter 6. The opinion on several items concerning authoritarian statements (6 ordered levels respectively) are considered as response variables and the socio-demographic variables named above and the five aspects of experienced deprivation answered on an ordinal scale (6 levels) are predictors. In the next chapter, ordinal regression models are presented.

Scale	**Name**	**Expression**
interval-scale	*age*	years
	bija	years of education
binary	*gender*	female/male
	ostd_ohne_b	0 (Western Germany and Berlin)/ 1 (Eastern Germany)
	migration	0 (no migration background) / 1 (oneself or (grand)parents immigrated)
ordinal	*aut_andere**	1 (strongly agree) / 2 / 3 / 4 / 5 / 6 (strongly disagree)
	*aut_gewohnt**	1 (strongly agree) / 2 / 3 / 4 / 5 / 6 (strongly disagree)
	*aut_beherrschen**	1 (strongly agree) / 2 / 3 / 4 / 5 / 6 (strongly disagree)
	*aut_staerkere**	1 (strongly agree) / 2 / 3 / 4 / 5 / 6 (strongly disagree)
	*aut_neues**	1 (strongly agree) / 2 / 3 / 4 / 5 / 6 (strongly disagree)
	*aut_eltern**	1 (strongly agree) / 2 / 3 / 4 / 5 / 6 (strongly disagree)
	ben_gender	1 (Never) / 2 (Rarely) / 3 (Sometimes) / 4 (Often) / 5 (Very often) / 6 (Always or almost always)
	ben_sozfin	1 (Never) / 2 (Rarely) / 3 (Sometimes) / 4 (Often) / 5 (Very often) / 6 (Always or almost always)
	ben_gew	1 (Never) / 2 (Rarely) / 3 (Sometimes) / 4 (Often) / 5 (Very often) / 6 (Always or almost always)
	ben_beh	1 (Never) / 2 (Rarely) / 3 (Sometimes) / 4 (Often) / 5 (Very often) / 6 (Always or almost always)
	ben_rel	1 (Never) / 2 (Rarely) / 3 (Sometimes) / 4 (Often) / 5 (Very often) / 6 (Always or almost always)

Table 2.1: Variables in the final data set. Response variables are marked with *.

Chapter 3

Ordinal Response Models

3.1 Basic Idea

Given a categorical response variable $Y \in \{1, ..., k\}$ with k ordered categories, ordinal response models are the adequate choice for regression analysis. The two main model types are the cumulative and the sequential regression model. Both will be described in this chapter, starting with the sequential model. The following explanations are based on Tutz (2012) and Tutz and Gertheiss (2016).

3.2 The Sequential Model

Sequential models assume a baseline and in order to reach a new category, all previous categories have to be gone through in ascending order. This regression type models the transition from one category to the next. A typical example is pain which is measured in ordered categories from "no pain" over "medium pain" to "strong pain". The assumption of the successive transition poses a theoretical constraint on the data eligible. For the survey questions under consideration such a constraint is hardly applicable: It would be difficult to define a basic characteristic for authoritarian attitudes. Would it start at birth or build on a philosophical assumption about the universal basic nature of human beings, which would have to be found in one of the extremes in order to meet the requirements of an ascending traversal? Discussing these issues is outside of the scope of this thesis and as a more appropriate model is available, the sequential regression will only roughly be

described. In the general model

$$P(Y = r|Y \geq r, X) = F(\beta_{0r} + X^\top \boldsymbol{\beta_r}), \quad r = 1, ..., k.$$

effects are category-specific. A global effect $\boldsymbol{\beta}$ poses a sub form of the above described version and would only imply varying intercepts $\beta_{01}, ..., \beta_{0k}$ for the categories. The other main type of ordinal regression model is the cumulative model.

3.3 The Cumulative Model

Cumulative models are based on the assumption that the realisations of Y are based on an underlying latent variable $\tilde{Y}$. This is a valid assumption for questions about attitudes, as it can be assumed that people's opinions have more subtle gradations than in a questionnaire can be listed. Also, no baseline has to be assumed. The cumulative model was popularised by McCullagh (1980) after earlier versions were proposed by other authors (cf. Snell (1964); Walker and Duncan (1967); Williams and Grizzle (1972)). Let

$$\tilde{Y} = -X^\top \boldsymbol{\beta} + \epsilon$$

with ϵ following a continuous distribution F. The response variable Y is assigned to category r iff the latent variable's value lays between two thresholds on the latent scale:

$$Y = r \quad \iff \quad \beta_{0r-1} < \tilde{Y} \leq \beta_{0r}.$$

The thresholds $-\infty = \beta_{00} < \beta_{01} < ... < \beta_{0k} = \infty$ also serve as the category-specific intercept. This can easily be seen when looking at the probability of Y to fall at least into category r:

$$\begin{aligned} P(Y \leq r|X) &= P(-X^\top \boldsymbol{\beta} + \epsilon \leq \beta_{0r}) \\ &= P(\epsilon \leq \beta_{0r} + X^\top \boldsymbol{\beta}) \\ &= F_\epsilon(\beta_{0r} + X^\top \boldsymbol{\beta}). \end{aligned}$$

The approach splits the model into binary parts $1, ..., r-1$ and $r, ..., k$ for each category, which enables collapsing other categories below or above r. The probability for Y falling

into category r is derived via

$$\begin{aligned} P(Y = r|X) &= P(Y \leq r|X) - P(Y \leq r-1|X) \\ &= F_\epsilon(\beta_{0r} + X^\top \boldsymbol{\beta}) - F_\epsilon(\beta_{0r-1} + X^\top \boldsymbol{\beta}). \end{aligned}$$

The final model is thus dependent on $X, \boldsymbol{\beta}, \beta_{00}, ..., \beta_{0k}$ and the choice of F, but no longer on the latent $\tilde{Y}$. F is also called the link function.

One of the most common link functions chosen due to its easy interpretation is the logistic distribution

$$F(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)}$$

yielding

$$P(Y \leq r|X) = \frac{\exp(\beta_{0r} + X^\top \boldsymbol{\beta})}{1 + \exp(\beta_{0r} + X^\top \boldsymbol{\beta})} \quad \equiv \quad \log\left[\frac{P(Y \leq r|X)}{P(Y > r|X}\right] = \beta_{0r} + X^\top \boldsymbol{\beta}.$$

The parameters can be interpreted as the (logistic) odds to fall at most into category r over the odds to fall into category $r+1$ or higher. For each category, the probabilities are dichotomized into those below or equal the threshold β_{0r} and above it. This model is also called the proportional odds model. The name can be deduced from the following characteristic: Given the ratio of the probability of $Y \leq r$ to the probability of $Y > r$ (odds) of two different populations X and $\tilde{X}$, the relation between those two odds stays the same (proportional) over all categories:

$$\frac{P(Y \leq r|X)/P(Y > r|X)}{P(Y \leq r|\tilde{X})/P(Y > r|\tilde{X})} = \frac{\exp(\beta_{0r} + X^\top \boldsymbol{\beta})}{\exp(\beta_{0r} + \tilde{X}^\top \boldsymbol{\beta})} = \exp((X - \tilde{X})^\top \boldsymbol{\beta}). \tag{3.1}$$

In other words, if the cumulative odds $(P(Y \leq r|X)/P(Y > r|X)$ in population X are for example twice the cumulative odds in population $\tilde{X}$, this cumulative odds ratio holds for all categories. In the same way the relation between two categories r and s is independent of the population's covariates:

$$\log\left[\frac{P(Y \leq r|X)/P(Y > r|X)}{P(Y \leq s|X)/P(Y > s|X)}\right] = \log\left[\frac{\exp(\beta_{0r} + X^\top \boldsymbol{\beta})}{\exp(\beta_{0s} + X^\top \boldsymbol{\beta})}\right] = \beta_{0r} - \beta_{0s}. \tag{3.2}$$

In order to use this model, the implied assumptions (3.1) and (3.2) must hold. This means,

that

$$\boldsymbol{\beta} = \boldsymbol{\beta_1} = ... = \boldsymbol{\beta_k}$$

must hold. For two populations, the cumulative odds ratio should not depend on the category, and this must hold for all p predictors involved. The corresponding linear null hypothesis

$$H_0 : \beta_{1j} = ... = \beta_{kj}, \quad j = 1, ..., p$$

can be tested with the likelihood ratio test, the Wald test or the score test.

3.4 Comparison

The two most prominent types of ordinal regression models have been presented. The sequential model is not applicable to the given response variables, as the assumption that categories are traversed in ascending order does not hold. Cumulative models on the other hand follow the approach of an underlying continuous variable $\tilde{Y}$. This is a valid assumption that can be made for questions about opinions, as it can be assumed that people's attitudes are more gradual than in a questionnaire can be realised. The proportional odds model has a very intuitive interpretation. As long as its assumptions hold, it is thus the most attractive option for regression analysis of survey data on opinions and is selected for this analysis. Now that the correct model choice for ordered categorical responses has been outlined, proper treatment of ordinal dependent variables is described.

Chapter 4

Penalization Terms

4.1 Basic Idea

Employing ordinal covariates requires other methods than those for linear or binary covariates. Be it dummy or split coding, the ordered categories of a predictor are included in the model with several parameters, usually one dummy parameter per category. They then constitute a group of parameters belonging to the same predictor and are therefore called grouped parameters or grouped variables. The usual way to estimate the parameters' values is via maximum-likelihood optimisation yielding $\hat{\boldsymbol{\beta}}_{\boldsymbol{ML}}$. However this method does not take into account the grouping of the variables, so it may select some categories and exclude others or estimate very different values for adjacent categories. For ordinal data though, a smooth transition between adjacent categories can be assumed (Tutz and Gertheiss, 2016). In order to account for the relationship between dummies of ordered categorical data and to consider the grouping, a penalization term can be added to the regression model (Tutz and Gertheiss, 2016).

The general form of a penalized log-likelihood for a parameter vector $\boldsymbol{\beta}$

$$l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + J_\lambda(\boldsymbol{\beta})$$

adds a penalization term $J_\lambda(\boldsymbol{\beta})$ to the usual log-likelihood before solving. λ is a tuning parameter and fixed via iterative computation (Tutz, 2012, p. 145-146).

One general penalization term called bridge estimator

$$J_\lambda(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} |\boldsymbol{\beta_j}|^\gamma$$

with $\boldsymbol{\beta_j} = (\beta_{j1}...\beta_{jk_j})^T$ was derived by Frank and Friedman (1993) and comprises for $\gamma = 1$ the LASSO penalization and for $\gamma = 2$ the ridge penalization. They are both smoothing penalties with no grouping property in the first place. Both classes will be discussed in the following sections and then further developed in order to account for grouping from ordinal variables. Subsequently, other penalization terms for grouped variables are presented that can not be derived from the bridge estimator. The most appropriate penalty terms are intended to be compared, but they are not available for implementation yet. Therefore, only one penalty is employed, as is discussed in the final section.

4.2 Ridge Penalization

The ridge penalization term (Hoerl and Kennard, 1970) has the form

$$J_\lambda(\boldsymbol{\beta}) = \lambda \sum_{j=1}^{p} \boldsymbol{\beta_j}^2$$

and transforms the usual maximization problem of ordinal least squares (OLS) into

$$\begin{aligned}\hat{\boldsymbol{\beta}}_R &= \underset{\beta}{\operatorname{argmin}} \left[(Y - X\boldsymbol{\beta})^\top (Y - X\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^\top \boldsymbol{\beta}\right] \\ &= (X^\top X + \lambda I)^{-1} X^\top Y.\end{aligned}$$

The shrinkage parameter $\lambda \geq 0$ is a meta parameter. For $\lambda > 0$ it adds a small amount on the diagonal of $X^\top X$ which makes the matrix always invertible. The variance is smaller than the standard OLS Variance, as

$$\sigma^2(X^\top X + \lambda I)^{-1} < \sigma^2(X^\top X)^{-1}, \quad \lambda > 0$$

shows. For $\lambda = 0$ the ridge regression yields the standard OLS-estimator and both variances are the same. It can be shown that $\hat{\boldsymbol{\beta}}_R$ is a biased estimator (Fahrmeir et al., 2009, p. 172).

A generalization of the ridge penalty as a first-order difference penalty is given by Gertheiss et al. (2021) as

$$J_j(\boldsymbol{\beta_j}) = \lambda_j \sum_{l=2}^{k_j} (\beta_{jl} - \beta_{j,l-1})^2.$$

and a second-order difference penalty that can be used to penalize derivations from linearity

as

$$J_j(\boldsymbol{\beta_j}) = \lambda_j \sum_{l=2}^{k_j} (\beta_{j,l+1} - 2\beta_{jl} + \beta_{j,l-1})^2.$$

Note that here λ_j is indexed to vary for each variable x_j, thus the maximization problem turns into

$$l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - \sum_{j=1}^{p} J_j(\boldsymbol{\beta_j}).$$

For smooth effects of a categorical predictor, Gertheiss and Tutz (2009) showed that the ridge type penalties for differences decreases the mean squared error of estimates strongly. Ridge penalization in general reduces the influence of variables, but can not shrink them to exactly zero. This is a property the LASSO penalization has (Tutz, 2012, p. 149).

4.3 LASSO Penalization

The first one to propose the LASSO, an acronym for *Least Absolute Shrinkage and Selection Operator*, was Tibshirani (1996). As above mentioned, for $\gamma = 1$ the LASSO-penalized OLS-estimator is

$$\hat{\boldsymbol{\beta}}_L = \underset{\beta}{\operatorname{argmin}} \; [(Y - X\boldsymbol{\beta})^\top (Y - X\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\boldsymbol{\beta}_j| \;].$$

Due to $|\boldsymbol{\beta}_j|$ it cannot be computed via an explicit formula but there exist several algorithms for its numerical optimisation (Yuan and Lin, 2006). It yields a biased estimator (Hoyer, 2018). The LASSO can estimate parameters to be exactly zero (Tutz and Gertheiss, ming, p. 9) and therefore exclude the respective variables from the model. Categorical and ordinal predictors are usually transformed into dummy variables, so that in order to exclude the predictor behind the corresponding dummies, they all have to be set to zero simultaneously. This is implemented in the group LASSO.

4.4 Group LASSO Penalty

The selection or exclusion of grouped variables, in this case groups of dummy variables belonging to one predictor j, can be obtained via employment of the penalization term

$$J_j \boldsymbol{\beta}_j = \sqrt{k_j} ||\boldsymbol{\beta}_j||_2.$$

$||\boldsymbol{\beta}_j||_2 = (\beta_{j1}^2 + ... + \beta_{jk_j}^2)^{1/2}$ represents the l_2-norm of the parameters of the jth predictor (Tutz and Gertheiss, ming, p. 9). The group LASSO is a sparsity-inducing penalty, referring to the property to exclude variables and thus create sparser models. This is especially helpful in contexts with many possible predictors, for example genetics. Two penalty terms that have been studied especially in the latter context, are the minimax convex penalty and smoothly clipped absolute deviation penalty which will be discussed in the following.

4.5 MCP and SCAD Penalty

The minimax convex penalty (MCP) (Zhang, 2010) is another sparsity-inducing penalty defined on [0, ∞):

$$J(\beta) = \begin{cases} \lambda\beta - \frac{\beta^2}{2a}, & \text{if } \beta \leq a\lambda \\ \frac{a\lambda^2}{2}, & \text{if } \beta > a\lambda \end{cases}$$

with $\lambda \geq 0$ and $a > 1$. A one-dimensional β is assumed here. First, the penalization rate is the same as the LASSO but it then relaxes it more and more until $\beta > a\lambda$, where the penalization rate turns zero. This can be illustrated by the first derivative (Breheny and Huang, 2011):

$$J'(\beta) = \begin{cases} \lambda - \frac{\beta}{a}, & \text{if } \beta \leq a\lambda \\ 0, & \text{if } \beta > a\lambda \end{cases}$$

The smoothly clipped absolute deviation (SCAD) penalty (Fan and Li, 2001) is similar to the MCP and can be presented by the first derivation

$$J'(\beta) = \sum_{j=1}^{p} \lambda \left\{ I(\beta \leq \lambda) + \frac{(a\lambda - \beta)_+}{(a-1)\lambda} I(\beta > \lambda) \right\}$$

for some $a > 2$ and a one-dimensional $\beta > 0$. This penalty function corresponds to a quadratic spline function with knots at λ and $a\lambda$. Both MCP and SCAD can be adapted to grouped parameters as well (Breheny and Huang, 2011; Huang et al., 2012; Ogutu and Piepho, 2014) by applying it to the sub-vectors $\boldsymbol{\beta}_j$ (Tutz and Gertheiss, ming, p. 9).

A comparison of group LASSO, group MCP and group SCAD among others in an application to genomic prediction found that all methods had a relatively high predictive accuracy and may be employed for selection decisions (Ogutu and Piepho, 2014). The difference is that in genomic prediction $n \ll p$ whereas in the sociological question at hand

$n > p$.

The overview over different penalty terms for penalized ordinal regression which allows for smoothing and selection of ordinal grouped variables identified group LASSO, group MCP and group SCAD as promising. Therefore, the performances of proportional odds models with these three penalization terms are to be compared. The software for this undertaking is reviewed in the next section.

4.6 Software Implementation

An overview over available software packages for the implementation in `R` (R Core Team, 2021) revealed several options for ordinal Regression (`ordinal` (Christensen, nd), `ordinalgmifs` (Archer et al., 2014), `VGAM` (Yee, 2010), `mvord` (Hirk et al., 2020), `ordinalNet` (Wurm et al., 2021), mgcv (Wood, 2017)) as well as the mentioned penalty terms (`grpreg` (Breheny and Huang, 2015), `ncvreg` (Breheny and Huang, 2011), `grplasso` (Meier, 2020), `penalized` (Goeman et al., 2018), `ordPens` (Hoshiyar and Gertheiss, 2021)). The combination of cumulative logit models with penalization terms narrows the available packages down fundamentally: only one package offers a penalized cumulative logit model with ordinal predictors, namely a recent extension of `ordPens` (Hoshiyar and Gertheiss, 2021) which allows for a new constructor function for smooth terms in `mgcv`'s `gam()` (Wood, 2011) via `s(..., bs = "ordinal")` (Gertheiss et al., 2021). This extension allows for first- and second-order generalized ridge penalties. In the documentation of the `gam()` function is stated that, although gam stands for generalized additive models, the term is "taken to include any quadratically penalized GLM" (Wood, nd). This applies to the ridge-type penalties.

As until now no other options are available, the ordinal smoothing penalty via `ordPens` and `mgcv::gam()` is implemented, although the more preferable option would be group LASSO, group MCP and group SCAD due to their additional selection property, as described above.

The named combination of packages has been implemented only once by Gertheiss et al. (2021) for `family = gaussian` and `family = binomial`. For computation of a proportional odds model in `gam()` `family = ocat` can be used, but this has not been done before. Therefore, previous to running the model on real data, its properties are examined. This is realised with a scenario analysis based on simulated survey data.

Chapter 5

Method Analysis

5.1 Guiding Questions

The data is simulated using the R package `GenOrd` (Barbiero and Ferrari, 2015) which builds upon gaussian copulas to generate multivariate discrete random variables with a pre-specified correlation matrix. The package allows for independent definition of the marginal distributions of each variable and the correlation matrix. Correlation ρ in this thesis always refers to Spearman's rank correlation coefficient due to the ordinal nature of the data. Since the creation of larger correlation matrices is more complex, the focus here is on cases with three ordinal predictors. For a first insight into the model's behaviour, four different combinations of marginal distributions and correlations are tested. The simulations are conducted with the specifications depicted in table 5.1. Unless otherwise stated,

	Combination 1	Combination 2	Combination 3	Combination 4
y	uniform	uniform	uniform	*aut_andere* (AID:A)
x_1	bellshape	bellshape	exponential	exponential
x_2	uniform (7 categories)	uniform (7 categories)	uniform	uniform
x_3	uniform (7 categories)	uniform (7 categories)	uniform	bellshape
ρ_{y,x_1}	0.8	0.8	0.6	0.6
ρ_{y,x_2}	0.6	0.6	0.6	0.6
ρ_{y,x_3}	0.1	0	0	0
ρ_{x_1,x_2}	0.12	0.12	0.12	0.12
ρ_{x_1,x_3}	0.08	0	0	0
ρ_{x_2,x_3}	0.06	0	0	0

Table 5.1: First combinations of simulated data.

predictors and response are ordinal variables with 6 levels. The marginal distributions are

mostly adaptations of the exponential or normal distribution (see table 5.2). Once the

Shape of distribution	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Exponentially-shaped	1/63	2/63	4/63	8/63	16/63	32/63
Bell-shaped	1/12	2/12	3/12	3/12	2/12	1/12

Table 5.2: Given probabilities of exponentially- and bell-shaped frequency distributions for ordinal variables with six levels.

distribution of the variable *aut_andere* was used, which can be found in figure 5.4 on page 22. Given the data, the models are run for the first-order difference penalty specified by `m=1`. Their summaries are examined for significance and confidence intervals. Predictions are then generated and summed up in barplots showing the average estimated probability of each category given the known true category of the observation. As these simulations only pose a first orientation, they are not described in greater detail, but only the mayor aspects are named. The complete output is found in appendix B. Combination 1 shows very easily interpretable predictions always emphasising the true category with highest probabilities (figure 5.2). The least important predictor x_3 is still marked as highly significant. Confidence intervals are very small for x_1 and x_2 and estimates differ clearly from zero (figure 5.1). Estimates of x_3 are very close to zero and for most levels the confidence interval covers the zero. For combination 2, predictions are equally easy to interpret. x_1 and x_2 are highly significant, whereas x_3 is not. This is also mirrored in the plots of the smooth terms. Predictions for combination 3 look different. The predicted probabilities for the true category is still the highest, but the probabilities are more distributed over the other categories. Smooth terms are estimated to be similar to those in combination 2. In combination 4, predictions for two given true categories, namely 2 and 5, do not put the highest probability on the true category but on adjacent ones. The distribution of the the probabilities is less dispersed given the true category 6 than for the others. Significance of predictors is correctly marked and confidence intervals support this. Overall, the probabilities assigned to the different categories differ and although in most cases the true category is emphasised strongest, this does not always hold. Also, the predictors are always identified correctly, but a more thorough investigation would make these findings more generalisable, if confirmed for different scenarios. As the setup of the simulation allows for correlation specification but not for specification of regression parameters, the correct estimation of the parameters for the irrelevant predictor can also be checked. Questions that arise are:

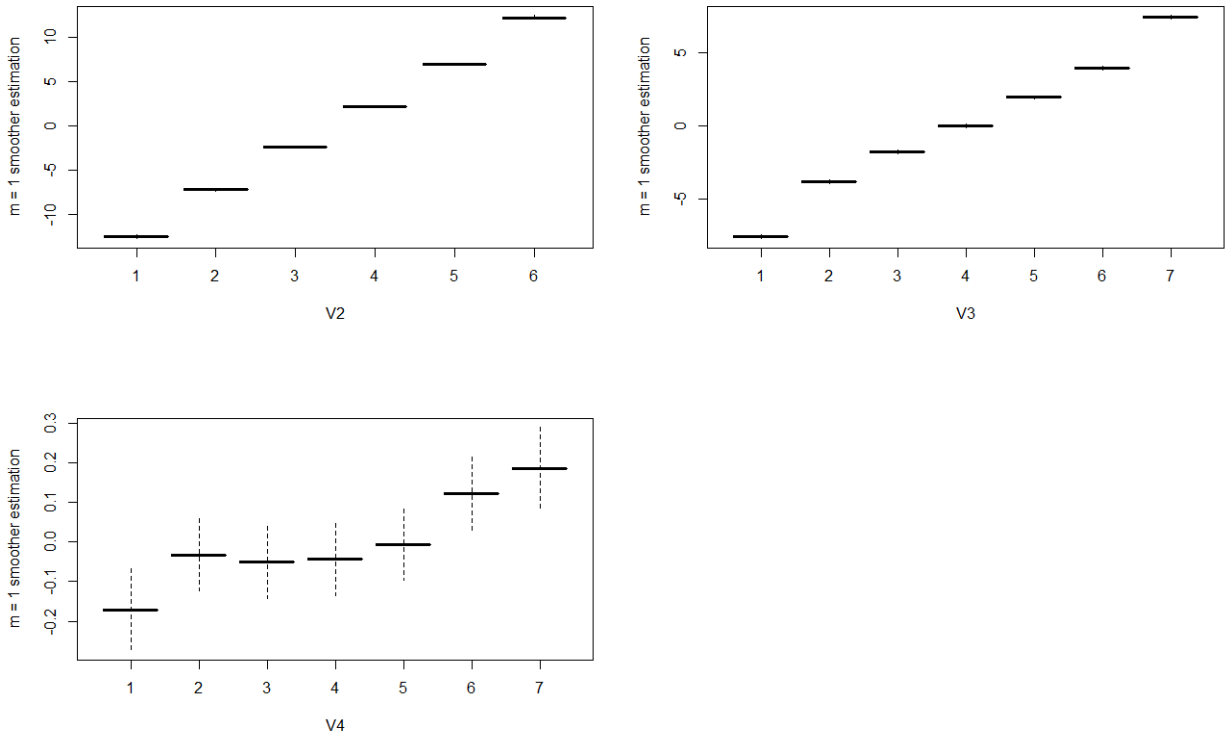


Figure 5.1: Estimated smooth terms for three covariates of combination 1.

I: Do predictions differ depending on the distribution of the response?

II: Do predictions differ depending on the distribution of the correlated predictors?

III: Are relevant and irrelevant predictors correctly identified?

IV: If predictors are not correctly identified: Is it due to their own marginal distribution, their own correlation strength or that of other variables?

As mentioned before, `ordPens` (Hoshiyar and Gertheiss, 2021) allows for first- and second-order penalties (set via `m=1` or `m=2`). In Gertheiss et al. (2021), `m=1` and `m=2` are compared for a `gam()` with continuous response and differences in the width of confidence intervals (better for `m=1`) and correctness of p-values (better for `m=2`) are found. To investigate how the choice of `m=1` and `m=2` influences the proportional odds model implemented in this thesis, the following question is additionally posed:

V: Do models with `m=1` and `m=2` differ in terms of accuracy of predictions and p-values?

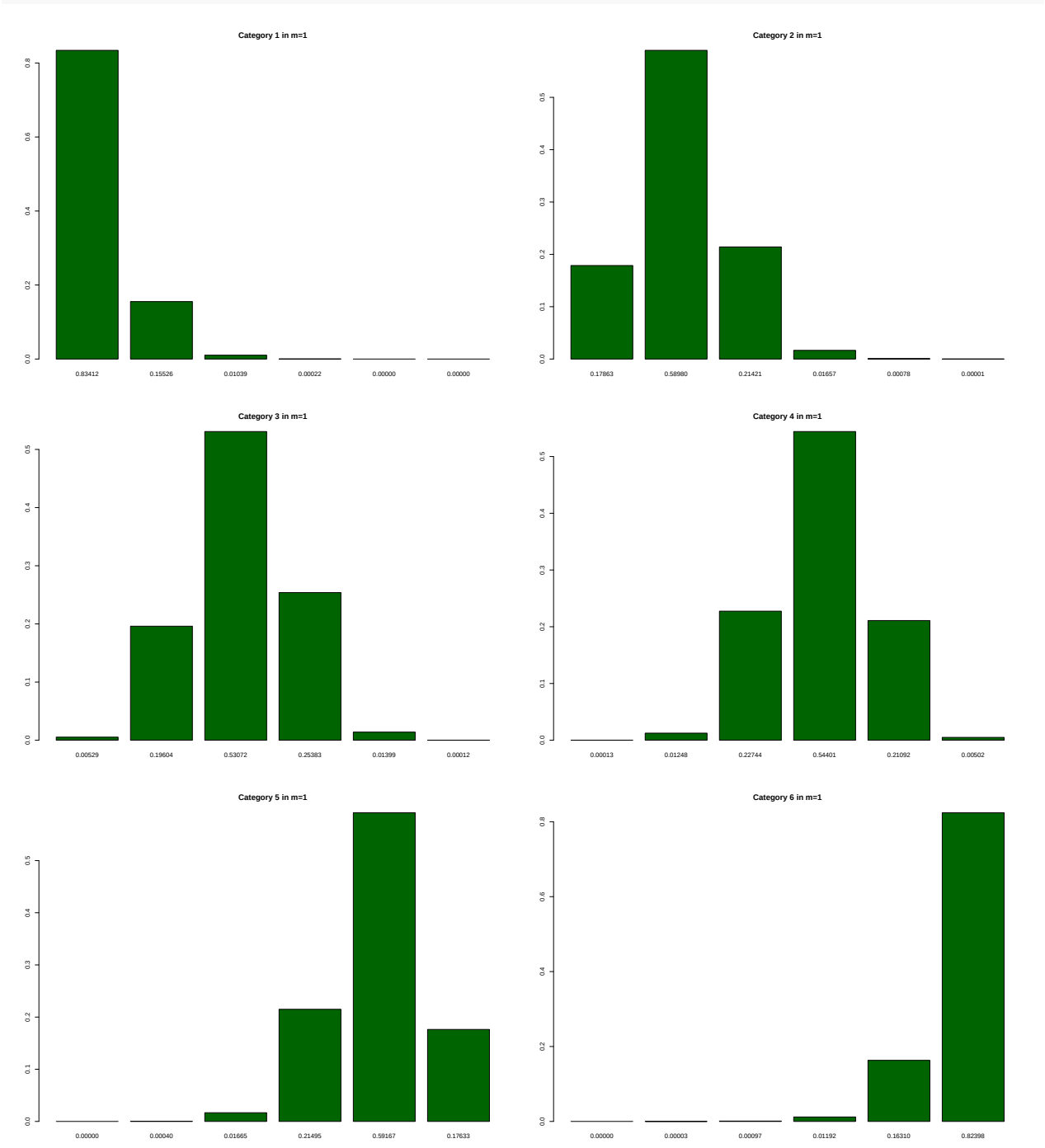


Figure 5.2: Mean estimated predictions given the true category of combination 1.

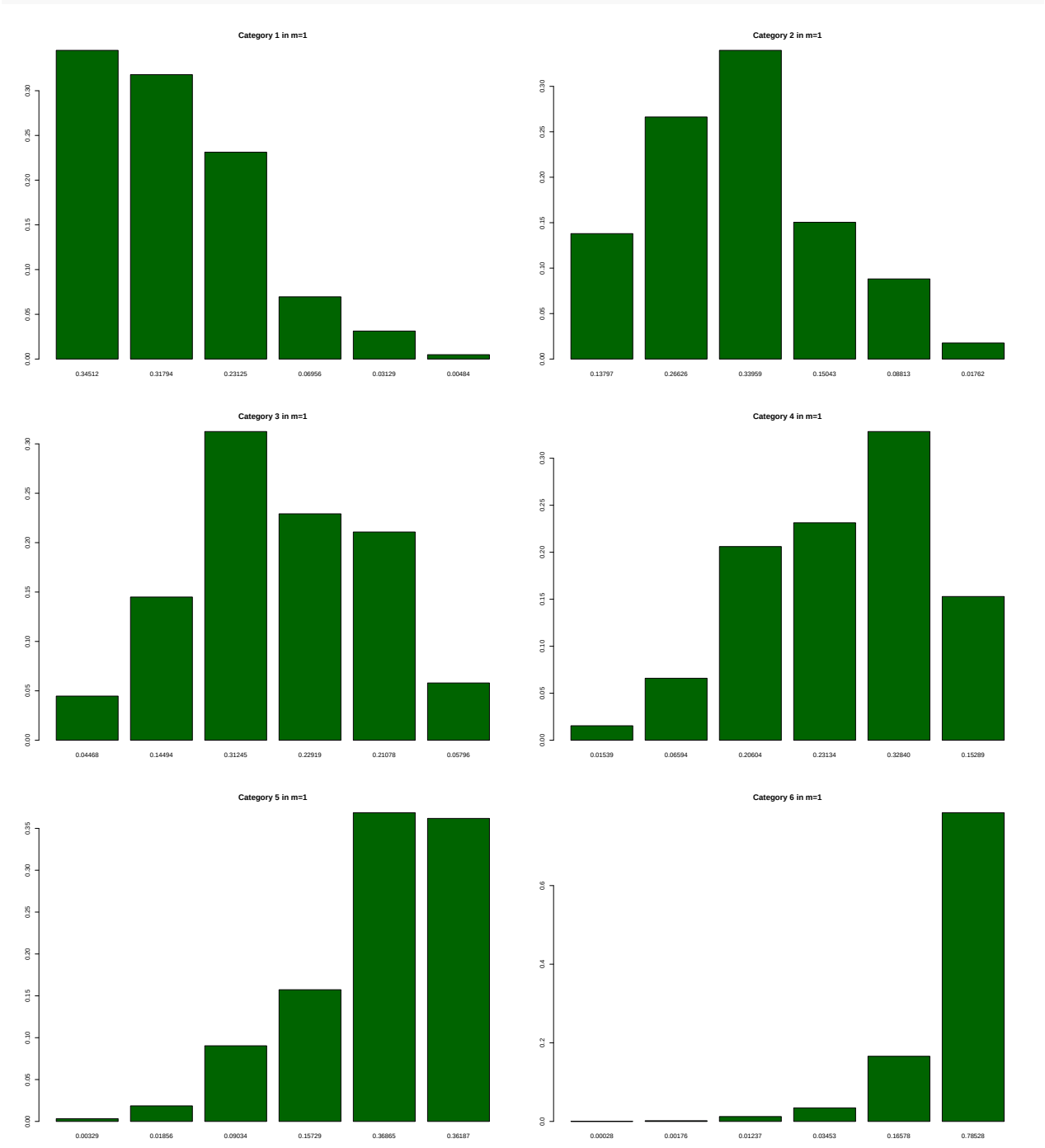


Figure 5.3: Mean estimated predictions given the true category of combination 4.

5.2 Derivation of Scenarios

In a next step, a systematic analysis of different scenarios is conducted in order to answer the questions posed. In the following, it is deduced which simulation variations are considered for each question and which tools are employed to answer them. In general, x_1 and x_2 are defined as relevant predictors with $\rho_{y,x_1} > 0$ and $\rho_{y,x_2} > 0$. x_3 is set as an irrelevant predictor ($\rho_{y,x_3} = 0$).

In order to elucidate question **I** different distributions of the response are compared. Six possible response variables were identified in the AID:A survey. An analysis of their distributions reveals roughly three categories (figure 5.4). Variables *aut_gewohnt* and *aut_eltern*

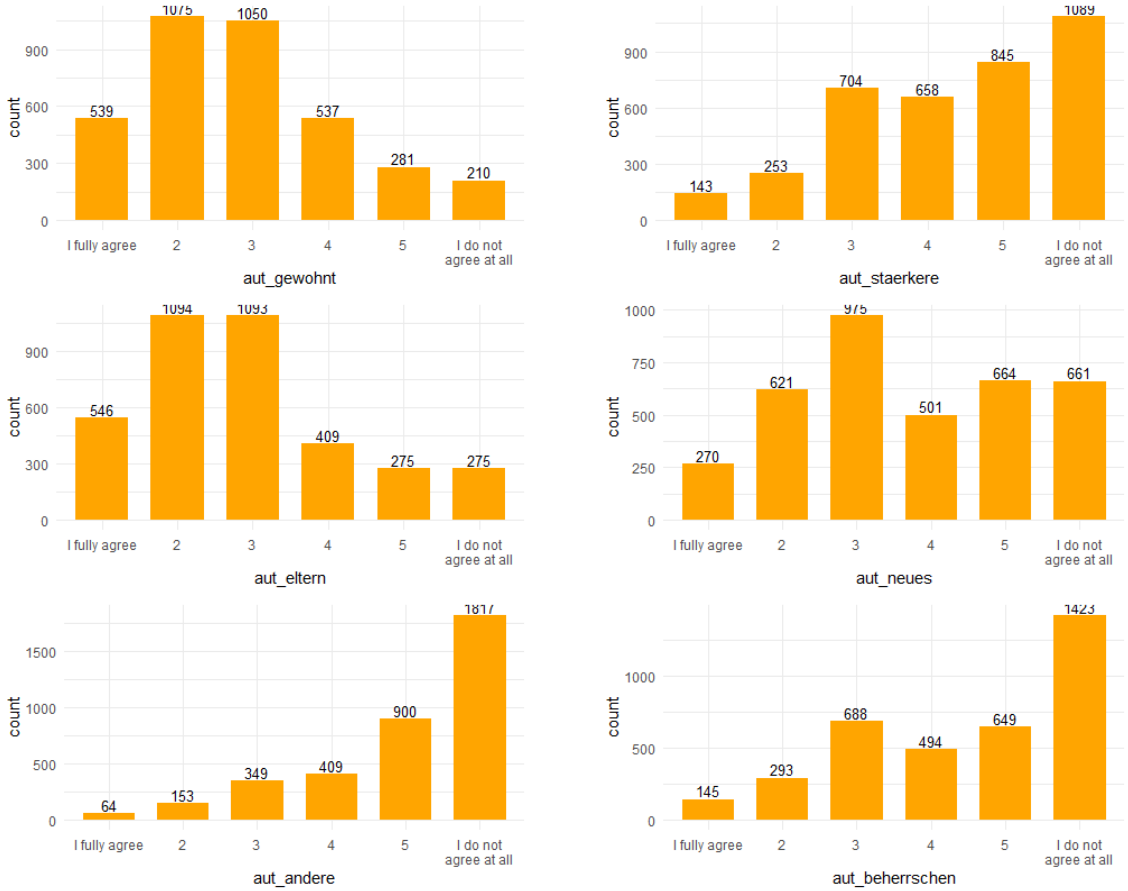


Figure 5.4: Marginal distriubtions of items on authoritarianism.

resemble a shifted normal distribution and *aut_andere* resembles an exponential distribution. The other three variables are better described by polynomials. To get the highest

diversity and at the same time to reference most common distributions, *aut_gewohnt* (bell-shaped, approximately normal distribution) and *aut_andere* (resembling an exponential distribution) are selected. Two types of boxplots are created to present the predictions. First, the prediction probability for each response category is visualised with boxplots differentiated by the true response category of the observation. Second, the estimated probabilities for each of the true categories given the estimated category are gathered in a plot.

In order to answer question **II** different distributions of the predictors have to be compared. The focus lays on x_1 which varies, whereas x_2 remains constant. The ordinal predictors at hand all display very similar exponential characteristics (see figure 5.5). The extreme skewness and thus poor variance can compromise results, so the least skewed are chosen, namely *ben_migration* (although not part of the regression model) and *ben_gender*. *ben_migration* is less skewed and taken as the distribution of x_2. x_1 in contrast should vary between different distributions. As the identified predictors do not comprise sufficient variety and an exponentially-like distribution would be desirable to stick to the idea of theoretically-founded choices, the distribution of response *aut_eltern* is chosen as second distribution of x_1. The same boxplots used for investigating question **I** are employed here.

In order to answer question **III** the p-values of the estimates of all three predictors are examined in the different scenarios. As for x_1 and x_2 no true regression parameters are known, confidence interval coverage is additionally analysed only for x_3. The Correlation ρ_{y,x_3} is constantly set to zero. Thus, p-values should not be significant, which would result in a low rejection rate, and the confidence interval should cover zero. To check this, the rates of significant p-values ($\alpha = 0.1$ and $\alpha = 0.05$) of the repeated simulations are plotted. For x_3 this value should keep the α level and for x_1 and x_2 it should be close to 1. The coverage rate of zero of the confidence interval of x_3 is depicted in a barplot.

In order to answer question **IV**, different distributions are compared. For x_1 they are already given above. x_2 is set constant over all combinations to reduce the number of scenarios to be conducted. For this step, only different distributions of x_3 are added. Its default distribution is a uniform distribution in order to keep any potential influence as uniform as possible. Divergent distributions are taken from the response or predictor x_1 to check whether the penalized proportional odds model is able to distinguish between the relevant and irrelevant predictors even if they share the same distribution. To reduce the number of scenarios, it is realised only for the bell-shaped distributions. Furthermore,

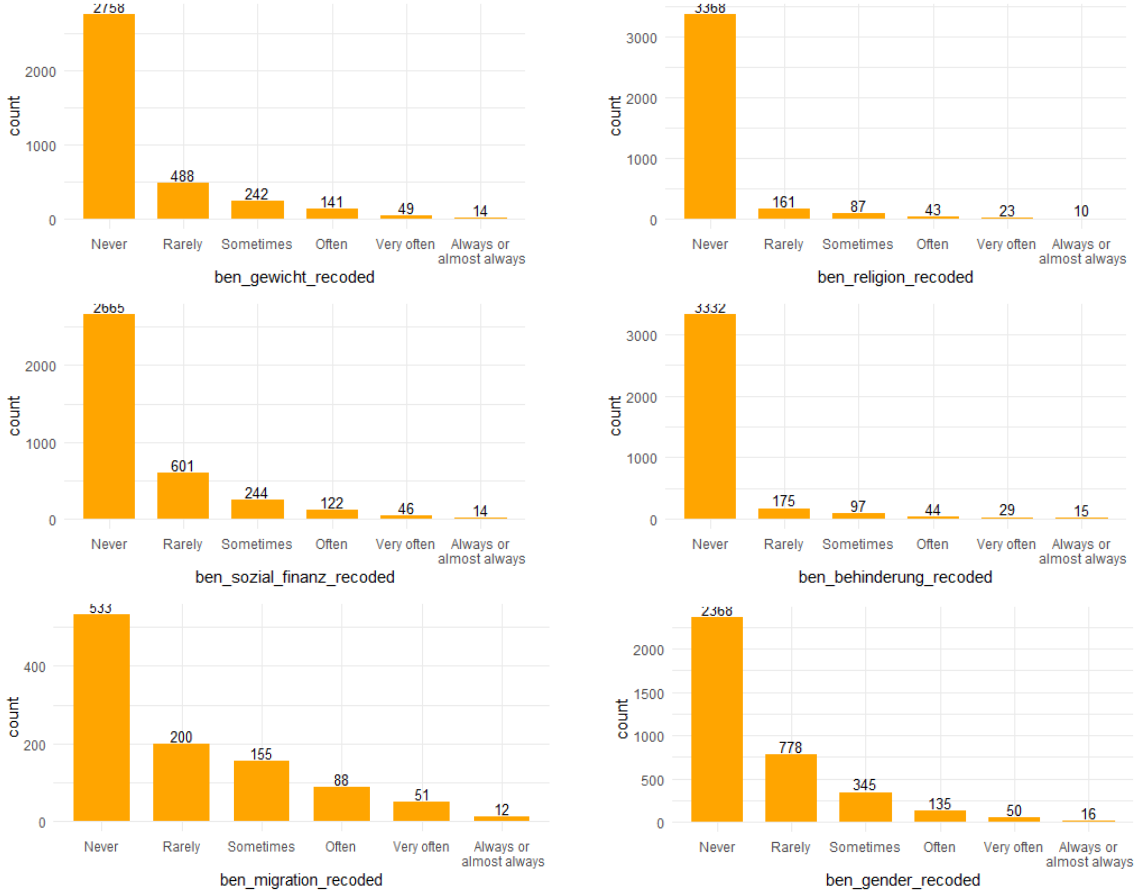


Figure 5.5: Marginal distriubtions of items on experienced deprivation.

the strength of the correlation between predictor x_1 and response will be varied. In order to determine sound values, the correlation matrix of the AID:A data is investigated (see appendix A). The highest overall correlation ($\rho = 0.5$) is between *age* and years of education (*bija*). The correlation within the possible response variables and the ordinal predictors respectively both round up to 0.2. The correlation between the response and the predictors is on average around zero. Therefore, ρ_{y,x_1} takes the values 0.5 and 0.2, whereas ρ_{y,x_2} stays at constant 0.2 and $\rho_{y,x_3} = 0$. All correlations between predictors are multiples of their respective correlation with y (e.g. $\rho_{x_1,x_2} = \rho_{y,x_1} * \rho_{y,x_2}$, procedure taken from Joubert and Langdell (2013)). In order to detect differences due to differing correlations and distributions, the boxplots described for question **I** are investigated.

In order to answer question **V**, all models will be run for `m=1` and `m=2` and the analysis

of p-values and confidence intervals (see question **III**) will be executed for both options and compared.

The different scenarios are depicted in table 5.6.

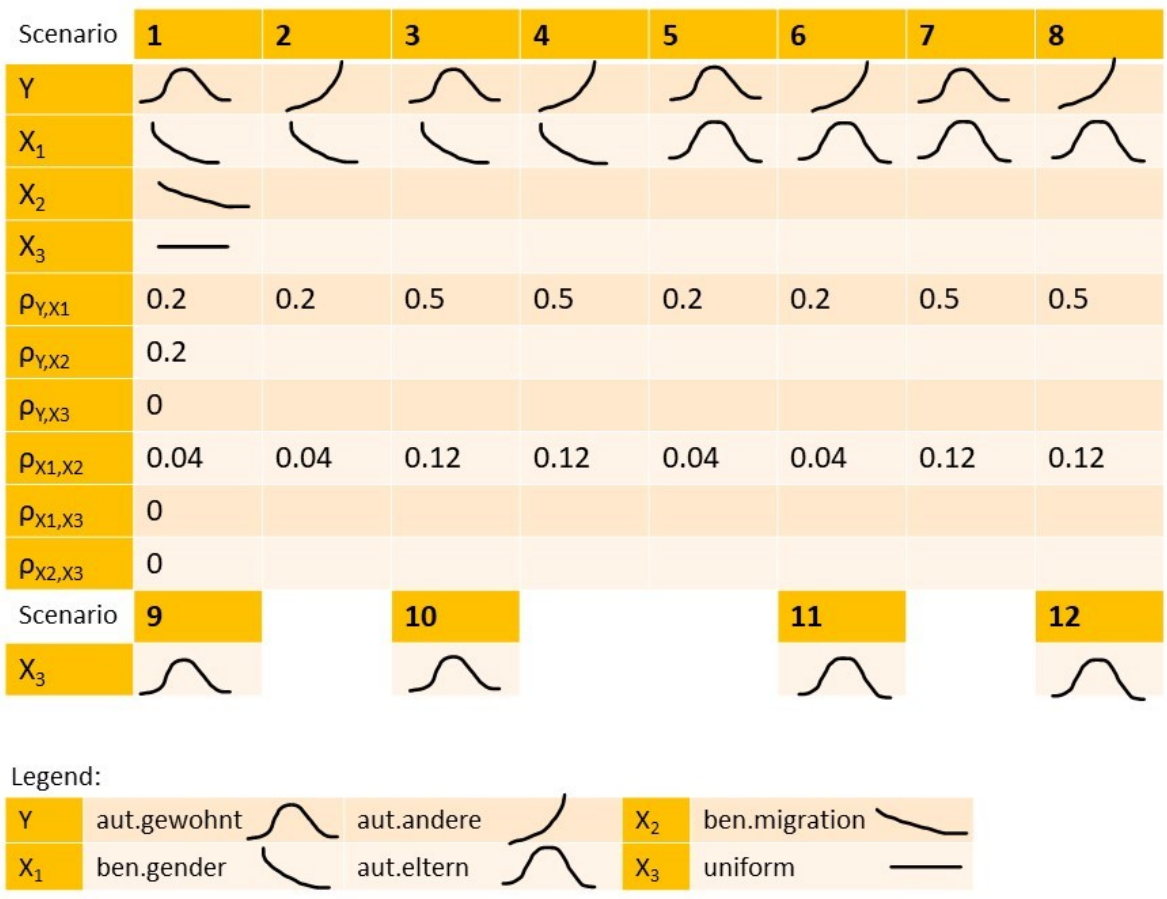


Figure 5.6: Overview over 12 scenarios analysed. Empty cells are filled with the value found in the left column. For scenarios 9-12 all parameters not mentioned are set as in the respective scenario above.

5.3 Method

The general workflow is the same for all scenarios and will be briefly presented. The comprehensive code can be found in appendix C. It consists of two parts; one for a single simulation in order to answer questions **I** and **II** concerning predictions and one for repeated simulations aiming at answering questions **III** to **V** involving p-values and confidence

intervals.

In the first part, the parameters for the simulation in `GenOrd` (Barbiero and Ferrari, 2015) are set and one simulated data set is created. For some seeds categories with very low marginal frequency are not occupied. Therefore, the seed is chosen in such a way that all categories are filled. Next, the distributions of the simulated variables are checked via Pearson's Chi-squared test (as described in Agresti (2007), p. 35) to fit the predefined marginals . Afterwards, models are computed in `gam()` with the `ordPens`' extension (Hoshiyar and Gertheiss, 2021) respectively for `m=1` and `m=2`. Following the procedure in Gertheiss et al. (2021), the restricted maximum likelihood estimator (`method = "REML"`) is employed.

```
m1_sim <- gam(V1 ~
                  s(V2, bs = "ordinal", m = 1)
                + s(V3, bs = "ordinal", m = 1)
                + s(V4, bs = "ordinal", m = 1),
                family=ocat(link="identity",R=6),
                method="REML",
                data=train)
```

Subsequently, the assumptions of the proportional odds model are checked with a likelihood ratio test. The proportional odds model (restricted model) is compared to a multinomial logistic model (unrestricted model) computed via `nnet::multinom` (Venables and Ripley, 2002) (proceeding as in Ford (2015)). Finally, predictions are made and two types of boxplots are created depicting predictions per category given the true or the estimated category respectively.

In the second part, data is simulated as above inside of a loop 100 times. Each time, the results of Pearson's Chi-squared test as well as the difference between the predefined correlation matrix and the correlation matrix of the simulated data are stored for later inspection. Models are computed, but this time estimates and their standard error as well as their p-values are stored for later analysis. In the documentation for `plot.gam()` it is stated that "[t]he function can not deal with smooths of more than 2 variables!" (Wood, 2021). For the present application, the values are confirmed to be correctly displayed in the plot (as confidence interval: $\pm 2 * se$), although standard error values stored in the `plot()` object are partly not explicable. Therefore, a separate prediction with the specification `type = "terms"` is run to obtain correct standard errors. The loop ends hereafter. Now,

rejection rates and confidence interval coverage rate are gathered in plots to answer the questions **III** to **V**.

5.4 Evaluation of Simulation and Modelling

Before going into detail on the findings, the simulation itself is evaluated and the modelling is assessed. The complete output for all scenarios can be found in appendix C.

The first characteristic inspected is the deviance of the simulated correlation matrix from the predefined correlation matrix. Overall, there are some deviations which are not centred around zero. Most of them reveal lower correlations in the simulation than specified beforehand. Furthermore, differences are found strongest and most often for ρ_{y,x_1}. Especially for given correlations of 0.5, simulation correlations are lower. It is possible that the high correlation in conjunction with the other specifications was not feasible for the algorithm. The highest contrast is of median 0.1 (scenario 4, figure 5.7), followed by 0.04 (scenario 8 and 12, figure 5.8) and -0.03 (scenario 10, figure 5.9). This means that for example for scenario 4 in most simulations ρ_{y,x_1} is around 0.4 and not 0.5. Remaining scenarios have lower median deviance from zero. For ρ_{y,x_2}, for those cases where there are deviations they are mostly of 0.01 (exception: scenario 6 with 0.02) with whiskers stretching $\pm$ 0.015, meaning that those ρ_{y,x_2} lie between 0.175 and 0.205 instead of being centred around 0.2. Deviation from specification for ρ_{y,x_3} is always centred around zero, as well as for correlations between predictors. Usually the dispersion of the differences does not exceed 0.03 or 0.02, while the upper and lower quartile cover a range of 0.005 to 0.01 around the median. In summary, most simulations show an acceptable behaviour concerning the correlation matrices. Exceptions with lower correlation than predefined are found for scenarios with specification $\rho_{y,x_1} = 0.5$. Still, correlations of minimum around 0.4 are sufficiently high to differ clearly from the specification $\rho_{y,x_1} = 0.2$ and thus allow for the intended comparison in correlation strength.

Pearson's Chi-squared test was conducted for each simulation, comparing predefined and simulated marginal distributions. The Results for y, x_1 and x_2 are overall satisfying, the distributions resemble each other well. Additionally, the boxplots reveal that values do not vary, neither over the simulations or over the scenarios. For x_3, two types of distributions are distinguished: uniform and bell-shaped. For the uniform distribution, the majority of simulations turns out well, but there are some few runs where the p-value is below 0.1 or 0.05 (see figure 5.10). Both bell-shaped marginals on the other hand show

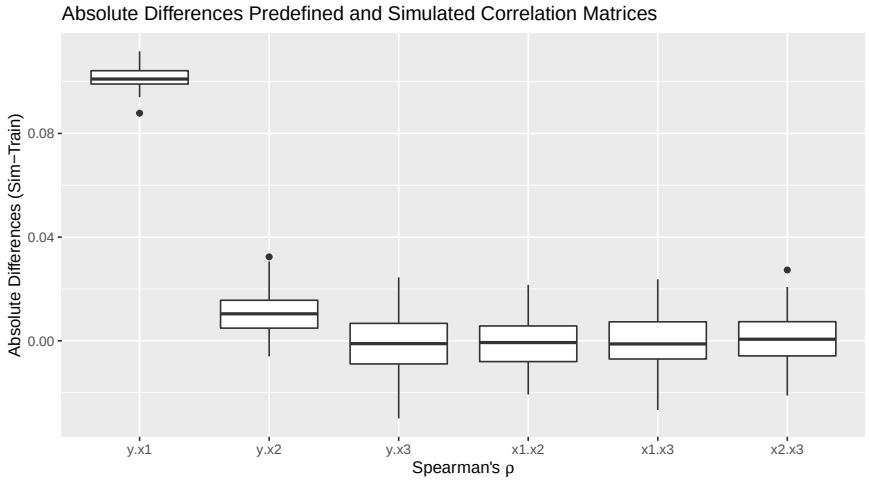


Figure 5.7: Differences between predefined and simulated correlation matrices for $nsim = 100$ for scenario 4.

the same results as y, x_1 and x_2 (see figure 5.11). `GenOrd` (Barbiero and Ferrari, 2015) has issues sampling an independent and uniformly distributed variable as part of a correlated data set.

Overall, it can be said that the simulations approximate the data fairly well, although the specification $\rho_{y,x_1} = 0.5$ is in some cases only realised as correlation around 0.4. The distributions are in the large majority satisfying.

In a next step, the model is assessed in two ways, though not via usual tools as `summary.gam()` or `gam.check()`. This is due to the novelty of the method. As the function in `ordPens` (Hoshiyar and Gertheiss, 2021) to use an ordinal smoothing penalty in `gam()` is very new, it is not certain that these tools work properly.

However, first the assumption of the proportional odds model is tested with a likelihood ratio test. It results that it is fulfilled for scenarios with $\rho_{y,x_1} = 0.2$ but not for those with $\rho_{y,x_1} = 0.5$ (see appendix C of respective scenario). Nevertheless, all scenarios will be analysed as described in section 5.2 in order to observe whether this has direct influence on estimates or predictions.

Now, the predictions are inspected visually. A good prediction would be expected to put emphasis on the category which corresponds to the true category. Clearest would be to compute the highest probability for the correct category as in combination 1 (see figure 5.2 on page 20) or at least show a shift in the estimated probabilities between observations with different true categories. In this simulation study, no such clear behaviour can be

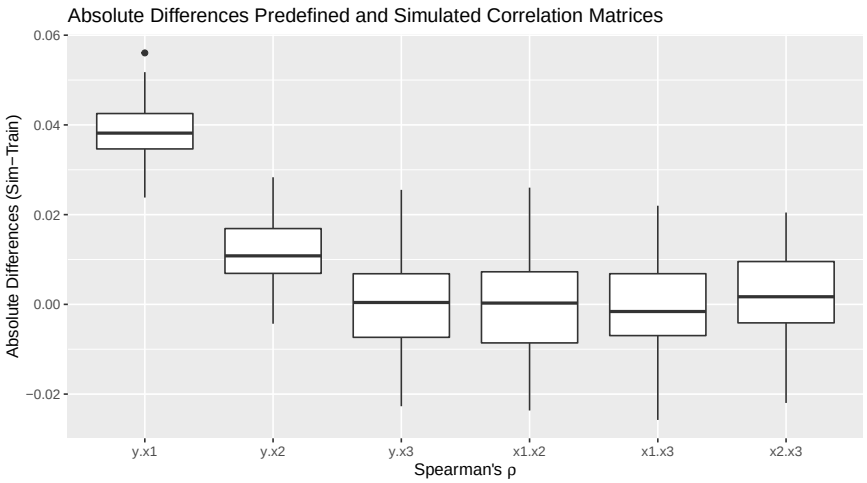


Figure 5.8: Differences between predefined and simulated correlation matrices for $nsim = 100$ for scenario 8.

observed. There are some small changes, but the category which is strongest represented in the distribution of the response turns out to be estimated as most likely in nearly all cases. Even though other categories usually express some small shifts depending on the true category, they usually do not have their highest median probability at their true category but the adjacent one or else. Thus, the predictive strength of the model remains overall poor although two variables with high correlations are given. The models for first and second order differences lead to very similar predictions, therefore only boxplots for `m=1` will be discussed.

The general evaluation of simulation and modelling shows that the simulation is good, although $\rho_{y,x_1} = 0.5$ is sometimes realised lower (minimum 0.4). Model assumptions are only fulfilled in half of the scenarios, which will be kept in mind for further analysis. Predictions are not very sensitive to true categories. How they change depending on the model specification will be discussed in the following section. The other questions posed in section 5.1 will be answered likewise. For ease of understanding, correlations and distributions will be referred to by their specified and not realised values.

5.5 Answering the Guiding Questions

I: Do predictions differ depending on the distribution of the response?

Whether predictions differ depending on the response's distribution is investigated by com-

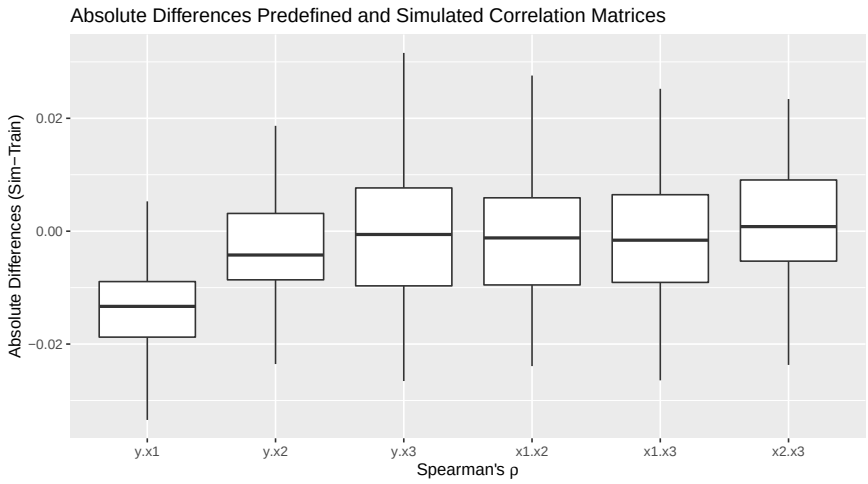


Figure 5.9: Differences between predefined and simulated correlation matrices for $nsim = 100$ for scenario 10.

paring those scenarios which only differ in the y distribution. Thus, scenario 1 vs. 2, 3 vs. 4, 5 vs. 6 and 7 vs. 8 are compared. In fact, differences can be found. They will be described exemplary for scenarios 3 and 4. Scenario 3 has a bell-shaped response distribution, whereas in scenario 4 it is steeply sloped. Both have an approximately exponential x_1 with $\rho_{y,x_1} = 0.5$. The comparison is easier for the boxplots showing the probabilities given the estimated category (see figures 5.12 and 5.13). In scenario 3, the dispersion of the predicted probabilities is higher for all categories than in scenario 4, except for category 6. Furthermore, in scenario 3 whiskers exist mostly on the upper and lower end, whereas in scenario 4 the probabilities are skewed such that there are nearly no upper whiskers. As mentioned, estimated category 6 is the only exception where dispersion of estimated probabilities is larger in scenario 4 than in scenario 3. In similar manner the other scenario pairs can be distinguished, with small differences depending on distribution and correlation strength of x_1. Generally, dispersion is higher given a bell-shaped response distribution than an approximately exponential one, except for the plot for the probabilities given estimated category 6. The question can thus be answered in the affirmative.

II: Do predictions differ depending on the distribution of the correlated predictors?

Now, scenarios 1 vs. 5, 2 vs. 6, 3 vs. 7 and 4 vs. 8 are compared, as they differ in the choice of x_1-distribution respectively. Scenarios 1 to 4 have an approximately exponential x_1-distribution, while scenarios 5 to 8 have a bell-shaped x_1-distribution. Differences

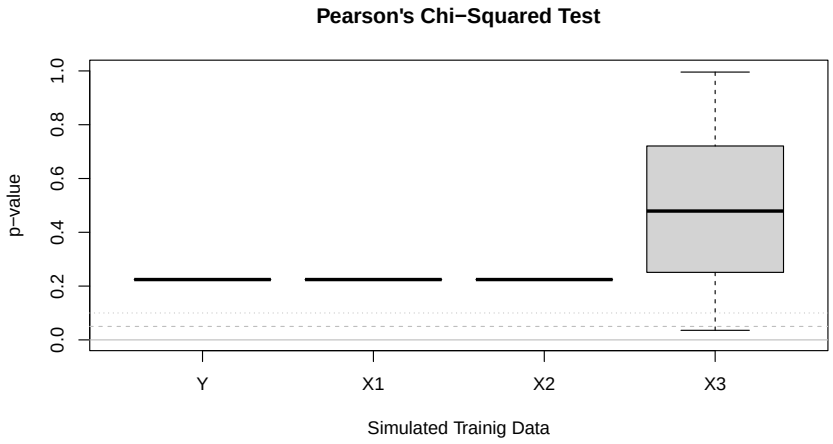


Figure 5.10: Pearson's Chi-squared test for $nsim = 100$ for scenario 8.

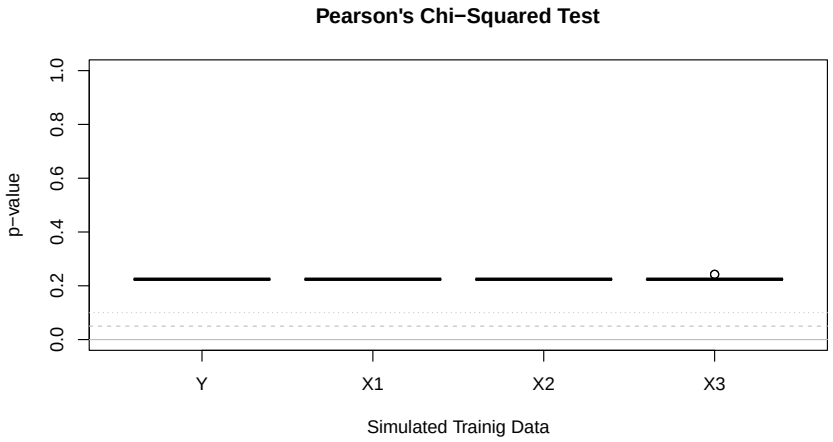


Figure 5.11: Pearson's Chi-squared test for $nsim = 100$ for scenario 12.

are a lot smaller, less systematic and more category-specific. The boxplots of estimated probabilities given the true category are investigated for the following comparison (see exemplary scenario 1 and 5 in figures 5.14 and 5.15). Those of scenarios 1 and 5 show small differences in the first three plots (thus given true categories 1 to 3): in scenario 1, dispersion and whiskers for categories 1 to 4 are more pronounced for one side. In scenario 5 these boxplots are more symmetrical. The comparison of scenarios 2 and 6 reveals that they differ for all plots except for the last one. In scenario 2, the distribution is more asymmetrical concerning the estimation of categories 1 to 4. In scenarios 3 and 7, patterns are more complex. For the first, third and fourth plot, differences are found for boxplots of categories 1 to 3 and for the last two plots, differences are found for categories 4 to 6.

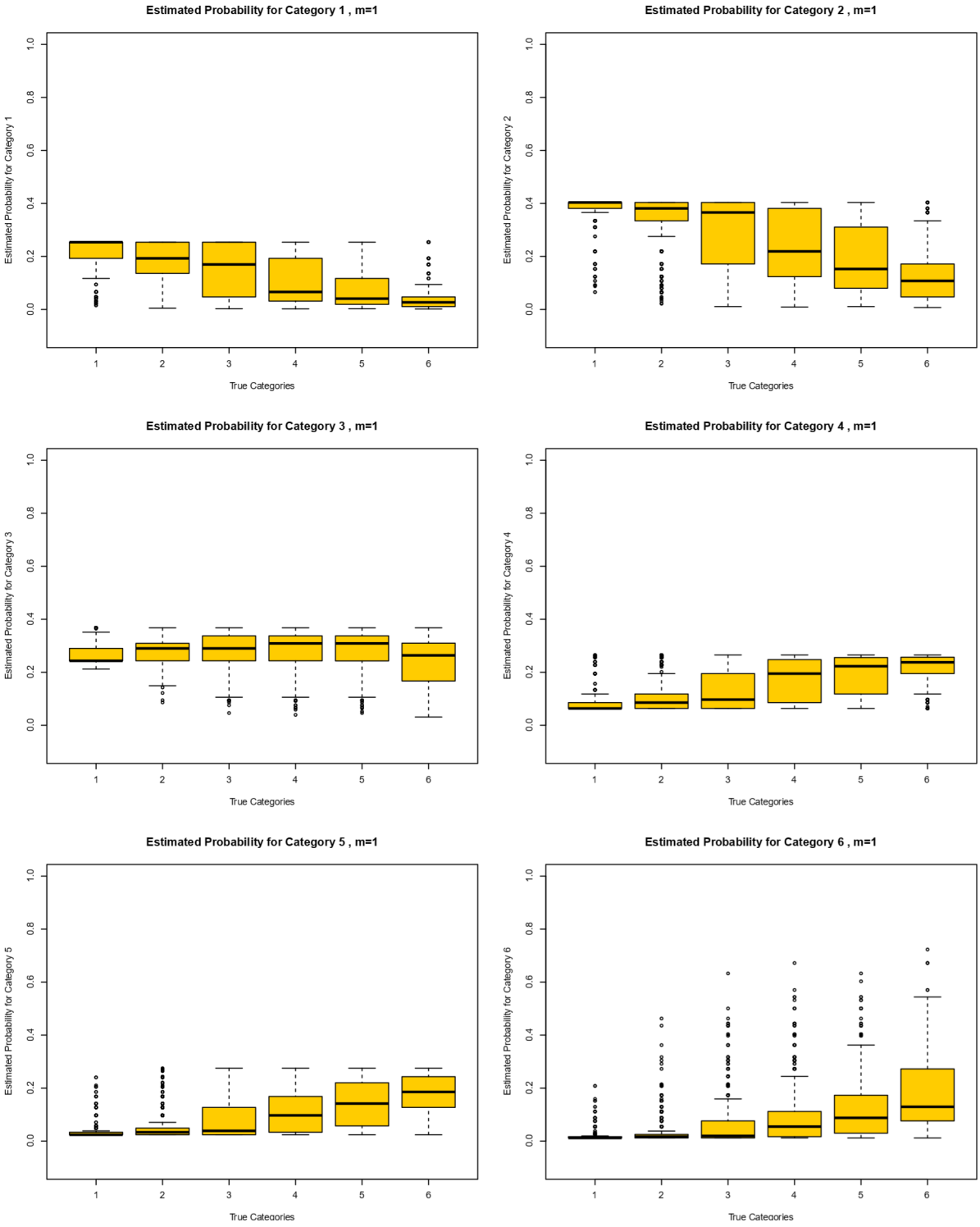


Figure 5.12: Boxplots of estimated predictions given the estimated category of scenario 3.

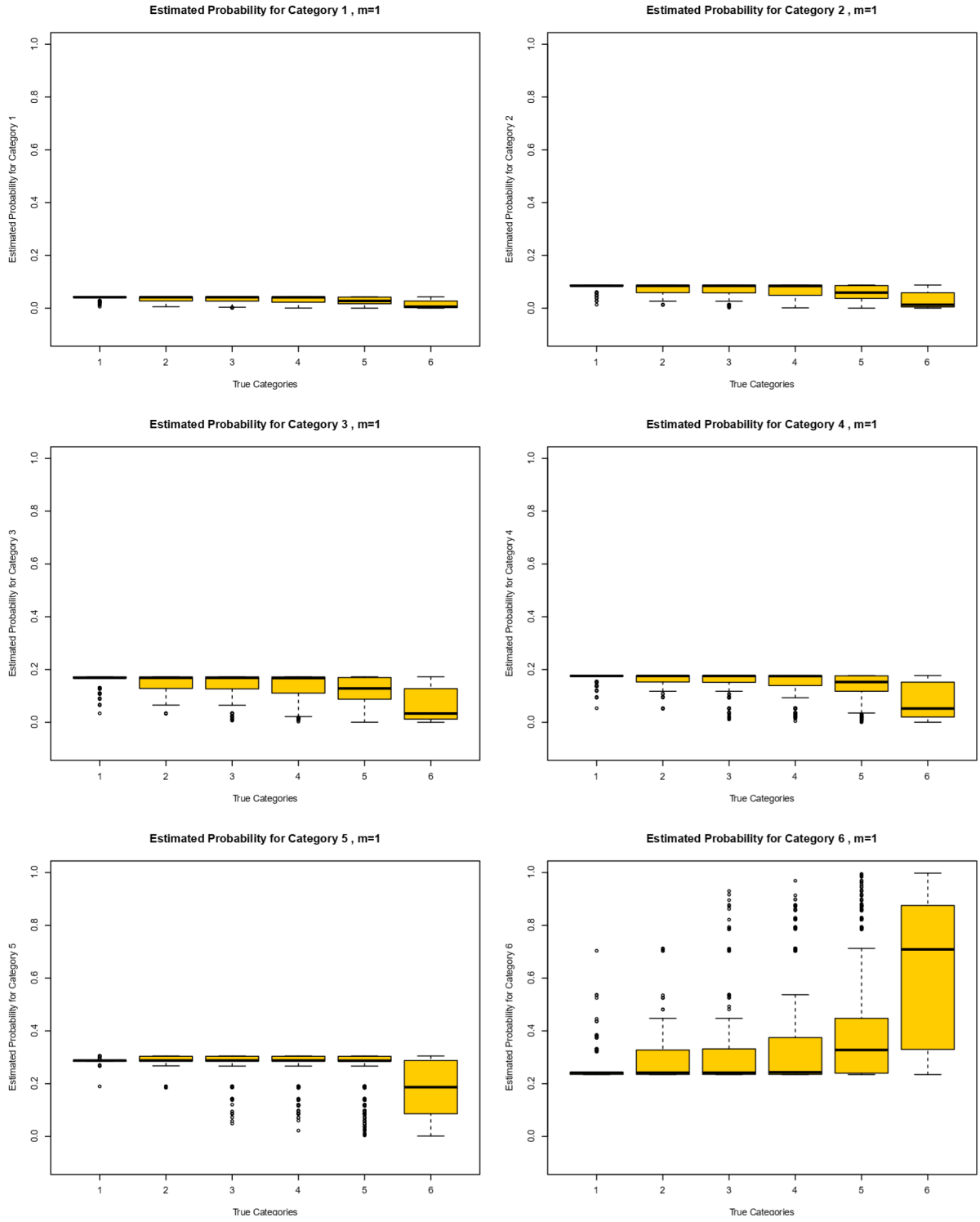


Figure 5.13: Boxplots of estimated predictions given the estimated category of scenario 4.

In the first plot, they show a higher dispersion in scenario 7 and in the third and fourth plot, they have a higher dispersion in scenario 3. In the last two plots, the boxplots of categories 4 to 6 display a higher dispersion in scenario 3. The comparison of the scenarios 4 and 8 exhibits a higher range of predicted values for the first five plots in scenario 8. In general, the x_1-distribution influences the distribution of probabilities for each category (symmetry and dispersion). Predictors with bell-shaped distribution tend towards a more symmetrical spread of predicted probabilities for a category whereas the asymmetrical, approximately exponential distribution tends to a more asymmetrical distribution of the predicted probabilities for a category (see especially scenarios 1, 2, 5 and 6). The question can be answered positively, although the influence of the distribution of predictor x_1 is less systematic than the distribution of y.

III: Are relevant and irrelevant predictors correctly identified?

For the two predictors with correlations different from zero, rejection rates are at around one (figure 5.16), which means they are constantly correctly classified as significant. The correlation ρ_{y,x_3} on the other hand is set to zero, so in a regression analysis this variable should not be significant. Therefore, p-values can be expected to be small. Figure 5.16 shows, that the rejection rates for $\alpha = 0.05$ and $\alpha = 0.1$ are in most scenarios higher than α. The α level of 0.05 is kept only for scenarios 11 and 12, where the x_3 distribution equals the bell-shaped distribution of predictor x_1. Furthermore, the α level of 0.1 is kept for model `m=2` for scenario 12. The strength of the other predictor's correlation seems not to affect this, as $\rho_{y,x_1} = 0.2$ in scenario 11 and $\rho_{y,x_1} = 0.5$ in scenario 12. In order to check the findings for a not bell-shaped distribution, another simulation is run for a 13th scenario (see appendix C.13) which resembles scenarios 3 and 10. The difference is that x_3 has the approximately exponential distribution of x_1. Here, the rejection rate does not hold the α level (figure 5.17). This means that the finding on scenarios 11 and 12 can not be generalised to other distributions than the bell-shaped distribution.

Overall, it can be concluded that p-values seem not to be reliable as they do not correctly identify the irrelevant variable x_3. The correlation of predictors x_1 and x_2 has no effect. It has to be said, however, that only scenarios where the predictor x_1 and the response had differently shaped distributions were analysed. On the other hand, the confidence intervals of x_3 cover the zero in at least 90% of all cases. Mostly, the 95% rate is also reached. Hence, the confidence intervals in this application are found to be trustworthy for irrelevant predictors.

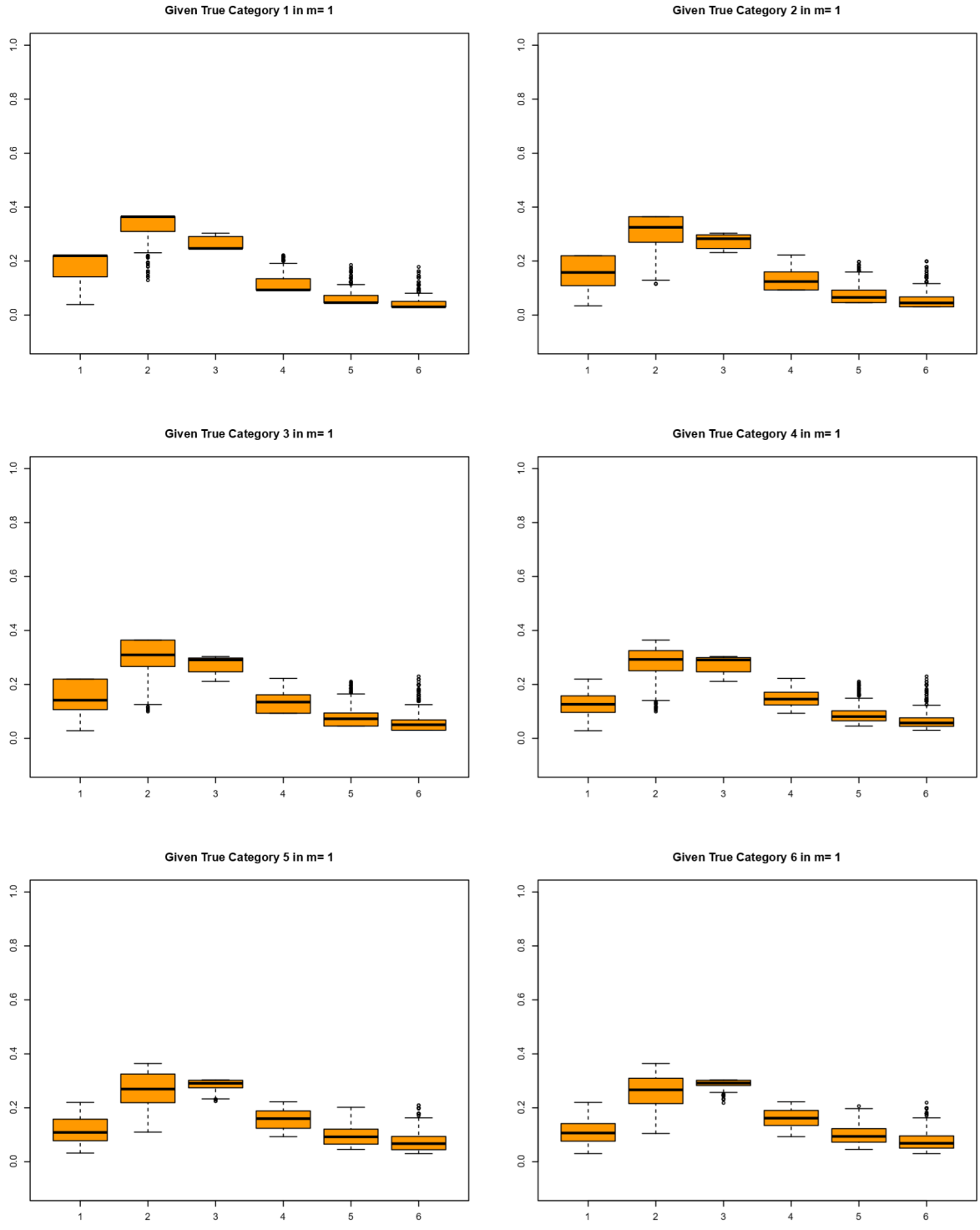


Figure 5.14: Boxplots of estimated predictions given the true category of scenario 1.

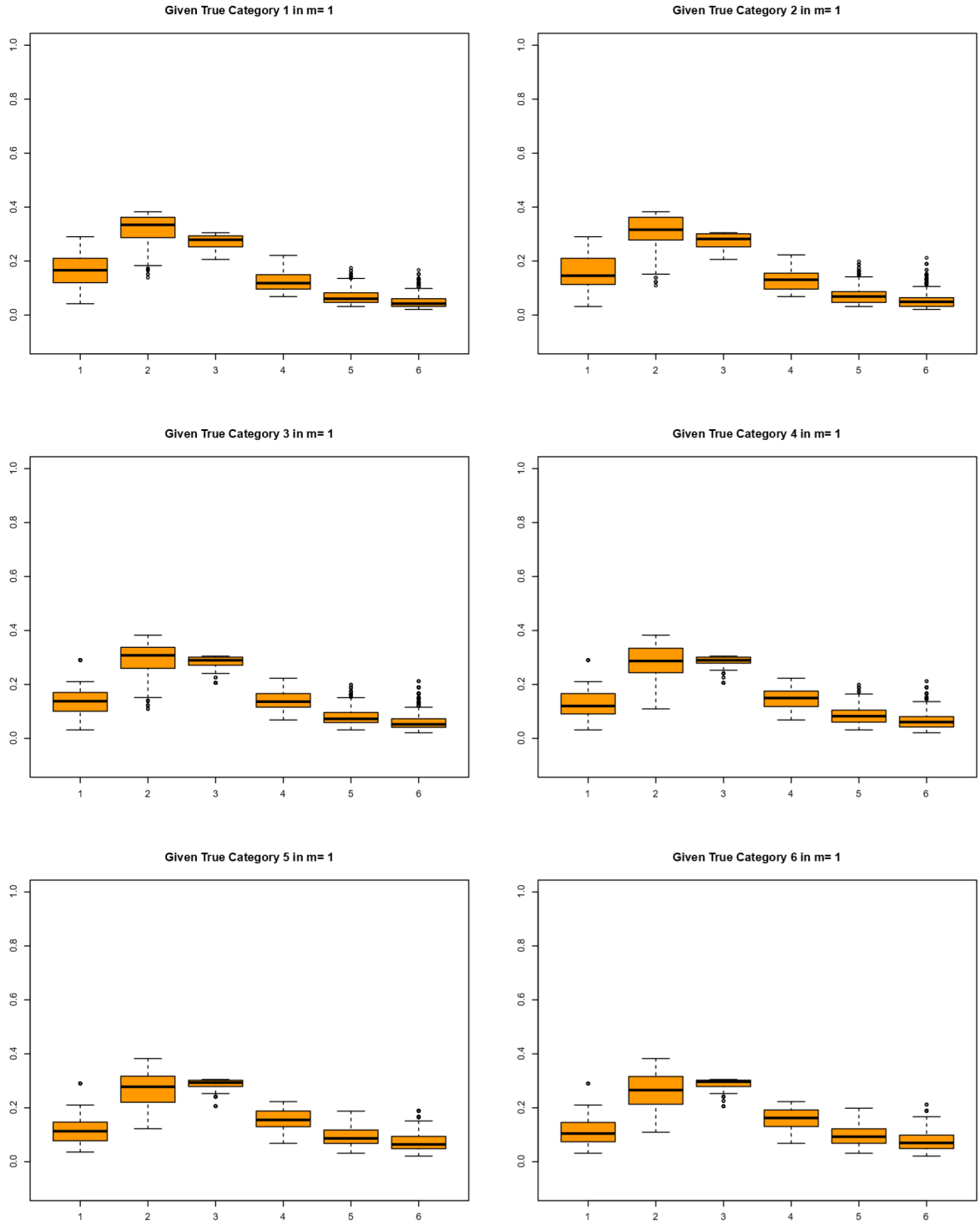


Figure 5.15: Boxplots of estimated predictions given the true category of scenario 5.

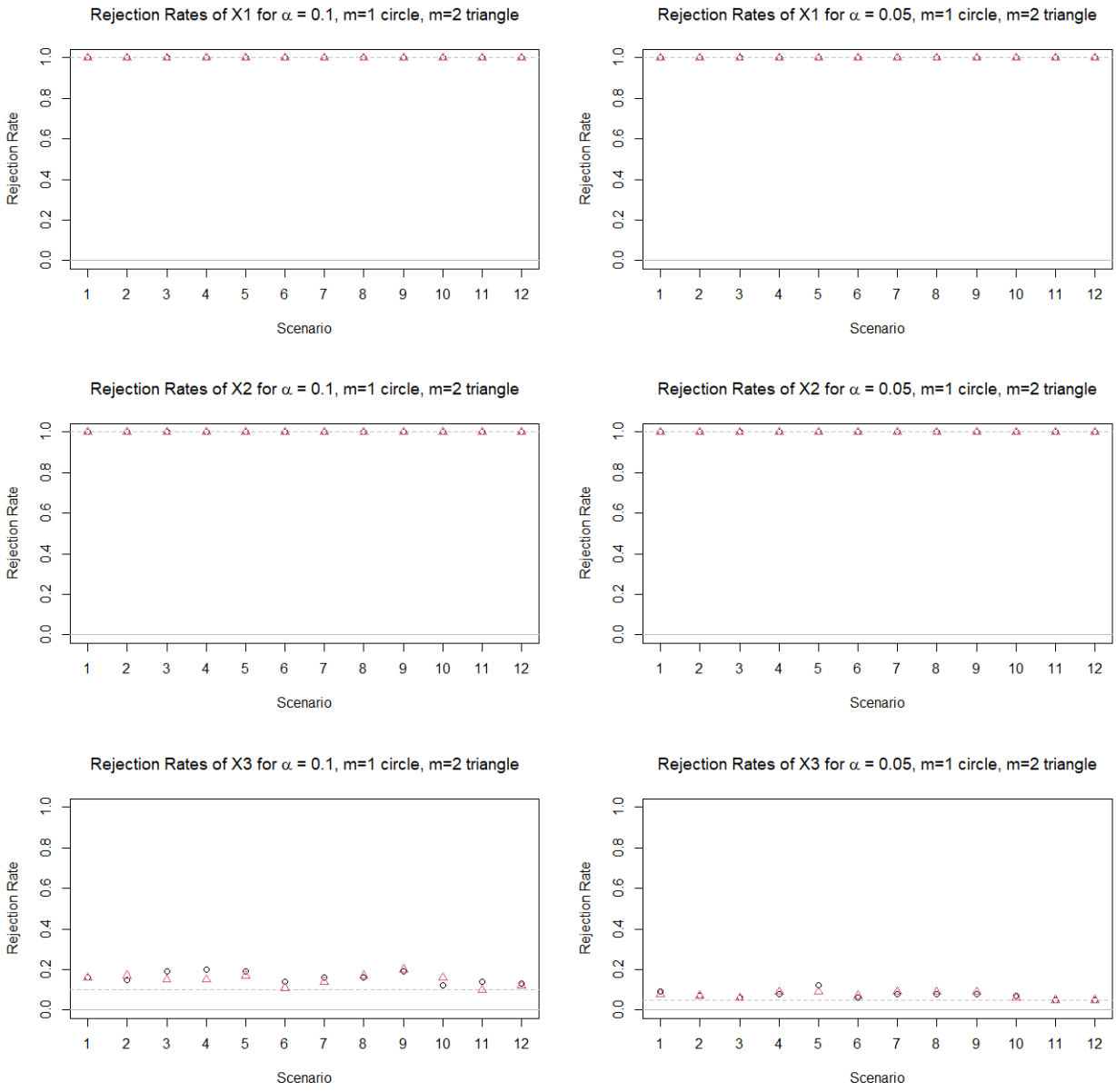


Figure 5.16: Rejection rates of each variable for `m=1` and `m=2` for all scenarios.

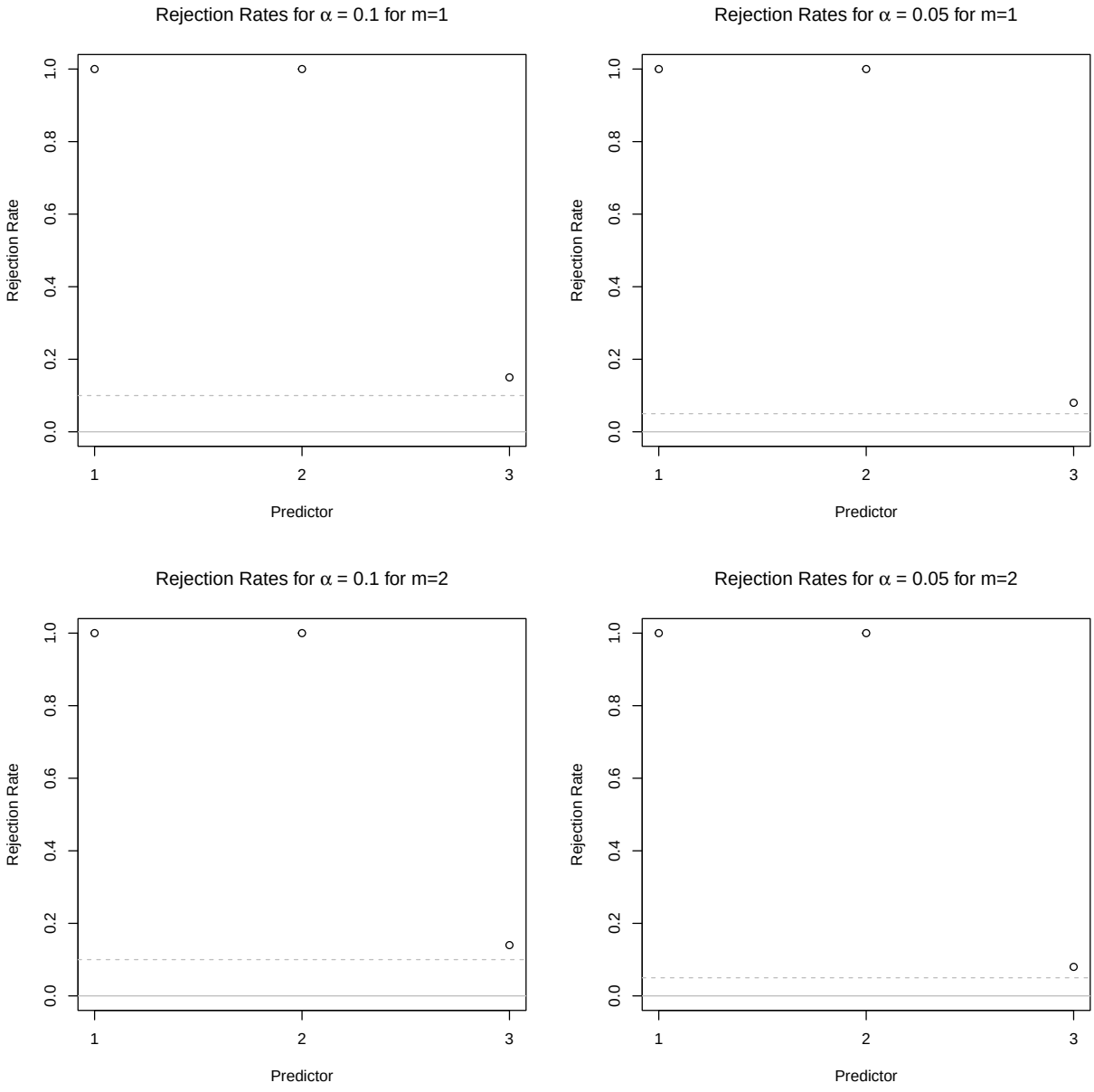


Figure 5.17: Rejection rates of all variables for $\alpha = 0.05$ and $\alpha = 0.1$ for `m=1` and `m=2` for scenario 13.

IV: If predictors are not correctly identified: Is it due to their own marginal distribution, their own correlation strength or that of other variables?

This question now refers only to x_3, as the other predictors were correctly classified. As depicted in the paragraph above, the marginal distribution of x_3 might only have an effect for some specific cases and the correlation strength of x_1 has no influence. Overall, it can be said that neither of the aspects has systematic effect on the correct identification of x_3 as not significant.

V: Do models with `m=1` and `m=2` differ in terms of accuracy of predictions and p-values?

The rejection rates of all predictors show no systematic difference between the two model types (see figure 5.16). The confidence intervals of x_3 on the other hand show a systematically lower coverage rate for `m=2`, although still keeping the 95%-rate in most cases (see exemplary 5.18, for all plots see appendix C). Furthermore, there is no systematic difference between those models where the assumption of the proportional odds model is fulfilled and those where it is not fulfilled.

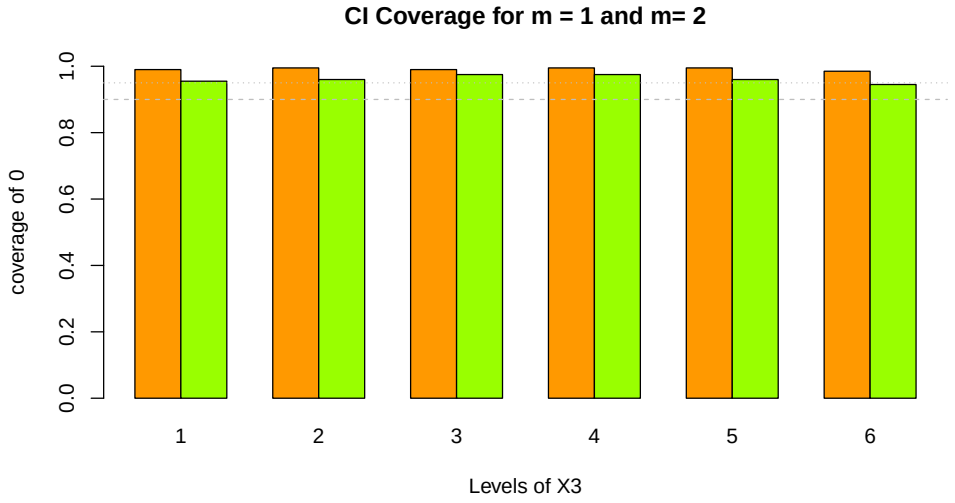


Figure 5.18: Confidene Interval Coverage for x_3 in Scenario 4.

5.6 Results

The scenario analysis conducted leads to several findings. First, the predictions are influenced by the response's distribution in a way that a bell-shaped distribution leads in tendency to more widespread prediction probabilities for the respective categories compared to an approximately exponential one. Also, the distribution of a relevant predictor has an impact on the prediction acuity insofar as that a predictor with a symmetrical bell-shaped distribution is more likely to come with a more symmetrical dispersion of prediction probabilities per category. Thus it can be concluded that distributions of response and predictor affect the prediction range of the categories.

Another focus of the scenario analysis is the classification of relevant and irrelevant predictors. For correlated covariates, rejection rates are high and show good results, but for the uncorrelated predictor p-values turn out to be in general not reliable as rejection rates are too high. The confidence intervals for the latter cover zero and perform well overall. For `m=1` the confidence intervals perform systematically better, while for `m=2` there are a few cases where the 95% coverage rate is a bit undercut. Changing the form of the marginal distribution of x_3 does generally not affect the model performance. The correlation strength of x_1 (ρ_{y,x_1} being 0.2 or 0.5) has no influence on the detection of the insignificant predictor x_3. Therefore, the `m=1` model is recommended for applications, as at least confidence intervals can be fully relied upon.

This is an interesting finding, as Gertheiss et al. (2021) found the inverse to be true for other `gam()` families (gaussian and binomial), where p-values were more reliable for `m=1` and confidence intervals for `m=2`.

In the next chapter, the sociological question is dealt with based on the results of the scenario analysis.

Chapter 6

Are Authoritarian Attitudes and Experienced Discrimination Linked?

6.1 Method

The data which has been prepared as described in chapter 2 is now used to model the six response variables. First, all missing rows of those variables used for the respective regression are excluded. Data sets now comprise between 3747 and 3791 observations each. The observations are split into a train and a test set (70% vs. 30%) and each model is computed based on the train set. As suggested in chapter 5, only `m=1` is implemented.

```
m1 <- gam(aut_neues ~ age
              + gender
              + bija
              + ostd_ohne_b
              + migration
              + s(ben_gewicht, bs = "ordinal", m = 1)
              + s(ben_religion, bs = "ordinal", m = 1)
              + s(ben_sozial_finanz, bs = "ordinal", m = 1)
              + s(ben_behinderung, bs = "ordinal", m = 1)
              + s(ben_gender, bs = "ordinal", m = 1),
              family=ocat(link="identity",R=6),
              method="REML",
              data=train)
```

The summary is printed for analysis and confidence interval boundaries are extracted separately. Subsequently, the proportional odds assumption is tested via a likelihood ratio test and finally predictions are made for the test set. They are visualised in the same types of plots as used in chapter 5. Additionally, barplots of the mean estimated probability are created. All output is found in appendix D.

6.2 Interpretation

General findings are that the assumptions of the proportional odds model are not fulfilled for any model. The predictions are quite insensitive to different true categories, as the plots of the predicted probabilities given the estimated categories reveal (see appendix D for the respective response). Barplots for example look the same, only small changes for *aut_neues*, *aut_eltern* and *aut_gewohnt* are visible where the probability ratio between two adjacent levels changes emphasising the higher one with ascending true category (levels 5 and 6 for the first, levels 2 and 3 for the latter).

As the assumptions are not fulfilled, estimates have to be treated with caution and the emphasis of the interpretation is on the confidence interval coverage, as chapter 5 showed they are reliable for irrelevant variables. Here, however, the real relevance of variables is unknown. Aided by the correlation matrix which shows correlations close to zero between ordinal predictors and the response variables (see appendix A and explanations in section 5.2), it will be assumed that the inverse is also true in this case: if confidence intervals of smooth terms cover zero, they really are insignificant. The interpretation of the models is exemplified in the following for one of the response variables, namely *aut_neues*. The `summary()` output gives:

```
##
## Family: Ordered Categorical(-1,0.41,1.61,2.16,3.08)
## Link function: identity
##
## Formula:
## aut_neues_integer ~ age + gender + bija + ostd_ohne_b + migration +
## s(ben_gew_ordered, bs = "ordinal", m = 1) + s(ben_rel_ordered,
## bs = "ordinal", m = 1) + s(ben_sozfin_ordered, bs = "ordinal",
## m = 1) + s(ben_beh_ordered, bs = "ordinal", m = 1) + s(ben_gender_ordered,
## bs = "ordinal", m = 1)
```

```
##
## Parametric coefficients:
##                Estimate Std. Error  z value Pr(>|z|)
## (Intercept)   1.3426925  0.1787374  7.512   5.82e-14 ***
## age           0.0001071  0.0078456  0.014   0.98911
## gender1  -0.3610960  0.0694870 -5.197   2.03e-07 ***
## bija          0.0360487  0.0127262  2.833   0.00462  **
## ostd_ohne_b  -0.0835187  0.0957019 -0.873   0.38283
## migration     0.0041718  0.0734287  0.057   0.95469
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                          edf   Ref.df  Chi.sq  p-value
## s(ben_gew_ordered)    0.946149     5  1.845  0.1184
## s(ben_rel_ordered)    0.001344     5  0.001  0.6846
## s(ben_sozfin_ordered) 1.433450     5  4.133  0.0356 *
## s(ben_beh_ordered)    2.371794     5  17.542 3.45e-05 ***
## s(ben_gender_ordered) 0.001879     5  0.000  0.8472
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Deviance explained = 0.821%
## -REML = 4563.9 Scale est. = 1 n = 2653
```

The level-specific intercepts are found in the second line. They are also the thresholds of the latent continuous $\tilde{y}$ and per default start at $\beta_{01} = -1$ for the transition from level 1 to level 2, next being $\beta_{02} = 0.41$ for the transition from level 2 to level 3 and so forth. Another intercept is given in the table of the parametric coefficients. This is usually not the case in proportional odds models and might be due to the combination of two packages in an unprecedented way. It is unclear what it stands for. It might have to be added to the level-specific intercepts. Next, the metric and binary variables are listed with their respective estimates. The (parametric) intercept, *gender* and *bija* are marked as significant but the general reliability of the parametric estimates is still to be investigated. Smooth terms are listed with their effective degrees of freedom and p-values. Details on the coefficients of

the ordinal covariates can be accessed via `gam.check()` which also outputs the plots of the confidence intervals (6.1). Note that `k'` is always 5, as we have six levels per predictor and the natural basis of the spline function equals the knots at each threshold between levels (Gertheiss et al., 2021). This is also why there are no p-values computed to evaluate the number of knots.

```
##
## Method: REML Optimizer: outer newton
## full convergence after 7 iterations.
## Gradient range [-0.0007233581,9.446607e-06]
## (score 4563.934 & scale 1).
## Hessian positive definite, eigenvalue range [0.0004179002,887.2551].
## Model rank = 31 / 31
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                          k'      edf  k-index p-value
## s(ben_gew_ordered)    5.00000 0.94615     NA     NA
## s(ben_rel_ordered)    5.00000 0.00134     NA     NA
## s(ben_sozfin_ordered) 5.00000 1.43345     NA     NA
## s(ben_beh_ordered)    5.00000 2.37179     NA     NA
## s(ben_gender_ordered) 5.00000 0.00188     NA     NA
```

The investigation of the smooth terms (figure 6.1) reveals confidence intervals covering zero for all coefficients except those of predictor *ben_beh* (deprivation due to disability or physical impairment). Interesting to see is that the finding of chapter 5 is reflected here: the p-value indicates *ben_sozfin* as significant, but ultimately all its confidence intervals cover zero. For *ben_beh*, estimates of levels 1, 3, 4, 5, 6 are indicated to deviate from zero. Table 6.1 displays these estimates. Effect size is not smaller than those of other variables,

Level 1	(Level 2)	Level 3	Level 4	Level 5	Level 6
0.04121539	(-0.23496546)	-0.38156444	-0.63970314	-0.87884511	-0.82433495

Table 6.1: Estimated coefficients for predictor *ben_beh*. The confidence interval of level 2 covers zero.

for example $age \in \{16, ..., 32\}$ with $\beta_{age} = -0.017$ yields values [-0.556576; -0.278288].

Thus, *ben_beh* looks like a variable which actually has an influence. But upon examining the marginal distribution of the predictor it appears not so certain anymore, as only very few people have experienced any discrimination due to this aspect (3332 with no experience vs. 360 of other levels). In category 2, for example, there are 119 people, in category 6 only 15. That is very few, considering they spread over six response categories. So few data points can hardly show a consistent pattern, and a look at the contingency table (table 6.2) supports this. It is clearly very few data compared to those for category 1.

response \predictor	1	2	3	4	5	6
1	156	8	10	6	5	4
2	388	26	7	7	7	2
3	632	38	17	11	2	1
4	317	16	11	2	0	0
5	436	25	9	1	1	2
6	475	13	6	7	2	3

Table 6.2: Contingency table of *aut_neues* and *ben_beh*

Regardless of the question on trustworthiness, the concrete interpretation of an estimate is now expounded. As for the ordinal covariates no reference category is set, it can be achieved by comparing different levels. An example is given for levels 1 and 2 of *ben_beh*. Equation (3.1) on page 10 states the assumption that the cumulative odds ratio of two populations X and $\tilde{X}$ is the same over all categories. For $r = 1$, (3.1) is

$$\frac{P(Y \leq 1|X)/P(Y > 1|X)}{P(Y \leq 1|\tilde{X})/P(Y > 1|\tilde{X})} = \frac{P(Y = 1|X)/(1 - P(Y = 1|X))}{P(Y = 1|\tilde{x})/(1 - P(Y = 1|\tilde{x}))}.$$

Let $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ be two observations which only differ in the level of *ben_beh*. For $\boldsymbol{x}$ let *ben_beh* take 1 and for $\tilde{\boldsymbol{x}}$ let *ben_beh* take 2. Here, the probability for $y = 1$ can be taken from a prediction made for each of the observations $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ yielding

$$\frac{P(y = 1|\boldsymbol{x})/(1 - P(y = 1|\boldsymbol{x}))}{P(y = 1|\tilde{\boldsymbol{x}})/(1 - P(y = 1|\tilde{\boldsymbol{x}}))} = 0.76.$$

If the assumption were true, the ratio should hold for any category. It can be checked for category $y = 2$ by the following calculation where the same procedure for obtaining

$P(y = 1|\boldsymbol{x})$ and $P(y = 2|\boldsymbol{x})$ is used as above.

$$\begin{aligned}
&\frac{P(y \leq 2|\boldsymbol{x})/P(y > 2|\boldsymbol{x})}{P(y \leq 2|\tilde{\boldsymbol{x}})/P(y > 2|\tilde{\boldsymbol{x}})} \\
&= \frac{(P(y = 1|\boldsymbol{x}) + P(y = 2|\boldsymbol{x}))/(1 - (P(y = 1|\boldsymbol{x}) + P(y = 2|\boldsymbol{x})))}{(P(y = 1|\tilde{\boldsymbol{x}}) + P(y = 2|\tilde{\boldsymbol{x}}))/(1 - (P(y = 1|\tilde{\boldsymbol{x}}) + P(y = 2|\tilde{\boldsymbol{x}})))} \\
&= 0.76
\end{aligned}$$

Cumulative odds ratios are computed correspondingly for the other categories, yielding always the same value. For *ben_beh* taking levels 1 and 2, the proportional odds assumption holds. Interpetation is that the cumulative odds in a population with *ben_beh*=1 are 0.76 the cumulative odds in a population with *ben_beh*=2 and this cumulative odds ratio holds for all categories of y. In other words, cumulative odds of a person who *never* has experienced deprivation due to disability or physical impairment are lower by three quarters compared to those of a person who *rarely* has experienced deprivation due to disability or physical impairment for any given attitude level on the item "New and unusual situations make me uncomfortable". Likewise, other estimates can be incorporated to compute further cumulative odds ratios of interest. As the likelihood ratio test leads to rejecting the proportional odds assumption, the cumulative odds ratio has to be checked for each case individually. This is out of the scope of this thesis, but the general procedure has been outlined.

6.3 Results

Now that the model itself has been discussed, the sociological question will be answered. Is there a connection between the attitude on the statement "New and unusual situations make me uncomfortable" and experienced deprivation? Mostly, there is no influence detected by the model when looking at the confidence intervals. Only for the predictor *ben_beh* they differ from zero. But the data quality is not very high, as there is few variance. Therefore, the other types of experienced deprivation can be discarded, leaving only experienced deprivation due to disability or physical impairment for further investigation, for example via bootstrapping or oversampling.

As for the other response variables, no ordinal predictors stand out. When considering the six coefficients of each of the ordered categorical predictors, there is usually only one level of one or two ordinal covariates with a confidence interval not covering zero per

model. For these coefficients, the interval mostly reaches close to zero. As stated earlier (see section 4.2), ridge type penalties have no selection properties, but can only push coefficients very close to zero. It depends thus on the researcher's evaluation whether or not to exclude a variable. The overall picture in this case leads to the conclusion that experienced deprivation is not linked significantly to authoritarian attitudes. Further investigation might be made for the item *aut_neues* concerning the influence of experienced deprivation due to disability or physical impairment.

In order to investigate whether authoritarian attitudes can be explained by experienced deprivation, a penalized proportional odds model was run. Results show few influence of the predictors and it can be concluded that there is no significant connection between the two aspects. For one item, further investigation might be fruitful, namely the item "New and unusual situations make me uncomfortable" (*aut_neues*) and the experienced deprivation due to disability or physical impairment (*ben_beh*). In a next step, the conducted penalized ordinal regression is compared to one of the common but inappropriate methods in social sciences to deal with ordinal data, namely linear regression.

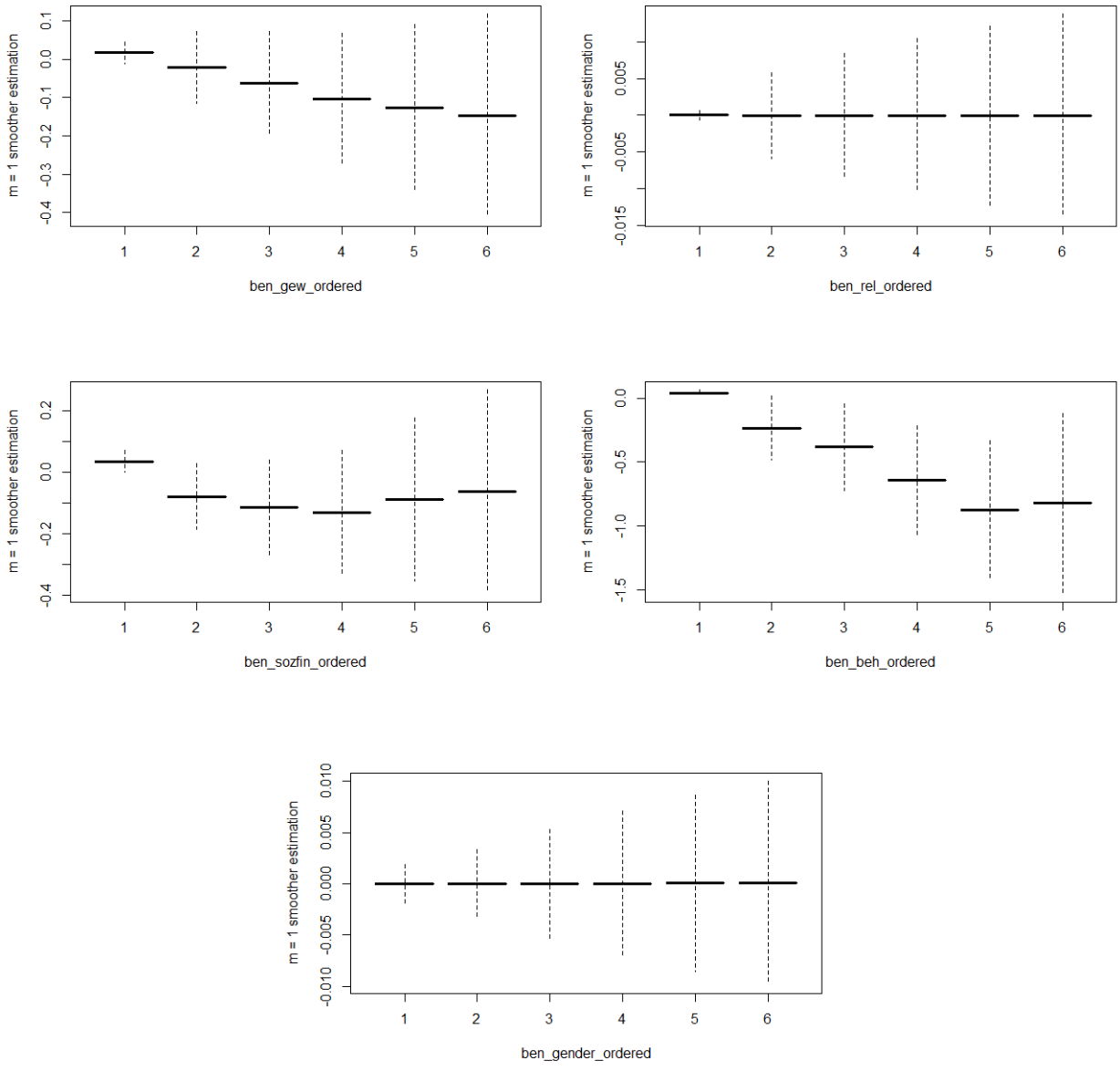


Figure 6.1: Estimates of smooth terms for *aut_neues*.

Chapter 7

Penalized Cumulative Logistic Regression vs. Linear Regression

7.1 Method

The linear model is based on the same variables as the proportional odds model, but treats response and predictors as metric.

```
lm <- lm(aut_gewohnt ~ age
                     + geschlecht
                     + bija
                     + ostd_ohne_b
                     + migration
                     + ben_gender
                     + ben_sozial_finanz
                     + ben_gewicht
                     + ben_behinderung
                     + ben_religion, train)
```

In order to compare the two models, some decision rules have to be applied, because neither of them predicts one category per observation. For the proportional odds model, probabilities for each category are given. Requirement of the rule for the ordinal model is that the distribution of the train set is mapped onto the test set predictions. Starting from the full test set with their predicted probabilities for all categories, the following steps are conducted over categories 1 to 5:

1. Determine, how many observations have to be allocated to category C:
   $s_C = \text{share in train set} * n_{testset}$

2. Sort observations of the test set in descending order of the probabilities predicted for category C

3. The upper s_C observations are assigned to C and excluded from the test set

All remaining observations are allocated to level 6, leaving a margin for small rounding errors (not more than one or two observations less than predefined in s_6).

For the linear model predictions are assigned by rounding the values to the nearest integer.

7.2 Results

Following the approach of machine learning, performance is compared albeit the assumptions not being fulfilled. Table 7.1 shows the accuracy of all six models. The accuracy of the proportional odds model is always higher than of the linear model, for the latter being as good as random allocation. Thus, although the assumptions of the proportional odds model are not fulfilled, it performs better - which might be due to the fact that the assumptions of the linear model are even less fulfilled. Using penalized ordinal regression compared to a common social science method is worth the effort for this application with survey data.

	aut_beherrschen		aut_eltern		aut_staerkere	
	POM	LM	POM	LM	POM	LM
Overall	0.29688889	0.1865402	0.24911661	0.18192564	0.22940655	0.1664394
Category 1	0.06666667	0.0000000	0.21264368	0.00000000	0.07894737	0.0000000
Category 2	0.16666667	0.0000000	0.34939759	0.18373494	0.07058824	0.0000000
Category 3	0.22026432	0.0000000	0.29179331	0.89057751	0.17224880	0.0000000
Category 4	0.14285714	0.5785714	0.12931034	0.01724138	0.17500000	0.6350000
Category 5	0.17703349	0.5406699	0.07608696	0.00000000	0.26515152	0.3636364
Category 6	0.50483092	0.0000000	0.12359551	0.00000000	0.32732733	0.0000000

	aut_gewohnt		aut_andere		aut_neues	
	POM	LM	POM	LM	POM	LM
Overall	0.24250441	0.16679795	0.35237258	0.16739336	0.1987687	0.1613917
Category 1	0.24712644	0.00000000	0.00000000	0.00000000	0.1263158	0.0000000
Category 2	0.27794562	0.02719033	0.04255319	0.00000000	0.1600000	0.0000000
Category 3	0.33003300	0.97359736	0.16346154	0.00000000	0.2828283	0.1077441
Category 4	0.17714286	0.00000000	0.12121212	0.01515152	0.1212121	0.8606061
Category 5	0.06818182	0.00000000	0.26618705	0.98920863	0.2352941	0.0000000
Category 6	0.04761905	0.00000000	0.52994555	0.00000000	0.1704545	0.0000000

Table 7.1: Comparison of accuracy of proportional odds model (POM) and linear model (LM) of six response variables on authoritarianism.

Chapter 8

Conclusion

There is little focus on ordinal regression in social sciences although ordered categorical data play a huge role, for example in surveys. The aim of this thesis is to present correct methods by means of a practical application. For this purpose, the sociological question has been examined as to whether authoritarian attitudes can be explained by one's own experience of discrimination. The basis for the analysis was a survey with ordinal dependent and independent variables. The statistical tools to model this question were presented. For ordinal response variables, ordinal regression models were explained concluding that a cumulative logistic regression, also called proportional odds model, is the adequate model for this undertaking. To account for the ordered categorical predictors, penalization terms can be employed. It was argued that the preferable type should be able to deal with grouped variables and have a selection property. Three penalties were identified to fulfil this requisite. An overview over software packages showed that a proportional odds model with these penalties is not available yet for the statistical open-source software R (R Core Team, 2021). A recent extension of the package `ordPens` (Hoshiyar and Gertheiss, 2021) allows for first- and second-order generalised ridge type penalties in `mgcv::gam()` (Wood, 2011). Using a proportional odds model was unprecedented for this method and its behaviour had to be examined before it could be used on real data. To this aim, a scenario analysis was conducted based on simulated data. It was concluded that using first-order differences is recommended. The rejection rates of p-values were proven to be too high. The confidence intervals for uncorrelated predictors on the other hand covered zero reliably. Therefore, they were identified as the better measurement. Based on these findings, the initial sociological question was pursued. Applying the model, no significant connection was found between authoritarian attitudes and experienced deprivation. For the response item "New

and unusual situations make me uncomfortable" and the predictor "experienced deprivation due to disability or physical impairment" further investigations have been indicated. Finally, the predictions of the penalized proportional odds model were compared to a linear regression, a common but inadequate method in social sciences to model ordinal data. Decision rules were applied to both outcomes to obtain single categories as predictions. A comparison of accuracy showed that the penalized ordinal regression performs better than the linear model in all cases. It could be concluded that for this application with survey data, the effort to familiarise oneself with this new method was worth it.

A caveat to these results is that the predictive power is surprisingly low, even for highly correlated variables as in the scenario analysis. In case of the survey, this might also have to do with the low variance predictors.

Moreover, the assumption of the proportional odds model was often not fulfilled. This might be because the test of the assumption grows sensitive with small cell counts and the contingency table for one pair of variables showed that there even are empty cells. Ultimately, it is not clear whether the assumptions were truly unfulfilled, or whether it was again due to the very skewed distribution of the predictors.

Apart from issues due to the data, some other aspects could be further analysed. For example the predicted probabilities were visualised and interpreted via boxplots in the scenario analysis, but they turned out to be very dense in information. Adding barplots depicting only the average probabilities might simplify interpretation. They can be found in the appendix

Also, data simulation was realised with a package which allows for adjustment of the correlation between predictors and this could be deepened more in future analyses, because in survey data it is very likely to have highly correlated predictors, for example when they belong to a same scale.

The final comparison of linear and ordinal model was based on different prediction rules for each model. This juxtaposition could be modified by using more similar prediction rules, for example by applying the same distribution conserving rule to the linear model as applied to the ordinal regression and compare it to random distribution-conserving allocation.

Beyond modifications to this study, there are open questions regarding the behaviour of the `ordPens` (Hoshiyar and Gertheiss, 2021) extension. The outputs of the model need further investigation to assure which characteristic values and plots can be consulted for model assessment. For example the explained deviance specified in the summaries was extremely low (for the simulations around 3% or 13%, for the survey data less than 1%).

The values could not be interpreted as it was unclear whether these values actually indicate a lower explained deviance for one model compared to the other.

Future analysis of the method should also verify the assumption made that confidence intervals for significant estimators are trustworthy inasmuch as they do not cover zero. This was assumed here, which might ultimately be incorrect. For this application it would not have greater drawbacks, however, as it applied to only one variable which then was found to be so skewed that results were not very trustworthy anyway. One option to realise this would be to simulate data via a regression model. This would allow to check estimation performance for significant coefficients. The evaluation could also be extended to other than ordinal covariates and might shed some light on the second intercept listed under the parametric coefficients.

Bibliography

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., and Sandford, R. N., editors (1950). *The Authoritarian Personality*. Harper, New York.

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, NJ, 2nd edition.

Archer, K. J., Hou, J., Zhou, Q., Ferber, K., Layne, J. G., and Gentry, A. E. (2014). ordinalgmifs: An R Package for Ordinal Regression in High-dimensional Data Settings. *Cancer Informatics*, 13:187–195.

Asbrock, F., Sibley, C. G., and Duckitt, J. (2010). Right-Wing Authoritarianism and Social Dominance Orientation and the Dimensions of Generalized Prejudice: A Longitudinal Test. *European Journal of Personality*, 24(4):324–340.

Atkinson, L. (1988). The Measurement-statistics Controversy: Factor Analysis and Subinterval Data. *Bulletin of the Psychonomic Society*, 26(4):361–364.

Baier, D., Pfeiffer, C., Rabold, S., Simonson, J., and Kappes, C. (2010). Kinder und Jugendliche in Deutschland. Gewalterfahrungen, Integration und Medienkonsum: Zweiter Forschungsbericht zum gemeinsamen Forschungsprojekt des Bundesministeriums des Innern und des KFN.

Barbiero, A. and Ferrari, P. A. (2015). GenOrd: Simulation of Discrete Random Variables with Given Correlation Matrix and Marginal Distributions. `https://CRAN.R-project.org/package=GenOrd`. Last accessed January 6, 2022.

Blasius, J. and Baur, N. (2019). Multivariate Datenstrukturen. In Baur, N. and Blasius, J., editors, *Handbuch Methoden der empirischen Sozialforschung*, pages 1379–1400. Springer VS, Wiesbaden.

Brähler, E. and Decker, O., editors (2018). *Flucht ins Autoritäre: Rechtsextreme Dynamiken in der Mitte der Gesellschaft : die Leipziger Autoritarismus-Studie 2018*. Forschung psychosozial. Psychosozial-Verlag, Gießen.

Breheny, P. and Huang, J. (2011). Coordinate Descent Algorithms for Nonconvex Penalized Regression, With Applications to Biological Feature Selection. *The Annals of Applied Statistics*, 5(1):232–253.

Breheny, P. and Huang, J. (2015). Group Descent Algorithms for Nonconvex Penalized Linear and Logistic Regression Models With Grouped Predictors. *Statistics and Computing*, 25(2):173–187.

Christensen, R. H. B. (n.d.). Cumulative Link Models for Ordinal Regression with the R Package ordinal. `https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf`. Last accessed January 6, 2022.

Cribbs, S. E. and Austin, D. M. (2011). Enduring Pictures in our Heads: the Continuance of Authoritarianism and Racial Stereotyping. *Journal of black studies*, 42(3):334–359.

Deutsches Jugendinstitut (2022). Forschungsdatenbank DJI-Studien. `https://surveys.dji.de/`. Last accessed January 6, 2022.

Fahrmeir, L., Kneib, T., and Lang, S. (2009). *Regression: Modelle, Methoden und Anwendungen*. Statistik und ihre Anwendungen. Springer-Verlag, Berlin, Heidelberg, 2nd edition.

Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Ford, C. (05.10.2015). Fitting and Interpreting a Proportional Odds Model. `https://data.library.virginia.edu/fitting-and-interpreting-a-proportional-odds-model/`. Last accessed January 8, 2022.

Frank, I. E. and Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, 35(2):109–135.

Gertheiss, J., Scheipl, F., Lauer, T., and Ehrhardt, H. (2021). Statistical Inference for Ordinal Predictors in Generalized Linear and Additive Models with Application to Bronchopulmonary Dysplasia. `https://arxiv.org/pdf/2102.01946`. Last accessed January 6, 2022.

Gertheiss, J. and Tutz, G. (2009). Penalized Regression with Ordinal Predictors. *International Statistical Review*, 77(3):345–365.

Goeman, J. J., Meijer, R. J., and Chaturvedi, N. (2018). Penalized: L1 (Lasso and Fused Lasso) and L2 (Ridge) Penalized Estimation in GLMs and in the Cox Model. `https://cran.rstudio.com/web/packages/penalized/index.html`. Last accessed January 6, 2022.

Heitmeyer, W., editor (2002). *Deutsche Zustände: Folge 1*, volume 2290 of *Edition Suhrkamp*. Suhrkamp Verlag, Frankfurt am Main, 1st edition.

Heitmeyer, W., editor (2012). *Deutsche Zustände Folge 10*, volume 2647 of *Edition Suhrkamp*. Suhrkamp Verlag, 4th edition.

Hirk, R., Hornik, K., and Vana, L. (2020). mvord : An R Package for Fitting Multivariate Ordinal Regression Models. *Journal of Statistical Software*, 93(4):1–41.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67.

Hoshiyar, A. and Gertheiss, J. (2021). ordPens: Selection, Fusion, Smoothing and Principal Components Analysis for Ordinal Variables. `https://CRAN.R-project.org/package=ordPens`. Last accessed January 6, 2022.

Hoyer, A. (2018). Generalisierte Regressionsmodelle: Lecture in the Winter Semester 2018/2019 at Ludwig-Maximilians-Universität München.

Huang, J., Breheny, P., and Ma, S. (2012). A Selective Review of Group Selection in High-Dimensional Models. *Statistical Science*, 27(4):481–499.

Jöreskog, K. G. and Moustaki, I. (2001). Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research*, 36(3):347–387.

Joubert, P. and Langdell, S. (2013). Modelling: Mastering the Correlation Matrix. *The Actuary*, (9).

Kuger, S., Prein, G., Linberg, A., Hoffmann-Recksiedler, C., Herz, A., Gille, M., Berngruber, A., Bernhardt, J., Pötter, U., Zerle-Elsässer, C., Steiner, C., Zimmermann, J., Quellenberg, H., Walper, S., Rauschenbach, T., Maly-Motta, H., Schickle, V., Naab, T., Guglhör-Rudan, A., Langmeyer, A., Tran, K., Gaupp, N., Milbradt, B., Heintz-Martin, V., Gerum, M., Entleitner-Phleps, C., and Deutsches Jugendinstitut (2020). Aufwachsen in Deutschland: Alltagswelten 2019 (AID:A 2019). `https://surveys.dji.de/index.php?m=msw,0&sID=118`. Last accessed January 6, 2022.

Mansel, J. and Spaiser, V. (2013). *Ausgrenzungsdynamiken: In welchen Lebenslagen Jugendliche Fremdgruppen abwerten*. Konflikt- und Gewaltforschung. Beltz Juventa, Weinheim.

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127.

Meier, L. (2020). Package grplasso. `https://cran.r-project.org/web/packages/grplasso/grplasso.pdf`. Last accessed January 6, 2022.

Oesterreich, D. (2005). Flight into Security: A New Approach and Measure of the Authoritarian Personality. *Political Psychology*, 26(2):275–298.

Ogutu, J. O. and Piepho, H.-P. (2014). Regularized Group Regression Methods for Genomic Prediction: Bridge, MCP, SCAD, Group Bridge, Group Lasso, Sparse Group Lasso, Group MCP and Group SCAD. *BMC Proceedings*, 8(Suppl 5):S7.

R Core Team (2021). R: A Language and Environment for Statistical Computing. `https://www.R-project.org/`. Last accessed January 6, 2022.

Rippl, S., Kindervater, A., and Seipel, C. (2000a). Die autoritäre Persönlichkeit: Konzept, Kritik und neuere Forschungsansätze. In Rippl, S., Seipel, C., and Kindervater, A., editors, *Autoritarismus*, pages 13–32. VS Verlag für Sozialwissenschaften, Wiesbaden.

Rippl, S. and Seipel, C. (2018). Modernisierungsverlierer, Cultural Backlash, Postdemokratie: Was erklärt rechtspopulistische Orientierungen? *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 70(2):237–254.

Rippl, S., Seipel, C., and Kindervater, A., editors (2000b). *Autoritarismus: Kontroversen und Ansätze der aktuellen Autoritarismusforschung*. VS Verlag für Sozialwissenschaften, Wiesbaden.

Snell, E. J. (1964). A Scaling Procedure for Ordered Categorical Data. *Biometrics*, 20(3):592.

Thomas, W. I. and Thomas, D. S. (1928). *The Child in America : Behavior Problems and Programs*. Alfred A. Knopf, New York.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tutz, G. (2012). *Regression for Categorical Data*, volume 34 of *Cambridge series on statistical and probabilistic mathematics*. Cambridge University Press, Cambridge.

Tutz, G. and Gertheiss, J. (2014). Rating Scales as Predictors–the old Question of Scale Level and Some Answers. *Psychometrika*, 79(3):357–376.

Tutz, G. and Gertheiss, J. (2016). Regularized Regression for Categorical Data. *Statistical Modelling: An International Journal*, 16(3):161–200.

Tutz, G. and Gertheiss, J. (Forthcoming). Regularization and Predictor Selection for Ordinal and Categorical Data. In Kateri, M. and Moustaki, I., editors, *Trends and Challenges for Categorical Data Analysis*. Springer.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition.

Walker, S. H. and Duncan, D. B. (1967). Estimation of the Probability of an Event as a Function of Several Independent Variables. *Biometrika*, 54(1-2):167–179.

Wang, L. L., Watts, A. S., Anderson, R. A., and Little, T. D. (2014). Common Fallacies in Quantitative Research Methodology. In Little, T. D., editor, *The Oxford handbook of quantitative methods*, Oxford Library of Psychology, pages 718–758. Oxford Univ. Press, New York.

Weigelt, I. (2020). Scale on Authoritarianism in AID:A Survey: Oral Conversation.

Williams, O. D. and Grizzle, J. E. (1972). Analysis of Contingency Tables Having Ordered Response Categories. *Journal of the American Statistical Association*, 67(337):55.

Wood, S. N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 73(1):3–36.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2nd edition.

Wood, S. N. (2021). plot.gam function - RDocumentation. `https://www.rdocumentation.org/packages/mgcv/versions/1.8-36/topics/plot.gam`. Last accessed January 6, 2022.

Wood, S. N. (n.d.). gam() documentation. `https://www.rdocumentation.org/packages/mgcv/versions/1.8-38/topics/gam`. Last accessed January 11, 2022.

Wurm, M. J., Rathouz, P. J., and Hanlon, B. M. (2021). Regularized Ordinal Regression and the ordinalNet R Package. *Journal of Statistical Software*, 99(6):1–42.

Yee, T. W. (2010). The VGAM Package for Categorical Data Analysis. *Journal of Statistical Software*, 32(10):1–34.

Yuan, M. and Lin, Y. (2006). Model Selection and Estimation in Regression With Grouped Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zhang, C.-H. (2010). Nearly Unbiased Variable Selection Under Minimax Concave Penalty. *The Annals of Statistics*, 38(2):894–942.

Zick, A., Küpper, B., Krause, D., and Berghan, W. (2016). *Gespaltene Mitte - feindselige Zustände: Rechtsextreme Einstellungen in Deutschland 2016*. Dietz, Bonn.

Appendix A

Table of Contents of Separate Appendix

Contents

Appendix B

Table of Contents of Electronic Appendix

Files and Folders Included in the Electronic Appendix

1_clean_data.R
This file reads in the survey data and extracts the relevant variables.
It then creates clean variable names, excludes all persons who are
not target persons and those under 14. It saves the data in an .RDS-file.

2_recode_data.R
This file recodes the ordinal predictors in reversed order. Some variables
are changed in type and new variables are created for migration, background
and residency. Observations with the third option for gender ("none of the
above mentioned") are excluded.
An .RDS-file is created.
People older than 33 are excluded and a second .RDS-file is created.

3_correlations.R
Behaviour of gender==3 and age==33 is inspected. A correlation table of all
relevant variables is created.

4_first_impressions.Rmd
This file runs four different combinations of marginal distributions

and correlation coefficients. A pdf is created.

5_*.Rmd
These files create each a simulation scenario and their evaluation tools.
The rejection rate of the p-values is saved in an .RDS file.
For each scenario, a PDF with the output is created.

6_rejection_rates_all_variations.Rmd
This file combines the rejection rates of the predictors of all scenarios and describes them graphically in a PDF.

7_aut_*.Rmd
These files run a proportional odds model for the respective response named in the filename. It is evaluated. A linear model based on the same variables is run and compared to the proportional odds model. A PDF with the output is created.

Appendix.pdf
The appendix as usually found in the end of a thesis. It contains all PDF files created by the files in the electronic appendix.

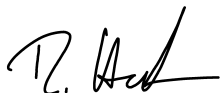
Data
This Folder contains the following subfolders.

clean data
This folder contains the following sub-folder.

p-values
This folder contains the rejection rates of the p-values as computed by the scenarios in 5_*.RDS.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die aus fremden veröffentlichten oder nicht veröffentlichten Quellen, wörtlich oder sinngemäß übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.



Ruben Hartmann, München, den 13. Januar 2022