



University of Abomey-Calavi

Faculty of Agronomic Sciences



Laboratoire de Biomathématiques &
d'Estimations Forestières



Relative performance of popular coding structures for contrasts analysis

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Statistics, major Biostatistics

Presented and defended by

Baro Nadejda SERO

Supervisor

Prof. Romain Lucas GLELE KAKAI

Co-supervisor

Dr. Micheline GBEHA

Members of the jury

President: Prof. Carlos OGOUYANDJOU

Supervisor: Prof. Romain L. GLELE KAKAI

Examiner: Prof. Orou G. GAOUE

Funded by the RUFORUM Research Grant - RU 2015 GRG – 135



January 2018

Table of Contents

Certification	i
Dedication	ii
Acknowledgements.....	iii
List of tables.....	iv
List of figures.....	v
List of abbreviations	vi
Abstract	vii
Résumé.....	viii
Introduction.....	1
1 Background.....	3
1.1 Generalities about contrasts	3
1.1.1 Definition	3
1.1.2 Planned and unplanned contrasts	3
1.1.3 Orthogonal and non-orthogonal contrasts.....	4
1.2 Grand mean and Sample mean.....	4
1.3 Selection of the coding structure	5
1.4 Correction of Type I error inflation.....	7
1.4.1 Šidàk correction	7
1.4.2 Bonferroni correction.....	7
1.4.3 Monte-Carlo technique correction	8
2 On the use and interpretation of the popular coding structures in R statistical software	9
2.1 Data and methods	9
2.2 Coding structures for non-trend contrasts	9
2.2.1 Dummy coding.....	9
2.2.2 Effect coding	11
2.2.3 Contrast coding	12
2.2.4 Weighted effect coding	13
2.3 Coding structure for trend contrasts or polynomial contrasts.....	15
3 Simulation study	16
3.1 Simulation design.....	16

3.1.1	Factors of interest.....	16
3.1.2	Simulation plan	17
3.1.3	Data analysis	19
3.2	Results	21
3.2.1	Per-contrast relative performance of the coding structures	21
3.2.2	Overall relative performance of the coding structures.....	23
3.3	Discussion	27
4	Application in Ecology	28
	Conclusion	29
	References.....	30
	Appendices.....	31
	Appendix 1. Logistic regression odds ratio for dummy coding with balanced data.....	31
	Appendix 2. Logistic regression odds ratio for dummy coding with unbalanced data.....	31
	Appendix 3. Logistic regression odds ratio for effect coding	32
	Appendix 4. Logistic regression odds ratio for reverse helmert	32
	Appendix 5. Coefficients of gamma models for dummy coding with balanced data	33
	Appendix 6. Coefficients of gamma models for dummy coding with unbalanced data	33
	Appendix 7. Coefficients of gamma models for effect coding	34
	Appendix 8. Coefficients of gamma models for reverse helmert coding.....	34
	Appendix 8. Coefficients of gamma models for weighted effect coding.....	35

Certification

I hereby certify that the master thesis entitled “**Relative performance of some popular categorical predictors’ coding structures for contrasts analysis**” is a record of independent research work achieved by **Baro Nadejda SERO** (Reg.No: 11055412) under my supervision and submitted to the University of Abomey-Calavi (UAC), Faculty of Agronomic Science (FSA) in partial fulfilment for the award of the Master degree in Statistics, major Biostatistics.

Supervisor

Prof. Romain L. GLELE KAKAÏ
Biostatistics and Forest Modeling, FSA/UAC

Dedication

To
my life-coach, God,
my beloved parents,
Jean-Pierre SERO & Florence DOKO.

Acknowledgements

To my first scientific coach, my supervisor Prof. Romain L. GLELE KAKAÏ, Director of the Laboratoire de Biomathématiques et d'Estimations Forestières. Please, receive my deepest gratitude without any reserve for your patience, letting me explore research and sowing the love for statistics and programming.

I would also like to thank the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) for the financial and technical support provided and all the team that worked on this research project especially Prof. Achille ASSOGBADJO, Dr. Valère SALAKO, Dr. Rodrigue IDOHOU and Guillaume HOUNSOU-DINDIN.

A very special gratitude goes out to Prof. Orou GAOUE, Wilfried BONOU, Frederic TOVISSODE and Donald AKOGOU who have provided assistance, guidance, suggestions, corrections and light throughout the conception and also for solving the many programming issues encountered.

With a special mention to Eugide AHIVODJI, Lévis, Audrey and Ghislaine ALLE, Candide, Gladstone and Grace HOUNGNIBO, Mouckaïla SALOUF for providing moral, emotional, financial, technical and medical help throughout tough and good times.

My eternal cheerleader, mummy, sister and friend Benedicta ESSUON: thank you for the everyday.

And finally, last but by no means least, to all my close friends, whom I consider my second family for caring and supporting me all these years.

Thanks for all your encouragement!

List of tables

Table 1. Summary of the most common categories of contrast coding	12
Table 2. Summary of the factors to be tested	16
Table 3. Levels ratio for unbalanced data according to the number of levels.....	17
Table 4. Values assigned by default to the parameter β considered for all the coding structures.	18
Table 5. Levels of confidence used per comparison	18
Table 6. Coding structures under investigation in R.....	19
Table 7. Terms of models and their meaning for all the tables of the section Results.....	21
Table 8. ANOVA on a logistic model to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc) on the occurrence of performance of the coding structures. Only probability values are recorded in the table	22
Table 9. ANOVA on a gamma model to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc) on the per-contrast performance of the coding structures. Only probability values are recorded in the table.	22
Table 10. ANOVA on a logistic model to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc) on the overall performance of the coding structures. Only probability values are recorded in the table	23
Table 11. Logistic regression model for the overall performance of Dummy coding for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc).....	24
Table 12. Logistic regression model for the overall performance of Dummy coding for unbalanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc).....	24
Table 13. Logistic regression model for the overall performance of Effect coding for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc).....	25
Table 14. Logistic regression model for the overall performance of Reverse Helmert contrasts for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc)	26
Table 15. Logistic regression model for the overall performance of Weighted effect coding for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc)	26
Table 16. Linear model to test the effects of the compost dose on the height of seedlings	28
Table 17. Logistic regression model for the per-contrast performance of dummy coding for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc)	31

Table 18. Logistic regression model for the per-contrast performance of dummy coding for unbalanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc).....	31
Table 19. Logistic regression model for the per-contrast performance of effect coding for unbalanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc).....	32
Table 20. Logistic regression model for the per-contrast performance of reverse helmert coding for unbalanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)	32
Table 21. GLM with gamma structure for errors to test the effects effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of the dummy coding (balanced design).	33
Table 22. GLM with gamma structure for errors to test the effects of size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of dummy coding (unbalanced design).	33
Table 23. GLM with gamma structure for errors to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of effect coding.	34
Table 24. GLM with gamma structure for errors to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of reverse helmert.	34
Table 25. GLM with gamma structure for errors test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of weighted effect coding.	35

List of figures

Figure 1. Decision tree for determining the appropriate coding structure (adapted from Davis, 2010)... 6

List of abbreviations

WEC	: Weighted effect coding
EC	: Effect coding
DC	: Dummy coding
RH	: Reverse Helmert
ANOVA	: Analysis of Variance
GLM	: Generalized linear model
P	: Probability value
E	: Effect size
n	: Sample size
p	: Number of levels of the factor
mu	: Mean of the distribution of the response variable
mc	: Method of correction for type I error inflation
df	: Degree of freedom

Abstract

Coding structures are employed to allow the use of categorical variables in multiple regression analysis and other more sophisticated models. The first part of this paper provides a basic guide for coding schemes implementation R statistical software. In fact, building the popular coding structures in R is easily realized using some embedded functions such as *contr.treatment*, *contr.sum*, *contr.helmert*, *contr.wec* and *contr.poly* for dummy, effect, reverse helmert, weighted effect coding and polynomial contrasts respectively. Also, the interpretation of their output is straightforward and only requires knowing the type of comparison being realized, the coefficients of assignment and their sign. On the other hand, the second part of this work evaluated the relative performance of the popular coding structures: dummy, effect, reverse helmert, weighted effect coding, using a Monte-Carlo simulation. The effects of the effect size, the sample size, the number of levels of the factor, the type of distribution of the response variable (normality against moderate non-normality) and the correction method for Type I error inflation were checked using a per-contrast (on individual contrasts) performance criterion and an overall (all the contrasts simultaneously) performance criterion. Simulations revealed that the correction method used for Type I error inflation had no effects on dummy, effect, reverse helmert and weighted effect coding per-contrast performance. This performance was only affected by the effect size, the number of levels and the type of distribution for all the structures. Furthermore, the overall performance of dummy, effect, reverse helmert and weighted effect coding was varying in function of the effect size, the number of levels and the type of the distribution. The correction method had a very slight effect on weighted effect coding and no influence on the other schemes. Globally, the correction method has no influence on the coding schemes performance while the effect size greatly affected the performance. The performance of these techniques was also associated with data following a normal distribution. No specific pattern was portrayed for the number of levels and the sample size. In spite of the fact that all these coding techniques do not imply the same type of comparisons and do not have the same internal structure, weighted effect coding was the least influenced as compared to the others.

Key words: contrasts, dummy coding, effect coding, weighted effect coding, simulation, performance

Résumé

Les structures de codification sont utilisées afin de permettre l'inclusion des variables catégorielles pour la régression multiple et d'autres modèles plus sophistiqués. La première partie de ce document vise à fournir un guide basique à l'utilisation des structures de codification dans le logiciel statistique R. En effet, la définition des structures populaires de codification dans R est facilement réalisée en utilisant certaines fonctions intégrées telles que *contr.treatment*, *contr.sum*, *contr.helmert*, *contr.wec* et *contr.poly* respectivement pour les codifications dummy, de l'effet, d'helmert inversée, des effets pondérés et des contrastes polynômiaux. De même, l'interprétation de leurs résultats est directe et ne nécessite que la connaissance du type de comparaison réalisé, de la technique d'affectation des coefficients et leurs signes. La seconde partie quant à elle a évalué la performance relative des techniques de codification dummy, de l'effet, d'helmert inversée et des effets pondérés, en utilisant une simulation de Monte-carlo. L'influence de la taille de l'effet, de la taille de l'échantillon, du nombre de niveaux du facteur, du type de distribution de la variable réponse (normalité contre non normalité) et de la méthode de correction pour l'inflation de l'erreur de Type I a été apprécié à l'aide du critère de performance par contraste (contrastes individuels) et du critère de performance globale (tous les contrastes simultanément). Les simulations ont révélé que la méthode de correction utilisée pour l'inflation de l'erreur de type I n'avait aucun effet sur les performances par contraste des schémas de codification dummy, de l'effet, d'helmert inversée et des effets pondérés. Cette performance n'était affectée que par la taille de l'effet, le nombre de niveaux et le type de distribution pour toutes les structures. De plus, la performance globale des codifications dummy, de l'effet, d'helmert inversée et des effets pondérés variait en fonction de la taille de l'effet, du nombre de niveaux et du type de distribution. La méthode de correction avait un effet très léger sur la structure des effets pondérés et aucune influence sur les autres schémas. Globalement, la méthode de correction n'avait aucun effet sur les techniques de codification alors que la taille de l'effet influençait grandement la performance. La performance de ces structures était aussi associée aux données suivant une distribution normale. Aucun modèle particulier n'a été décrit pour le nombre de niveaux et la taille de l'échantillon. En dépit du fait que toutes ces techniques de codification n'impliquent pas les mêmes types de comparaisons et n'ont pas la même structure interne, la structure de codification des effets pondérés était la moins influencée par rapport aux autres.

Mots clés : contrastes, codification dummy, codification de l'effet, codification des effets pondérés, simulation, performance

Introduction

In many fields of science, economy and arts, the most frequent research question has been about identifying the effect of a factor on an outcome or variable of interest. The first statistical method widely used to address that question has been the standard analysis of variance (ANOVA) which provides the Fisher's statistic (F) that reflects all possible differences between the means of the groups analyzed (Abdi, 2010). Another famous approach is multiple regression analysis (MRA) which has spread over the last decades because of its ability of integrating a large set of ANOVA models (Davis, 2010; Wendorf, 2004). However, including a categorical factor in an MRA requires the use of coding structures to insure the interpretability of the results (Wendorf, 2004). Once the primary question of determining a factor's impact on an outcome is solved, the next problem is about how and where these effects arise. This new issue has found a useful tool in the multiple comparison methods (MCMs), which are used to investigate differences between pairs of population means or, more generally, between subsets of population means using sample data (Rafter, 2002). A very common MCM related to the coding structures required for an MRA is contrasts analysis. Therefore, coding schemes have been linked to ANOVA, MRA, generalized linear models and linear mixed effects models. The simple fact of using these statistical models implies using coding schemes or contrasts.

The notion of contrasts is often assimilated to comparisons, since a contrast can be defined as a linear combination of coefficients that allow to make comparisons or look for a trend in means (Rafter, 2002; Davis, 2010). Various fields have been using these contrasts techniques. Indeed, in pharmaceutical studies, treatment-control contrasts have been used to find the optimal number of replications and optimal experimental design (Majumdar, 2016). Meanwhile, in drug development and dose response studies; pairwise and Helmert contrasts were involved in the identification of the minimum effective dose (MED, the lowest dose level with an effect that exceeds that of the zero dose control) and allowed the comparison of many active treatments with one control treatment (Show-Li, 2007). A study, in designs of experiment with linear mixed effects models, has compared several methods for constructing a confidence interval on contrasts of fixed effects in a balanced three-factor mixed factorial design with one fixed effect and two random effects (Coombs, 2003).

Nonetheless, the use of coding schemes, their interpretations and the pattern of their selection have been fairly reviewed and discussed (Davis, 2010; Thompson, 2006). Some literature has been available to help select the desired contrasts. However, contrasts behavior in front of some statistical parameters like sample size, the non-normality, the number of levels of the factors and even the correction method of Type I error inflation are also important issues for a concerned researcher. This issue is poorly discussed in the available literature on contrasts. Actually, sample size is a parameter which affects the power of tests and varies from one experiment to another depending on aims, means and research area. Moreover, normality is one of the assumptions needed for contrasts analysis (Rafter, 2002). This suggests that non-normality could affect contrasts performance. Plus, the factors considered can involve many levels to compare. Thus, many comparisons involve a possible inflation of Type I error. This latter is known to be an

historical issue for MCMs (Abdi, 2007). Though, the literature suggests and offers a few approaches for correcting this problem. Three methods have been suggested: Šidák equation, Bonferroni equation and Monte Carlo approach (Abdi & Williams, 2010). Besides, modeling implies determining the effect of a categorical predictor on the response variable which implicate means differences. This suggests that the size or scale of these effects could impact the comparisons made. These key points would allow a better understanding of these structures and provide guidance on which coding structure to use regarding the sample size and the number of levels of factors considered.

The present document is the result of a two years training in a Master of Science in Statistics, major Biostatistics. It is subdivided in two main parts. The first component aims at providing guidance on how to define and interpret the common coding structures available in R statistical software. Specifically, it gives the basic knowledge for understanding the techniques, how to choose the suitable one for data, use R embedded functions and packages to apply them and interpret the output. Meanwhile, the main objective of the second is to assess the relative performance of some popular coding structures for contrasts analysis regarding the sample size, the number of levels of factors, the method of correction for Type I error inflation, the effect size and the distribution of the response variable (moderate non-normality and normality). Its specific objectives are (1) assess the relative per-contrast performance of dummy, effect and weighted effect coding structures, polynomial and reverse helmert contrasts in terms of the sample size, the number of levels of factors, the method of correction for Type I error inflation, the effect size and the distribution of the response variable (moderate non-normality and normality) in case of unbalanced or balanced design; and (2) assess the overall performance of dummy, effect and weighted effect coding structures, polynomial and reverse helmert contrasts in terms of the sample size, the number of levels of factors, the method of correction for Type I error inflation, the effect size and the distribution of the response variable (moderate non-normality and normality) in case of unbalanced or balanced design.

1 Background

1.1 Generalities about contrasts

1.1.1 Definition

A contrast is a linear combination of coefficients that allows to make comparisons or to look for a trend in means (Rafter, 2002; Davis, 2010). For a factor with p number of levels, the number of contrasts is $p-1$. There exist two types: trend and non-trend contrasts. Non-trend contrasts investigate the differences among groups or levels of predictors. This first group includes: dummy, effect, weighted effect and contrast coding. Conversely, trend contrasts seek for a trend of the means or how the means react across levels of the predictor.

Example 1. For a factor representing groups of people with four levels, contrasts can be designed as follows:

Levels	Contrast coefficients		
	C1	C2	C3
Group 1	-1	-1	3
Group 2	-1	0	-1
Group 3	0	1	-1
Group 4	2	0	-1

Generally, coefficients are either negative or positive and taking integers as coefficients. However, they can be decimals according to the coding structure used (for examples refer to weighted effect coding and polynomial contrasts). The groups having the same sign are in the same chapter or bigger group. The group having 0 as coefficient is not considered in the comparison. Here, C1 code for a comparison between the mean of group 4 and the mean of groups 1 and 2. C2 represents the comparison between the first two groups whereas C3 compares the first group to the other groups.

1.1.2 Planned and unplanned contrasts

Contrasts can be grouped into planned or *a priori* contrasts and unplanned or post hoc contrasts (Abdi & Williams, 2010). *A priori* or planned contrasts are selected before running the experiment. Generally, they reflect the few hypotheses the experimenter wants to test and are defined based on literature review. However, *a posteriori* or unplanned contrasts are decided after the experiment has been run. Characteristic of a virgin field investigation, they represent all the possible contrasts that can be made explicitly or not and ensure that unexpected results are reliable available (Davis, 2010; Thompson 2006; Abdi & Williams, 2010).

1.1.3 Orthogonal and non-orthogonal contrasts

Orthogonal contrasts are uncorrelated contrasts which are individually interpretable. Two contrasts are orthogonal or independent if the sum of the products of their coefficients is null. The main advantage is that each linear contrast of a set of orthogonal contrasts can be analyzed independently from the others. Also, the fact that they are uncorrelated helps minimizing the type I error rate.

The opposite holds for non-orthogonal contrasts which are correlated and messy to interpret. They are more complex and assessing the importance of a given contrast conjointly with the other contrasts is the main issue. Generally, orthogonal contrasts are preferred.

Example 2. Considering the example above (Example 1), the orthogonality of contrasts can be checked as follows

Levels	Contrast coefficients			Product		
	C1	C2	C3	C1*C2	C1*C3	C2*C3
Group 1	-1	-1	3	1	-3	-3
Group 2	-1	0	-1	0	1	0
Group 3	0	1	-1	0	0	-1
Group 4	2	0	-1	0	-2	0
Sum	0	0	0	1	-4	-4

The sum of the products of each pair of contrasts is not null hence the contrasts involved are not orthogonal.

1.2 Grand mean and Sample mean

The grand mean or unweighted mean is the mean of all the means of the levels of a given factor. It is computed for balanced designs and used as the intercept for effect and contrast coding structures. Conversely, the sample mean or weighted mean, used for unbalanced designs, is the mean which is obtained by dividing the sum of all the observations by the sample size.

$$Grand\ Mean = \frac{m_1 + m_2 + \dots + m_p}{p} \quad (Eq. 1)$$

Where p is the number of groups or levels of the factor and m_1 to m_p the means of the groups 1 to p respectively.

$$Sample\ Mean = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (Eq. 2)$$

Where n is the number of observations and $x_1 \dots x_n$ the observations of the unbalanced data.

Example 3. For a factor with four levels, some measurements are recorded for each group subject (G1 to G4). The two types of means are computed as follows:

$$\text{Grand Mean} = \frac{4.636 + 4.909 + 2.273 + 5.364}{4} = 4.295$$

$$\text{Sample Mean} = \frac{7 + 9 \dots + 4 + 0 + 9 \dots + 4 + 4 + 8 \dots 5 + 10 + 1}{25} = \frac{96}{24} = 4$$

	Balanced design				Unbalanced design			
Observations	G1	G2	G3	G4	G1	G2	G3	G4
1	7	0	4	10	7	0	4	10
2	9	9	8	1	9	9	8	1
3	2	5	1	9	2	5	1	-
4	1	9	0	2	1	9	0	-
5	1	4	1	2	1	4	1	-
6	2	1	0	10	2	-	0	-
7	4	5	0	4	4	-	0	-
8	9	7	5	6	9	-	5	-
9	4	4	2	8	4	-	-	-
10	6	7	3	6	-	-	-	-
11	6	3	1	1	-	-	-	-
Mean	4.636	4.909	2.273	5.364	4.333	5.400	2.375	4.500
Grand mean	4.295				-			
Sample mean	-				4			

1.3 Selection of the coding structure

The popular contrasts coding schemes are: dummy coding, effect coding, contrast coding, polynomial contrasts and weighted effect coding. They are detailed in the next section of this document. A decision tree has been proposed for selecting the type of contrasts to use (Figure 1). This tree decision from Davis (2010) has been modified to add weighted effect coding which has recently been developed (Grotenhuis, 2016). It is essential to consider the literature available on the research topic, decide if we are interested in trends or means comparisons and construct the comparisons to verify (Figure 1). Plus, it is vital to the researcher to know which questions should be answered.

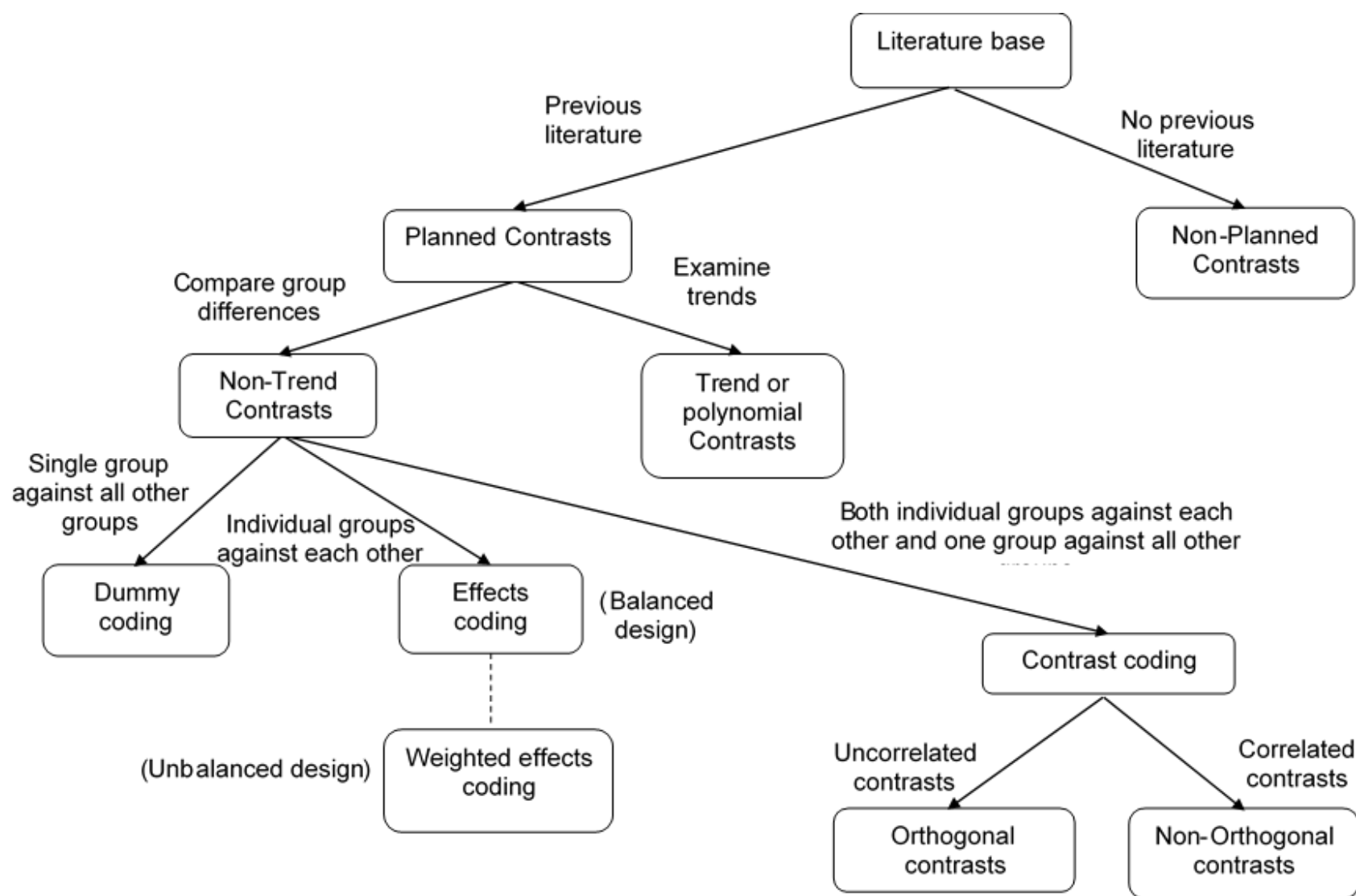


Figure 1. Decision tree for determining the appropriate coding structure (adapted from Davis, 2010)

1.4 Correction of Type I error inflation

The Type I error is the probability of the event “reject the null hypothesis when it is true”. Generally, this probability is called *level of significance* or α (*alpha*) and set to 0.05 or 5 %. This event is assumed to be rare. A large number of tests on the same dataset increase its occurrence one could find significance where there is not. This phenomenon is called inflation. Some correction methods have been developed to solve this problem. Two different meanings can be assigned to α (Abdi, 2007):

- the probability of making a Type I error for one given test, testwise alpha or alpha per test denoted $\alpha[PT]$;
- the probability of making at least one Type I error for a family of tests, familywise or experimentwise alpha or alpha per family of tests represented by $\alpha[PF]$.

The probability of making at least one Type I error for a family of c contrasts is given by:

$$\alpha[PF] = 1 - (1 - \alpha[PT])^c \quad (\text{Eq.3})$$

Where C is the number of contrasts (number of levels of the factor - 1). Some correction methods for independent tests are briefly presented below.

1.4.1 Šidák correction

It shows that $\alpha[PT]$ values must be adapted for each test to reach a given $\alpha[PF]$. Its formula is derived from the previous formula (Eq.1).

- For dependent tests (non-orthogonal contrasts)

$$\alpha[PF] \leq 1 - (1 - \alpha[PT])^c \quad (\text{Eq.4})$$

- For independent tests (orthogonal contrasts)

$$\alpha[PT] = 1 - (1 - \alpha[PF])^{1/c} \quad (\text{Eq.5})$$

1.4.2 Bonferroni correction

This formula also is a simpler approximation of the Šidák equation. Generally, the two methods yield close values. Bonferroni approximation is given by:

- For dependent tests (non-orthogonal contrasts)

$$\alpha[PF] < c \alpha[PT] \quad (\text{Eq.6})$$

- For independent tests (orthogonal contrasts)

$$\alpha[PT] \approx \frac{\alpha[PF]}{c} \quad (\text{Eq.7})$$

Bonferroni and Šidák equations are linked to each other by the inequality given as:

- For dependent tests (non-orthogonal contrasts)

$$\alpha[PF] \leq 1 - (1 - \alpha[PT])^c < c \alpha[PT] \quad (\text{Eq.8})$$

- For independent tests (orthogonal contrasts)

$$\alpha[PT] = 1 - (1 - \alpha[PF])^{\frac{1}{c}} \geq \alpha[PT] \approx \frac{\alpha[PF]}{c} \quad (\text{Eq.9})$$

1.4.3 Monte-Carlo technique correction

This technique consists of running a simulated experiment many times using random data, with the aim of obtaining a pattern of results showing what would happen just on the basis of chance. The formulas used here are:

$$\alpha[PT] = \frac{\text{number of contrasts having reached significance}}{\text{total number of contrasts}} \quad (\text{Eq.10})$$

$$\alpha[PF] = \frac{\text{number of families with at least 1 Type I error}}{\text{total number of families}} \quad (\text{Eq.11})$$

2 On the use and interpretation of the popular coding structures in R statistical software

2.1 Data and methods

Three datasets are used in this section. The first dataset, *bcmort*, from the package *ISwr* (Dalgard, 2015) is a Danish study on the effect of screening for breast cancer. The number of breast cancer has been reported for women in different age groups (6 groups) and cohort. Four cohorts were collected. The “study group” consists of the population of women in the appropriate age range in Copenhagen and Frederiksberg after the introduction of routine mammography screening. The “national control group” consisted of the population in the parts of Denmark in which routine mammography screening was not available. These two groups were both collected in the years 1991-2001. The “historical control group” and the “historical national control group” are similar cohorts from 10 years earlier (1981-1991), before the introduction of screening in Copenhagen and Frederiksberg. The study group comprises the entire population, not just those accepting the invitation to be screened. It is a balanced data that will be used for all the coding structures except weighted effect coding.

The unbalanced data is an embedded R dataset, *chickwts* in the package *datasets* (R Core Team, 2017). The data present the results of an experiment conducted to measure and compare the effectiveness of various feed supplements on the growth rate of chickens. Newly hatched chicks were randomly allocated into six groups, and each group was given a different feed supplement. Their weights in grams after six weeks are given along with feed types. The particular case of trend or polynomial contrasts require a quantitative factor with equally spaced levels. The dataset *coking* of the package *ISwR*. It contains the time to coking in an experiment where the oven width and temperature were varied.

For simplicity purposes, a linear additive model is used for all the coding structures with the function *lm* of the package *stats* (R Core Team, 2017). The coding structures can be specified by three different methods. Though, only one is presented in this document, the specification inside the model. The others can be found in the global version of the R code used. However, the output is the same for all of them.

2.2 Coding structures for non-trend contrasts

2.2.1 Dummy coding

Also called treatment contrasts, it is the most common structure used and the basic coding structure in R. It tests the effects of one group (base or reference) against all other groups, for instance a control group against multiple treatment groups. The coding only uses 1 and 0. The reference should be well defined and allow useful comparisons. This scheme is used for balanced and unbalanced data.

- **Advantages:** Interpretation is simple. It especially works well with nominal and more specifically dichotomous data.
- **Limitations:** Its ability to make interpretations is limited and can result erroneous tests of significance for multi-factor designs.

R code and output

```
# To see the contrasts the software uses for the factors, just use the following code
contrasts(age) ## 1 represents the group that is compared to the reference
```

	55-59	60-64	65-69	70-74	75-79
50-54	0	0	0	0	0
55-59	1	0	0	0	0
60-64	0	1	0	0	0
65-69	0	0	1	0	0
70-74	0	0	0	1	0
75-79	0	0	0	0	1

```
## As dummy coding also called treatment contrasts is the default, no need to specify
model.dc <- lm(bc.deaths ~ age , data = bcmort)
## model.dc <- lm(bc.deaths ~ age , contrasts = list(age = "contr.treatment"), data =
bcmort) ## the argument "contrasts" allows to specify the desired structure directly in
the model, here contr.treatment is the function needed.

summary(model.dc)
```

	Estimate	Std.Error	t value	Pr(> t)	
(Intercept)	56	101.22	0.553	0.587	ns
age55-59	183	143.15	1.278	0.217	ns
age60-64	225.5	143.15	1.575	0.133	ns
age65-69	239.75	143.15	1.675	0.111	ns
age70-74	230.75	143.15	1.612	0.124	ns
age75-79	64.25	143.15	0.449	0.659	ns

*Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant.*

Interpretation

For dummy, effect and weighted coding structures, the estimates except the intercept represent the difference between the mean of the group considered and the intercept. The null hypothesis of the equivalent t test states that this difference is equal to 0. If the probability value (last column in the output) is greater than 0.05 then the difference is significantly equal to 0. Thus, there is no significant difference between the terms involved in that difference of means. Each column given by displaying contrasts represent the terms of the summary output. For example, the column "55-59" of contrasts(age) is represented by "age55-59" in the summary of the model.

Here, the intercept represents the mean of the reference group here 50-54. No probabilities are significant meaning that there is no significant difference between each of the other 5 groups and the reference 50-54. The mortality of the other groups is then not different from the group of women between 50 and 54 years old.

2.2.2 Effect coding

The coding process is the same as in dummy coding except that the last group receives a “-1” for all contrasts. It is used with balanced data. Also, this structure is very similar to contrast coding (next structure), identical in the two-category situation. The purpose is to compare each group mean to the grand mean of all the groups.

- **Advantages:** It allows to test mean differences between groups by using simple contrasts. Interpretation is easy.
- **Limitations:** It only tests differences between simple contrasts and does not allow to test hypotheses for complex contrasts.

R code and output

```
## Define the effect coding structure only for age. cohort keeps the dummy coding
contr.ec <- contr.sum(6) ## 6 is the number of levels of the factor
## Assign the matrix as contrasts of the factor age
contrasts(age) <- contr.ec
print(contr.ec) ## to visualize the coefficients
```

	55-59	60-64	65-69	70-74	75-79
50-54	1	0	0	0	0
55-59	0	1	0	0	0
60-64	0	0	1	0	0
65-69	0	0	0	1	0
70-74	0	0	0	0	1
75-79	-1	-1	-1	-1	-1

```
## Insert the new coding variable in the model
model.ec1 <- lm(bc.deaths ~ age, data = bcmort, contrasts = list(age = "contr.sum"))
summary(model.ec1)
```

	Estimate	Std.Error	t value	Pr(> t)	
(Intercept)	213.21	41.32	5.16	6.6E-05	***
age1	-157.21	92.4	-1.701	0.106	ns
age2	25.79	92.4	0.279	0.783	ns
age3	68.29	92.4	0.739	0.469	ns
age4	82.54	92.4	0.893	0.383	ns
age5	73.54	92.4	0.796	0.436	ns

*Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant*

Interpretation

The intercept is the grand mean of all the groups of age which is significantly non null. The probabilities are not significant implying that each of the groups is not significantly different from the grand mean. The breast cancer mortality of each group is not different from the average mortality.

2.2.3 Contrast coding

This structure is an extension of dummy coding used in balanced designs which tests specific hypotheses of interest between group means. The sum of the contrast coefficients is null. It involves two types of contrasts: orthogonal and non-orthogonal. This structure comprises other structures, the most common are presented in Table 1.

- **Advantages:** Interpretation of main effects and interactions is fairly straightforward and accurate.
- **Limitations:** For unbalanced designs or non-orthogonal designs, interpretations can become confounded even if ways to deal with it exists.

Table 1. Summary of the most common categories of contrast coding

Contrast	Use
Simple	Pairwise contrasts. Compare means of each level with the mean of a reference category (any level, but usually first or last level)
Repeated contrast	Pairwise contrasts. Compare the mean of each level to the mean of the immediately following level.
Helmert	Compare the mean of each level with the mean of the subsequent levels starting by the last, the penultimate level with the last, the level before penultimate with the average of the last and penultimate, and so on.
Reverse Helmert	Reverse of Helmert contrast. Compare the mean of the second level with the first, the third with the average of the first two, and so on.

R code and output

```
## Example of reverse Helmert contrasts
model.cc.rh <- lm(bc.deaths ~ age, contrasts = list(age = "contr.helmert"), data = bcmort)
summary(model.cc.rh)
```

	Estimate	Std.Error	t value	Pr(> t)	
(Intercept)	213.21	41.32	5.16	6.6E-05	***
age1	91.5	71.57	1.278	0.217	ns
age2	44.67	41.32	1.081	0.294	ns
age3	25.9	29.22	0.886	0.387	ns
age4	13.74	22.63	0.607	0.551	ns
age5	-18.59	18.48	-1.006	0.328	ns

contrasts(age)	55-59	60-64	65-69	70-74	75-79
50-54	-1	-1	-1	-1	-1
55-59	1	-1	-1	-1	-1
60-64	0	2	-1	-1	-1
65-69	0	0	3	-1	-1
70-74	0	0	0	4	-1
75-79	0	0	0	0	5

*Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant*

Interpretation

The intercept is the grand mean of all the groups of age which is significantly non null. The advantage of contrast coding is the ability to deal with complex contrasts. Generally, the groups having coefficients of the same sign are in the same big group and are compared to those in the other group having the opposite sign. This holds for contrasts coding structures and polynomial contrasts.

Here, the term “age1” shows that there is no significant difference between mortalities of women in 50-54 and 55-59 years old. Meanwhile, “age2” reveals the same conclusion for the women in 60-64 and the average of the previous two groups. “age3” brings out the same conclusion too but for women in 65-69 and the average of the three preceding groups. This is the process followed till the last comparison.

2.2.4 Weighted effect coding

Variation of effect coding, it has been designed for groups with unequal size or unbalanced designs. The effect for each category represents the deviation of that category from the weighted mean (which corresponds to the sample mean).

- **Advantages:** It allows comparisons for unbalanced data. It is a type of dummy coding to facilitate the inclusion of categorical variables in generalized linear models (GLM).
- **Limitations:** It does not test complex contrasts like contrast coding.

R code and output

```
contr.wec.feed <- contr.wec(feed, omitted = "casein")
print(contr.wec.feed)
```

	horsebean	linseed	meatmeal	soybean	sunflower
1	-0.8333333	-1	-0.916667	-1.1667	-1
2	1	0	0	0	0
3	0	1	0	0	0
4	0	0	1	0	0
5	0	0	0	1	0
6	0	0	0	0	1

```
model.wec <- lm(weight ~ feed, contrasts = list(feed = contr.wec.feed), data = chickwts)
summary(model.wec)
```

	Estimate	Std.Error	t value	Pr(> t)	
(Intercept)	261.31	6.51	40.143	< 2E-16	***
feedhorsebean	-101.11	16.08	-6.289	3.1E-08	***
feedlinseed	-42.56	14.43	-2.949	0.00443	**
feedmeatmeal	15.6	15.2	1.026	0.30867	ns
feedsoybean	-14.88	13.13	-1.133	0.26139	ns
feedsunflower	67.61	14.43	4.684	1.5E-05	***

*Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant.*

Interpretation

The intercept here represents the sample mean. The estimates are the deviations of the groups means from that sample mean. The group specified in omitted in the code is like the reference group for dummy coding. Although, the intercept is different from the dummy coding intercept (sample mean against reference group mean). All the feed types groups except the meat meal were significantly different from the sample mean which weight was the sample mean plus almost 16.

2.3 Coding structure for trend contrasts or polynomial contrasts

The target here is to check if there is any trend or test whether the means across the levels form a certain pattern such as a line or a parabola. The scheme requires our predictor to quantitative with levels equally spaced.

- **Advantages:** It allows checking for effects of treatment or any other type of data that may perform in a manner that is not simply on a line.
- **Limitations:** The interpretation can be tricky.

R code and output

```
contr.poly(3)
      .L      .Q
4  -0.7071068  0.40825
8  -7.8505E-17 -0.8165
12  0.7071068  0.40825

model.poly <- lm(time ~ width, contrasts = list(width = "contr.poly", temp = "contr.poly"),
               data = chickwts)
summary(model.poly)
```

	Estimate	Std.Error	t value	Pr(> t)	
(Intercept)	6.0333	0.3117	19.355	5.1E-12	***
width.L	4.5137	0.5399	8.36	5E-07	***
width.Q	-0.3878	0.5399	-0.718	0.484	ns

*Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant.*

Interpretation

The intercept, the grand mean of all the groups of age, is significantly non null. “.L” and “.Q” are “linear” and “quadratic” respectively. Here, the trend across the groups is linear. This implies that the average coking time needed by the ovens of width 4 and 8 together is significantly different from the one needed by the oven of width 12.

3 Simulation study

3.1 Simulation design

3.1.1 Factors of interest

Among the popular coding techniques mentioned, dummy coding, effect coding, weighted effect coding, and reverse Helmert contrasts were taken into account. Šidák and Bonferroni correction methods were used because of their easier application than Monte-Carlo. All the factors considered are listed in Table 2.

Balanced data (where the levels have the same sample size) was generated and applied for dummy coding, effect coding and reverse helmert. Meanwhile, unbalanced data was created for dummy coding and weighted effect coding. For this latter, we only considered that the first two levels were different from the others with the first being the most weighted (Table 3). The smallest sample size was set to 48 in order to have at least 8 replications for each level of the factor. Then, a step of 24 was used to get a sequence from 48 to 168. The number of levels and sample size modalities had been selected using the least common multiple property.

The robustness of the contrasts structures against non-normality has been checked by varying the distribution of the response variable. The particular case of moderate non-normality was obtained through a Chi-square distribution with low degrees of freedom (Table 2). In addition, a standard normal distribution was generated as a control for comparison purposes.

Table 2. Summary of the factors to be tested

Factors	Levels
Number of levels of the factor (p)	3, 4, 6
Sample size (n)	48, 72, 96, 144, 168
Effect size (E)	1, 8, 20, 50
Distribution of the response variable (μ)	$\chi^2(5)$, $\chi^2(10)$, $\chi^2(20)$, $\mathcal{N}(0,1)$
Correction for Type I error inflation (mc)	- no correction - Šidák - Bonferroni

The first three factors are quantitative while the remaining are qualitative.

Number of combinations: Data generation used the combination of number of levels (4), sample size (5), distribution type of the response variable (4) and effect size (4). This yields a total of **320** configurations.

Number of replications: 3000 random generations.

Table 3. Levels ratio for unbalanced data according to the number of levels

Number of levels (p)	Level ratio
$p = 3$	3:2:1
$p = 4$	3:2:1:1
$p = 6$	3:2:1:1:1:1

3.1.2 Simulation plan

The simulation design was implemented according to the following steps:

Step 1. A linear model was considered, in matrix notation as follows:

$$Y = \beta_0 + X\beta + \varepsilon \quad (\text{Eq.10})$$

Where Y is the response or dependent variable, β_0 the intercept, β the vector of coefficients, X the matrix of observations of the independent variable and ε the residuals.

The independent variable x was generated using a multinomial distribution with parameters the proportions given by the level ratio for each number of levels (Table 4) and the different sample sizes i.e. $x \sim MNOMIAL(q_1, q_2, \dots, n_1)$. X was obtained using x .

Step 2. The residuals were generated by creating vectors of data following 3 different types of distributions. The case of moderate non-normality has been checked using chi-square distributions with degrees of freedom 5, 10 and 20 respectively ($\chi^2(5)$, $\chi^2(10)$, $\chi^2(20)$) while the normality case was represented by a standard normal distribution ($\mathcal{N}(0,1)$). For a chi-square distribution, the mean and the degree of freedoms are the same.

Step 3. As contrasts make comparisons among means of levels of the factor, it is essential to know by default which levels are significantly different or not in order to check the performance of the structure. This prior helped in finding if the contrasts were really detecting the differences purposefully set by default. Thus, some prior values were assigned to the coefficients vector (Table 4). Indeed, it allowed to define by default the differences or comparisons where probabilities were expected to be significant. Furthermore, β was a function of the effect size (E). This latter was varying in order to check if the results obtained could be influenced by the size of the coefficients in β .

Table 4. Values assigned by default to the parameter β considered for all the coding structures.

Number of levels	β
$p = 3$	$[-E, -E, E]^*, [-E, 0, E]**$
$p = 4$	$[-E, -E, 0, 2E]$
$p = 6$	$[-3E, -3E, 0, 2E, 2E, 2E]$

The notation $[-E, -E, E]$ has been used to represent a vector which has a number of elements equal to the number of levels considered. Also, the notation $2E$ represents E multiplied by 2. Plus, the values have been defined to match the comparison or contrasts of each coding structure. Defining only one parameter for all the coding structures for $p = 3$ was quite tricky. Thus, two parameters were defined. The vector marked with “*” was used for reverse helmert and the one followed by “**” for dummy coding (Table 4).

Step 4. To ensure that the probability of rejecting the null hypothesis when it is true is at most 0.05, the familywise level of significance was set to 0.05 so $\alpha[PF] = 0.05$. Then, the level of significance per test or comparison was computed using specific equations (as defined in the first section, Background, of the present document, Table 5).

Table 5. Levels of confidence used per comparison

	No correction	Bonferroni	Šidák
Level of significance $\alpha[PT]$	$\alpha[PT] = 0.05$	$\alpha[PT] \approx \frac{\alpha[PF]}{c}$	$\alpha[PT] = 1 - (1 - \alpha[PF])^{1/c}$

c is the number of contrasts involved in the analysis.

Step 5. For the 4 types of the response variable’s distribution, β_0 took as default value 0. Moreover, Y was computed using Eq.10 for standard normal distribution. For the chi-square distributions, the response was generated following the equation:

$$\begin{aligned}
 Y &= X\beta + \chi^2(\beta_0) \\
 Y &= \beta_0 + X\beta - \mu + \chi^2(\mu) \\
 Y &= (\beta_0 - \mu) + X\beta + \chi^2(\mu)
 \end{aligned} \tag{Eq.11}$$

Where μ is the degree of freedom/mean of the chi-square distribution.

Step 6. Each coding structure was applied to either balanced or unbalanced data (Table 7). Note that dummy coding was used for the two types (balanced and unbalanced).

Step 7. For each contrast coding, the true significant probabilities (probabilities detecting the significant contrasts as set by default) were checked comparing the probabilities from the

simulation to the one known by default (Step 3). For a given structure, the number of these probabilities was stored and denoted m .

Two performance criteria have been defined and used: the per-contrast performance k_1 and the overall performance of the coding structure k_2 .

- The per-contrast performance criterion k_1 is defined as the ratio between the number of true significant probabilities and the number of contrasts. The formula is as follows:

$$k_1 = \frac{m}{p - 1} \quad (\text{Eq.12})$$

Where m is the number of true significant probabilities (probability < level of significance as default) for the type of contrast, p the number of levels then $p - 1$ the number of contrasts.

k_1 lies between the 0 and 1. When it is close to 1, it implies that the contrast coding performs well at the level of individual contrasts or comparisons. The opposite suggests that the contrast used is not performant.

- The overall performance criterion k_2 is made to take only two values: 0 if the coding structure has all the true significant probabilities for each of its contrasts, 1 otherwise. Thus, the structure globally performs well if $k_2 = 1$ and does not if $k_2 = 0$. As this criterion takes into account all the contrasts together, the level of significance to use is the familywise level of significance. However, the two other levels mentioned in Step 4 was also used just to check if any difference raises when the familywise error becomes smaller.

Step 8. The process of computing performance criteria for each combination of the factors being tested was replicated 3000 times. This output was used to compute the mean of each performance criterion. They are presented in a table for each combination of sample size, number of levels, effect size and the distribution of the response variable.

Table 6. Coding structures under investigation in R

Coding scheme	Type of data	Syntax in R	Package
Dummy	Balanced and unbalanced	<i>contr.treatment</i>	<i>stats</i> (R Core Team, 2017)
Effect coding	Balanced	<i>contr.sum</i>	<i>stats</i> (R Core Team, 2017)
Weighted effect coding	Unbalanced	<i>contr.wec</i>	<i>wec</i> (Grotenhuis,2016)
Reverse Helmert	Balanced	<i>contr.helmert</i>	<i>stats</i> (R Core Team, 2017)

3.1.3 Data analysis

Each performance criterion was independently modeled for each coding structure. Moreover, these models only assessed the additive effects of the sample size (n), the number of

levels (p), the effect size (E), the distribution of the response variable (μ) and the correction method for type I error inflation (mc). The per-contrast performance criterion ($k1$) was modeled following two steps because of the nature of the data. Indeed, $k1$ was constituted of proportions but with many zeros. The first step used a generalized linear model with a binomial structure for errors (using *glm*, package *stats* (R Core Team, 2017)) to deal with the many zeros in the data in order to know whether the structure is performant or not regarding the predictors. Then, the part of the data non null (with $k1 \neq 0$) was analyzed using a generalized linear model with a gamma structure for errors. However, weighted effect coding per-contrast performance was only modeled using the second step. Meanwhile, a generalized linear model with a binomial family was used to assess the additive effects of the predictors involved in the study on the overall performance criterion ($k2$).

The simplification of the models was done using a stepwise selection based on Akaike Information Criterion (AIC) actually with the function *stepAIC* of the package *MASS* (Venables, 2002). Furthermore, the best model selection was also based on the AIC using the package *bbmle* (Bolker & R Core Team, 2017). Furthermore, an ANOVA on the best model was used to assess the global effects of the predictors.

3.2 Results

The results have been presented with many tables from the output of the statistical models used. The terms of these models or predictors in the tables considered are coded as showed in Table 7.

Table 7.Terms of models and their meaning for all the tables of the section Results

Predictors	Meaning
Intercept	Effect size = 1, sample size = 48, number of levels = 3, Normal standard distribution (mean = 0), method of correction = no correction
E8	Effect size = 8
E20	Effect size = 20
E50	Effect size = 50
n72	Sample size = 72
n96	Sample size = 96
n144	Sample size = 144
n168	Sample size = 168
p4	Number of levels of the factor = 4
p6	Number of levels of the factor = 6
mu5	Chi-square distribution with mean = 5
mu10	Chi-square distribution with mean = 10
mu20	Chi-square distribution with mean = 20
mcbonf	Method of correction = Bonferroni
mcsid	Method of correction = Šidák

3.2.1 Per-contrast relative performance of the coding structures

The per-contrast relative performance tells about the success of the coding technique in being right at finding the real output (difference or not) of its individual contrasts. Its assessment was done in two steps because of the great number of zeros observed in the per-contrast criterion except for weighted effect coding which was only analyzed in the second step.

The first step revealed that the probability of effect and reverse helmert coding being performant for each of their individual contrasts was influenced by the effect size, the sample size, the number of levels and the type of the distribution ($P > 0.05$; Table 8). Although, dummy coding probability, when used in unbalanced design, was not affected by any of the predictors ($P < 0.05$; Table 8), it was varying depending on the effect size, the sample size and the type of the distribution when applied to balanced data ($P > 0.05$; Table 8). Moreover, the correction method used for Type I error inflation has no effect on all the structures probability of being performant ($P < 0.05$; Table 8).

Table 8. ANOVA on a logistic model to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of the coding structures. Only probability values are recorded in the table

Coding structure	Design type	E	n	p	μ	mc
Dummy coding	balanced	2.82e-14 ***	0.008 **	-	8.04e-04 ***	-
Dummy coding	unbalanced	-	-	-	-	-
Effect coding	balanced	8.44e-07 ***	0.014 *	0.000 ***	1.71e-04 ***	-
Reverse helmert	balanced	3.37e-06 ***	0.001 ***	0.008 **	5.23e-04 ***	-

Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

On the other hand, the second step showed that the method of correction has no effect on the coding structures performances as compared to the effect size, number of levels and the type of distribution which were influencing the per-contrast performances variation of all the coding structures (Table 9). Only, the performances of dummy and reverse helmert schemes were depending on the sample size (Table 9).

Table 9. ANOVA on a gamma model to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the per-contrast performance of the coding structures. Only probability values are recorded in the table.

Coding structure	Design type	E	n	p	μ	mc
Dummy coding	balanced	< 2.2e-16 ***	2.53e-05 ***	0.01822 *	< 2.2e-16 ***	-
Dummy coding	unbalanced	< 2.2e-16 ***	-	0.000 ***	< 2.2e-16 ***	-
Effect coding	balanced	< 2.2e-16 ***	-	0.000 ***	< 2.2e-16 ***	-
Reverse helmert	balanced	< 2.2e-16 ***	1.72e-05 ***	0.000 ***	< 2.2e-16 ***	-
Weighted effect coding	unbalanced	< 2e-16 ***	-	< 2e-16 ***	0.00154 ***	-

Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

3.2.2 Overall relative performance of the coding structures

The overall relative performance informs on the ability or success of the coding scheme in finding the significant differences/similarities for contrasts or comparisons provided that they are really different/similar. Then, the structure is either performant or not.

For all the coding structures except weighted effect coding, the performance of the coding structures while considering all the contrasts involved (k_2) was mainly affected by the effect size, the number of levels and the type of the distribution ($P > 0.05$; Table 10). Plus, the correction method used for Type I error inflation has no effect on this performance ($P > 0.05$ or not included in the best model; Table 10). Weighted effect coding was only affected by the correction method and the number of levels ($P < 0.05$; Table 10).

Table 10. ANOVA on a logistic model to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (mu) and the method of correction (mc) on the overall performance of the coding structures. Only probability values are recorded in the table

Coding structure	Design type	E	n	p	mu	mc
Dummy coding	balanced	< 2.2e-16 ***	0.111 ns	0.002 ***	1.38E-15 ***	0.105 ns
Dummy coding	unbalanced	3.22E-05 ***	0.000 ***	< 2.2e-16 **	1.63E-08 ***	-
Effect coding	balanced	< 2.2e-16 ***	0.000 ***	0.000 ***	1.03E-13 ***	-
Reverse helmert	balanced	< 2.2e-16 ***	-	0.000 ***	1.10E-10 ***	-
Weighted effect coding	unbalanced	0.133 ns	0.074 ns	0.000 ***	-	0.041 *

Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

- **Dummy coding**

Dummy coding (DC), for unbalanced and balanced designs, with an effect size greater than 1 was slightly more likely to be performant than the one with the effect size equal 1 (Table 11 and 12). The probability of DC (balanced) being performant when used with moderate non-normal data was lesser than the one when dealing with normal data while this was the opposite for DC (unbalanced) (Table 11 and 12). The structure was performing lesser for $n = 86$ and a little more for $p = 6$ as compared to the references $n = 48$ and $p = 3$ respectively. Meanwhile, DC (unbalanced) with $n = 168$ and $p = 6$ was more successful than the reference.

Table 11. Logistic regression model for the overall performance of Dummy coding for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	0.783	0.783	0.688	0.515	0.824	0.036	*
E8	2.070	2.070	0.888	0.827	0.831	0.000	***
E20	2.854	2.854	0.846	0.808	0.868	0.000	***
E50	2.854	2.854	0.846	0.808	0.868	0.000	***
n72	-0.230	-0.230	0.443	0.304	0.580	0.448	ns
n86	-0.781	-0.781	0.314	0.202	0.450	0.008	**
n144	-0.084	-0.084	0.477	0.333	0.624	0.758	ns
n168	-0.230	-0.230	0.443	0.304	0.580	0.448	ns
p4	-0.113	-0.113	0.472	0.358	0.587	0.635	ns
p6	-0.783	-0.783	0.314	0.224	0.418	0.001	***
mu5	-2.063	-2.063	0.113	0.062	0.188	0.000	***
mu10	-2.063	-2.063	0.113	0.062	0.188	0.000	***
mu20	-2.188	-2.188	0.101	0.055	0.171	0.000	***
mcbonf	0.421	0.421	0.604	0.483	0.706	0.068	ns
mcsid	0.421	0.421	0.604	0.483	0.706	0.068	ns

Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

Table 12. Logistic regression model for the overall performance of Dummy coding for unbalanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	-5.237	-5.237	0.005	0.001	0.022	0.000	***
E8	1.815	1.815	0.860	0.705	0.844	0.000	***
E20	3.217	3.217	0.861	0.888	0.887	0.000	***
E50	3.217	3.217	0.861	0.888	0.887	0.000	***
n72	0.000	0.000	0.500	0.258	0.741	1.000	ns
n86	-0.814	-0.814	0.307	0.134	0.552	0.123	ns
n144	-0.684	-0.684	0.335	0.150	0.585	0.185	ns
n168	-4.002	-4.002	0.018	0.004	0.060	0.000	***
p4	-15.806	-15.806	0.000	0.000	1.000	0.887	ns
p6	6.788	6.788	0.888	0.886	1.000	0.000	***
mu5	-3.066	-3.066	0.045	0.013	0.123	0.000	***
mu10	-2.265	-2.265	0.084	0.031	0.230	0.000	***
mu20	-2.854	-2.854	0.050	0.015	0.135	0.000	***

*Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant*

- **Effect coding**

The current scheme was performing better for greater effect sizes, bigger sample sizes ($n = 144$ and $n = 168$) and the other number of levels as compared to the references ($E = 1$, $n = 48$, $p = 3$ respectively). Although, EC used with normal data was doing better than when used with moderate non-normal data (Table 13).

Table 13. Logistic regression model for the overall performance of Effect coding for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	-1.407	0.433	0.187	0.084	0.362	0.001	**
E8	4.146	0.386	0.884	0.868	0.883	0.000	***
E20	4.858	0.430	0.882	0.883	0.887	0.000	***
E50	4.858	0.430	0.882	0.883	0.887	0.000	***
n72	0.461	0.365	0.613	0.438	0.766	0.206	ns
n86	0.255	0.358	0.563	0.381	0.723	0.476	ns
n144	2.168	0.447	0.887	0.788	0.856	0.000	***
n168	1.356	0.404	0.785	0.641	0.887	0.001	***
p4	1.557	0.331	0.826	0.716	0.803	0.000	***
p6	0.620	0.281	0.650	0.514	0.768	0.033	*
mu5	-2.278	0.433	0.083	0.041	0.188	0.000	***
mu10	-3.018	0.441	0.047	0.018	0.101	0.000	***
mu20	-2.218	0.432	0.088	0.043	0.188	0.000	***

*Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant*

- **Reverse helmert coding**

Reverse helmert contrasts associated with all the other effect size levels was more likely to perform well than when using the reference level. Plus, normal data related more to a good performance than the other type of distributions. This trend was the same for $p = 3$ as compared to $p = 6$ (Table 14).

Table 14. Logistic regression model for the overall performance of Reverse Helmert contrasts for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	0.521	0.285	0.627	0.482	0.748	0.068	ns
E8	2.183	0.260	0.888	0.843	0.837	0.000	***
E20	2.745	0.280	0.840	0.801	0.865	0.000	***
E50	2.745	0.280	0.840	0.801	0.865	0.000	***
p4	0.087	0.241	0.522	0.405	0.636	0.718	ns
p6	-0.876	0.230	0.274	0.183	0.370	0.000	***
mu5	-1.688	0.303	0.155	0.081	0.246	0.000	***
mu10	-1.730	0.303	0.151	0.088	0.240	0.000	***
mu20	-1.603	0.303	0.168	0.088	0.265	0.000	***

Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

- **Weighted effect coding**

There was no significant difference between the performance of the other levels of the predictors as compared to the reference levels (Table 15).

Table 15. Logistic regression model for the overall performance of Weighted effect coding for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	-1.446	0.884	0.181	0.030	0.532	0.102	ns
E8	-0.813	0.834	0.307	0.053	0.722	0.384	ns
E20	-0.813	0.834	0.307	0.053	0.722	0.384	ns
E50	-18.500	0.882	0.000	-	1.000	0.887	ns
n72	-18.870	0.882	0.000	-	1.000	0.887	ns
n86	-18.870	1.115	0.000	-	1.000	0.887	ns
n144	0.470	1.115	0.615	0.188	0.831	0.632	ns
n168	0.470	4365.861	0.615	0.188	0.831	0.632	ns
p4	-20.018	4365.861	0.000	-	1.000	0.886	ns
p6	-20.018	4627.708	0.000	-	1.000	0.886	ns
mcbonf	-1.876	5287.168	0.122	0.007	0.470	0.076	ns
mcsid	-1.876	5287.168	0.122	0.007	0.470	0.076	ns

Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

3.3 Discussion

The present work showed that the method of correction for type I error inflation did not really impact the performance of the coding structures involved in the study. This could be explained by the few number of comparisons involved. Indeed, the current work only has a maximum of five contrasts to check. Although, the correction methods have been advised for multiple comparisons (Abdi, 2007), it may require a greater number of comparisons to picture their usefulness.

Conversely, the effect size has been the principal variation trigger for the two types of performance. This really portrays the fact that the contrasts coefficients are based on the differences between the means. The bigger the size, the more likely the structure tends to perform well for all the contrasts. In fact, the effect size which represents the magnitude and direction of the difference between two groups can also be a percentage or a correlation (Vacha-Hasse & Thompson, 2004; Durlak, 2009). It is linked to the probability values assessing its significance and provides specific information depending on the context of the study. Though, there is no direct relationship between them (Durlak, 2009).

There was no specific trend for the effect of the sample size. It may have an influence but still unclear. For instance, bigger sample sizes were linked to an overall performance for effect coding and reverse helmert structure. Also, only the per-contrast performance of dummy coding and reverse helmert were sample size dependent. The same remark holds for the number of levels, no specific trend occurred. It had an impact on the overall performance but the pattern was varying from one structure to another.

On the other hand, normal data has been associated with the overall performance of the structures. This was expected because the normality of data has been mentioned as an assumption for contrasts analysis (Rafter, 2002). However, for dummy coding applied to unbalanced data, the opposite trend was remarked. This could be justified by the difference in designs but the link between normality and the type of design is yet to be found.

Weighted coding has been the technique less subject to the factors variation as compared to the others. This suggests that it is robust enough in its structure to perform well. Its robustness noticed in our study may confirm its application in generalized linear model (Grotenhuis, 2016). This implies that it really does not require normality as an assumption. This structure has existed for decades but was less cited than other schemes such as dummy coding (Grotenhuis, 2016). Also, its availability in R software is very recent.

4 Application in Ecology

An experiment that examine growth and leaf biomass from seedlings of baobab was conducted in the three agro-ecological zones of Benin. One of the researchers were interested in discovering the effects of the dose of compost used on the height of seedlings, one of the growth parameter. The data and is constituted of 5400 observations and the factor “Dose” has four levels expressed in tons per hectare (t/ha): 0 t/ha, 10 t/ha, 20 t/ha and 30 t/ha.

The researchers were interested in comparing the doses to a control dose which is 0 t/ha. The specific hypotheses were:

- The height of seedlings in the plots enriched with 10 t/ha of compost was bigger than those not enriched (0 t/ha)
- The height of seedlings in the plots enriched with 20 t/ha of compost was bigger than those not enriched (0 t/ha)
- The height of seedlings in the plots enriched with 30 t/ha of compost was bigger than those not enriched (0 t/ha)

The appropriate distribution for the data considered here is normal distribution. According to our simulation results, the suitable coding structure can be applied without using a correction method. Here, the number of levels, the distribution type of the response variable and the sample size are known. Plus, the effect size cannot be manipulated.

The questions of interest can be answered using dummy coding. Then, a linear model had been used and yielded the results showed in Table 16.

Table 16. Linear model to test the effects of the compost dose on the height of seedlings

	Estimate	Standard error	Prob.
Dose = 0 t/ha (reference)	6.901	0.124	< 2e-16 ***
Dose = 10 t/ha	0.235	0.181	0.193 ns
Dose = 20 t/ha	0.147	0.176	0.406 ns
Dose = 30 t/ha	0.763	0.174	1.12e-05 ***

*Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant*

The results showed that there was no significant difference between the height of seedlings where the plots were enriched with the doses 10 and 20 t/ha and the control group (Table 16). Conversely, the height of the plants where the compost dose was 30 t/ha was significantly different. Overall, the average height of seedlings was greater in the plots enriched with compost than those who were not enriched (Table 16). From a practical standpoint, this implies that the effect of the compost is really noticeable for a dose of 30 t/ha.

Conclusion

Coding structures are very useful for contrasts analysis under various types of models. In R statistical software, they are defined using many codes and embedded functions. Each of these structures has its benefits and limitations. However, the choice of the scheme to use depends mainly on the literature available and the type of comparison to make. The key to their interpretation is the understanding of the type of comparison being realized, the coefficients assignment and their sign. Contrast coding has been reported as the most sophisticated coding but dummy coding is still used because of the ease of interpretation. The output of our simulation study showed that the correction method for type I error inflation does not impact the coding structures' overall performance and per-contrast performance. Conversely, the effect size affected these performances. The current study brings an overview on the performance of the most used coding structures but this topic could be addressed deeper. For instance, another work may be useful in finding out a potential range or threshold for applying the correction methods recommended in literature.

References

- Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics*, 3, 103-107.
- Abdi, H., & Williams, J., (2010). Contrast analysis, *Encyclopedia of Research Design*. In Neil Salkind (Ed.), Thousand Oaks, CA: Sage
- Bolker, B. and R Core Team (2017). bbmle: Tools for General Maximum Likelihood Estimation. R package version 1.0.18. <https://CRAN.R-project.org/package=bbmle>.
- Dalgaard, P. (2015). ISwR: Introductory Statistics with R. R package version 2.0-7. <https://CRAN.R-project.org/package=ISwR>.
- Davis, M. J. (2010). Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures. *Journal of Data Science*, 8(1), 61-73.
- Durlak, J. A. (2009). How to Select, Calculate, and Interpret Effect Sizes. *Journal of Pediatric Psychology*, 34(9) pp. 917–928, 2009. doi:10.1093/jpepsy/jsp004
- Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., & König, R. (2016). When size matters: advantages of weighted effect coding in observational studies. *International Journal of Public Health*, 1-5.
- Lenth, R. V. (2017). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 0.8.1. <https://CRAN.R-project.org/package=emmeans>
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Thompson, B. (2006). Foundations of behavioral statistics: An insight-based approach. Guilford Press.
- Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret effect sizes. *Journal of Counseling Psychology*, 51, 473–481.
- Venables, W. N. & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-85457-0.
- Wendorf, C. A. (2004). Primer on multiple regression coding: Common forms and the additional case of repeated contrasts. *Understanding Statistics*, 3(1), 47-57.

Appendices

Appendix 1. Logistic regression odds ratio for dummy coding with balanced data

Table 17. Logistic regression model for the per-contrast performance of dummy coding for balanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	38.714	4320.188	1.000	0.000	1.000	0.883	ns
E8	20.618	3276.371	1.000	0.000	-	0.885	ns
E20	20.618	3276.371	1.000	0.000	-	0.885	ns
E50	20.618	3276.371	1.000	0.000	-	0.885	ns
n72	-18.375	3210.824	0.000	-	1.000	0.885	ns
n86	-18.213	3210.824	0.000	0.000	1.000	0.885	ns
n144	-18.213	3210.824	0.000	0.000	1.000	0.885	ns
n168	-18.608	3210.824	0.000	-	1.000	0.885	ns
mu5	-18.462	2880.327	0.000	-	1.000	0.885	ns
mu10	-18.462	2880.327	0.000	-	1.000	0.885	ns
mu20	-18.718	2880.327	0.000	0.000	1.000	0.885	ns

*Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant*

Appendix 2. Logistic regression odds ratio for dummy coding with unbalanced data

Table 18. Logistic regression model for the per-contrast performance of dummy coding for unbalanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	6.578	1.001	0.888	0.884	1.000	0.000	**

*Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant*

Appendix 3. Logistic regression odds ratio for effect coding

Table 19. Logistic regression model for the per-contrast performance of effect coding for unbalanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	42.004	10156.408	1.000	0.000	-	0.887	ns
E8	22.148	7288.086	1.000	0.000	-	0.888	ns
E20	22.148	7288.086	1.000	0.000	-	0.888	ns
E50	22.148	7288.086	1.000	0.000	-	0.888	ns
n72	-20.565	7523.854	0.000	-	1.000	0.888	ns
n86	-21.725	7523.854	0.000	-	1.000	0.888	ns
n144	0.000	10640.478	0.500	0.000	1.000	1.000	ns
n168	-20.565	7523.854	0.000	-	1.000	0.888	ns
p4	21.710	5868.803	1.000	0.000	-	0.887	ns
p6	3.242	1.180	0.862	0.782	0.888	0.006	**
mu5	-21.106	6822.226	0.000	-	1.000	0.888	ns
mu10	-21.501	6822.226	0.000	-	1.000	0.887	ns
mu20	0.000	8648.085	0.500	0.000	1.000	1.000	ns

Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

Appendix 4. Logistic regression odds ratio for reverse helmert

Table 20. Logistic regression model for the per-contrast performance of reverse helmert coding for unbalanced designs with independent variables: effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc)

Predictors	Estimates	Standard error	Odds ratio	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	42.868	10501.030	1.000	0.000	-	0.887	ns
E8	21.818	7364.832	1.000	0.000	-	0.888	ns
E20	21.818	7364.832	1.000	0.000	-	0.888	ns
E50	21.818	7364.832	1.000	0.000	-	0.888	ns
n72	0.000	11183.260	0.500	0.000	1.000	1.000	ns
n86	-20.832	7807.760	0.000	-	1.000	0.888	ns
n144	0.000	11183.260	0.500	0.000	1.000	1.000	ns
n168	-21.568	7807.760	0.000	-	1.000	0.888	ns
p4	20.208	5788.768	1.000	0.000	-	0.887	ns
p6	-0.737	0.871	0.324	0.073	0.720	0.387	ns
mu5	0.000	8771.285	0.500	0.000	1.000	1.000	ns
mu10	-21.288	6808.348	0.000	-	1.000	0.888	ns
mu20	-20.562	6808.348	0.000	-	1.000	0.888	ns

Prob.: probability value; Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

Appendix 5. Coefficients of gamma models for dummy coding with balanced data

Table 21. GLM with gamma structure for errors to test the effects effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of the dummy coding (balanced design).

Predictors	Estimates	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	0.663	0.617	0.714	0.000	***
E8	1.644	1.547	1.747	0.000	***
E20	1.721	1.618	1.828	0.000	***
E50	1.721	1.618	1.828	0.000	***
n72	1.005	0.842	1.072	0.878	ns
n86	0.862	0.801	1.027	0.246	ns
n144	1.065	0.888	1.137	0.058	ns
n168	1.121	1.048	1.186	0.001	***
p4	0.883	0.834	1.034	0.504	ns
p6	0.818	0.872	0.867	0.001	**
mu5	0.804	0.758	0.853	0.000	***
mu10	0.785	0.740	0.833	0.000	***
mu20	0.764	0.720	0.811	0.000	***

Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

Appendix 6. Coefficients of gamma models for dummy coding with unbalanced data

Table 22. GLM with gamma structure for errors to test the effects of size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of dummy coding (unbalanced design).

Predictors	Estimates	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	0.661	0.628	0.686	0.000	***
E8	1.883	1.881	2.100	0.000	***
E20	2.027	1.823	2.136	0.000	***
E50	2.027	1.823	2.136	0.000	***
p4	0.844	0.802	0.887	0.011	*
p6	0.863	0.825	0.802	0.000	***
mu5	0.770	0.731	0.811	0.000	***
mu10	0.745	0.707	0.785	0.000	***
mu20	0.735	0.688	0.775	0.000	***

Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

Appendix 7. Coefficients of gamma models for effect coding

Table 23. GLM with gamma structure for errors to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of effect coding.

Predictors	Estimates	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	0.640	0.605	0.676	0.000	***
E8	1.7146534	1.6217601	1.8127856	2.76E-66	***
E20	1.7484382	1.6535876	1.8486446	4.15E-70	***
E50	1.7484382	1.6535876	1.8486446	4.15E-70	***
p4	1.0700655	1.0204688	1.1220547	5.28E-03	**
p6	0.8654055	0.8204813	1.0124872	1.47E-01	ns
mu5	0.8488508	0.8045743	0.8878887	7.83E-08	***
mu10	0.7761377	0.7343865	0.8202738	8.88E-18	***
mu20	0.8076778	0.7646105	0.8531684	3.73E-14	***

Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

Appendix 8. Coefficients of gamma models for reverse helmert coding

Table 24. GLM with gamma structure for errors to test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of reverse helmert.

Predictors	Estimates	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	0.648	0.602	0.700	0.000	***
E8	1.678	1.578	1.784	0.000	***
E20	1.737	1.633	1.847	0.000	***
E50	1.737	1.633	1.847	0.000	***
n72	1.011	0.846	1.081	0.743	ns
n86	0.887	0.823	1.055	0.702	ns
n144	1.088	1.018	1.163	0.013	*
n168	1.138	1.064	1.218	0.000	***
p4	0.886	0.836	1.038	0.582	ns
p6	0.888	0.843	0.837	0.000	***
mu5	0.813	0.765	0.863	0.000	***
mu10	0.784	0.738	0.833	0.000	***
mu20	0.785	0.738	0.834	0.000	***

Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant

Appendix 9. Coefficients of gamma models for weighted effect coding

Table 25. GLM with gamma structure for errors test the effects of effect size (E), sample size (n), number of levels (p), type of distribution (μ) and the method of correction (mc) on the occurrence of performance of weighted effect coding.

Predictors	Estimates	lower 85%CI limit	upper 85% CI limit	Prob.	
Intercept	10.433	8.283	11.737	0.000	***
E8	0.470	0.423	0.521	0.000	***
E20	0.534	0.480	0.585	0.000	***
E50	0.540	0.485	0.601	0.000	***
p4	0.708	0.650	0.773	0.000	***
p6	0.606	0.558	0.658	0.000	***
mu5	1.148	1.051	1.256	0.002	**
mu10	1.164	1.065	1.273	0.001	***
mu20	1.148	1.052	1.256	0.002	**

*Asterisks represent the significance level for each term of the model: *** significant at 0.1%; ** significant at 1 %; * significant at 5 %; ns not significant*