A Graphical Diagnostic Test for Two-Way Contingency Tables

Un prueba gráfica de diagnóstico para tablas de contingencia de doble entrada

Jorge Iván Vélez[1,2,3,a], Fernando Marmolejo-Ramos[4,b],
Juan Carlos Correa[3,5,c]

[1]The Arcos-Burgos Group, John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia

[2]Neuroscience Research Group, University of Antioquia, Medellín, Colombia

[3]Research Group in Statistics, Department of Statistics, National University of Colombia, Medellín, Colombia

[4]Gösta Ekman Laboratory, Department of Psychology, Stockholm University, Frescati Hagväg, Stockholm, Sweden

[5]Department of Statistics, National University of Colombia, Medellín, Colombia

Abstract

We propose and illustrate a new graphical method to perform diagnostic analyses in two-way contingency tables. In this method, one observation is added or removed from each cell at a time, whilst the other cells are held constant, and the change in a test statistic of interest is graphically represented. The method provides a very simple way of determining how robust our model is (and hence our conclusions) to small changes introduced to the data. We illustrate via four examples, three of them from real-world applications, how this method works.

Key words: Contingency Tables, Diagnostics, Statistical Graphics.

Resumen

Proponemos e ilustramos un nuevo método gráfico para realizar análisis de diagníístico en tablas de contingencia de doble entrada. En este método, se adiciona o remueve una observación de cada celda a la vez mientras las demas se mantienen constantes, y el cambio en un estadíístico de interés se representa gráficamente. El método proporciona una manera simple de determinar cuán robusto es nuestro modelo (y por lo tanto nuestras conclusiones) cuando se introducen pequeños cambios en los datos. Ilustramos cómo funciona el método con cuatro ejemplos, tres de ellos con datos reales.

Palabras clave: diagnósticos, gráficos estadíísticos, tablas de contingencia.

[a]Ph.D. Scholar. E-mail: jorge.velez@anu.edu.au

[b]Postdoctoral Fellow. E-mail: fernando.marmolejo.ramos@psychology.su.se

[c]Associate Professor. E-mail: jcorrea@unal.edu.co

1. Introduction

In applied statistics, the analysis of contingency tables (CTs) is one of the most common tasks performed by statisticians on a daily basis. CTs display the frequency distribution of two or more categorical variables and, depending on how many variables are considered for analysis, they are called two-way CTs or multi-way CTs. Information presented as CTs is far from uncommon in the literature, and includes applications in a wide variety of areas such as psychology (Iossifova & Marmolejo-Ramos 2013), genetics (Kamish 1988, Clarke, Anderson, Pettersson, Cardon, Morris & Zondervan 2011, Dickhaus, StraSSburger, Schunk, Morcillo-Suarez, Illig & Navarro 2012), demography (Carlier & Ewing 1992, Cung 2013) and the social sciences (Wickens 1969). In the context of two-way CTs, categorical data is traditionally displayed as shown in Table 1. In order to analyse this type of data, it is necessary to use a statistical model M that describes how the categorical variables X and Y are stochastically related. Once model M is fitted, the statistical inference begins, and we will therefore, be able to draw conclusions about the data at hand (Harrell 2001, Chapter 5).

As an example, let us assume that gender and the presence/absence of a particular disease are our two variables of interest. In this case, we have a two-way contingency table which would display how many males and females developed the disease, as well as how many of them did not (i.e., a 2×2 CT). Using Table 1 as our testbed, we can note that the number of categories per variable is I and J, respectively: $I \times J$ is the number of possible combinations (also called cells) that will result after crossing X and Y, n_{ij} is the number of individuals in the ith category of X and the jth category of Y ($i = 1, 2, \ldots, I, j = 1, 2, \ldots, J$), and $n = n_{11} + n_{12} + \cdots + n_{IJ}$ is the total number of observations (or individuals) in the sample.

TABLE 1: A two-way contingency table in which a total of $I \times J$ cells are present.

Variable X	Variable Y			
	1	2	$\cdots$	J
1	n_{11}	n_{12}	$\cdots$	n_{1J}
2	n_{21}	n_{22}	$\cdots$	n_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}

Models for CTs often include log-linear (Agresti 2002, Chapters 8-9), logistic regression (Hosmer & Lemeshow 1989), and GSK (Grizzle, Starmer & Koch 1969) models. After fitting model M to an observed CT, it is crucial to validate the model and determine how robust the results are (Agresti 2002, MacCullagh 2002). Although several approaches have been proposed in the literature (Snee 1977, Marcus & Elias 1998, Kleijnen 1999, Geweke 2007), in this paper we will focus on validation via diagnostics.

Diagnostics is an area in statistics that started with the seminal work in regression analysis by Belsey, Kuh & Welsch (1980), which has received a lot of attention over the past three decades, and has rapidly extended to many

other areas. For CTs in particular, several diagnostic methods emulating those in log-linear models have been proposed (Lustbader & Moolgavkar 1985, Tsujitani & Koch 1991, Andersen 1992), as well as graphical methods (Genest & Green 1987, Friendly 1994, Friendly 1995, Friendly 1999) and methods for detecting outlier cells (Fuchs & Kennet 1980, Simonoff 1988). Unfortunately, graphical methods are not very popular since they have not been fully integrated into any statistical software, are not easy to understand, and/or because the observed CT contains so few observations that the user feels it is unnecessary to go further into an exploratory (graphical) analysis.

In the CT literature, one of the most frequently used diagnostics tests is the addition (or elimination) of one observation to each of the $I \times J$ cells at a time, followed by the calculation of a test statistic under model M (Andersen 1992). By doing so, changes between the test statistic calculated with the full data and then after adding (or eliminating) observations can be determined and those changes tell us how important this observation is to the fitted model. If removing some observations from one cell makes us reject the model, it suggests that the model is not robust enough. Although the process of the addition (or elimination) of observations to specific cells can be extended to more than one cell at the time, some challenges still emerge.

Here, we propose a graphical diagnostic test (GDT) that allows us to determine the quality and robustness of the fitted model M without the aforementioned complexities, and that emulates those tests used for linear models. Furthermore, we also provide an easy-to-use implementation in R (R Core Team 2015) to perform the test in order to visualise the effect that adding/eliminating k observations to/from the (ij)-th cell of the original CT has on the model (see Appendix A). Three real and three simulated examples are presented for illustration purposes.

2. Proposed Method

We propose the following procedure to study the contribution of a single observation to the model by removing (or adding) one observation each time to every cell the CT, whilst other cells are held constant. This process is then repeated for each cell. If the model is rejected or accepted, and this contradicts the conclusion reached when no modification was introduced, it suggests that our model is not robust enough. However, if we eliminate many observations of a single cell and our conclusion does not change, we can conclude that our model is robust to (not necessarily small) changes in that cell.

2.1. Notation

From now on, we will consider the notation in Table 2. Observe that even though the test statistic is represented by T and no particular distribution has been

assumed, it is straightforward to generalise it with any test statistic applicable to two-way CTs when a hypothetical model M is considered[1].

Table 2: Notation used in our alternative test for contingency tables.

Notation	Description
M	Hypothetical model
$\mathbf{N}$	Observed two-way contingency table
$\mathbf{N}_{ij}$	Number of observations in the (ij)th cell
$\mathbf{N}_{(ij)-k}$	$\mathbf{N}$ after removing k observations from the (ij)–th cell
$\mathbf{N}_{(ij)+k}$	$\mathbf{N}$ after adding k observations to the (ij)–th cell
T_M	Test statistic calculated after fitting M to $\mathbf{N}$
$T_{M(ij)-k}$	Test statistic calculated after fitting M to $\mathbf{N}_{(ij)-k}$
$T_{M(ij)+k}$	Test statistic calculated after fitting M to $\mathbf{N}_{(ij)+k}$

To illustrate the proposed notation, let us suppose we are interested in testing independence between two categorical variables using $\mathbf{N}$. Thus, it is possible to define T as the Pearson χ^2 statistic given by

$$\chi^2_M = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(\mathbf{N}_{ij} - \mathbf{E}^M_{ij})^2}{\mathbf{E}^M_{ij}}, \tag{1}$$

with $\mathbf{E}^M_{ij}$ being the expected value of the (ij)-th cell of $\mathbf{N}$ under model M (independence). The calculation of the $\chi^2_{M(ij)-k}$ and $\chi^2_{M(ij)+k}$ statistics are similar, and this has been left as an exercise for the reader[2]. Under the null hypothesis, χ^2_M follows a χ^2 distribution with $(I-1)(J-1)$ degrees of freedom, i.e., $\chi^2_M \sim \chi^2_{(I-1)(J-1)}$.

Another possibility when working with CTs is to use a likelihood ratio test (LRT), denoted as G^2, to determine whether two categorical variables are independent. Under model M, this test uses the likelihood of the data under the null hypothesis relative to the maximum likelihood, i.e.,

$$G^2_M = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\mathbf{N}_{ij} \log \frac{\mathbf{N}_{ij}}{\mathbf{E}^M_{ij}} \right), \tag{2}$$

with $\mathbf{N}_{ij}$ and $\mathbf{E}^M_{ij}$ as previously defined. Under the null hypothesis of independence the two categorical variables, $G^2 \sim \chi^2_{(I-1)(J-1)}$. The test statistics $G^2_{M(ij)-k}$ and $G^2_{M(ij)+k}$ are similarly defined.

An important observation when testing independence is that the degrees of freedom of the test statistic under the null hypothesis are not affected by the test

---

[1]Although model selection is not in the scope of this paper, one approach to determine whether model M is appropriate or not for the contingency table is to use statistics such as the Likelihood Ratio test (LRT) to compare two models M_1 and M_2, or to use the log-likelihood function and select the model with the highest value.

[2]Although there is no specific criterion for selecting the value of k, it can only be, at the most, equal to the value of a particular cell when observations are being removed from every cell at the time. This is to guarantee that no negative entries are going to be present in the transformed contingency table.

statistic that is being used, the number of k observations added to the (ij)-the cell of $\mathbf{N}_{ij}$, or by removing the same number of observations from that cell. This is true and holds for any CT since the degrees of freedom only depend on the dimension of the table.

2.2. Procedure

Beginning with the information in the (ij)th cell, we propose the following procedure to perform graphical diagnostics on the CT:

1. Under model M, calculate $T_{M(ij)-\mathbf{N}_{ij}}, T_{M(ij),-\mathbf{N}_{ij}+1}, \ldots, T_{M(ij)-1}, T_M$ (see Table 2 for more details).

2. Determine the p-value of the test as

$$p_M = \begin{cases} 1 - F(T_{M(ij)-k}) & \text{if } k \text{ observations are removed} \\ 1 - F(T_M) & \text{for the original data} \\ 1 - F(T_{M(ij)+k}) & \text{if } k \text{ observations are added} \end{cases} \quad (3)$$

   with F as the cumulative distribution function of the test statistic under the null hypothesis.

3. Plot p_M against the number of added/removed observations k.

Steps 1-3 are repeated as many times as the number of cells in the CT. For instance, when the CT is a 2×2 table, and it is of interest to evaluate how robust the model M is by initially adding or removing k observations from each cell of the original data, the resulting plot for each condition (adding or removing up to k observations) will be four-lined. Although it is somewhat expected that a gradual change in both the p-value and the test statistic will occur as observations are added and removed one at a time, it might be the case that adding (or removing) just one observation completely changes the result obtained with the original data set.

3. Examples

In this section we will illustrate the proposed GDT with six examples, three of them from the literature.

3.1. Real Data

Example 1. Hand ability and gender. Table 3(a) shows the number of left- and right-handed children, by gender, in a sample of 476 elementary school children whose teachers were asked about which hand their students used more skilfully (Correa 2002). The question of interest is whether hand ability and gender are independent.

TABLE 3: Contingency tables to exemplify the graphical diagnostic test.

(a) Hand ability and gender (Correa 2002).

Gender	Hand ability	
	Left-handed	Right-handed
Boys	79	202
Girls	57	138

(b) Polygraph evaluation (Simonoff 2003).

True group	Classified group	
	Guilty	Innocent
Guilty	6	2
Innocent	4	8

(c) Hand preference and brain injury (Vlachos et al. 2013).

Brain injury	Hand preference		
	Left-handed	Mixed-handed	Right-handed
No brain injury	42	34	472
With brain injury	4	9	67

For this data, the χ^2-based test of independence gives $\chi^2_M = 0.0263$ and $p_M = 0.8712$, from which we conclude that these variables can be considered independent. As shown in Figure 1(a) adding slightly more than 20 observations to cell (2,1) (i.e., left-handed girls), whilst leaving the other cells intact, would have shown that gender and handedness are not independent. A similar result would have been obtained if slightly more than 20 observations would have been removed from that very same cell or from the cell representing the left-handed boys exclusively (see Figure 1(b)). Hence, the independence model is robust enough and our conclusion reliable.

Example 2. Polygraph evaluation. Simonoff (2003, pp. 221) presented the results of a polygraph evaluation applied to twenty individuals as shown in Table 3(b). The LRT for testing independence produces $G^2_M = 3.4539$ and $p_M = 0.0631$. Using a type I error probability of 5%, the hypothesis of independence is not rejected. But how robust is this result? Figures 1(c)-(d) show how fragile the independence model is. In fact, adding just one observation to cells (1,1) or (2,2), or removing one observation from cells (1,2) or (2,1) would have changed our conclusion.

Example 3. Hand preference and brain injury. The brain pathology hypothesis argues that brain injury leads to hand preference. Vlachos et al. (2013) found no evidence to support such an association (see Table 3(c)). This finding s corroborated using the G^2 statistic, which gives $G^2_M = 3.031$ and $p_M = 0.21969$. However, if three more cases of mixed-handed participants with brain injury had been added (i.e., cell (2,2)) whilst leaving the other cells intact, an association between brain injury and handedness would have been found. Similarly, if three or four participants had been removed from cell (2,1), other things being equal, a significant result would have been found. Notice that only adding more than 10 observations to cell (2,1) would have also led to a significant result (see Figure 1(e)-(f)).
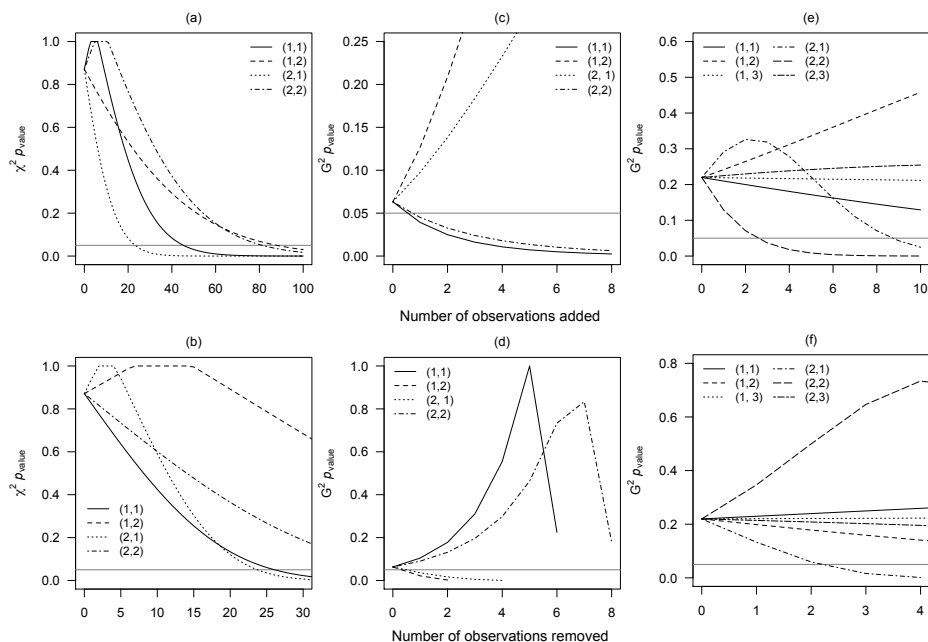
FIGURE 1: Graphical diagnostic test for Examples 1, 2, and 3. In Example 1, more than 20 observations need to be (a) added or (b) removed from two cells in order to reach statistical significance. In Example 2, (c) adding or (d) removing a couple of observations in a couple of cells is enough to obtain a significant result. In the last example, three observations would need to be (e) added to a cell or (f) removed from another cell to reach a significant result. The grey horizontal line corresponds to a type I error probability of 5%.

All these three cases exemplify how our GDT is highly valuable in testing hypotheses in the light of specific theories and concepts under investigation. Despite this, the method will output the results of the p-value associated with the test statistic under use (e.g., G^2 and χ^2), but it is up to the researcher to seek interpretations based on the data at hand and the underlying theories (i.e., those specific to the phenomena being studied).

3.2. Simulated Data

Simulating potential two-way CT situations that can occur in actual research is a rather cumbersome task since various variables would need to be considered. For instance, one can take into consideration the levels in each of the categories, the total number of observations, the percentage of observations assigned to each cell, the test statistic used to analyse the CT[3] and the effect of missing data on the test statistic (Correa & Vélez 2014).

---

[3]To study the independence of categorical data, the G^2 and χ^2 tests, the Fisher's exact test, or Barnard's test could be used. If the data is assumed to be dependent, then the McNemar's test or the Stuart-Maxwell might be suitable alternatives.
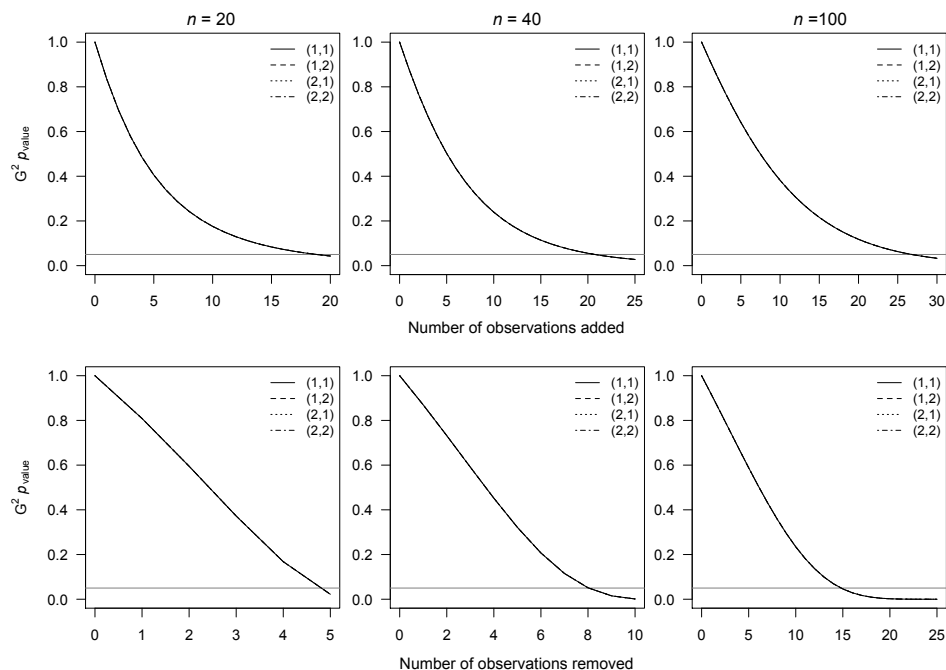
FIGURE 2: Graphical diagnostic test for three simulated 2×2 CTs varying the sample size n. The percentage of observations assigned to each cell is 25% in all cases. The grey horizontal line corresponds to a type I error probability of 5%.

Thus, in order to give the researcher a feel for what our method can offer, we only present the case of 2×2 CTs in which the percentage of observations per cell was kept equal (i.e., each cell had the same proportion of observations) and only the total amount of observations varied ($n = 20, 40$ and 100). Also, we assumed the data to be independent and used the G^2 statistic.

In a 2×2 CT with just five observations per cell, much more than five observations need to be added to any cell or one cell, needs to be emptied in order to reject the independence model (first column, Figure 2). When a 2×2 CT has 10 observations per cell, leaving one cell with just one observation would suffice to reject independence between the categories. Alternatively, adding more than 10 observations to any of the cells would show an association (second column, Figure 2). When $n = 100$, there seems to be more freedom in the number of observations that need to be added or removed in each cell in order to find an association between the categories. That is, just above 25 observations would need to be added to any cell or just 15 or 16 observations would need to be removed from any cell in order to reject independence (third column, Figure 2).

These results simply confirm that it is more advantageous to have CTs with a large number of observations in order to remove or add a number of observations that is proportionally small in relation to the cell size. Hence, this approach shows

how much a model of perfect independence needs to change in order to show an association between the two categories and that the CT's sample size plays a key role in this process.

4. Discussion

One of the main tasks of any researcher consists in validating the models built to account for data, and models fitted to two-way contingency tables are not the exception. In this paper, we have presented a graphical method to verify the robustness of a fitted model to a two-way contingency table. We strongly believe that this method is useful in model assessment and data analysis, easy to interpret and fully adaptable to any data represented in the form of two-way CTs. The GDT proposed here is useful in checking the robustness of a CT model that is suitable for categorical data analyses methods. As shown in the simulated CTs, going from an independence to an association model requires adding or removing a number of observations that is very disproportional as the sample size of the CT shrinks.

As mentioned in the simulation section, many two-way CT cases could be considered for a simulation study; however, given the multiplicity of factors and levels, that is a rather complex task. Having said this, we believe the options presented in that section need to be tackled to better understand how the proposed method behaves under different circumstances. However, we reiterate that those simulations are valid merely for statistical purposes. In other words, researchers almost never have CTs that resemble a fully independent model; on the contrary, most, if not all, real CTs seek to demonstrate an association between the categories of interest. Precisely because of this reason, we believe the proposed graphical test is helpful in assessing potential explanatory models underlying the categorical data under study.

Despite the virtues of our graphical test, some of its weaknesses need to be acknowledged. First, in its current form, our implementation of the method can only deal with two-way CTs, so further work is needed to extend it to multi-way CTs. In the hypothetical case of two-way CTs stratified by a third variable with S categories (e.g., socioeconomic status), performing our GDT for each of the resulting S two-way CTs and plotting them side-by-side is straightforward. Another possibility would be to split the CT and perform the GDT on each resulting two-way CT using different aesthetics to distinguish each cell (i.e., each cell in the CT would be represented by S lines).

Second, the current implementation of the GDT removes or adds one observation at a time from each cell in the CT. Thus, the total number of lines being displayed will be $c = I \times J$, and for large values of I and J the resulting plot might be specially crowded. Even though the user is free to make the necessary changes to have an aesthetically appealing statistical graphic, we suggest using the test when no more than ten cells are present in the two-way CT. However, if the data is presented in a multi-way CT, then the two possibilities previously described could be a starting point. A further improvement would require that sets of pairs, triplets or p-tuplets of cells are simultaneously evaluated whilst observations are

added to and/or removed from them. Although this is not computationally challenging, its graphical representation is. The main difficulty is how to represent the $c!/\{(c-p)!\,p!\}$ total number of possible p-tuples. When $p=2$, a 3D graphical representation is suitable, but for $p>2$ the alternatives are more scarce. We plan to tackle these issues in the near future, and provide an easy-to-use implementation in R.

Acknowledgements

References

Agresti, A. (2002), *Categorical Data Analysis*, 2 edn, Jhon Wiley & Sons, Hoboken, New Jersey, New York.

Andersen, E. B. (1992), 'Diagnostics in Categorical Data Analysis', *Journal of the Royal Statistical Society, Series B* **54**(3), 781–791.

Belsey, D. A., Kuh, E. & Welsch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York.

Carlier, A. & Ewing, G. (1992), 'Longitudinal analysis of contingency tables with application to demographic data', *Computational Statistics* **7**, 329–353.

Clarke, G. M., Anderson, C. A., Pettersson, F. H., Cardon, L. R., Morris, A. P. & Zondervan, K. T. (2011), 'Basic statistical analysis in genetic case-control studies', *Nature Protocols* **6**(2), 121–133.

Correa, J. C. (2002), *Un estudio de habilidad manual en niñs en edad escolar en Antioquia, Colombia*, Unpublished manuscript.

Correa, J. C. & Vélez, J. I. (2014), 'Una nota de cuidado sobre el efecto de datos parcialmente faltantes en la prueba de independencia χ^2', *Comunicaciones en Estadística* **7**(2), 189–199.

Cung, C. (2013), Crime and Demographics: An Analysis of LAPD Crime Data, M. sc. thesis, UCLA, Department of Statistics, Los Angeles, California.

Dickhaus, T., StraSSburger, K., Schunk, D., Morcillo-Suarez, C., Illig, T. & Navarro, A. (2012), 'How to analyze many contingency tables simultaneously in genetic association studies', *Statistical Applications in Genetics and Molecular Biology* **11**(4), 1544–6115.

Friendly, M. (1994), 'Mosaic Displays for Multi-Way Contingency Tables', *Journal of the American Statistical Association* **89**(425), 190–200.

Friendly, M. (1995), 'Conceptual and Visual Models for Categorical Data', *The American Statistician* **49**(2), 153–160.

Friendly, M. (1999), 'Extending Mosaic Displays: Marginal, Conditional, and Partial Views of Categorical Data', *Journal of Computational and Graphical Statistics* **8**(3), 373–395.

Fuchs, C. & Kennet, R. (1980), 'A Test for Detecting Outlying Cells in the Multinomial Distribution and Two-Way Contingency Tables', *Journal of the American Statistical Association* **75**(370), 395–398.

Genest, C. & Green, P. E. J. (1987), 'A Graphical Display of Association in Two-Way Contingency Tables', *The Statistician* **36**, 371–380.

Geweke, J. (2007), 'Bayesian Model Comparison and Validation', *The American Economic Review* **97**(2), 60–64.

Grizzle, J. E., Starmer, C. F. & Koch, G. G. (1969), 'Analysis of Categorical Data by Linear Models', *Biometrics* **25**(3), 489–504.

Harrell, Jr., F. E. (2001), *Regression modelling strategies: with applications to linear models, logistic regression, and survival analysis*, 1 edn, Springer Science, New York.

Hosmer, D. & Lemeshow, S. (1989), *Applied Logistic Regression*, Jhon Wiley & Sons, New York.

Iossifova, R. & Marmolejo-Ramos, F. (2013), 'When the body is time: spatial and temporal deixis in children with visual impairments and sighted children', *Research in Developmental Disabilities* **34**(7), 2173–2184.

Kamish, H. J. (1988), 'Contingency table estimation of genetic parameters and disease risks', *Statistics in Medicine* **7**(5), 591–600.

Kleijnen, J. (1999), Validation of Models: Statistical Techniques and Data Availability, Discussion Paper 1999-104, Tilburg University, Center for Economic Research.
*http://ideas.repec.org/p/dgr/kubcen/1999104.html

Lustbader, E. D. & Moolgavkar, S. H. (1985), 'A Diagnostic Statistic for the Score Test', *Journal of the American Statistical Association* **80**(390), 375–379.

MacCullagh, P. (2002), 'What is a Statistical Model?', *The Annals of Statistics* **30**(5), 1225–1310.

Marcus, A. H. & Elias, R. W. (1998), 'Some Useful Statistical Methods for Model Validation', *Environmental Health Perspectives* **106**, 1541–1550.

R Core Team (2015), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*http://www.R-project.org/

Simonoff, J. S. (1988), 'Detecting Outlying Cells in Two-Way Contingency Tables Via Backwards-Stepping', *Technometrics* **30**(3), 339–345.

Simonoff, J. S. (2003), *Analyzing Categorical Data*, Springer, New York.

Snee, R. D. (1977), 'Validation of Regression Models: Methods and Examples', *Technometrics* **19**(4), 415–428.

Tsujitani, M. & Koch, G. G. (1991), 'Residual Plots for Log Odds Ratio Regression Models', *Biometrics* **47**(3), 1135–1141.

Vlachos, F., Avramidis, E., Dedousis, G., Katsigianni, E., Ntalla, I., Giannakopoulou, M. & Chalmpe, M. (2013), 'Incidence and Gender Differences for Handedness among Greek Adolescents and Its Association with Familial History and Brain Injury', *Research in Psychology and Behavioral Sciences* **1**(1), 6–10.

Wickens, T. D. (1969), *Multiway Contingency Tables Analysis for the Social Sciences*, 1 edn, Psychology Press, United States.

Wickham, H. & Chang, W. (2015), *devtools: Tools to Make Developing R Packages Easier*. R package version 1.7.0.
*http://CRAN.R-project.org/package=devtools

Appendix A. Graphical diagnostic test in R

The R code to perform the GDT based on the model of independence for contingency tables can be downloaded from here here. Installing the `devtools` package by Wickham & Chang (2015) is required. Two functions are presented to perform the GDT as previously described: `gdt_G2()` and `gdt_chi2()`, which use the G^2 and χ^2 statistics, respectively (see §2.1 for more details). The following arguments are taken by both functions:

`thetable`	2×2 contingency table.
`ylim`	Limits of the y-axis. By default it is the interval `c(0,1)`.
`type`	Defines whether to `"add"` or `"remove"` observations, respectively.
`k`	Number of observations to be added/removed. When `type = "remove"`, the maximum value for `k` is that for the corresponding cell.
`alpha`	Type I error probability of the test. By default 5%.
`lin.col`	Color of the horizontal line. By default `lin.col = 2` (red).
`...`	Additional arguments passed to `plot`. See `?plot` in R for more details.