Multinomial Restricted Unfolding for Multivariate Binary Data

Machado Ladeira, Mariana Rita

Faculteit der Sociale Wetenschappen

Universiteit Leiden

Multinomial Restricted Unfolding for Multivariate Binary Data

Master's Thesis

Mariana Rita Machado Ladeira

Master's Thesis Methodology and Statistics Master

Methodology and Statistics Unit, Institute of Psychology,

Faculty of Social and Behavioral Sciences, Leiden University

Date: November 2021

Student number: s2593777

Supervisor: Dr. Mark de Rooij

Abstract

Multivariate binary data with multiple binary response variables arise in many areas of research, including biology, psychology, medicine, dentistry, and other empirical sciences. In such data, the effect of a predictor on the response variable and the effect of a predictor on the association structure between the response variables is of interest. Multinomial Restricted Unfolding (MRU) is a probabilistic multidimensional unfolding model that can be used to analyse multicategory response variables in the presence of predictors. In this thesis, we investigated an extension of the MRU model to analyse multivariate binary data focusing on how diagnoses of depressive and anxiety disorders are influenced by personality traits and how the association between two disorders is affected by these personality traits. We compared the results using usual and squared Euclidean distances for the main effects and associations MRU models.

We have demonstrated that MRU models using squared and usual Euclidean distances can be used to analyse multivariate binary data, representing well the changes in log odds and the changes in log odds ratio. Our results indicated that the MRU models using squared Euclidean distances are more straightforward and easier to be interpreted than those using usual Euclidean distance. However, despite the more complicated interpretation, the model using the usual Euclidean distance is more flexible, which might lead to a better model fit

Regarding the change in log odds of having GAD, the main effects model results indicated a constant change in slope between different pairs of categories that represent GAD. For the associations model, the parallel lines indicate that the change in the slope was constant within a pair of categories that represents GAD but not the same for different pairs of categories. When usual Euclidean distances were used, the change in slope was not constant for both main effects and associations models. Regarding the interpretation rules to express the change in the log odds ratio, the main effects model showed that the association structure, does not dependent on the value of the predictor variable. However, for the associations model, the log odds ratio is dependent on the value of the predictor variable, in which a constant change in slope is shown. When using usual Euclidean distances to explore the association structure between two response variables, the change in slope was not constant for both models.

Table of Contents

1 Introduction

1.1 Multivariate Binary Data

Multivariate binary data with multiple binary response variables arise in many areas of research, including biology, psychology, medicine, dentistry, and other empirical sciences. One example of this type of data can be encountered in the study from Komac and Sajn (2001) that investigates the soil pollution and the presence of heavy metals in the urban area of Ljubljana. Thirty-six heavy metals were measured in 97 soil samples, and the eight most hazardous heavy metals (i.e., Cadmium, Cobalt, Chromium, Copper, Nickel, Lead, Zinc, and Mercury) were used to analyse the soil pollution. These eight variables were dichotomized according to whether the measurement of each specific heavy metals exceeds the limit established by Slovenian legislation. Another example can be encountered in the Netherlands Study of Depression and Anxiety (NESDA), in which data were collected on depression and anxiety disorders, aiming to understand the association between personality traits and comorbidity of anxiety and depressive disorders (Spinhoven, De Rooij, Heiser, Penninx, & Smit, 2009).

In multivariate binary data, each of the multiple responses variables denotes a binary measurement observed for each individual i and may, for example, indicate the presence or absence of a disorder in a patient or a metal in a sample. Multivariate binary data can be considered as a case of multicategory response data, in which the categorical response variable has multiple possible values. For instance, a multivariate variable with three binary response categories might be thought of as a single response variable with $2^3 = 8$ categories. When working with multicategory or multinomial models, people are often interested in (1) the marginal relation of the predictor variable on the log odds of Y1xxxx against Y0xxxx, where xxxx is any combination of zeros and one, (2) the effect of the predictor variable on the odds ratio of Y11xxx, Y10xxx, Y01xxx and Y00xxx, which represents the association between two binary response variables. The standard model to analyse multicategory response variables is Multinomial Logistic Regression, which is a model for the joint probabilities.

The major disadvantage of this model is that the model uses many parameters $(P + 1)$ x $(C - 1)$, therefore it is difficult to interpret.

De Rooij (2009) introduced the multinomial restricted unfolding (MRU) model for multicategory response variables in the presence of predictors, which was called ideal point classification (IPC) model at that time. The MRU is a probabilistic multidimensional unfolding model that in maximum dimensionality becomes equal to Multinomial Logistic Regression. The MRU model is similar to ideal point discriminant analysis (IPDA) proposed by Takane, Bozdogan, and Shibayama (1987). Both the IPDA and MRU models are multidimensional unfolding (MDU) based categorization algorithms (De Leeuw, 2005). The MRU model specifies the likelihood function employing the joint probabilities and uses squared Euclidean distances (Worku & De Rooij, 2017). Recently, De Rooij and Busing (paper in progress) developed a new algorithm for estimating an IPC/MRU model with distances instead of squared distances.

Worku and De Rooij (2017) considered the MRU model for bivariate binary variables, that is, a multicategory with $C = 4$ categories in the response variable, and linked each of the responses to a dimension in two-dimensional Euclidean space. Worku and De Rooij (2017) demonstrated that the initial MRU model either represented the marginal or the association structure well, in the sense of bias. Hence, they expanded this model by changing the multinomial likelihood function in accordance with the methods of Bahadur (1961) and Lipsitz, Laird, and Harrington (1990). In comparison to other approaches for jointly modeling the marginal and association structures, this alternative parameterization offers the benefit of dimension reduction and, when the dimensionality is 2, the result can be visualized in a two dimensional biplot, which facilitates model interpretation.

Worku and De Rooij (2017) concluded that it is not easy to extend this model to analyse multivariate binary responses. This is due to the fact that the likelihood function must provide both the pairwise and higher order association structure parameters, which makes the model complicated and hence difficult to estimate. In addition, they stated that this model can be impractical for analysing multivariate binary data, due to the assumption that in a 2-dimensional space the response variable should be linked to one dimension, which will make it impossible to find unbiased marginal coefficients. However, for this thesis project, we will assume that the response variable does not need to correspond to one dimension.

1.2 Research Question

For this theses project we will investigate the properties of the MRU model to analyse multivariate binary data. Our focus is to research how diagnoses of depressive and anxiety disorders are influenced by personality traits and how the association between two disorders is affected by these personality traits. We will extend the MRU model by adding a constraint on the class points, making them a function of main effects and pairwise associations of the underlying binary variables. By imposing this structure, we expect to obtain general interpretation rules to express the marginal effect of a predictor on the response variable and the effect of a predictor on the association between the response variables in a model for the joint probabilities. Moreover, we will compare the MRU model results using usual and squared Euclidean distances. Additionally, we will analyse the order of association among response variables separately. We will begin with a model that just considers the main effects, and then we will extend the analyse to a model that includes the main effects and pairwise associations. To simplify, we will refer to these models as main effects and associations models, respectively. This thesis is organized as follows. Section 2 presents information regarding distance models. Section 3 describes the empirical data used to illustrate the MRU model and to derive the algebraic formulas. In addition the statistical analyses are explained in this section. Section 4 presents the results of our analyses. In section 5 we conclude the thesis with a discussion and suggestions for future research.

2 Distance Models

2.1 Multidimensional Scaling and Unfolding

Assume that dissimilarity data has been collected for a set of n objects or individuals, with a value of dissimilarity assigned to each pair. A common example of dissimilarity data can be seen in a list with distances between cities. The objective of multidimensional scaling (MDS) is to obtain a arrangement of points in a space, typically Euclidean, in which each point corresponds to one of the objects or individuals and the distances between pairs of points closely match the original dissimilarities between the pairs of objects (Cox & Cox, 2008).

MDS is a technique to provide a visual representation of proximities (i.e., object similarity or object dissimilarity) between pairs of objects (Kruskal & Wish, 1978). With dissimilarity data, a greater proximity measure value indicates similar pairs of objects, whereas in similarity data, a higher proximity measure value indicates more similar pairs of objects. Multidimensional Unfolding (MDU) is a similar technique that uses preference data to find distances between individuals and objects in space, typically Euclidean, that are as close to a set of proximities as possible (De Leeuw, 2005). A common example of preference data arises when participants are presented with a collection of designs and are asked which design they prefer.

2.2 Ideal Point Discriminant Analysis

Takane, Bozdogan, and Shibayama (1987) presented the Ideal Point Discriminant Analysis (IPDA) as a procedure for discriminant analysis with predictor variables in a mixed measurement level. IDPA is a classification approach and uses multidimensional unfolding for classification purposes (De Rooij, 2007; Worku, 2008). The proximity is determined in the IPDA model by an indicator matrix that corresponds to the multinomial responses. The model is closely related to the canonical discriminant analysis, however it does not make any assumptions about the multivariate normality of the explanatory variables (De Rooij, 2007; Worku, 2008; De Rooij, 2009). IPDA, in maximum dimensionality, is equivalent to multinomial logistic regression.

Takane, Bozdogan and Shibayama (1987) defined the probability for the c-th category in the IPDA model as

$$\pi_c(\mathbf{x}_i) = \frac{m_c \exp(\theta_{ic})}{\sum_{c'} m_{c'} \exp(\theta_{ic'})}$$

where θ_{ic} is

$$\theta_{ic} = -d^2(\mathbf{u}_i, \mathbf{v}_c) = -\sum_m (u_{im} - v_{cm})^2,$$

in which $u_{im} = \sum_m x_{ip} b_{pm}$ and m_c is a bias parameter for category c. $\mathbf{v}_c$ is an M vector for the coordinates of category c and $\mathbf{u}_i$ represents the M vector with coordinates for person i in M dimensional Euclidean space.

Takane (1998) analysed some visualization aspects of the IPDA model and concluded that the bias parameters reduce model interpretation, since the decision boundaries are shifted away from the class with the largest bias term. De Rooij (2009) revisited the IPDA model emphasizing the visualization of the model and concluded that, in maximum dimensionality, the effect of the bias parameters on the model fit is nil. In other words, the model without bias parameters has the same fit to the data as the model with the bias terms. This new parameterization was called Multinomial Restricted Unfolding (IPC at that time).

2.3 Multinomial Restricted Unfolding model

The MRU model is a probabilistic multidimensional unfolding model and is equivalent to the IPDA model, except that it does not include the bias parameter. The proximity is determined in the MRU model by an indicator matrix that corresponds to the multinomial responses. In the MRU model the conditional joint probabilities are modelled using the Euclidean distance between two points in a Euclidean space of dimensionality M (De Rooij, 2009).

To illustrate the MRU model we will use the following notation: $i = 1, \ldots, n$ for individuals, $p = 1, \ldots, P$ for predictor variables, $m = 1, \ldots, M$ as an indicator for the dimensions. Observations of the responses are put in the $n \times 1$ vector $\mathbf{y}$. $\mathbf{v}_c$ is an M vector for the coordinates of category c. These will be collected in the matrix $\mathbf{V}$. With the MRU model we can impose a restriction on $\mathbf{V}$, such as $\mathbf{V} = \mathbf{ZG}$. With this restriction on the coordinates for the class points, we may analyse the order of association among response variables separately, in which main effects or main effects and associations are represented.

$\mathbf{Z}$ is an indicator matrix that corresponds to the multinomial responses and has C rows and the number of columns depends on the order of association. If we are considering only the main effects, matrix $\mathbf{Z}$ has $r+1$ columns, where r is the number of response variables. When we incorporate the main effects and the associations, $\mathbf{Z}$ has $(r+1)+\frac{r(r-1)}{2}$ columns. For the NESDA data, the two $\mathbf{Z}$ matrices are given in Appendix A. $\mathbf{G}$ represents a $Z \times M$ matrix with regression weights for the class points. $\mathbf{X}$ is an $n \times P$ matrix with values for the predictor variables for all individuals. This matrix collects observations x_{ip}. We assume, without loss of generality, that the predictor variables are centered. $\mathbf{B}$ represents a $P \times M$ matrix with regression weights for the external variables and $\mathbf{u}_i$ represents the M vector with coordinates for person i in M dimensional Euclidean space.

We are interested in a model defined as

$$\pi_c(\mathbf{x}_i) = \frac{\exp(\theta_{ic})}{\sum_{c'} \exp(\theta_{ic'})}$$

where θ_{ic} is

$$\theta_{ic} = -d^f(\mathbf{u}_i, \mathbf{v}_c),$$

where $f \in \{1, 2\}$ and $\mathbf{u}_i = \mathbf{x}_i^\top \mathbf{B}$. The π_{ic} can be considered probabilities or simply estimated values. The probabilities are inversely related to the relative distance. The closer by a class point for a person positioned somewhere in the Euclidean space, the higher the probability.

3 Application

The data used to illustrate the MRU model and to derive the algebraic formulas were drawn from the Netherlands Study of Depression and Anxiety (NESDA), which aimed at examining the effect of personality traits on the possibility of developing mental disorders (Spinhoven et al., 2009). A sub sample of the original data was used for the analysis. This sub sample considers only the primary care recruitment setting and comprises N = 1598 subjects aged from 18 to 65 years (Mean = 45.8; S.D. = 12.1). Approximately 69% were female and the average number of years of education achieved was 12.4 (SD = 3.4). Regarding the disorders, 25.3% of the participants have major depressive disorder (MDD), 7.0% have dysthmia (D), 11.7% have generalized anxiety disorder (GAD), 17.1% have social anxiety disorder (SP), and 22.5% have panic disorder (PD).

The mru2 function was used to fit the data (see Appendix B). This function implements the MRU model to fit multivariate binary data using usual Euclidean distances and the results may be shown graphically using biplots. The order of association among responses can be specified as a function argument. For this thesis, we will interpret the outputs of this function as if it is the same for a distance or squared distance model. These can be plotted in M dimensions as a scatterplot. As with standard biplots, the predictor variables can be expressed as biplot axes. The variable axis for variable p is simply defined as a collection of points with changing X_p values, with all other variables set to zero. By completing parallelograms, the positions of the objects may be determined from the variable axes. The two variable axes intersect at the mean values of the two centered predictor variables. The classes can be represented in the biplots as points.

4 Results

The findings of the MRU model employing usual Euclidean distance to analyse the subsample of the NESDA data are displayed in the biplots shown below. Neuroticism (N), extraversion (E), agreeableness (A), conscientiousness (C), openness to experience (O), education (Edu), gender, and age are shown in the biplots by solid variable axes, with the label of the variable printed on the positive side of the variable axis. Each of the possible 32 combination of disorders is represented by a green point, and their labels are listed in the following order: dysthmia (D), major depressive disorder (MDD), generalized anxiety disorder (GAD), social phobia (SP), and panic disorder (PD). In each of these class coordinates, a 0 indicates the absence of the condition and a 1 indicates its existence. For example 00100 refers to the existence of GAD and absence of all other disorders. Figure 1 displays the solution for the two-dimensional main effects model. Figure 2 gives the solution of the two-dimensional associations model. The algebraic formulae that describe the influence of a single predictor on a single response variable as well as the effect of a single predictor on the pairwise association structure between two response variables are discussed in detail in the upcoming log odds and log odds ratio sections.

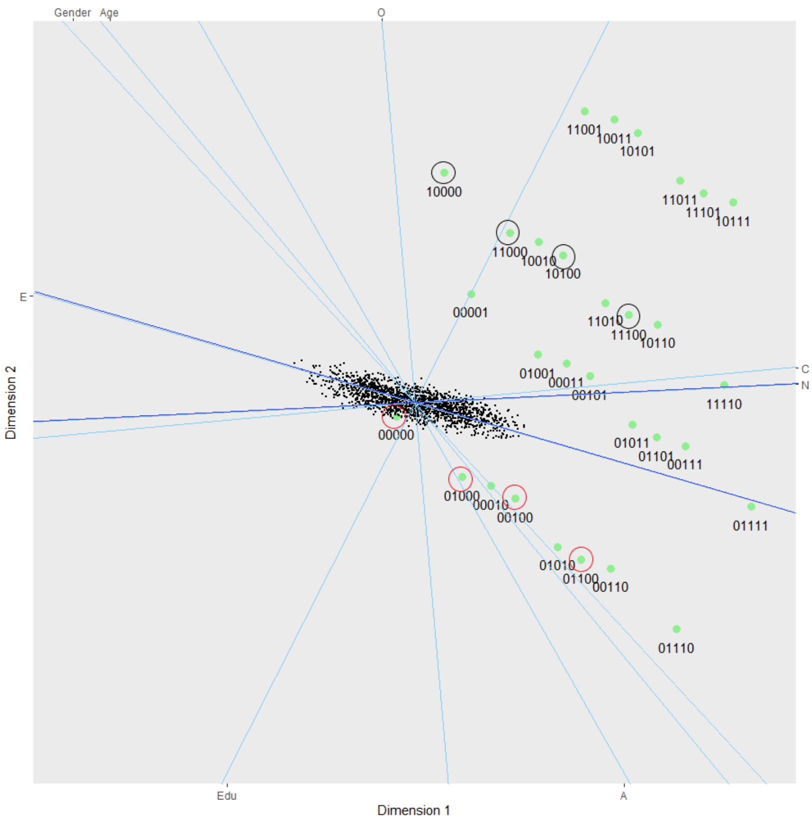

Figure 1: MRU for the main effects model fitted to a subset of the NESDA data in a two dimension solution. The axis lines represent the following predictors: Neuroticism (N), extraversion (E), agreeableness (A), conscientiousness (C), openness to experience (O), education (Edu), gender and age. The dark blue lines are used to highlight the N and E predictors. The classes coordinates are displayed using green dots and they represent the disorders in the following order: Dysthmia (D), major depressive disorder (MDD), generalized anxiety disorder (GAD), social phobia (SP), and panic disorder (PD). The 0 denotes the absence and 1 the presence of each of the five disorders. The classes coordinates marked in red represent four categories that only differ in MDD and GAD, such as absence of both disorders (00000), presence of MDD (01000), presence of GAD (00100) and co-morbidity (01100). The classes coordinates marked in black represent another set of four categories that only differ in MDD and GAD, such as: Presence of D (10000), presence of D and MDD (11000), presence of D and GAD (10100), and co-morbidity of all three disorders (11100). The angle formed by the extraversion and neuroticism axes indicates a negative correlation, with a higher level of neuroticism and a lower level of extraversion being associated with a greater likelihood of developing one or more disorders.

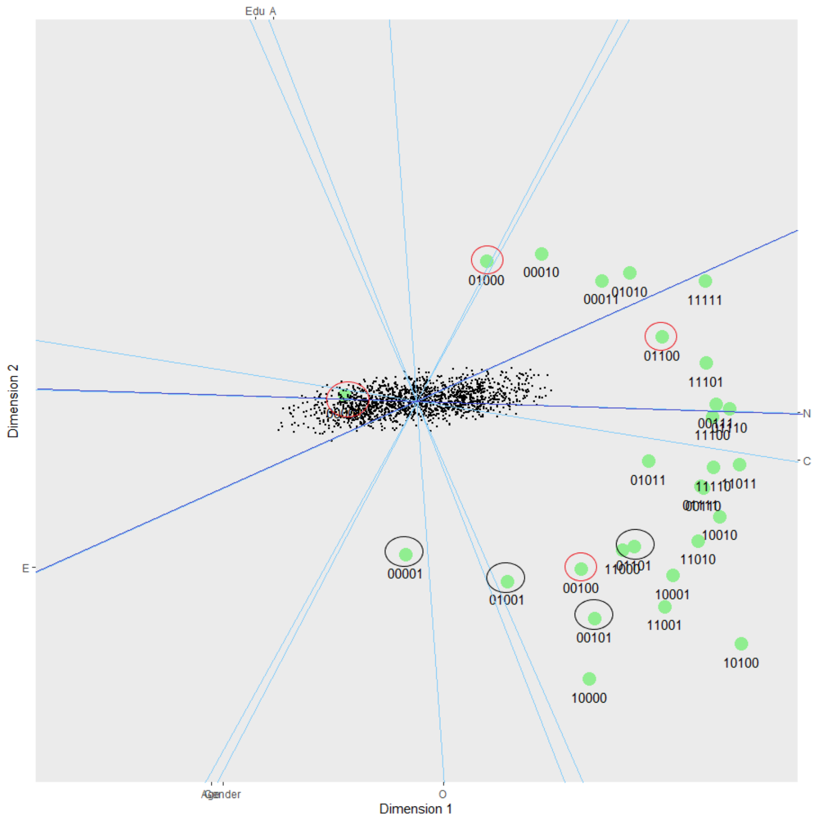

Figure 2: MRU for the associations model fitted to a subset of the NESDA data in a two dimension solution. The axis lines represent the predictors and the dark blue lines are used to distinguish the N and E from the other predictors. The class coordinates are displayed using green dots and the red circles represent four categories that only differ in MDD and GAD, such as: Absence of both disorders (00000), presence of MDD (01000), presence of GAD (00100) and the co-morbidity (01100). The classes coordinates marked in black represent another set of four categories that only differ in MDD and GAD: Presence of D (10000), presence of D and MDD (11000), presence of D and GAD (10100), and co-morbity of all three disorders (11100). The angle produced by the extraversion and neuroticism axes suggests a negative correlation, with increased neuroticism and decreased extraversion being associated with a larger probability of developing one or more disorders.

4.1 Log Odds

To explore the change in log odds, we compared individuals who vary in only one disorder but are otherwise identical. When focusing on GAD we have the following response pattern $\gamma\sigma0\beta\delta$ and $\gamma\sigma1\beta\delta$ where γ, σ, β,and δ are constants (either 0 or 1). Therefore, the pairs 00000 & 00100, 10000 & 10100, and 10101 & 10001, among others, are all representative of GAD.

We were interested in determining if the log odds of being in the first category a ($\gamma\sigma1\beta\delta$) versus the log odds of being in the second category b ($\gamma\sigma0\beta\delta$) vary in the same way for the different pairs of categories at various levels of neuroticism. Additionally, we wanted to determine if this change in log odds is dependent on another predictor's value. Hence, we intended to determine if this relationship changes for different values of extraversion in the main effects and associations models.

The log odds function for any two categories a and b, where category a represents the presence of GAD and category b represents the absence of GAD, can be defined as:

$$\lambda(\mathbf{x}) = \log\left(\frac{\pi_a(\mathbf{x}_i)}{\pi_b(\mathbf{x}_i)}\right) = d^f(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_b) - d^f(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_a) \tag{1}$$

where $f \in \{1, 2\}$.

4.1.1 Main Effects Model

From the general log odds function (Equation 1), when we would like to focus on neuroticism, which is the second predictor in our data set, we may write

$$u_{im} = \sum_p x_{ip}b_{pm} = \sum_{p\neq 2} x_{ip}b_{pm} + x_{i2}b_{2m} = u_{\bullet m} + x_{i2}b_{2m},$$

where $u_{\bullet m} = \sum_{p\neq 2} x_{ip}b_{pm}$.

When we would like to focus on the log odds of GAD, which is the third response variable in our data set, we may write

$$v_{cm} = g_{0m} + \sum_r z_{cr}g_{rm} = g_{0m} + \sum_{r\neq 3} z_{cr}g_{rm} + z_{c3}g_{3m} = v_{\bullet m} + z_{c3}g_{3m},$$

where $v_{\bullet m} = g_{0m} + \sum_{r\neq 3} z_{cr}$. We would like to compare two classes a and b, that only differ in GAD. Category a represents the presence of GAD, thus $z_{c3} = 1$, and category b represents the absence of GAD, consequently $z_{c3} = 0$. Therefore, the difference in the class coordinates is defined by the regression weights g_{3m}.

Squared Distance Model

When exploring this relationship for a model using squared Euclidean distance ($f = 2$), Equation 1 yields the following:

$$\lambda(\mathbf{x}_i) = d^2(\mathbf{u}_i, \mathbf{v}_b) - d^2(\mathbf{u}_i, \mathbf{v}_a),$$

which may be written as

$$\lambda(\mathbf{x}_i) = \sum_m (u_{im}^2 - 2u_{im}v_{bm} + v_{bm}^2) - \sum_m (u_{im}^2 - 2u_{im}v_{am} + v_{am}^2)$$

and after some manipulation (see Appendix C)

$$\lambda(\mathbf{x}_i) = \sum_m v_{bm}^2 - v_{am}^2 + 2u_{im}(v_{am} - v_{bm}).$$

Now, using the additive definitions of the coordinates for the subject points and the class points we get $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + g_{3m}$, and $v_{bm} = v_{\bullet m}$. Replacing these variables with their respective values into the equation above results in

$$\lambda(\mathbf{x}_i) = \sum_m (v_{\bullet m})^2 - (v_{\bullet m} + g_{3m})^2 + 2(u_{\bullet m} + x_{i2}b_{2m})((v_{\bullet m} + g_{3m}) - v_{\bullet m}),$$

which may be simplified into

$$\lambda(\mathbf{x}_i) = \sum_m -g_{3m}^2 - 2v_{\bullet m}g_{3m} + 2(u_{\bullet m} + x_{i2}b_{2m})(g_{3m}).$$

To demonstrate the effect of a unit change in neuroticism on the log odds of GAD, we can compare two persons where the first has $x_{12} = q$ and the second has $x_{22} = q + 1$.

$$\lambda(\mathbf{x}_1) = \sum_m -g_{3m}^2 - 2v_{\bullet m}g_{3m} + 2(u_{\bullet m} + qb_{2m})(g_{3m}),$$

$$\lambda(\mathbf{x}_2) = \sum_m -g_{3m}^2 - 2v_{\bullet m} g_{3m} + 2(u_{\bullet m} + (q+1)b_{2m})(g_{3m}).$$

Their difference $(\lambda(\mathbf{x}_2) - \lambda(\mathbf{x}_1))$ in log odds is twice the inner product of b_{2m} and g_{3m}. Therefore, with every unit change in neuroticism the log odds of GAD changes with

$$2\sum_m b_{2m} g_{3m}.$$

This difference is illustrated in Figure 3, where two distinct sets of categories representing the log odds of GAD are shown to demonstrate whether the slopes are invariant to changes in γ, σ, β,and δ. This figure is based on the solution shown in Figure 1, assuming it results from a squared distance model. The changes in the log odds of GAD for the pairs 00100 and 00000 vs 10100 and 10000 are plotted as a function of neuroticism and extraversion.

On the left hand side of Figure 3 is shown that for both pairs of categories (i.e., 00100 and 00000 vs 10100 and 10000) that represent the log odds of GAD the change in log odds of developing GAD is constant over a range of neuroticism levels for fixed values of extraversion, given the other variables are held constant. In other words, the slopes for the pairs 00100 and 00000 vs 10100 and 10000 are the same, but the intercepts are different. On the right hand side of Figure 3 is demonstrated that the change in log odds of developing GAD for both pairs of categories is constant across different levels of extraversion for fixed values of neuroticism, given the other variables are held constant. Likewise, it is possible to see that both pairs that represent the log odds of GAD have the same slope, but different intercepts. Moreover, similar to what is seen in Figure 1, Figure 3 shows that higher levels of neuroticism and lower levels of extraversion are associated with a greater likelihood of developing GAD.

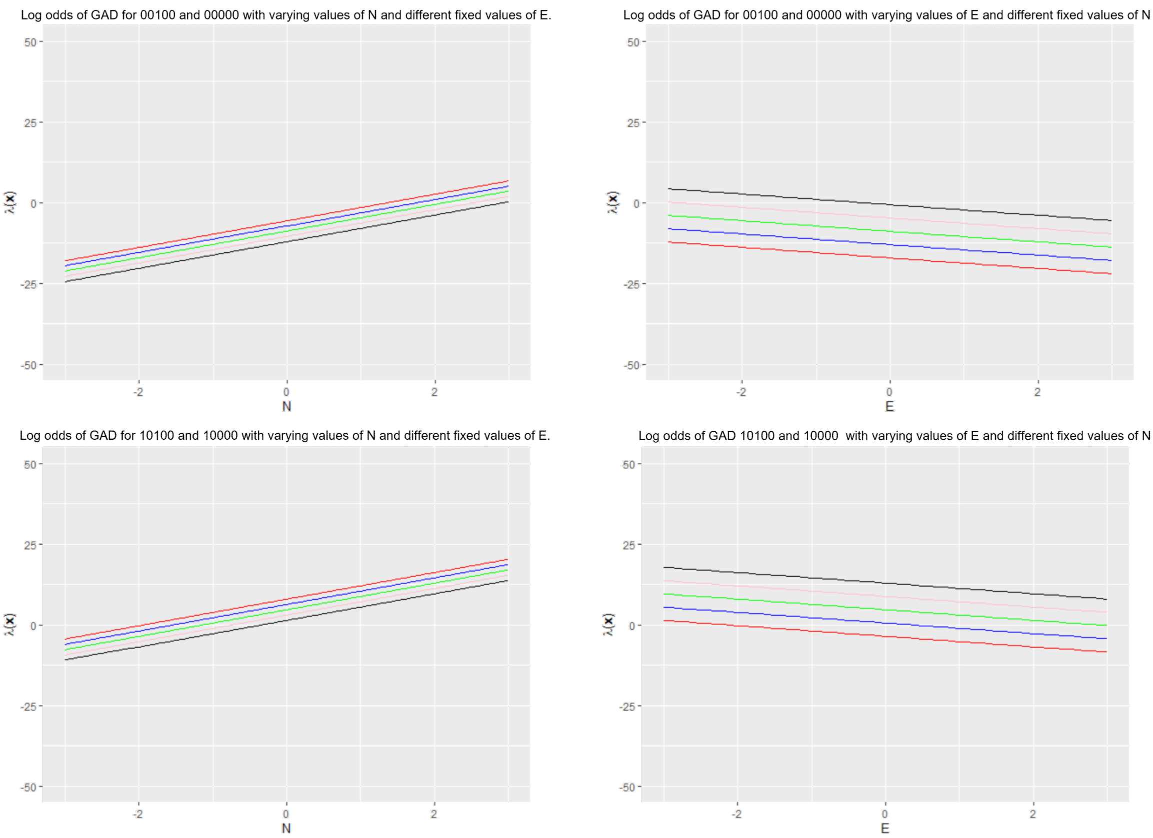

Figure 3: Log odds of developing GAD for the pairs 00100 (GAD) and 00000 (no disorder) vs 10100 (Dysthmia and GAD) and 10000 (Dysthmia) as a function of neuroticism and extraversion levels in a main effect model using squared Euclidean distances. On the left hand side, the log odds of developing GAD as a function of neuroticism levels with given fixed values of extraversion is displayed. One can see that the change in log odds of developing GAD is constant over a range of neuroticism levels for both pairs. On the right hand side the log odds of developing GAD as a function of extraversion with given fixed values of neuroticism is shown. One can see that the change in log odds of developing GAD is constant across different levels of extraversion for fixed values of neuroticism. It is also shown that higher levels of neuroticism and lower levels of extraversion are associated with a greater likelihood of developing GAD for both pairs.

Usual distance model

When analysing the same relationship for a model using usual Euclidean distance ($f = 1$), the general log odds function (Equation 1) yields

$$\lambda(\mathbf{x}_i) = d^1(\mathbf{u}_i, \mathbf{v}_b) - d^1(\mathbf{u}_i, \mathbf{v}_a),$$

which may be written in the following way

$$\lambda(\mathbf{x}_i) = \sqrt{\sum_m (u_{im} - v_{bm})^2} - \sqrt{\sum_m (u_{im} - v_{am})^2}.$$

Now we may substitute the coordinates for the subject point and class point with their respective values in the equation above, resulting in

$$\lambda(\mathbf{x}_i) = \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - g_{3m})^2}.$$

As the equation above cannot be further simplified, there is no simple and direct algebraic equation to explain the difference between person 1 and person 2, that differ one unit on the neuroticism. Despite no simple equation to illustrate this relationship, it is possible to be calculated, given the values for each variable. Figure 4 illustrates the log odds of GAD as a function of neuroticism and extraversion values for two different pairs of categories that represent GAD. These results are based on Figure 1, assuming it is a result of using usual distances.

In contrast to what is seen for the model using squared Euclidean distance, Figure 4 shows that, when using standard Euclidean distance to fit the data the change in slope is not constant across different levels of neuroticism and extraversion for both pairs of categories that represent GAD. In other words, the slopes and intercepts for the pairs 00100 and 00000 vs 10100 and 10000 are not the same. Despite the fact that for categories 10100 and 10000 the changes in the slope are not as clear as for categories 00100 and 00000, this difference becomes more evident when the vertical axis ranges from -1 to 3 (see Appendix C). Furthermore, similar to what is seen in Figures 1 and 3, Figure 4 indicates that increased neuroticism and decreased extraversion values are associated with a greater likelihood of having GAD.

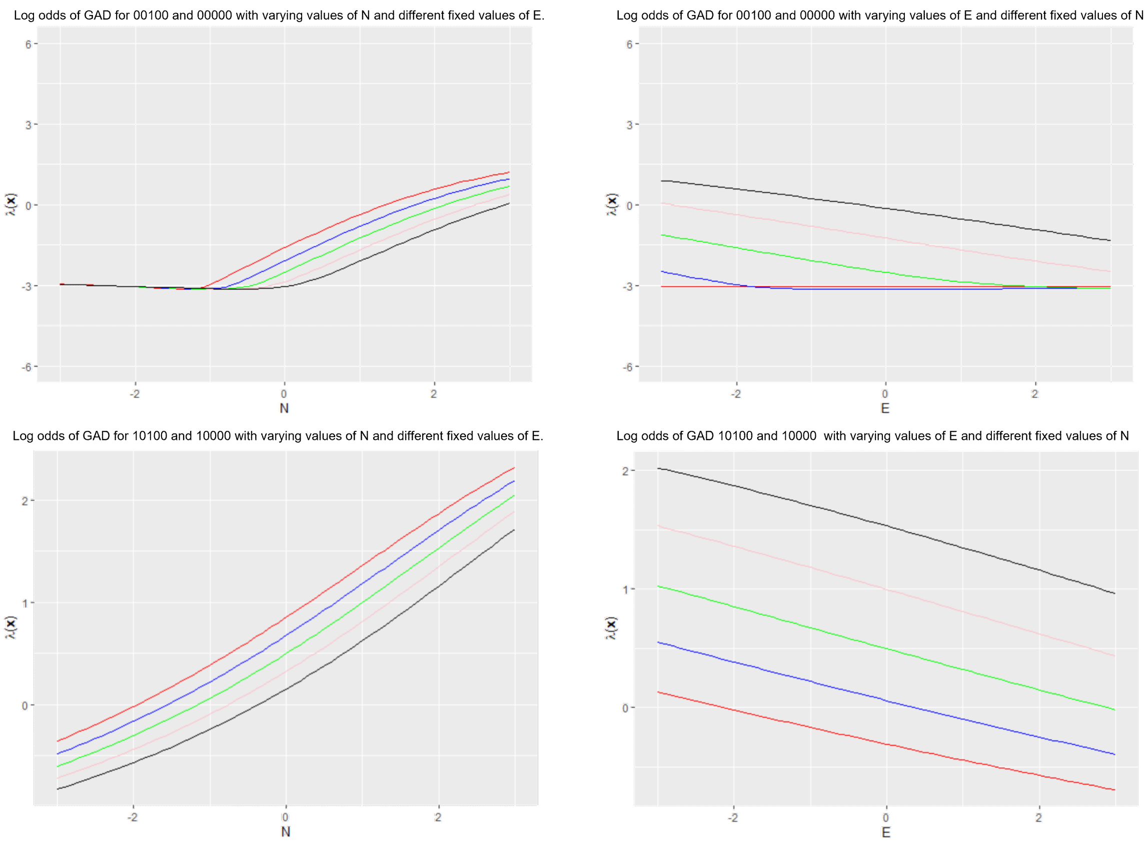

Figure 4: Log odds of developing GAD for the pairs 00100 (GAD) and 00000 (no disorder) vs 10100 (Dysthmia and GAD) and 10000 (Dysthmia) as a function of neuroticism and extraversion levels in a main effect model using usual Euclidean distances. On the left hand side, the log odds of developing GAD as a function of neuroticism levels with given fixed values of extraversion is displayed. One can see that the change in log odds of developing GAD is not constant over a range of neuroticism levels for both pairs. On the right hand side the log odds of developing GAD as a function of extraversion with given fixed values of neuroticism is shown. One can see that the change in log odds of developing GAD is not constant across different levels of extraversion for fixed values of neuroticism. The figure also indicates that higher levels of neuroticism and lower levels of extraversion are associated with a greater likelihood of developing GAD for both pairs.

4.1.2 Associations Model

In this section, the results for the model considering the main effect and pairwise associations are described in terms of usual and squared Euclidean distances. Using Equation 1 as reference, when we would like to focus on neuroticism, which is the second predictor on our data set, we may write

$$u_{im} = \sum_p x_{ip}b_{pm} = \sum_{p \neq 2} x_{ip}b_{pm} + x_{i2}b_{2m} = u_{\bullet m} + x_{i2}b_{2m},$$

where $u_{\bullet m} = \sum_{p \neq 2} x_{ip}b_{pm}$.

Similarly, we also have an equation for the class points $\mathbf{v}_c$ of any class c, which now includes the pairwise associations between the response variables

$$v_{cm} = g_{0m} + \sum_r z_{cr} g_{rm} + \sum_r \sum_{t>r} z_{ct} z_{cr} g_{trm}.$$

When we would like to focus on the log odds of GAD, which is the third response variable in our data set, we may write

$$v_{cm} = g_{0m} + \sum_{r \neq 3} (z_{cr} g_{rm}) + z_{c3} g_{3m} + \sum_{r \neq 3} \sum_{t>r} z_{ct} z_{cr} g_{trm} + \sum_t z_{ct} z_{c3} g_{t3m} = v_{\bullet m} + z_{c3} g_{3m} + \sum_t z_{ct} z_{c3} g_{t3m},$$

where $v_{\bullet m} = g_{0m} + \sum_{r \neq 3} z_{cr} g_{rm} + \sum_{r \neq 3} \sum_{s>r} z_{cs} z_{cr} g_{trm}$. We would like to compare two classes, say class a and b, that only differ in one response variable (GAD). Without GAD, which is represented by category b we have $z_{c3} = 0$, with GAD (category a) we have $z_{c3} = 1$. Therefore, the difference in the coordinates is defined by the regression weights $g_{3m} + \sum_t z_{ct} g_{t3m}$.

Squared Distance Model

From the general log odds function (Equation 1), the model using squared Euclidean distance ($f = 2$) gives

$$\lambda(\mathbf{x}_i) = d^2(\mathbf{u}_i, \mathbf{v}_b) - d^2(\mathbf{u}_i, \mathbf{v}_a),$$

and after some manipulation (see Appendix C)

$$\lambda(\mathbf{x}_i) = \sum_m v_{bm}^2 - v_{am}^2 + 2u_{im}(v_{am} - v_{bm}).$$

Now, applying the additive definitions of the coordinates for the subject points and the class points, in which $u_{im} = u_{\bullet m} + x_{i2} b_{2m}$, $v_{am} = v_{\bullet m} + g_{3m} + \sum_t z_{ct} g_{t3m}$, and $v_{bm} = v_{\bullet m}$, we have the following

$$\lambda(\mathbf{x}_i) = \sum_m (v_{\bullet m})^2 - (v_{\bullet m} + g_{3m} + \sum_t z_{ct} g_{t3m})^2 + 2(u_{\bullet m} + x_{i2} b_{2m})(v_{\bullet m} + g_{3m} + \sum_t z_{ct} g_{t3m} - v_{\bullet m}),$$

which may be simplified into

$$\lambda(\mathbf{x}_i) = \sum_m -g_{3m}^2 - \sum_t (z_{ct} g_{t3m})^2 + 2g_{3m}(u_{\bullet m} - v_{\bullet m} + x_{i2} b_{2m}) + 2(x_{i2} b_{2m} - g_{3m} - v_{\bullet m} + u_{\bullet m}) \sum_t z_{ct} g_{t3m}.$$

To illustrate the effect of one unit change in neuroticism on the log odds of GAD we can compare two people, where the first has $x_{12} = q$ and the second has $x_{22} = q + 1$.

$$\lambda(\mathbf{x}_1) = \sum_m -g_{3m}^2 - \sum_t (z_{ct} g_{t3m})^2 + 2g_{3m}(u_{\bullet m} - v_{\bullet m} + q b_{2m}) + 2(q b_{2m} - g_{3m} - v_{\bullet m} + u_{\bullet m}) \sum_t z_{ct} g_{t3m},$$

$$\lambda(\mathbf{x}_2) = \sum_m -g_{3m}^2 - \sum_t (z_{ct} g_{t3m})^2 + 2g_{3m}(u_{\bullet m} - v_{\bullet m} + (q+1) b_{2m}) + 2((q+1) b_{2m} - g_{3m} - v_{\bullet m} + u_{\bullet m}) \sum_t z_{ct} g_{t3m}.$$

Their difference ($\lambda(\mathbf{x}_2) - \lambda(\mathbf{x}_1)$) shows that, with every unit change in neuroticism the log odds of GAD changes with

$$2 \sum_m g_{3m} b_{2m} + 2 \sum_t z_{ct} g_{t3m} b_{2m}.$$

This relationship is illustrated in Figure 5, in which two different pairs of categories that represent the log odds of GAD are displayed to illustrate whether the slopes are invariant to changes in γ, σ, β,and δ. Figure 5 is based on Figure 2, assuming squared distances. Likewise, Figure 5 displays the log odds of GAD for the pairs 00100 and 00000 vs 10100 and 10000 as a function of neuroticism and extraversion. Consistently with the findings for the main effects model, these figures show that the slope is constant for different values of neuroticism and extraversion within a pair of categories that represents GAD. However the slope is not the same when we compare 00100 and 00000 vs 10100 and 10000, as now the z_{ct} value, which represents the association, differs based on the pair we are examining. Furthermore, analogous to what is observed in Figure 2, higher levels of neuroticism and lower levels of extraversion are related to higher probability of having GAD.

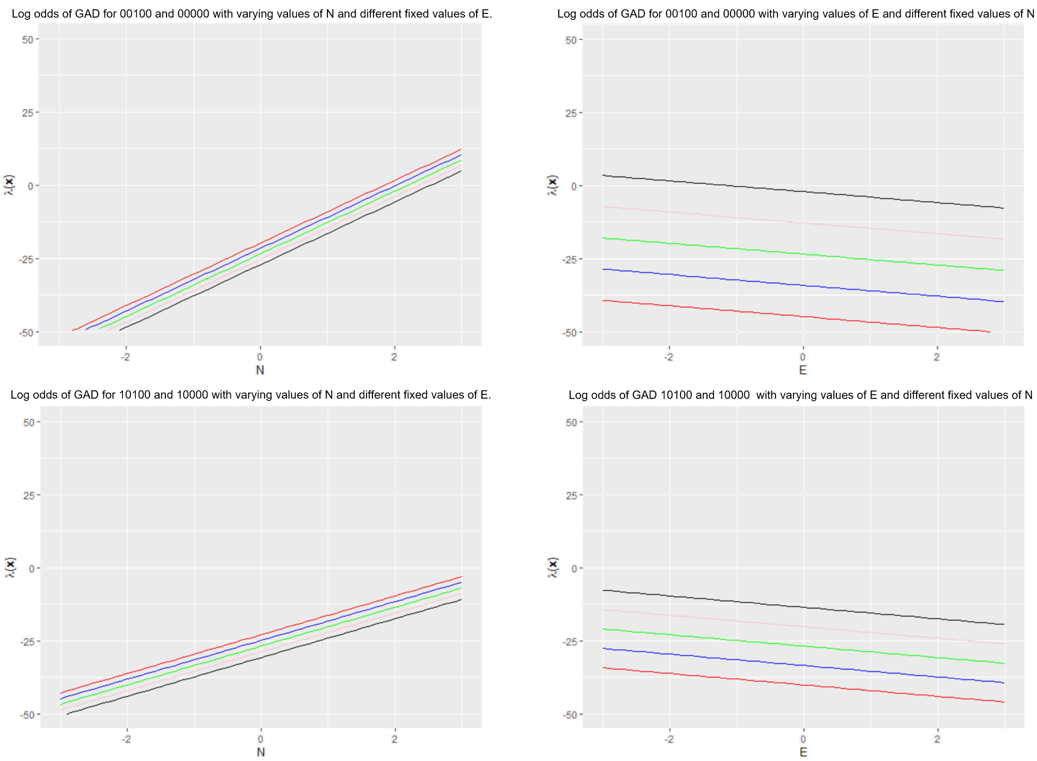

Figure 5: Log odds of developing GAD for the pairs 00100 (GAD) and 00000 (no disorder) vs 10100 (Dysthmia and GAD) and 10000 (Dysthmia) as a function of neuroticism and extraversion levels in the associations model using squared Euclidean distances. On the left hand side, the log odds of developing GAD as a function of neuroticism levels with given fixed values of extraversion displayed. One can notice that the change in log odds of developing GAD is constant over a range of neuroticism levels within a pair of categories that represents GAD. However the change in the slope is different when comparing 00100 and 00000 vs 10100 and 10000, as now the z_{ct} has different values. Likewise, the log odds of developing GAD as a function of extraversion with given fixed values of neuroticism is shown on the right hand side. One can observe that the change in log odds of developing GAD is constant across different levels of extraversion for fixed values of neuroticism for the pair 00100 and 00000, but different to what is seen for the pair 10100 and 10000. It is also shown that higher levels of neuroticism and lower levels of extraversion are associated with a greater likelihood of developing GAD for both pairs.

Usual Distance Model

When analysing the same relationship for a model using usual Euclidean distance ($f = 1$), the general log odds function (Equation 1) yields

$$\lambda(\mathbf{x}_i) = d(\mathbf{u}_i, \mathbf{v}_b) - d(\mathbf{u}_i, \mathbf{v}_a),$$

which may be written in the following way:

$$\lambda(\mathbf{x}_i) = \sqrt{\sum_m (u_{im} - v_{bm})^2} - \sqrt{\sum_m (u_{im} - v_{am})^2}.$$

Using the additive definitions of the coordinates for the subject points and the class points, we have $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + g_{3m} + \sum_t z_{ct}g_{t3m}$, and $v_{bm} = v_{\bullet m}$. Substituting these variables with their respective values into the equation above results in

$$\lambda(\mathbf{x}_i) = \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - g_{3m} - \sum_t z_{ct}g_{t3m})^2}.$$

Similar to what we found for the main effect model using usual Euclidean distance model, the equation above cannot be further simplified. Therefore, there is no straightforward algebraic equation to explain the difference between person 1 and person 2, that differ one unit on neuroticism. This difference is possible to be estimated given the values for the variables. Figure 6 represents the log odds of GAD as a function of neuroticism and extraversion values for two different pairs of categories (i.e., 00000 and 00100 vs 10000 and 10100). These results are based on Figure 2, assuming usual distances.

In contrast to what is seen for the model using squared Euclidean distance, Figure 6 indicates that the change in the log odds of GAD for these two pairs of categories is not constant across different levels of neuroticism and extraversion. In other words, the slope and intercept for the pairs 00100 and 00000 vs 10100 and 10000 are not the same. This is similar to what is observed for the main effect model using usual Euclidean distances. In addition, as observed in Figure 2, higher levels of neuroticism and lower levels of extraversion are related to higher probability of having GAD.

4.2 Log Odds Ratio

The second goal of this thesis was to provide interpretation rules to express the influence of a single predictor on the pairwise association structure between two response variables. To analyse the

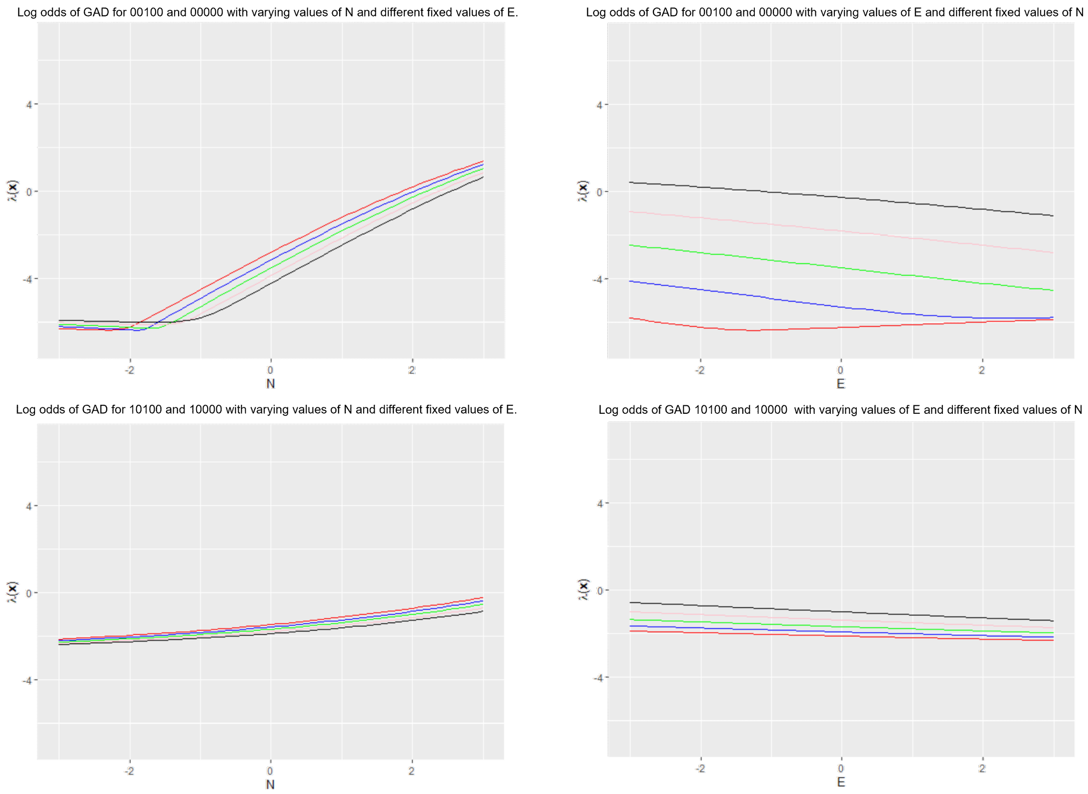

Figure 6: Log odds of developing GAD for the pairs 00100 (GAD) and 00000 (no disorder) vs 10100 (Dysthmia and GAD) and 10000 (Dysthmia) as a function of neuroticism and extraversion levels in the associations model model using usual Euclidean distances. On the left hand side, the log odds of developing GAD as a function of neuroticism levels with given fixed extraversion values is displayed. It is possible to see that the change in the log odds of developing GAD is not constant over a range of neuroticism levels for both pairs. On the right hand side the log odds of developing GAD as a function of extraversion with given fixed values of neuroticism is displayed. It is clear from the plots that the changes in log odds of developing GAD are not constant across different levels of extraversion for fixed values of neuroticism for both pairs. In addition, the figure also shows that higher levels of neuroticism and lower levels of extraversion are associated with a greater likelihood of developing GAD for both pairs.

change in odds ratio, we will compare individuals who vary in two disorders, while being identical in all other disorders.

When we examine the log odds ratio between neuroticism and two disorders, such as MDD and GAD, there are six distinct sets of four points associated to this specific odds ratio (i.e., with or without dysthmia; with or without social phobia and with or without panic disorder). When we are focusing on MDD and GAD the set of four categories that only differ in those disorders can be written as $\gamma 00\beta\delta$, $\gamma 10\beta\delta$, $\gamma 01\beta\delta$, and $\gamma 11\beta\delta$, where γ, β,and δ are constants (either 0 or 1). Therefore, the set 00000, 00100, 01000, 01100 or the set 10000, 10100, 11000, 11100, among others, are all representative for MDD - GAD disorders.

First, let us define the log odds ratio function for the four categories a, b, c, and d that represent the log odds ratio of MDD and GAD

$$\theta(\mathbf{x}) = \log\left(\frac{\frac{\pi_a(\mathbf{x}_i)}{\pi_b(\mathbf{x}_i)}}{\frac{\pi_c(\mathbf{x}_i)}{\pi_d(\mathbf{x}_i)}}\right) = \log\left(\frac{\pi_a(\mathbf{x}_i)\cdot\pi_d(\mathbf{x}_i)}{\pi_b(\mathbf{x}_i)\cdot\pi_c(\mathbf{x}_i)}\right),$$

which may be written as

$$\theta(\mathbf{x}) = -d^f(\mathbf{x}_i^\top\mathbf{B}, \mathbf{v}_a) + d^f(\mathbf{x}_i^\top\mathbf{B}, \mathbf{v}_b) + d^f(\mathbf{x}_i^\top\mathbf{B}, \mathbf{v}_c) - d^f(\mathbf{x}_i^\top\mathbf{B}, \mathbf{v}_d), \tag{2}$$

where $f \in \{1, 2\}$ and with $\mathbf{u}_i = \mathbf{x}_i^\top\mathbf{B}$. Category a represents a person with MDD and GAD ($\gamma 11\beta\delta$), category b represents the absence of MDD and the presence of GAD ($\gamma 01\beta\delta$), category c indicates the presence of MDD and absence of GAD ($\gamma 10\beta\delta$), and category d indicates no disorder ($\gamma 00\beta\delta$).

4.2.1 Main Effects Model

In this section, the results for the main effects model are described in terms of squared and usual Euclidean distances. Similarly to the log odds, we have an equation for the class points $\mathbf{v}_c$ of any class c

$$v_{cm} = g_{0m} + \sum_r z_{cr} g_{rm}.$$

When we would like to focus on the log odds ratio of major depressive disorder and generalized anxiety disorder, we may write

$$v_{cm} = g_{0m} + \sum_r z_{cr} g_{rm} = g_{0m} + \sum_{r \neq \{2,3\}} z_{cr} g_{rm} + z_{c3} g_{3m} + z_{c2} g_{2m} = v_{\bullet m} + z_{c3} g_{3m} + z_{c2} g_{2m}.$$

We would like to compare four classes, say class a, b, c and d that only differ in two response variables such as GAD and MDD. Category a represents a person with MDD and GAD, then we have $z_{c2} = 1$ and $z_{c3} = 1$. For category b, without MDD and with GAD, $z_{c2} = 0$ and $z_{c3} = 1$.

For category c, with MDD and without GAD, we have $z_{c2} = 1$ and $z_{c3} = 0$. Lastly, for category d without MDD and GAD we have $z_{c2} = 0$ and $z_{c3} = 0$.

Squared Distance Model

When exploring this relationship for a model using squared Euclidean distance ($f = 2$), Equation 2 yields

$$\theta(\mathbf{x}) = -d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_a) + d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_b) + d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_c) - d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_d),$$

which can be written as follows

$$\theta(\mathbf{x}) = -\sum_m (u_{im} - v_{am})^2 + \sum_m (u_{im} - v_{bm})^2 + \sum_m (u_{im} - v_{cm})^2 - \sum_m (u_{im} - v_{dm})^2.$$

After some manipulation we get

$$\theta(\mathbf{x}) = \sum_m 2u_{im}(v_{am} - v_{bm} - v_{cm} + v_{dm}) - v_{am}^2 + v_{bm}^2 + v_{cm}^2 - v_{dm}^2.$$

According to the additive definitions of the coordinates for the subject points and the class points, $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + g_{2m} + g_{3m}$, and $v_{bm} = v_{\bullet m} + g_{3m}$, $v_{cm} = v_{\bullet m} + g_{2m}$, $v_{dm} = v_{\bullet m}$. Replacing these variables in the algebraic equation above and working out some simplifications we get (see Appendix D)

$$\theta(\mathbf{x}) = \sum_m -2g_{2m}g_{3m}.$$

This demonstrates that the odds ratio is not dependent on the value of neuroticism. Therefore, for any of the distinct sets that represent the log odds ratio of MDD and GAD we obtain the same relationship between neuroticism and the log odds ratio of MDD and GAD. This finding is demonstrated in Figure 7, which depicts the log odds ratio of MDD and GDD as a function of neuroticism and extraversion for two sets of categories that represent this log odds ratio (i.e., first set: 01100, 00100, 01000, 00000, second set: 11100, 10100, 11000, 10000). The figure shows that the log odds ratio of MDD and GAD is equal at each degree of neuroticism and extraversion for both set of categories. Therefore, the log odds ratio is not conditional on the value of the predictors.

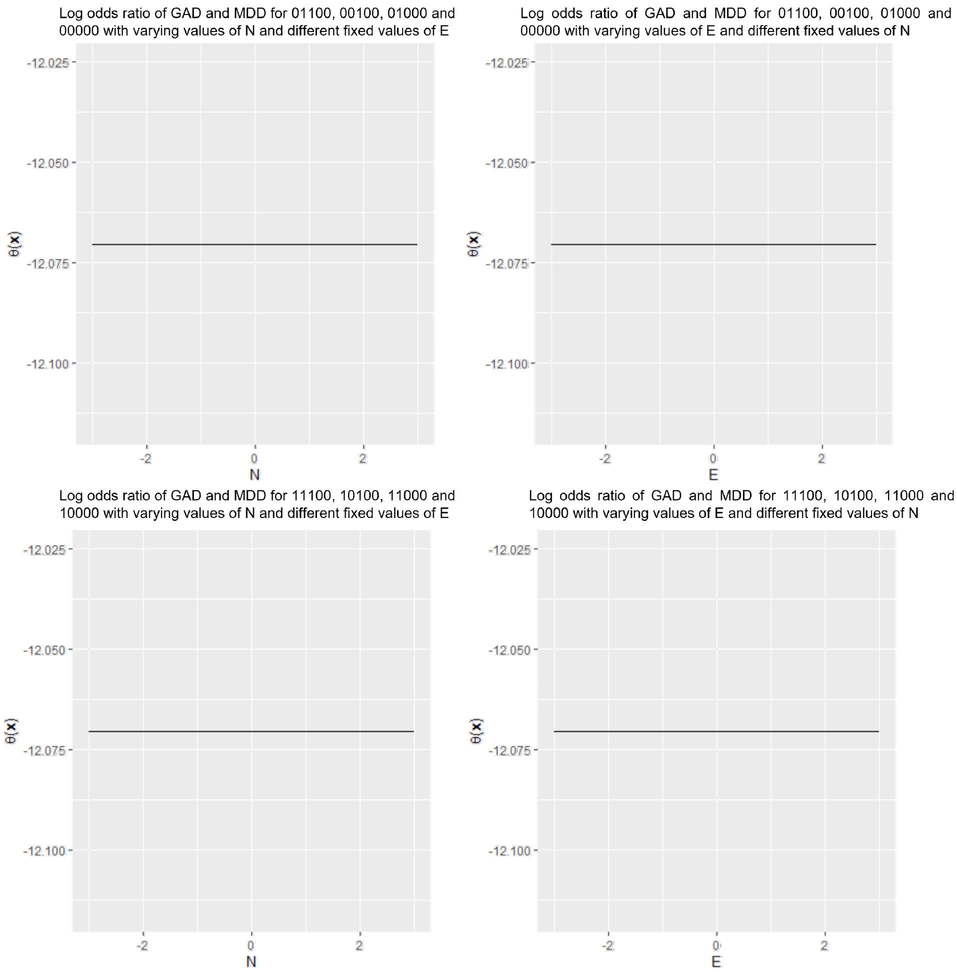

Figure 7: Log odds ratio of MDD and GAD as a function of neuroticism and extraversion values for the main effects model using squared Euclidean distance for two sets of categories that represent this log odds ratio (i.e., first set: 01100, 00100, 01000, 00000, second set: 11100, 10100, 11000, 10000). The log odds ratio of MDD and GAD is equal at each degree of neuroticism and extraversion for both set of categories. This indicates that the log odds ratio of MDD and GAD is not conditional on the value of neuroticism and extraversion.

Usual Distance Model

When analysing the same relationship for a model using usual Euclidean distance ($f = 1$), the general log odds ratio function (Equation 2) yields

$$\theta(\mathbf{x}) = -d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_a) + d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_b) + d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_c) - d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_d),$$

which may be written as follows

$$\theta(\mathbf{x}) = -\sqrt{\sum_m (u_{im} - v_{am})^2} + \sqrt{\sum_m (u_{im} - v_{bm})^2} + \sqrt{\sum_m (u_{im} - v_{cm})^2} - \sqrt{\sum_m (u_{im} - v_{dm})^2}.$$

Once more we would like to compare four classes (a, b, c and d) that only differ in two response variables (GAD and MDD). Category a represents a person with MDD and GAD, then we have $z_{c2} = 1$ and $z_{c3} = 1$. For category b, without MDD and with GAD, $z_{c2} = 0$ and $z_{c3} = 1$. For category c, with MDD and without GAD, we have $z_{c2} = 1$ and $z_{c3} = 0$. Lastly, for category d without MDD and GAD we have $z_{c2} = 0$ and $z_{c3} = 0$.

Using the additive definitions of the coordinates for the subject points and the class points, we get $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + g_{2m} + g_{3m}$, and $v_{bm} = v_{\bullet m} + g_{3m}$, $v_{cm} = v_{\bullet m} + g_{2m}$, $v_{dm} = v_{\bullet m}$. Replacing these values in the equation above gives the following

$$\theta(\mathbf{x}) = -\sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - g_{2m} - g_{3m})^2} + \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - g_{3m})^2} + \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - g_{2m})^2} - \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m})^2}.$$

Similar to what is observed for the log odds using usual Euclidean distance, the equation above cannot be further simplified. Therefore, there is no straightforward interpretation rule that can be translated into an algebraic formula to explain the difference between person 1 and person 2, that differ only one unit on neuroticism. Despite the difficulty to obtain these interpretation rules to explain the log odds ratio of two disorders, it is possible to be calculated, given the values for each variable. Figure 8 illustrated the log odds ratio of MDD and GDD as a function of neuroticism and extraversion for two set of categories that represent this log odds ratio (i.e., first set: 01100, 00100, 01000, 00000, second set: 11100, 10100, 11000, 10000). These results are based on Figure 1, assuming usual distances.

In contrast to what is seen for the model using squared Euclidean distance, Figure 8 indicates the log odds ratio of MDD and GAD is dependent on the value of neuroticism and extraversion.

The change in slope is not constant for both sets of categories that represent the log odds ratio of MDD and GAD. Therefore, a difference of one unit in neuroticism and/or extraversion have different effects on the log odds ratio of MDD and GAD depending on the value of other predictors. Moreover, neuroticism showed a negative effect on the log odds ratio, indicating that the association between the two disorders weakened as a person's level of neuroticism increased. Whereas, extroversion had a positive effect on the log odds ratio, in which the association between the two disorders increased when the level of extraversion for a given person grew.

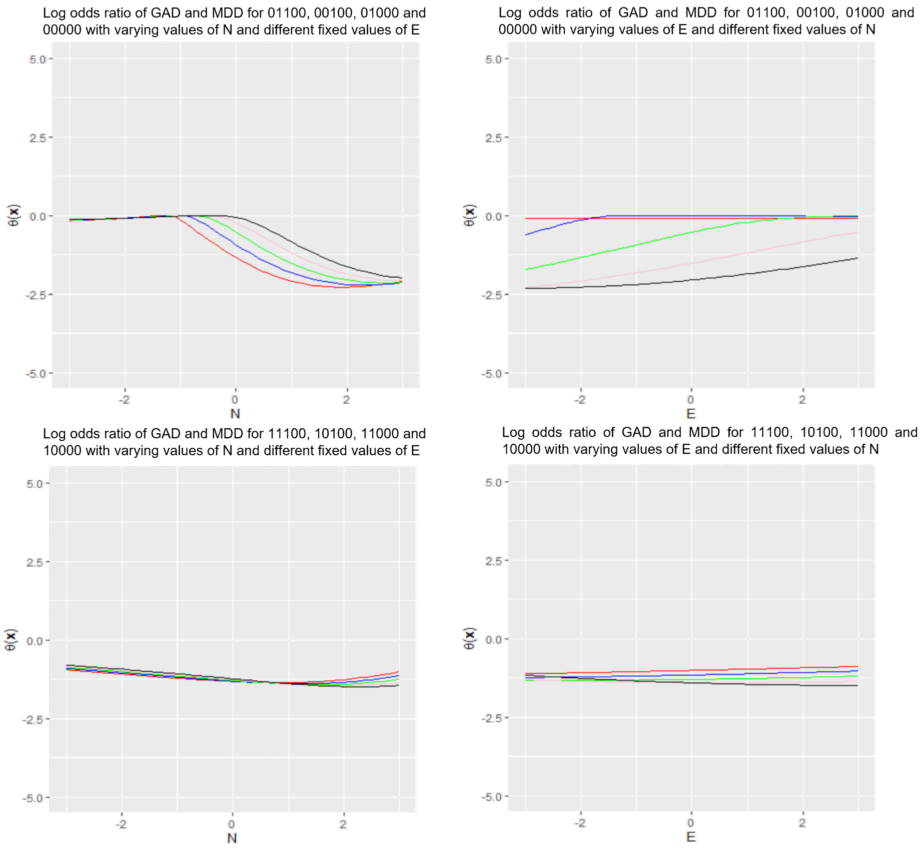

Figure 8: Log odds ratio of MDD and GAD as a function of neuroticism and extraversion values for the main effects model using usual Euclidean distance for two sets of categories that represent this log odds ratio (i.e., first set: 01100, 00100, 01000, 00000, second set: 11100, 10100, 11000, 10000). The change in slope for the log odds ratio of MDD and GAD for both sets of categories varies at different levels of neuroticism and extraversion. Furthermore, neuroticism presented a negative effect on the log-odds ratio, implying that the association between the two disorders weakened as a person's level of neuroticism grew. Whereas, extroversion had a positive effect on the log odds ratio, in which the association between the two disorders increased when the level of extraversion for a given person increased.

4.2.2 Associations Model

In this section, the results of a model considering the main effects and association are described in terms of squared and usual Euclidean distances. From the general log odds ratio function (Equation 2), we have an equation for the class points $\mathbf{v}_c$ of any class c, which now includes the pairwise association between the response variables.

$$v_{cm} = g_{0m} + \sum_r z_{cr} g_{rm} + \sum_r \sum_{t>r} z_{ct} z_{cr} g_{trm}.$$

When we would like to focus on the log odds ratio of GAD (third response variable) and MDD (second response variable), we may write

$$v_{cm} = g_{0m} + \sum_{r \neq \{2,3\}} (z_{cr} g_{rm}) + z_{c3} g_{3m} + z_{c2} g_{2m} + \sum_{r \neq \{2,3\}} \sum_{t>r} z_{ct} z_{cr} g_{trm} + \sum_{t \neq \{2,3\}} z_{ct} z_{c2} g_{t2m} + \sum_{s \neq \{2,3\}} z_{cs} z_{c3} g_{s3m}.$$

This can be written as

$$v_{cm} = v_{\bullet m} + z_{c3} g_{3m} + z_{c2} g_{2m} + \sum_{t \neq \{2,3\}} z_{ct} z_{c2} g_{t2m} + \sum_{s \neq \{2,3\}} z_{cs} z_{c3} g_{s3m},$$

where $v_{\bullet m} = g_{0m} + \sum_{r \neq \{2,3\}} z_{cr} g_{rm} + \sum_{r \neq \{2,3\}} \sum_{s>r} z_{cs} z_{cr} g_{trm}$.

We would like to compare four classes a, b, c and d that only differ in those response variables (GAD and MDD). Category a represents a person with MDD and GAD, then $z_{c2} = 1$ and $z_{c3} = 1$. Category b represents a person without MDD and with GAD, thus $z_{c2} = 0$ and $z_{c3} = 1$. Category c represents a person with MDD and without GAD, resulting in $z_{c2} = 1$ and $z_{c3} = 0$. Lastly, for category d, which represents a person without MDD and GAD we have $z_{c2} = 0$ and $z_{c3} = 0$.

Squared Distance Model

When analysing this relationship for a model using squared Euclidean distance (f= 2), equation 2 gives

$$\theta(\mathbf{x}) = -d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_a) + d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_b) + d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_c) - d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_d),$$

which can be simplified into

$$\theta(\mathbf{x}) = \sum_m 2u_{im}(v_{am} - v_{bm} - v_{cm} + v_{dm}) - v_{am}^2 + v_{bm}^2 + v_{cm}^2 - v_{dm}^2.$$

Considering the additive definitions of the coordinates for the subject points and the class points we have $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + g_{3m} + g_{2m} + \sum_{t \neq \{2,3\}} z_{ct}z_{c2}g_{t2m} + \sum_{s \neq \{2,3\}} z_{cs}z_{c3}g_{s3m} + z_{c2}z_{c3}g_{10m}$, and $v_{bm} = v_{\bullet m} + g_{3m} + \sum_{s \neq \{2,3\}} z_{cs}z_{c3}g_{s3m}$, $v_{cm} = v_{\bullet m} + g_{2m} + \sum_{t \neq \{2,3\}} z_{ct}z_{c2}g_{t2m}$, $v_{dm} = v_{\bullet m}$.

After solving the equation above using the aforementioned definitions of coordinates, we can compare two persons that only differ one unit in neuroticism, where the first has $x_{12} = q$ and the second has $x_{22} = q + 1$. This shows that with every unit change in neuroticism the log odds ratio of MDD and GAD changes with

$$2\sum_m b_{2m}g_{10m}.$$

Contrary to the results for the main effect model, the association model indicates that the log odds ratio does depend on the value of the predictor. This can be seen in Figure 9 that represents the log odds ratio of MDD and GAD for two different set of four categories as a function of neuroticism and extraversion values.

Figure 9 indicates that the slope for the log odds ratio for the two set of four categories is constant for different values of neuroticism and extraversion. In other words, one can see that the slope for the set 01100, 00100, 01000 and 00000 and for the set 11100, 10100, 11000 and 10000 is the same, but the intercepts are different. Furthermore, neuroticism had a negative effect on the log odds ratio whereas extraversion had a positive effect, indicating that the association between the two disorders became weaker when, for a given person, the level of neuroticism increased, and extraversion decreased.

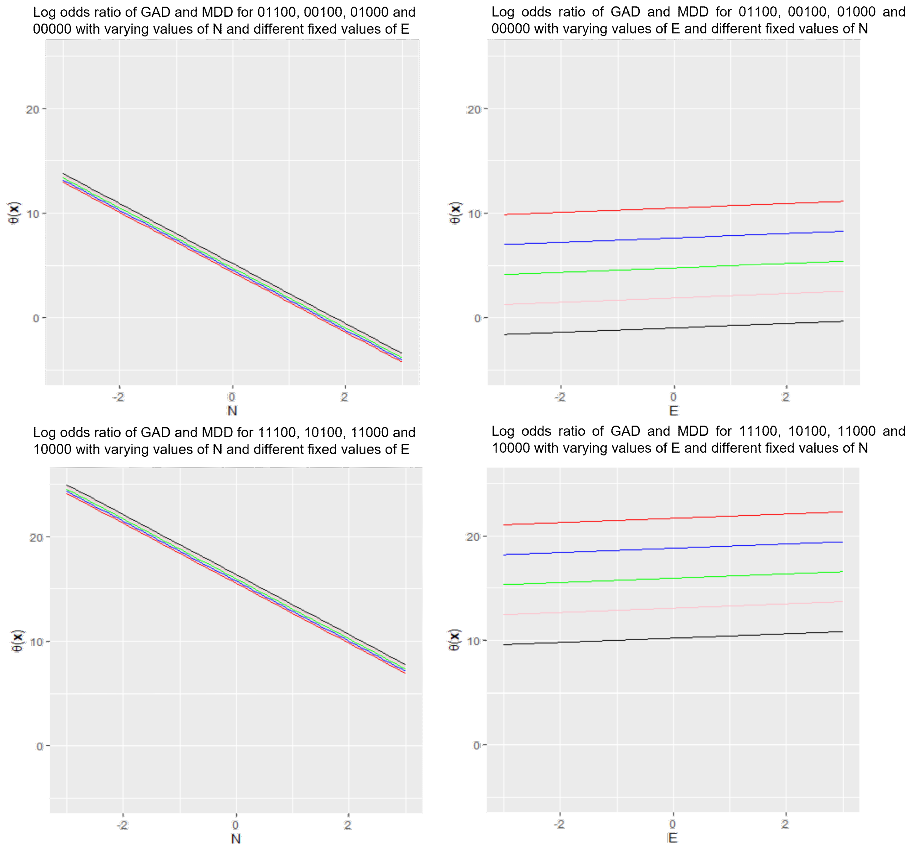

Figure 9: Log odds ratio of MDD and GAD as a function of neuroticism and extraversion values for the associations model using squared Euclidean distance for two sets of categories that represent this log odds ratio (i.e., first set: 01100, 00100, 01000, 00000, second set: 11100, 10100, 11000, 10000). At each level of neuroticism and extraversion, the change in log odds ratio of GAD and MDD is constant for both set of categories. Neuroticism has a negative effect and extraversion has a positive effect on the log odds ratio of MDD and GAD.

Usual Distance Model

When analysing the same relationship for a model using usual Euclidean distance ($f = 1$), the general log odds function (Equation 2) yields

$$\theta(\mathbf{x}) = -d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_a) + d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_b) + d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_c) - d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_d),$$

which can be written as

$$\theta(\mathbf{x}) = -\sqrt{\sum_m (u_{im} - v_{am})^2} + \sqrt{\sum_m (u_{im} - v_{bm})^2} + \sqrt{\sum_m (u_{im} - v_{cm})^2} - \sqrt{\sum_m (u_{im} - v_{dm})^2}.$$

We could replace all the values for the coordinates for the subject points and the class points in the equation above. However this will result in a long and difficult equation which cannot be further simplified (see Appendix D for more details). Therefore, when using usual Euclidean distances for a model considering main effects and associations, there is an absence of a straightforward interpretation rule which can be translated into an algebraic formula. However, this difference in the log odds ratio between person 1 and person 2 that differ only one unit on the predictor variable 2 can be estimated, given the values for each variable.

Figure 10 represents the log odds ratio of MDD and GDD as a function of neuroticism and extraversion values for two set of categories that represent this log odds ratio (01100, 00100, 01000, 00000 and 11100, 10100, 11000, 10000). These results are based on Figure 2, assuming usual distances. In contrast to what is seen for the associations model using squared Euclidean distance, Figure 10 indicates the change in the slope for the log odds ratio of MDD and GAD is not constant for both sets of categories. Therefore, a difference in one unit in neuroticism and/or extraversion have different effects on the log odds ratio of MDD and GAD depending on the value of the predictors.

Furthermore, neuroticism showed a negative effect on the log-odds ratio, indicating that the association between the two disorders weakened as a person's level of neuroticism increased. Whereas, extroversion had a positive effect on the log odds ratio, in which the association between the two disorder increased when the level of extraversion for a given person grew.

5 Discussion

In this thesis we aimed to describe (1) the marginal effect of a predictor on a response variable and (2) the effect of a predictor on the association between the response variables in a model for the joint probabilities. Regarding the former objective, the effect of Neuroticism and Extraversion on

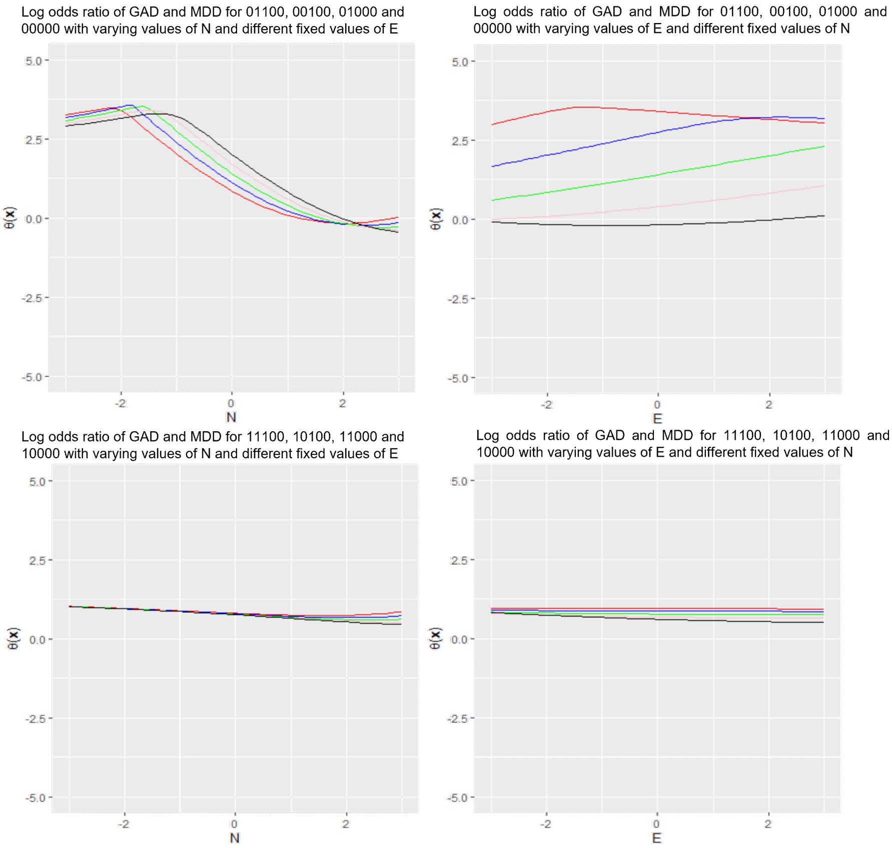

Figure 10: Log odds ratios of MDD and GAD as a function of neuroticism and extraversion values for the MRU model fitted to the NESDA data using usual Euclidean distances and considering the main effects and associations. Similar to what is observed for the model only with the main effects, for a model that incorporates the main effects and associations, the log odds ratio of MDD and GAD for the 01100, 00100, 01000, and 00000 categories changes at different levels of neuroticism and extraversion. In addition, similar to what is seen in figure 2, neuroticism and extraversion indicated a negative correlation, in which extraversion presented a negative effect on the log-odds ratio, implying that the association between the two disorders weakened as a person's level of extraversion grew. Whereas, neuroticism had a positive effect on the log odds ratio, in which the association between the two disorder increased when the level of neuroticism for a given person increased.

the change in the log odds of generalized anxiety disorder was investigated, intending to find general rules to explain how the log odds of a given disorder is affected if the value of one personality trait changes. Concerning the latter, the effect of neuroticism and extraversion on the change in the log odds ratio between generalized anxiety disorder and major depressive disorder was studied, aiming to discover general rules to describe how changes in the values of personality traits impact the log odds ratio between two disorders. We compared how the changes in log odds and log odds ratio occur in two MRU models using usual and squared Euclidean distances. First, we considered only the main effects model, later we extended it to a model that incorporates both the main effect and pairwise associations.

Figure 1 and Figure 2 illustrate the MRU models' findings, showing the influence of predictors on the outcome variables for our subset of the NESDA data to the main effect and association models, respectively. Figures 1 and 2 indicate that, for both models, extraversion and neuroticism have opposite effects, in which a higher level of neuroticism and a low level of extraversion are related to higher probability of one or more disorders. In line with these results, numerous studies have established a positive correlation between neuroticism and depression and anxiety disorders. In contrast, extraversion has been found to correlate negatively with depression and, more particularly, social phobia (Clark, Watson & Mineka, 1994; Bienvenu et al., 2001, Bienvenu & Stein, 2003; Malouff, Thorsteinsson, Schutte, 2005). In addition, several studies have found that individuals with two or more psychiatric disorders had a higher degree of neuroticism and a lower level of extraversion than those with a single condition (Bienvenu et al., 2001; Cuijpers, van Straten, Donker, 2005).

Regarding the interpretation rules to express the change in log odds, overall, the MRU models using squared Euclidean distance have shown to be more straightforward and easier to be interpreted than those using usual Euclidean Distance. For the log odds of having GAD, the algebraic formulae for the main effects model indicate a constant change in slope between different pairs of categories that represent GAD. This constant change in slope for the main effects model is also represented in Figure 3 by parallel lines, indicating that the log odds of having GAD versus the log odds of not having GAD vary in the same way for the different pairs of categories at various levels of neuroticism.

For the associations model, the parallel lines indicate that the change in the slope was constant within a pair of categories that represents GAD but not the same for different pairs of categories as indicated in Figure 5. This implies that the log odds of having GAD does not change in the same way for various pairs of categories at different levels of neuroticism and extraversion. This is due to the fact that z_{ct}, which represents the pairwise association between response variables, has different values based on the pair we are examining.

Moreover, for both main effects and associations models, the more the neuroticism, the greater the log odds of GAD; conversely, the lower the value of extroversion, the greater the log odds of GAD. These results corroborate what was shown in Figures 1 and 2, where greater values of neuroticism are associated with greater log odds of GAD, whereas higher extraversion values are associated with reduced log odds of GAD.

Furthermore, when using squared Euclidean distances, the log odds can be expressed as a function of the regression weights for the predictor variable and the regression weights for the class points. Therefore, for the main effects model, with every unit increase in one predictor variable x_{ip}, the log odds of having one disorder changes by $2\sum_m g_{rm}b_{pm}$. Similarly, for the model that includes the main effects and pairwise association, with every unit increase in one predictor variable x_i, the log odds of having one disorder changes by $2\sum_m g_{rm}b_{pm} + 2\sum_t z_{ct}g_{trm}b_{pm}$, in which a term for the association is included.

When usual Euclidean distances were used to indicate the effect of a single predictor on one response variable, the change in slope was not constant for both models. Figure 4 suggest that, for the main effect model, a difference in one unit in neuroticism has different effect on the log odds of developing GAD depending on the value of neuroticism. Furthermore, its effect varies according to the value of extraversion, implying that different levels of neuroticism will lead to different changes in extraversion. Similar occurs to the association model as seen in Figure 6, in which a change in one unit in neuroticism will lead to different effect on the log odds of developing GAD depending on the value of neuroticism. In addition, its effect varies depending on the value of extraversion. Similar to what was seen for the models using squared distance, when using usual distances, main effect and association models indicate that greater log odds of GAD are related to higher neuroticism and lower extraversion values. These results are in line to what was seen in Figures 1 and 2, where greater values of neuroticism and lower values of extraversion are associated with greater log odds of GAD.

Regarding the interpretation rules to express the change in the association structure between two response variables, our results have demonstrated that models using squared Euclidean distance are more straightforward and easier to be interpreted than those employing usual Euclidean distance.

While considering the main effects model, our findings indicate that the association structure, here represented by the log odds ratio, does not dependent on the value of the predictor variable. This is demonstrated on the algebraic equation for the log odds ratio of two disorders for any four categories and in Figure 7, in which a constant and a single line is represented. This indicates that the log odds of two disorders, in an main effect model using squared Euclidean distances, is identical for different values of one predictor variable.

However, as expected, for the associations model the log odds ratio is dependent on the value of the predictor variable. This is demonstrated in Figure 9 and in the algebraic equation, in which a constant change in slope is shown. This constant change in slope can be seen in the parallel lines, indicating that the association between the two disorders change in the same way for the different pairs of categories at various levels of neuroticism. Figure 9 also shown that at any level of extraversion, a certain change in neuroticism will lead to a constant change in extraversion, meaning that the association structure between the two disorders do not depend on the value of extraversion. The association establishes a connection between the two binary response variables. That is, it tells whether the chance of the second response variable occurring increases/decreases in reaction to an increase in the probability of the first response variable occurring, and vice versa (Worku, 2018). For the associations model, neuroticism had a negative effect on the log odds ratio whereas extraversion had a positive effect, which implies that the association between the two disorders became weaker when for a given person, the level of neuroticism increased, and extraversion decreased.

When using usual Euclidean distances to explore the association structure between two response variables, the change in slope was not constant for both models. Figure 8 suggests that, for the main effect model, a difference in one unit in neuroticism has different effect on the log odds ratio of MDD and GAD depending on the value of neuroticism. Furthermore, its effect varies according to the value of extraversion, implying that different levels of neuroticism will lead to different changes in extraversion. Similar occurs to the association model as seen in Figure 10, in which a change in one unit in neuroticism led to different change on the log odds ratio of MDD and GAD depending on the value of neuroticism and extraversion.

In this thesis, we investigated the properties of MRU models. We have demonstrated that MRU models using squared and usual Euclidean distances can be employed to analyse multivariate binary data, representing well the changes in log odds and the changes in log odds ratio. The MRU models using squared Euclidean distances are more straightforward and easier to be interpreted than those using usual Euclidean distance. However, despite the more complicated interpretation of biplots, the model utilizing the standard Euclidean distance has a higher degree of freedom to fit the data, which implies greater sensitivity to changes in predictor values. Because this model's representation of log odds and log odds ratio is more flexible than the model that uses squared Euclidean distances, it might lead to a better fit. We conclude this thesis with a suggestion for future researchers to investigate the change in log odds and in log odds ratio for a model considering also the third order association among the response variables.

References

Agresti, A. (2002). Categorical data analysis (2nd ed.). New York: Wiley

Bahadur, R. R. (1961). A representation of the joint distribution of responses to dichotomous items. In H. Solomon (Ed.), *Studies in item analysis and prediction* (pp. 158–176). Stanford, CA: Stanford University Press.

Bienvenu O. J., Brown, C, Samuels, J. F., Liang, K. Y., Costa, P. T., Eaton, W. W., Nestadt, G. (2001). Normal personality traits and comorbidity among phobic, panic, and major depressive disorders. *Psychiatry Research, 102*(1), 73–85.

Bienvenu, O. J., Stein, M. B. (2003). Personality and anxiety disorders. *Journal of Personality Disorders, 17*, 139–51.

Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103*(1), 103–116.

Cox M., Cox T. (2008) Multidimensional Scaling. In: Handbook of Data Visualization. Springer Handbooks Comp.Statistics. Springer, Berlin, Heidelberg.

Cuijpers, P., van Straten, A., Donker, M. (2005). Personality traits of patients with mood and anxiety disorders, *Psychiatry Research, 13*(3), 229-237.

De Leeuw, J. (2005). Multidimensional unfolding. In B. S. Everitt & D. C. Howelll (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1289–1294). Hoboken: Wiley.

De Rooij, M. (2009). Ideal Point Discriminant Analysis with a Special Emphasis on Visualization. *Psychometrika, 74,* 317–330.

Hoekstra, H. A, Ormel J., De Fruyt F. (1996). De NEO-PI-R/NEO-FFI; Big Five Persoonlijkheidsvragenlijsten; Handleiding. Lisse: Swets and Zeitlinger; [Manual of the Dutch version of the NEO-PI-R/NEO-FFI]

Komac, M., Sajn, R. (2001). Polluted or nonpolluted - a fuzzy logic approach determing soil pollution. *Geological Survey of Slovenia.*

Kruskal, J.B. and Wish, M. (1978) Multidimensional Scaling. Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-011, Sage Publications, Newbury Park.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalised linear models. *Biometrika*, *73*, 13–22.

Liang, K. Y., Zeger, S. L., & Qaqish, B. (1992). Multivariate regression analysis of categorical data. *Journal of the Royal Statistical Society, Series B*, *54*, 3–40.

Lipsitz, S. R., Laird, N. M., & Harrington, D. P. (1990). Maximum likelihood regression methods for paired binary data. *Statistics in Medicine*, *9*, 1517–1525.

Malouff, J. M. Thorsteinsson, E. B. & Schutte, N. S. (2005). The Relationship Between the Five-Factor Model of Personality and Symptoms of Clinical Disorders: A Meta-Analysis. *Journal Psychopathology and Behavioral Assessment 27,* 101-114.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. London: Chapman and Hall.

Spinhoven, P., De Rooij, M., Heiser, W., Penninx, B. W. J. H., & Smit, J. (2009). The Role of Personality in Comorbidity Among Anxiety and Depressive Disorders in Primary Care and Speciality Care: A Cross-Sectional Analysis. *General Hospital Psychiatry, 31,* 470–477.

Takane, Y., Bozdogan, H., & Shibayama, T. (1987). Ideal point discriminant analysis. *Psychometrika*, *52*, 371–392.

Takane, Y. (1998). Visualization in ideal point discriminant analysis. In J. Blasius & M. J. Greenacre (Eds.), *Visualization of categorical data* (pp. 441–459). Academic press: New York.

Ter Smitten M. H, Smeets R. M. W, Van den Brink W. (1998). Composite International Diagnostic Interview (CIDI), version 2.1, 12 months. Amsterdam: World Health Organization [in Dutch].

Worku, H.M., and De Rooij, M. (2017), "Properties of Ideal Point Classification Models for Bivariate Binary Data", *Psychometrika, 82,* 308-328.

Worku, H. M. (2018). Distance models for analysis of multivariate binary data. (Doctoral dissertation, Leiden University, Leiden, The Netherlands). Retrieved from http://hdl.handle.net/1887/67140.

Appendix

Appendix A

Example of matrix $\mathbf{Z}$ for the NESDA data considering only the main effects. This matrix has C rows and $r+1$ columns.

	(Intercept)	D	M	G	S	P
00000	1	0	0	0	0	0
00001	1	0	0	0	0	1
00010	1	0	0	0	1	0
00011	1	0	0	0	1	1
00100	1	0	0	1	0	0
00101	1	0	0	1	0	1
00110	1	0	0	1	1	0
00111	1	0	0	1	1	1
01000	1	0	1	0	0	0
01001	1	0	1	0	0	1
01010	1	0	1	0	1	0
01011	1	0	1	0	1	1
01100	1	0	1	1	0	0
01101	1	0	1	1	0	1
01110	1	0	1	1	1	0
01111	1	0	1	1	1	1
10000	1	1	0	0	0	0
10001	1	1	0	0	0	1
10010	1	1	0	0	1	0
10011	1	1	0	0	1	1
10100	1	1	0	1	0	0
10101	1	1	0	1	0	1
10110	1	1	0	1	1	0
10111	1	1	0	1	1	1
11000	1	1	1	0	0	0
11001	1	1	1	0	0	1
11010	1	1	1	0	1	0
11011	1	1	1	0	1	1
11100	1	1	1	1	0	0
11101	1	1	1	1	0	1
11110	1	1	1	1	1	0
11111	1	1	1	1	1	1

Example of matrix $\mathbf{Z}$ for the NESDA data incorporating the main effects and the pairwise associations between the response variables. This matrix has C rows and $r+1+\frac{r(r-1)}{2}$ columns.

	(Intercept)	D	M	G	S	P	D:M	D:G	D:S	D:P	M:G	M:S	M:P	G:S	G:P	S:P
00000	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
00001	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
00010	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
00011	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1
00100	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
00101	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
00110	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
00111	1	0	0	1	1	1	0	0	0	0	0	0	0	1	1	1
01000	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
01001	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0
01010	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0
01011	1	0	1	0	1	1	0	0	0	0	0	1	1	0	0	1
01100	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0
01101	1	0	1	1	0	1	0	0	0	0	1	0	1	0	1	0
01110	1	0	1	1	1	0	0	0	0	0	1	1	0	1	0	0
01111	1	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1
10000	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10001	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0
10010	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0
10011	1	1	0	0	1	1	0	0	1	1	0	0	0	0	0	1
10100	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
10101	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1	0
10110	1	1	0	1	1	0	0	1	1	0	0	0	0	1	0	0
10111	1	1	0	1	1	1	0	1	1	1	0	0	0	1	1	1
11000	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
11001	1	1	1	0	0	1	1	0	0	1	0	0	1	0	0	0
11010	1	1	1	0	1	0	1	0	1	0	0	1	0	0	0	0
11011	1	1	1	0	1	1	1	0	1	1	0	1	1	0	0	1
11100	1	1	1	1	0	0	1	1	0	0	1	0	0	0	0	0
11101	1	1	1	1	0	1	1	1	0	1	1	0	1	0	1	0
11110	1	1	1	1	1	0	1	1	1	0	1	1	0	1	0	0
11111	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Appendix B

```r
################################### mru2 ################################################
#' Multinomial Restricted Unfolding using Euclidean Distance
#'
#'  With this function it is possible to add restrictions on the class coordinates
#'
#'@param   X : N by P matrix with predictor variables
#'         Y : N by R matrix with dichotomous response variables 0 and 1
#'         M : dimensionality (defaults to 2)
#'         coding : dummy or effect coding
#'         start : random or linear discriminant analysis
#'         ord : illustrated(1 = main effects; 2 = main effects and pairwise associations)
#'         G is an indicator matrix
#'         B is a matrix with regressions weights for the external variables
#'         V is a matrix with the coordinates of category c
#'         U is a matrix with the coordinates for person i

mru2 = function(X, Y, M = 2, ord = 2, coding = "dummy",
                start = "lda", maxiter1 = 1e8, maxiter2 = 30,
                tol1 = 1e-6, tol2 = 1e-3, DEVCRIT = 1.0e-6,
                STRCRIT = 1.0e-12, DISCRIT = 1.0e-12, EPSCRIT = 2.5e-13,
                DIVCRIT = -1.0e-08, trace = F){
  library(nnet)
  library(Rfast)
  library(microbenchmark)
  library(Rcpp)
  n = nrow(X);P = ncol(X);R = NCOL(Y)

   # Create G
  profile = matrix(NA, n, 1)
  for(i in 1:n){profile[i, 1] = paste(Y[i,], collapse = "")}
  G = class.ind(profile) # Generates a class indicator function from a given factor

   # Creates matrix A which will be used to create Z
  A = unique(Y)
  Aprofile = matrix(NA, nrow(A),1)
  for(i in 1:nrow(A)){Aprofile[i, 1] = paste(A[i,], collapse = "")}
  Ai = t(class.ind(Aprofile))
  aidx = rep(NA, nrow(A)); for(j in 1:nrow(A)){aidx[j] = which(Ai[j,] == 1, arr.ind = T)}
  A = A[aidx,]
  A = as.data.frame(A)
  rownames(A) = colnames(G)
  colnames(A) = colnames(Y)

  # Creates Z based on the order of associations
  if (ord == 1){
    Z = model.matrix(~ ., data = A)}
```

```r
else if (ord == 2){
  Z = model.matrix(~ .^2, data = A)}
else if (ord == 3){
  Z = model.matrix(~ .^3, data = A)}
else if(ord == R){
  Z = diag(nrow(A))}
if(coding == "effect"){Z[Z==0] = -1}

## 1. starting values from a generalized singular value decomposition (get V and B)
if(start == "random"){
  B = matrix( runif( ncol( X ) * M ), ncol( X ), M )
  V = matrix( runif( ncol( G ) * M ), ncol( G ), M )
  U = X %*% B
}
else if (start == "lda"){
##  starting values from joint linear discriminant analysis:
Dy = t(G) %*% G
Dzy = t(Z) %*% Dy %*% Z
eig.dzy = eigen(Dzy)
Dzy.invsqrt = eig.dzy$vectors %*% diag(sqrt(1/eig.dzy$values))
#
U = t(X) %*% G %*% Z %*% Dzy.invsqrt
V = (1/n) * t(X) %*% X
#
out.evd = eigen(V)
V2 = out.evd$vectors %*% diag(sqrt(1/out.evd$values))
#
out.svd = mysvd(t(U) %*% V2, M = M)
B1 = V2 %*% out.svd$v
B2 = solve(Dzy) %*% t(Z) %*% t(G) %*% X %*% B1
#
U = X %*% B1
V = Z %*% B2}

## 2. With starting values compute pi (current probabilities)
theta = - dista(U, V)  # Distance between vectors and a matrix
Pi = sweep(exp(theta), 1, rowsums(exp(theta)),`/`)  # Probabilities

## 3. Compute initial Deviance
D = -2*sum(G*log(Pi))
deviance = rep(NA, 1e6)

## 4. START ITERATIONS
dif = 1; iter = 0
while(dif > tol1 & iter < maxiter1){
  # Compute Delta based on current distances and pi
  iter = iter + 1

  # main majorization step
  delta = - ( theta + 4*(G - Pi))

  # minimize majorizing function
  result = rmdu.neg2(delta, M, X, B1, Z, B2, maxiter = maxiter2, tol = tol2)
```

```r
    B1 = result$B1
    B2 = result$B2
    theta = -result$Theta

    # Compute new pi
    Pi = sweep(exp(theta), 1, rowsums(exp(theta)),`/`)

    # Compute new deviance
    deviance[iter] = -2*sum(G*log(Pi))

    # Check convergence
    if(iter > 1){
      dif = (deviance[iter-1] - deviance[iter])/deviance[iter]
      if(trace){
        cat(iter, deviance[iter-1], deviance[iter], dif, "\n")
      }}}

  ## matrix rotation
  U = X %*% B1
  R = eigen(t(U) %*% U)$vectors

  B1 = B1 %*% R
  rownames(B1) = colnames(X)
  B2 = B2 %*% R

  result = list(Pi = Pi, G = G, X = X,
    Z = Z, B1 = B1, B2 = B2, U = U, V = Z %*% B2,
    deviance = deviance[1:iter],iter = iter)

  return(result)}


mysvd = function(A, M){
  # an SVD procedure that keeps matrices
  I = nrow(A)
  J = ncol(A)
  svd.out = svd(A, nu = M, nv = M)
  U = matrix(svd.out$u, I, M)
  V = matrix(svd.out$v, J, M)
  D = diag(svd.out$d, M, M)
  result = list(u = U, v = V, d = D)

  return(result)}


################################ rmdu.neg2 ###########################################
#' Restricted Multidimensional Unfolding for partly negative dissimilarities
#'
#' Model based on Heiser (1991), Psychometrika.
#'
#' In this model both row and column points are restricted.
#' U = Xb
#' V = ZC
```

```r
#' STRCRIT: stress criterion, to terminate the whole procedure
#' DISCRIT: distance criterion, to check whether distance is usable
#' EPSCRIT: small positive number, when the previous distance is not usable,
#'          choosing EPS smaller than the stopping criterion used to terminate
#'          the process. (beta = 2*EPS if abs(delta_(ij)) <= 1)
#' DIVCRIT: divergent criterion, check whether there is divergent during updating
#' epsilion = CRIT/4 # declare epsilion such that epsilion < STRCRIT

rmdu.neg2 = function( delta,
                      M = 2, X, B1,
                      Z, B2,maxiter = 30,
                      tol = 1.0e-12, DISCRIT = 1.0e-12,
                      EPSCRIT = 2.5e-13, DIVCRIT = -1.0e-08){

  # get n and C from delta
  n = nrow( delta )
  C = ncol( delta )

  # get dim P from X
  P = ncol(X)

  # prepare for updates
  U = X %*% B1
  V = Z %*% B2

  # updating new distance and calculate squared of the Euclidean norm
  D = dista( U, V )
  fold = sum( ( delta - D )^2 )

  # register negative index
  logicN = delta < 0

  iter = 0; dif = 1
  while( dif > tol & iter < maxiter) {
    iter = iter + 1

    # register not usable index
    logic0 = D < DISCRIT

    # Calculate tildeU and tildeV
    A = delta / D

    # Deal with not usable S(o)
    A[logic0] = 0

    # Deal with negative S(N)
    A[logicN] = 0

    P = diag( rowsums( A ) )
    Q = diag( colsums( A ) )
    tildeU = P %*% U - A %*% V
    tildeV = Q %*% V - t(A) %*% U
```

```
  # initail weight matrix: identity matrix
  W = matrix( 1, n, C )

  # W for S(N) and S(+)
  W[logicN] = (D[logicN] + abs(delta[logicN])) / D[logicN]

  # W for S(N) and S(o)
  logicN0 = logicN & logic0
  W[logicN0] = (EPSCRIT + abs(delta[logicN0])^2) / EPSCRIT

  # construct matrix R and S based on W
  R = diag(rowSums(W), nrow = n, ncol = n) # original:C
  S = diag(colSums(W), nrow = C, ncol = C) # original:n

  # update B and V in each step
  B1 = chol2inv(chol(t(X) %*% R %*% X)) %*% (t(X) %*% tildeU + t(X) %*% W %*% V)
  U = X %*% B1

  B2 = chol2inv(chol(t(Z) %*% S %*% Z)) %*% (t(Z) %*% tildeV + t(Z) %*% t(W) %*% U)
  V = Z %*% B2

  # calculate distance
  D = dista( U, V )
  fnew = sum( ( delta - D)^2 )
  dif = 2.0 * ( fold - fnew ) / (fold + fnew )
  fold = fnew }

output = list( B1 = B1, B2 = B2,
  U = U, V = V, Theta = D,
  iter=iter,  stress = fnew)

return(output)}
```

Appendix C

Below the algebraic equations to explain the changes in log odds of generalized anxiety disorder using squared and usual Euclidean distances are described for the main effects and associations models.

Main Effects model

First, let us define the log odds function for any two categories a and b:

$$\lambda(\mathbf{x}) = \log\left(\frac{\pi_a(\mathbf{x}_i)}{\pi_b(\mathbf{x}_i)}\right) = d^s(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_b) - d^s(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_a)$$

where $s \in \{1, 2\}$ and with $\mathbf{u}_i = \mathbf{x}_i^\top \mathbf{B}$. Per dimension we can define the coordinate

$$u_{im} = \sum_p x_{ip} b_{pm}.$$

When we would like to focus on the second predictor we may write

$$u_{im} = \sum_p x_{ip} b_{pm} = \sum_{p \neq 2} x_{ip} b_{pm} + x_{i2} b_{2m} = u_{\bullet m} + x_{i2} b_{2m},$$

where $u_{\bullet m} = \sum_{p \neq 2} x_{ip} b_{pm}$. Now, we can compare two persons where the first has $x_{12} = q$ and the second has $x_{22} = q + 1$, a difference of 1 on predictor variable 2 to see what the effect of a unit change is on the log-odds.

Similarly, we also have an equation for the class points $\mathbf{v}_c$ of any class c. When we are focusing on the main effects model we have for example

$$v_{cm} = \gamma_{0m} + \sum_r z_{cr} \gamma_{rm}.$$

When we would like to focus on the log odds of the third response variable (say, Generalized Anxiety Disorder), we may write

$$v_{cm} = \gamma_{0m} + \sum_r z_{cr} \gamma_{rm} = \gamma_{0m} + \sum_{r \neq 3} z_{cr} \gamma_{rm} + z_{c3} \gamma_{3m} = v_{\bullet m} + z_{c3} \gamma_{3m}.$$

where $v_{\bullet m} = \gamma_{0m} + \sum_{r \neq 3} z_{cr}$. We like to compare two classes, say class a and b, that only differ in one response variable (GAD). Without GAD that is for category b we have $z_{c3} = 0$, with GAD (for category a) $z_{c3} = 1$. Therefore, the difference in the coordinates is defined by the regression weights γ_{3m}.

Squared Distance Model ($s = 2$)

Therefore, for the squared distance model we have the following

$$\lambda(\mathbf{x}_i) = d^2(\mathbf{u}_i, \mathbf{v}_b) - d^2(\mathbf{u}_i, \mathbf{v}_a)$$

$$\lambda(\mathbf{x}_i) = \sum_m (u_{im}^2 - 2u_{im} v_{bm} + v_{bm}^2) - \sum_m (u_{im}^2 - 2u_{im} v_{am} + v_{am}^2)$$

$$\lambda(\mathbf{x}_i) = \sum_m u_{im}^2 - 2u_{im} v_{bm} + v_{bm}^2 - u_{im}^2 + 2u_{im} v_{am} - v_{am}^2$$

$$\lambda(\mathbf{x}_i) = \sum_m v_{bm}^2 - v_{am}^2 - 2u_{im}v_{bm} + 2u_{im}v_{am}$$

$$\lambda(\mathbf{x}_i) = \sum_m v_{bm}^2 - v_{am}^2 + 2u_{im}(v_{am} - v_{bm})$$

Now, we start making use of the additive definitions of the coordinates for the subject points and the class points. That is $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + \gamma_{3m}$, and $v_{bm} = v_{\bullet m}$. Therefore,

$$\lambda(\mathbf{x}_i) = \sum_m (v_{\bullet m})^2 - (v_{\bullet m} + \gamma_{3m})^2 + 2(u_{\bullet m} + x_{i2}b_{2m})((v_{\bullet m} + \gamma_{3m}) - v_{\bullet m})$$

$$\lambda(\mathbf{x}_i) = \sum_m v_{\bullet m}^2 - (v_{\bullet m}^2 + \gamma_{3m}^2 + 2v_{\bullet m}\gamma_{3m}) + 2(u_{\bullet m} + x_{i2}b_{2m})(\gamma_{3m})$$

$$\lambda(\mathbf{x}_i) = \sum_m -\gamma_{3m}^2 - 2v_{\bullet m}\gamma_{3m} + 2(u_{\bullet m} + x_{i2}b_{2m})(\gamma_{3m})$$

Now, if we compare person 1 and person 2, that differ 1 unit on predictor variable 2, we get

$$\lambda(\mathbf{x}_1) = \sum_m -\gamma_{3m}^2 - 2v_{\bullet m}\gamma_{3m} + 2(u_{\bullet m} + qb_{2m})(\gamma_{3m})$$

$$\lambda(\mathbf{x}_2) = \sum_m -\gamma_{3m}^2 - 2v_{\bullet m}\gamma_{3m} + 2(u_{\bullet m} + (q+1)b_{2m})(\gamma_{3m})$$

So, that their difference $(\lambda(\mathbf{x}_2) - \lambda(\mathbf{x}_1))$ in log odds is $2b_{2m}\gamma_{3m}$. So, with every unit change in x_2 the log odds of the third disorder (GAD) changes with

$$2\sum_m b_{2m}\gamma_{3m}$$

Distance Model ($s = 1$)

For the model using usual Euclidean distances, we have the following:

$$\lambda(\mathbf{x}_i) = d^1(\mathbf{u}_i, \mathbf{v}_b) - d^1(\mathbf{u}_i, \mathbf{v}_a)$$

$$\lambda(\mathbf{x}_i) = \sqrt{\sum_m (u_{im} - v_{bm})^2} - \sqrt{\sum_m (u_{im} - v_{am})^2}$$

Making use of the additive definitions of the coordinates for the subject points and the class points, we have $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + \gamma_{3m}$, and $v_{bm} = v_{\bullet m}$. Therefore,

$$\lambda(\mathbf{x}_i) = \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - \gamma_{3m})^2}$$

Now, if we compare person 1 and person 2, that differ 1 unit on predictor variable 2, we get

$$\lambda(\mathbf{x}_1) = \sqrt{\sum_m (u_{\bullet m} + qb_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + qb_{2m} - v_{\bullet m} - \gamma_{3m})^2}$$

$$\lambda(\mathbf{x}_2) = \sqrt{\sum_m (u_{\bullet m} + (q+1)b_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + (q+1)b_{2m} - v_{\bullet m} - \gamma_{3m})^2}$$

So, that their difference ($\lambda(\mathbf{x}_2) - \lambda(\mathbf{x}_1)$) in log odds indicates that with every unit change in x_2 the log odds of the third disorder (GAD) changes with

$$\lambda(\mathbf{x}_2) - \lambda(\mathbf{x}_1) = \sqrt{\sum_m (u_{\bullet m} + (q+1)b_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + (q+1)b_{2m} - v_{\bullet m} - \gamma_{3m})^2} -$$

$$\left(\sqrt{\sum_m (u_{\bullet m} + qb_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + qb_{2m} - v_{\bullet m} - \gamma_{3m})^2}\right)$$

Associations Model

Similarly, we also have an equation for the class points $\mathbf{v}_c$ of any class c. When we are focusing on the associations model we have for example

$$v_{cm} = \gamma_{0m} + \sum_r z_{cr}\gamma_{rm} + \sum_r \sum_{t>r} z_{ct}z_{cr}\gamma_{trm}.$$

When we would like to focus on the log odds of the third response variable (say, Generalized Anxiety Disorder), we may write

$$v_{cm} = \gamma_{0m} + \sum_{r\neq 3}(z_{cr}\gamma_{rm}) + z_{c3}\gamma_{3m} + \sum_{r\neq 3}\sum_{t>r} z_{ct}z_{cr}\gamma_{trm} +$$

$$\sum_t z_{ct}z_{c3}\gamma_{t3m} = v_{\bullet m} + z_{c3}\gamma_{3m} + \sum_t z_{ct}z_{c3}\gamma_{t3m}$$

where $v_{\bullet m} = \gamma_{0m} + \sum_{r\neq 3} z_{cr}\gamma_{rm} + \sum_{r\neq 3}\sum_{s>r} z_{cs}z_{cr}\gamma_{trm}$. We like to compare two classes, say class a and b, that only differ in one response variable (GAD). Without GAD that is for category b we have $z_{c3} = 0$, with GAD (for category a) $z_{c3} = 1$. Therefore, the difference in the coordinates is defined by the regression weights $\gamma_{3m} + \sum_t z_{ct}\gamma_{t3m}$.

Squared Distance Model ($s = 2$)

Therefore, for the squared distance model we have the following

$$\lambda(\mathbf{x}_i) = d^2(\mathbf{u}_i, \mathbf{v}_b) - d^2(\mathbf{u}_i, \mathbf{v}_a)$$

Which can be simplified into

$$\lambda(\mathbf{x}_i) = \sum_m v_{bm}^2 - v_{am}^2 + 2u_{im}(v_{am} - v_{bm})$$

Making use of the additive definitions of the coordinates for the subject points and the class points, in which $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + \gamma_{3m} + \sum_t z_{ct}\gamma_{t3m}$, and $v_{bm} = v_{\bullet m}$. We have the following

$$\lambda(\mathbf{x}_i) = \sum_m (v_{\bullet m})^2 - (v_{\bullet m} + \gamma_{3m} + \sum_t z_{ct}\gamma_{t3m})^2 +$$

$$2(u_{\bullet m} + x_{i2}b_{2m})(v_{\bullet m} + \gamma_{3m} + \sum_t z_{ct}\gamma_{t3m} - v_{\bullet m})$$

$$\lambda(\mathbf{x}_i) = \sum_m v_{\bullet m}^2 - (v_{\bullet m}^2 + \gamma_{3m}^2 + \sum_t (z_{ct}\gamma_{t3m})^2 + 2v_{\bullet m}\gamma_{3m} + 2\gamma_{3m}\sum_t z_{ct}\gamma_{t3m} + 2v_{\bullet m}\sum_t z_{ct}\gamma_{t3m}) +$$

$$2(u_{\bullet m} + x_{i2}b_{2m})(v_{\bullet m} + \gamma_{3m} + \sum_t z_{ct}\gamma_{t3m} - v_{\bullet m})$$

$$\lambda(\mathbf{x}_i) = \sum_m -\gamma_{3m}^2 - \sum_t (z_{ct}\gamma_{t3m})^2 - 2v_{\bullet m}\gamma_{3m} - 2\gamma_{3m}\sum_t z_{ct}\gamma_{t3m} - 2v_{\bullet m}\sum_t z_{ct}\gamma_{t3m} + 2(u_{\bullet m} + x_{i2}b_{2m})(\gamma_{3m} + \sum_t z_{ct}\gamma_{t3m})$$

$$\lambda(\mathbf{x}_i) = \sum_m -\gamma_{3m}^2 - \sum_t (z_{ct}\gamma_{t3m})^2 - 2v_{\bullet m}\gamma_{3m} + 2u_{\bullet m}\gamma_{3m} + 2x_{i2}b_{2m}\gamma_{3m} + 2(x_{i2}b_{2m} - \gamma_{3m} - v_{\bullet m} + u_{\bullet m})\sum_t z_{ct}\gamma_{t3m}$$

$$\lambda(\mathbf{x}_i) = \sum_m -\gamma_{3m}^2 - \sum_t (z_{ct}\gamma_{t3m})^2 + 2\gamma_{3m}(u_{\bullet m} - v_{\bullet m} + x_{i2}b_{2m}) + 2(x_{i2}b_{2m} - \gamma_{3m} - v_{\bullet m} + u_{\bullet m})\sum_t z_{ct}\gamma_{t3m}$$

Now, if we compare person 1 and person 2, that differ 1 unit on predictor variable 2, we get

$$\lambda(\mathbf{x}_1) = \sum_m -\gamma_{3m}^2 - \sum_t (z_{ct}\gamma_{t3m})^2 + 2\gamma_{3m}(u_{\bullet m} - v_{\bullet m} + qb_{2m}) + 2(qb_{2m} - \gamma_{3m} - v_{\bullet m} + u_{\bullet m})\sum_t z_{ct}\gamma_{t3m}$$

$$\lambda(\mathbf{x}_2) = \sum_m -\gamma_{3m}^2 - \sum_t (z_{ct}\gamma_{t3m})^2 + 2\gamma_{3m}(u_{\bullet m} - v_{\bullet m} + (q+1)b_{2m}) + 2((q+1)b_{2m} - \gamma_{3m} - v_{\bullet m} + u_{\bullet m})\sum_t z_{ct}\gamma_{t3m}$$

Their difference ($\lambda(\mathbf{x}_2) - \lambda(\mathbf{x}_1)$) shows that, with every unit change in x_2 the log odds of the third disorder (GAD) changes with

$$2\sum_m \gamma_{3m}b_{2m} + 2\sum_t z_{ct}\gamma_{t3m}b_{2m}$$

Distance Model ($s = 1$)

For the distance model we have the following

$$\lambda(\mathbf{x}_i) = d(\mathbf{u}_i, \mathbf{v}_b) - d(\mathbf{u}_i, \mathbf{v}_a)$$

which can be written as follows:

$$\lambda(\mathbf{x}_i) = \sqrt{\sum_m (u_{im} - v_{bm})^2} - \sqrt{\sum_m (u_{im} - v_{am})^2}$$

Making use of the additive definitions of the coordinates for the subject points and the class points, in which $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + \gamma_{3m} + \sum_t z_{ct}\gamma_{t3m}$, and $v_{bm} = v_{\bullet m}$. We have the following

$$\lambda(\mathbf{x}_i) = \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - \gamma_{3m} - \sum_t z_{ct}\gamma_{t3m})^2}$$

Now, if we compare person 1 and person 2, that differ 1 unit on predictor variable 2, we get

$$\lambda(\mathbf{x}_1) = \sqrt{\sum_m (u_{\bullet m} + qb_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + qb_{2m} - v_{\bullet m} - \gamma_{3m} - \sum_t z_{ct}\gamma_{t3m})^2}$$

$$\lambda(\mathbf{x}_2) = \sqrt{\sum_m (u_{\bullet m} + (q+1)b_{2m} - v_{\bullet m})^2} - \sqrt{\sum_m (u_{\bullet m} + (q+1)b_{2m} - v_{\bullet m} - \gamma_{3m} - \sum_t z_{ct}\gamma_{t3m})^2}$$

Figure 1 describes the log odds of GAD for the pairs 10100 and 1000 when the vertical axis ranges from -1 to 3.

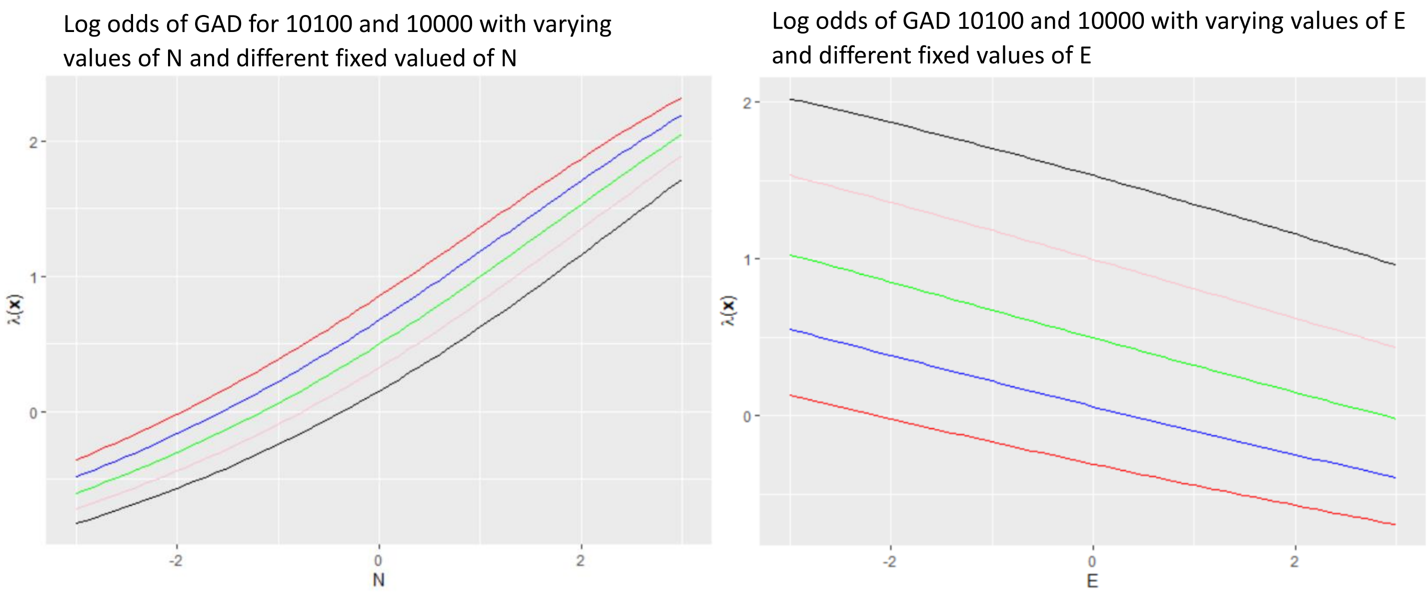

Figure 1: Log odds of developing GAD for the pairs 10100 (Dysthmia and GAD) and 10000 (Dysthmia) as a function of neuroticism and extraversion levels in a main effect model using usual Euclidean distances. On the left hand side, the log odds of developing GAD as a function of neuroticism levels is displayed. One can see that the change in log odds of developing GAD is not constant over a range of neuroticism levels for both pairs. On the right hand side the log odds of developing GAD as a function of extraversion with given fixed values of neuroticism is shown. One can see that the change in log odds of developing GAD is not constant across different levels of extraversion for fixed values of neuroticism. The figure also indicates that higher levels of neuroticism and lower levels of extraversion are associated with a greater likelihood of developing GAD for both pairs.

Appendix D

Below the algebraic equations to explain the changes in log odds ratio of generalized anxiety disorder and dysthmia using squared and usual Euclidean distances are described for the main effects and associations models.

Main Effects Model

First, let us define the log odds ratio function for the four categories a, b, c, and d:

$$\theta(\mathbf{x}) = \log\left(\frac{\frac{\pi_a(\mathbf{x}_i)}{\pi_b(\mathbf{x}_i)}}{\frac{\pi_c(\mathbf{x}_i)}{\pi_d(\mathbf{x}_i)}}\right) = \log\left(\frac{\pi_a(\mathbf{x}_i)\cdot\pi_d(\mathbf{x}_i)}{\pi_b(\mathbf{x}_i)\cdot\pi_c(\mathbf{x}_i)}\right)$$

$$\theta(\mathbf{x}) == -d^s(\mathbf{x}_i^\top\mathbf{B},\mathbf{v}_a) + d^s(\mathbf{x}_i^\top\mathbf{B},\mathbf{v}_b) + d^s(\mathbf{x}_i^\top\mathbf{B},\mathbf{v}_c) - d^s(\mathbf{x}_i^\top\mathbf{B},\mathbf{v}_d)$$

where $s \in \{1,2\}$ and with $\mathbf{u}_i = \mathbf{x}_i^\top\mathbf{B}$. Per dimension we can define the coordinate

$$u_{im} = \sum_p x_{ip}b_{pm}.$$

Category - a (MDD = 1; GAD = 1) - b (MDD = 0; GAD = 1) - c (MDD = 1; GAD = 0) - d (MDD = 0; GAD = 0)

Similarly to the log odds, we also have an equation for the class points $\mathbf{v}_c$ of any class c. When we are focusing on the main effects model we have for example

$$v_{cm} = \gamma_{0m} + \sum_r z_{cr}\gamma_{rm}.$$

When we would like to focus on the log odds ratio of the third response variable (say, Generalized Anxiety Disorder) and the second response variable (Major Depressive Disorder), we may write

$$v_{cm} = \gamma_{0m} + \sum_r z_{cr}\gamma_{rm} = \gamma_{0m} + \sum_{r\neq\{2,3\}} z_{cr}\gamma_{rm} + z_{c3}\gamma_{3m} + z_{c2}\gamma_{2m} = v_{\bullet m} + z_{c3}\gamma_{3m} + z_{c2}\gamma_{2m}.$$

We like to compare four classes, say class a, b, c and d that only differ in two response variables (GAD and MDD). Category a represents a person with MDD and GAD, then we have $z_{c2} = 1$ and $z_{c3} = 1$. For category b, without MDD and with GAD, $z_{c2} = 0$ and $z_{c3} = 1$. For category c, with MDD and without GAD, we have $z_{c2} = 1$ and $z_{c3} = 0$. Lastly, for category d without MDD and GAD we have $z_{c2} = 0$ and $z_{c3} = 0$.

Squared Distance Model ($s = 2$)

Therefore, for the squared distance model we have the following

$$\theta(\mathbf{x}) = -d^2(\mathbf{x}_i^\top\mathbf{B},\mathbf{v}_a) + d^2(\mathbf{x}_i^\top\mathbf{B},\mathbf{v}_b) + d^2(\mathbf{x}_i^\top\mathbf{B},\mathbf{v}_c) - d^2(\mathbf{x}_i^\top\mathbf{B},\mathbf{v}_d)$$

$$\theta(\mathbf{x}) = -\sum_m(u_{im} - v_{am})^2 + \sum_m(u_{im} - v_{bm})^2 + \sum_m(u_{im} - v_{cm})^2 - \sum_m(u_{im} - v_{dm})^2$$

$$\theta(\mathbf{x}) = -\sum_m (u_{im}^2 - 2u_{im}v_{am} + v_{am}^2) + \sum_m (u_{im}^2 - 2u_{im}v_{bm} + v_{bm}^2) + \sum_m (u_{im}^2 - 2u_{im}v_{cm} + v_{cm}^2) - \sum_m (u_{im}^2 - 2u_{im}v_{dm} + v_{dm}^2)$$

$$\theta(\mathbf{x}) = \sum_m -u_{im}^2 + 2u_{im}v_{am} - v_{am}^2 + u_{im}^2 - 2u_{im}v_{bm} + v_{bm}^2 + u_{im}^2 - 2u_{im}v_{cm} + v_{cm}^2 - u_{im}^2 + 2u_{im}v_{dm} - v_{dm}^2$$

$$\theta(\mathbf{x}) = \sum_m 2u_{im}(v_{am} - v_{bm} - v_{cm} + v_{dm}) - v_{am}^2 + v_{bm}^2 + v_{cm}^2 - v_{dm}^2$$

Now, we can make use of the additive definitions of the coordinates for the subject points and the class points. That is $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + \gamma_{2m} + \gamma_{3m}$, and $v_{bm} = v_{\bullet m} + \gamma_{3m}$, $v_{cm} = v_{\bullet m} + \gamma_{2m}$, $v_{dm} = v_{\bullet m}$. Therefore,

$$\begin{aligned}\theta(\mathbf{x}) = \sum_m &(2u_{\bullet m} + x_{i2}b_{2m})(v_{\bullet m} + \gamma_{2m} + \gamma_{3m} - (v_{\bullet m} + \gamma_{3m}) - (v_{\bullet m} + \gamma_{2m}) + v_{\bullet m}) - \\ &(v_{\bullet m} + \gamma_{2m} + \gamma_{3m})^2 + (v_{\bullet m} + \gamma_{3m})^2 + (v_{\bullet m} + \gamma_{2m})^2 - v_{\bullet m}^2\end{aligned}$$

$$\begin{aligned}\theta(\mathbf{x}) = \sum_m &(2u_{\bullet m} + x_{i2}b_{2m})(v_{\bullet m} + \gamma_{2m} + \gamma_{3m} - v_{\bullet m} - \gamma_{3m} - v_{\bullet m} - \gamma_{2m} + v_{\bullet m}) \\ &-(v_{\bullet m} + \gamma_{2m} + \gamma_{3m})^2 + (v_{\bullet m} + \gamma_{3m})^2 + (v_{\bullet m} + \gamma_{2m})^2 - v_{\bullet m}^2\end{aligned}$$

$$\theta(\mathbf{x}) = \sum_m (2u_{\bullet m} + x_{i2}b_{2m})(2v_{\bullet m} - 2v_{\bullet m} + \gamma_{2m} - \gamma_{2m} + \gamma_{3m} - \gamma_{3m}) - (v_{\bullet m} + \gamma_{2m} + \gamma_{3m})^2 + (v_{\bullet m} + \gamma_{3m})^2 + (v_{\bullet m} + \gamma_{2m})^2 - v_{\bullet m}^2$$

$$\theta(\mathbf{x}) = \sum_m -(v_{\bullet m} + \gamma_{2m} + \gamma_{3m})^2 + v_{\bullet m}^2 + 2v_{\bullet m}\gamma_{3m} + \gamma_{3m}^2 + v_{\bullet m}^2 + 2v_{\bullet m}\gamma_{2m} + \gamma_{2m}^2 - v_{\bullet m}^2$$

$$\theta(\mathbf{x}) = \sum_m -(v_{\bullet m} + \gamma_{2m} + \gamma_{3m})^2 + v_{\bullet m}^2 + 2v_{\bullet m}\gamma_{3m} + \gamma_{3m}^2 + 2v_{\bullet m}\gamma_{2m} + \gamma_{2m}^2$$

$$\theta(\mathbf{x}) = \sum_m -(v_{\bullet m}^2 + \gamma_{2m}^2 + \gamma_{3m}^2 + 2v_{\bullet m}\gamma_{2m} + 2v_{\bullet m}\gamma_{3m} + 2\gamma_{2m}\gamma_{3m}) + v_{\bullet m}^2 + 2v_{\bullet m}\gamma_{3m} + \gamma_{3m}^2 + 2v_{\bullet m}\gamma_{2m} + \gamma_{2m}^2$$

$$\theta(\mathbf{x}) = \sum_m -v_{\bullet m}^2 - \gamma_{2m}^2 - \gamma_{3m}^2 - 2v_{\bullet m}\gamma_{2m} - 2v_{\bullet m}\gamma_{3m} - 2\gamma_{2m}\gamma_{3m} + v_{\bullet m}^2 + 2v_{\bullet m}\gamma_{3m} + \gamma_{3m}^2 + 2v_{\bullet m}\gamma_{2m} + \gamma_{2m}^2$$

$$\theta(\mathbf{x}) = \sum_m -2\gamma_{2m}\gamma_{3m}$$

The final term does not depend on x_i, that is, the log odds ratio for the four categories a, b, c, and d is identical at each level of x.

Distance Model ($s = 1$)

Therefore, for the distance model we have the following

$$\theta(\mathbf{x}) = -d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_a) + d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_b) + d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_c) - d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_d)$$

$$\theta(\mathbf{x}) = -\sqrt{\sum_m (u_{im} - v_{am})^2} + \sqrt{\sum_m (u_{im} - v_{bm})^2} + \sqrt{\sum_m (u_{im} - v_{cm})^2} - \sqrt{\sum_m (u_{im} - v_{dm})^2}$$

We like to compare four classes, say class a, b, c and d that only differ in two response variables (GAD and MDD). Category a represents a person with MDD and GAD, then we have $z_{c2} = 1$ and $z_{c3} = 1$. For category b, without MDD and with GAD, $z_{c2} = 0$ and $z_{c3} = 1$. For category c, with MDD and without GAD, we have $z_{c2} = 1$ and $z_{c3} = 0$. Lastly, for category d without MDD and GAD we have $z_{c2} = 0$ and $z_{c3} = 0$.

Now, we can make use of the additive definitions of the coordinates for the subject points and the class points. That is $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + \gamma_{2m} + \gamma_{3m}$, and $v_{bm} = v_{\bullet m} + \gamma_{3m}$, $v_{cm} = v_{\bullet m} + \gamma_{2m}$, $v_{dm} = v_{\bullet m}$. Therefore,

$$\theta(\mathbf{x}) = -\sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - \gamma_{2m} - \gamma_{3m})^2} + \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - \gamma_{3m})^2} +$$
$$\sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m} - \gamma_{2m})^2} - \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m})^2}$$

This equation cannot be further simplified.

Associations Model

Similarly to the log odds, we also have an equation for the class points $\mathbf{v}_c$ of any class c. When we are focusing on the associations model we have for example

$$v_{cm} = \gamma_{0m} + \sum_r z_{cr}\gamma_{rm} + \sum_r \sum_{t>r} z_{ct}z_{cr}\gamma_{trm}.$$

When we would like to focus on the log odds ratio of the third response variable (say, Generalized Anxiety Disorder) and the second response variable (Major Depressive Disorder), we may write

$$v_{cm} = \gamma_{0m} + \sum_{r \neq \{2,3\}} (z_{cr}\gamma_{rm}) + z_{c3}\gamma_{3m} + z_{c2}\gamma_{2m} + \sum_{r \neq \{2,3\}} \sum_{t>r} z_{ct}z_{cr}\gamma_{trm} + \sum_{t \neq 2} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s \neq \{2,3\}} z_{cs}z_{c3}\gamma_{s3m}$$

$$v_{cm} = v_{\bullet m} + z_{c3}\gamma_{3m} + z_{c2}\gamma_{2m} + \sum_{t \neq \{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s \neq \{2,3\}} z_{cs}z_{c3}\gamma_{s3m} + z_{c2}z_{c3}\gamma_{10m}$$

where $v_{\bullet m} = \gamma_{0m} + \sum_{r \neq \{2,3\}} z_{cr}\gamma_{rm} + \sum_{r \neq \{2,3\}} \sum_{s>r} z_{cs}z_{cr}\gamma_{trm}$. We like to compare four classes, say class a, b, c and d that only differ in two response variables (GAD and MDD). Category a represents a person with MDD and GAD, then we have $z_{c2} = 1$ and $z_{c3} = 1$. For category b, without MDD and with GAD, $z_{c2} = 0$ and $z_{c3} = 1$. For category c, with MDD and without GAD, we have $z_{c2} = 1$ and $z_{c3} = 0$. Lastly, for category d without MDD and GAD we have $z_{c2} = 0$ and $z_{c3} = 0$.

Squared Distance Model ($s = 2$)

Therefore, for the squared distance model we have the following

$$\theta(\mathbf{x}) = -d^2(\mathbf{x}_i^\top\mathbf{B}, \mathbf{v}_a) + d^2(\mathbf{x}_i^\top\mathbf{B}, \mathbf{v}_b) + d^2(\mathbf{x}_i^\top\mathbf{B}, \mathbf{v}_c) - d^2(\mathbf{x}_i^\top\mathbf{B}, \mathbf{v}_d)$$

which can be simplified into

$$\theta(\mathbf{x}) = \sum_m 2u_{im}(v_{am} - v_{bm} - v_{cm} + v_{dm}) - v_{am}^2 + v_{bm}^2 + v_{cm}^2 - v_{dm}^2$$

Making use of the additive definitions of the coordinates for the subject points and the class points, in which $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + \gamma_{3m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m} + z_{c2}z_{c3}\gamma_{10m}$, and $v_{bm} = v_{\bullet m} + \gamma_{3m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m}$, $v_{cm} = v_{\bullet m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m}$, $v_{dm} = v_{\bullet m}$. Therefore

$$\begin{aligned}\theta(\mathbf{x}) = \sum_m 2(u_{\bullet m} + x_{i2}b_{2m})(v_{\bullet m} + \gamma_{3m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m} + z_{c2}z_{c3}\gamma_{10m}) - \\ (v_{\bullet m} + \gamma_{3m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m}) - (v_{\bullet m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m}) + v_{\bullet m}) - \\ (v_{\bullet m} + \gamma_{3m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m})^2 + \\ (v_{\bullet m} + \gamma_{3m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m})^2 + (v_{\bullet m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m})^2 - v_{\bullet m}^2\end{aligned}$$

which can be simplified into

$$\begin{aligned}\theta(\mathbf{x}) = \sum_m 2(u_{\bullet m} + x_{i2}b_{2m})(z_{c2}z_{c3}\gamma_{10m}) - (v_{\bullet m} + \gamma_{3m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m})^2 + \\ (v_{\bullet m} + \gamma_{3m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m})^2 + (v_{\bullet m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m})^2 - v_{\bullet m}^2\end{aligned}$$

Now, if we compare person 1 and person 2, that differ 1 unit on predictor variable 2, we get

$$\begin{aligned}\theta(\mathbf{x_1}) = \sum_m 2(u_{\bullet m} + qb_{2m})(z_{c2}z_{c3}\gamma_{10m}) - (v_{\bullet m} + \gamma_{3m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m})^2 + \\ (v_{\bullet m} + \gamma_{3m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m})^2 + (v_{\bullet m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m})^2 - v_{\bullet m}^2\end{aligned}$$

$$\begin{aligned}\theta(\mathbf{x_2}) = \sum_m 2(u_{\bullet m} + (q+1)b_{2m})(z_{c2}z_{c3}\gamma_{10m}) - (v_{\bullet m} + \gamma_{3m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m})^2 + \\ (v_{\bullet m} + \gamma_{3m} + \sum_{s\neq\{2,3\}} z_{cs}z_{c3}\gamma_{s3m})^2 + (v_{\bullet m} + \gamma_{2m} + \sum_{t\neq\{2,3\}} z_{ct}z_{c2}\gamma_{t2m})^2 - v_{\bullet m}^2\end{aligned}$$

Their difference ($\lambda(\mathbf{x}_2) - \lambda(\mathbf{x}_1)$) shows that, with every unit change in x_2 the log odds ratio changes with

$$2\sum_m b_{2m}z_{c2}z_{c3}\gamma_{10m}$$

Distance Model ($s = 1$)

Therefore, for the squared distance model we have the following

$$\theta(\mathbf{x}) = -d^1(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_a) + d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_b) + d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_c) - d^2(\mathbf{x}_i^\top \mathbf{B}, \mathbf{v}_d)$$

$$\theta(\mathbf{x}) = -\sqrt{\sum_m (u_{im} - v_{am})^2} + \sqrt{\sum_m (u_{im} - v_{bm})^2} + \sqrt{\sum_m (u_{im} - v_{cm})^2} - \sqrt{\sum_m (u_{im} - v_{dm})^2}$$

Making use of the additive definitions of the coordinates for the subject points and the class points, in which $u_{im} = u_{\bullet m} + x_{i2}b_{2m}$, $v_{am} = v_{\bullet m} + \gamma_{3m} + \gamma_{2m} + \sum_{t \neq \{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s \neq \{2,3\}} z_{cs}z_{c3}\gamma_{s3m} + z_{c2}z_{c3}\gamma_{10m}$, and $v_{bm} = v_{\bullet m} + \gamma_{3m} + \sum_{s \neq \{2,3\}} z_{cs}z_{c3}\gamma_{s3m}$, $v_{cm} = v_{\bullet m} + \gamma_{2m} + \sum_{t \neq \{2,3\}} z_{ct}z_{c2}\gamma_{t2m}$, $v_{dm} = v_{\bullet m}$. Therefore

$$\theta(\mathbf{x}) = -\sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - (v_{\bullet m} + \gamma_{3m} + \gamma_{2m} + \sum_{t \neq \{2,3\}} z_{ct}z_{c2}\gamma_{t2m} + \sum_{s \neq \{2,3\}} z_{cs}z_{c3}\gamma_{s3m} + z_{c2}z_{c3}\gamma_{10m}))^2} +$$

$$\sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - (v_{\bullet m} + \gamma_{3m} + \sum_{s \neq \{2,3\}} z_{cs}z_{c3}\gamma_{s3m}))^2} +$$

$$\sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - (v_{\bullet m} + \gamma_{2m} + \sum_{t \neq \{2,3\}} z_{ct}z_{c2}\gamma_{t2m}))^2} - \sqrt{\sum_m (u_{\bullet m} + x_{i2}b_{2m} - v_{\bullet m})^2}$$

This equation cannot be further simplified.