*Chilean Journal of Statistics*
Vol. 8, No. 1, April 2017, 53-65

Research Paper

# Extension of a graphical diagnostic test for contingency tables

Jorge I. Vélez[1,2], and Fernando Marmolejo-Ramos[3,*]

[1]Universidad del Norte, Barranquilla, Colombia,
[2]Genomics and Predictive Medicine Group, John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia,
[3]Department of Psychology, University of Adelaide, Adelaide, SA, Australia

(Received: July 8, 2016 · Accepted in final form: February 27, 2017)

**Abstract**

Contingency tables (CTs) are widely used in a variety of disciplines for displaying, presenting and analysing categorical data. Generally speaking, the analyses performed involve the study of potential associations between two or more categorical variables (also known as factors) using a statistical test. We recently proposed, exemplified and implemented a graphical diagnostic test (GDT) to study such associations based on two-way CTs diagtable. In this test, several observations are added (or removed) from each cell whilst the other cells are held constant, and a test statistic $T$ of interest is graphically represented. Provided that the initial CT can be reduced to a series of two-way CTs, here we present an extension of the GDT to allow the analysis of stratified CTs, and the evaluation of the effect of simultaneous changes in pairs of cells over $T$. Three examples are given, an implementation in the `R` language is presented, and future directions are discussed.

**Keywords:** Graphical diagnostic test · Contingency tables · Diagnostics · Statistical graphics · Data science.

## 1. Introduction

In applied statistics, we often deal with data sets that contain attributes, also called variables, that describe the characteristics of an object, individual or experimental unit. These variables can be of several types, one of which is *categorical* variables. Categorical variables can be seen as categories that reflect possible characteristics of the experimental unit such as gender, eye colour, political affiliation, severity of symptoms (i.e., none, mild, severe), consumer preferences, school attended or race (Agresti, 2002, Chapter 1).

Contingency tables (CTs), first introduced by Karl Pearson (Pearson, 1904, p. 34), are widely used in several disciplines, including data science, business intelligence, engineering and scientific research Wickens (1969); Kamish (1988); Agresti (2002); Iossifova and Marmolejo-Ramos (2013); Vélez et al. (2015). CTs are referred to as multiway or two-way CTs depending upon how many variables are considered for either displaying the frequency distribution of such variables, or for summarising the potential association between them

*Corresponding author. Email: fernando.marmolejoramos@adelaide.edu.au

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

using a statistical test to measure the departure from a hypothesised model $M$ which is believed to describe such an association (Agresti, 2002, Chapter 2). In its simplest form, model $M$ could correspond to the independence model where two categorical variables $X$ and $Y$ with up to $I$ and $J$ levels, respectively, are considered to be stochastically independent. A classical example in which the model of independence $M$ is rejected is that shown in Agresti (2002, Table 2.5) and reproduced in Table 1. Indeed, a $\chi^2$ test of independence gives $\chi^2_M = 18.1$ and a $p$-value of $p_M = 2.05 \times 10^{-5}$ from which it is concluded that the development of lung cancer is not independent from smoking habits (see also Example 3). Note, however, that this previous result gives no information about how robust the conclusion is as it is not possible to evaluate the effect - on the conclusion initially reached - of small changes in the entries of the CT. This effect has traditionally been assessed via diagnostic methods, an area of relative proliferation in statistics Belsey et al. (1980); Lustbader and Moolgavkar (1985); Genest and Green (1987); Tsujitani and Koch (1991); Andersen (1992); Friendly (1994), with the most frequently used approach being the elimination or addition of one observation at the time to each of the $I \times J$ entries of a two-way CT, followed by the calculation of a test statistic under model $M$.

Table 1. Lung Cancer and Smoking. Source: Doll and Hill (1950), and taken from Table 2.5 in Agresti (2002).

| Smoker | Lung Cancer | |
|---|---|---|
| | Yes (Case) | No (Control) |
| Yes | 688 | 650 |
| No | 21 | 59 |

Recently, Vélez et al. (2016) proposed, implemented and illustrated a graphical diagnostic test (GDT) for two-way CTs. Using this method, the authors are able to study the effect of small changes in the cells of the CT on a model of interest $M$ (i.e., independence model) when up to $k$ observations, one at a time, are added to or removed from every cell. When a specific model $M$ is fitted to the CT, this procedure determines how robust the model is and how confident we are about the conclusions drawn. In other words, the authors attempt, by using this method, to determine if the conclusion would change when small changes in the entries of the CT are introduced.

Despite showing how useful the GDT is, the authors acknowledged that more work needs to be done to extend the GDT to ($i$) stratified two-way CTs, and ($ii$) allow the possibility of introducing changes to more than one cell at the time. The first case refers to situations in which, given a third variable with $S$ categories (i.e., gender or socioeconomic strata), equal number of two-way CTs are constructed. The second topic implies an improvement of the GDT method such that it is possible to evaluate the effect, on model $M$, of simultaneously adding or removing up to $k$ $p$-tuples ($p = 2, 3, \ldots$) of observations from the corresponding cells in the CT. It has been argued that the implementation of such a procedure is not computationally challenging, but the graphical representation is. According to the authors, the main difficulty lies in representing the $c\,!\,/\{(c-p)\,!\,p\,!\}$ total number of possible $p$-tuples, where $c$ is the number of cells in the two-way CT diagtable.

In this paper we present an extension of the GDT, namely $e$GDT, that addresses the analysis of stratified two-way CTs and allows the quantification of simultaneous changes in pairs of observations on the statistical model of interest. We initially put forward the GDT proposed by Vélez et al. (2016). Further, we propose and describe the workflow that ultimately leads to the development of the $e$GDT, illustrate it with three examples, and present our implementation in `R` R Core Team (2016) statistical package. Finally, future directions of research as well as other potential applications of the $e$GDT are discussed.

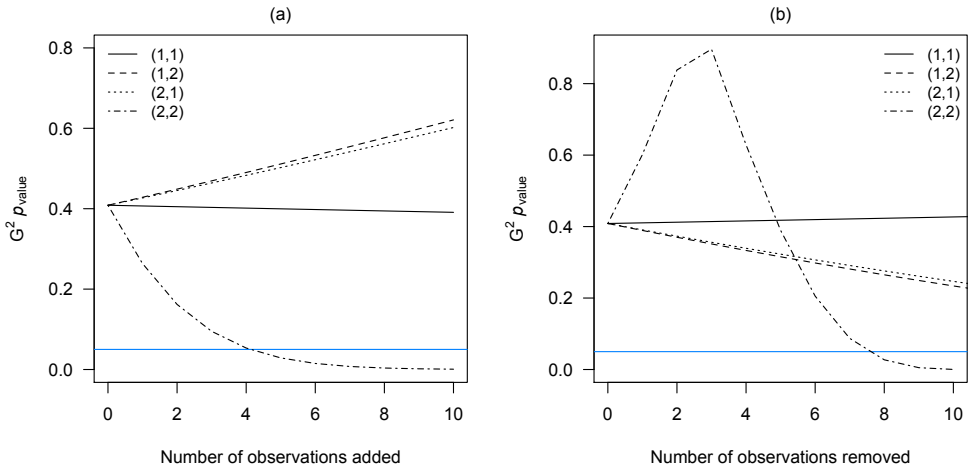


Figure 1. GDT for a two-way CT when up to 10 observations are (a) added or (b) removed to/from every cell. The entries and the corresponding number of observations are given by 833, 75, 82 and 10 for cells (1,1), (1,2), (2,1) and (2,2), respectively. The LRT of independence fives $G^2_M = 0.683$ and $p_M = 0.409$. The blue horizontal line corresponds to a 5% significance level.

## 2. Extended GDT

### 2.1 Background

Starting with a two-way CT, the GDT proposed by diagtable determines the change produced by a single cell on a hypothetical model $M$ when up to $k$ observations, one at the time, are added to or removed from each cell of the CT. Formally, the GDT can be seen as a three-step procedure to test whether two categorical variables are independent (null hypothesis). First, $k$ observations are added to (or removed from) the $(i, j)$th cell and the statistic $T_{M,(i,j),k}$ under model $M$ is calculated $(i = 1, 2, \ldots, I; j = 1, 2, \ldots, J)$. Secondly, the $p$-value of the test is calculated as $p_M = 1 - F(T_{M,(i,j),+k})$ when $k$ observations are added, and $p_M = 1 - F(T_{M,(i,j),-k})$ when $k$ observations are removed. Thirdly, $p_M$ is plotted as a function of $k$. In the expressions above, $F$ is the cumulative distribution function of the test statistic $T$.

Figure 1 shows the results of the GDT applied on a two-way CT using a likelihood ratio test (LRT) for the independence model, $M$. Here, the test statistic is calculated as $G^2_M = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\mathbf{N}_{ij}\log\frac{\mathbf{N}_{ij}}{\mathbf{E}^M_{ij}}\right)$, where $\mathbf{E}^M_{ij}$ is the expected value of the $(ij)$–th cell of the CT (denoted as $\mathbf{N}$). The calculation of $p$-values follows as $G^2 \sim \chi^2_{(I-1)(J-1)}$ under the null hypothesis of independence. Although the model of independence is not rejected $(G^2_M = 0.683, p_M = 0.409)$ when the full data is used, either adding four observations to or removing eight observations from $(2, 2)$ would have changed our conclusion entirely.

### 2.2 From GDT to *e*GDT

The GDT only deals with changes in one cell of the CT at a time. But, what would happen if more than one cell is simultaneously changed? And what would we do if stratified CTs are available? We tackle these questions using the *e*GDT provided that the original CT can be broken down into a series of smaller two-way CTs.

Before we move ahead and describe the *e*GDT in detail, let us introduce some notation. Throughout this paper we will denote $(a, b)$ as the $(a, b)$th cell of the two-way CT, and $u|v$ as the pair of entries $u$ and $v$ of the CT. In a $2 \times 2$ CT for example, one will have four cells and equal number of entries such that cells $(1, 1), (1, 2), (2, 1)$ and $(2, 2)$ are respectively equivalent to entries 1, 2, 3 and 4. We shall say that $T_{rs}$ is the test statistic when entries

$u$ and $v$ change $r$ and $s$ units, respectively ($r = 0, 2, \ldots, k; s = 0, 2, \ldots, k$). diagtable indicate that $k$ can take any integer value provided that the resulting entry in the CT is not negative. Additionally, we will say that $u^{+x} \,|\, v^{-y}$ denotes that $x$ observations are being added to entry $u$, and $y$ observations removed from entry $v$. For completeness, $c$ and $n$ denote the number of cells and the total number of observations, respectively.

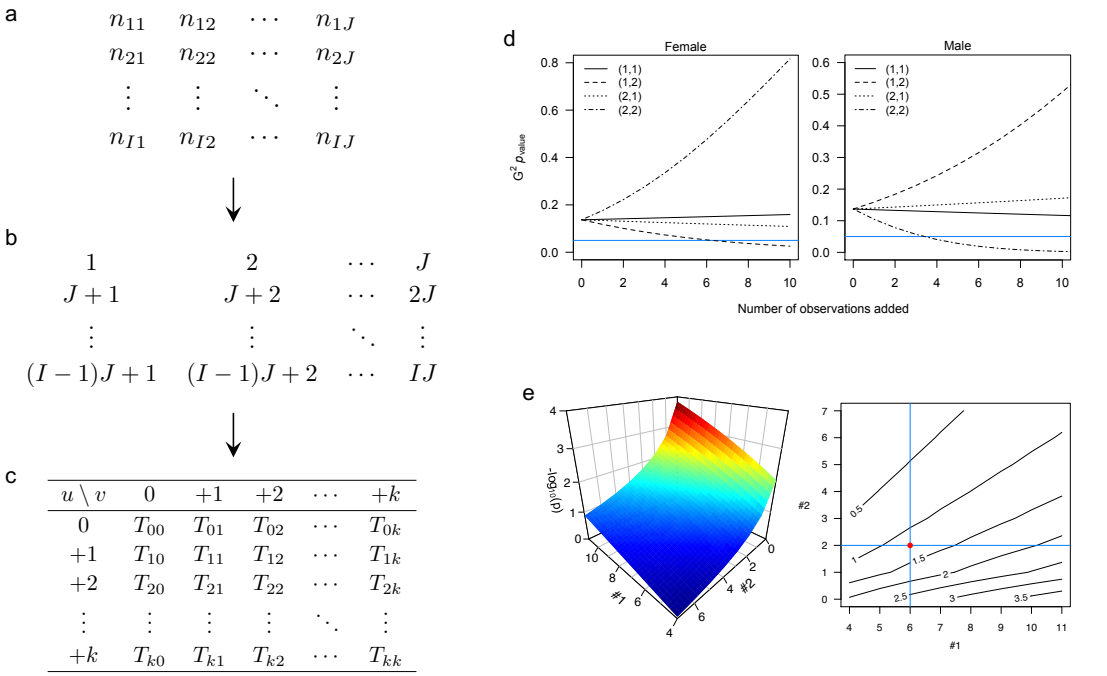


Figure 2. Workflow in the *e*GDT. Every cell in the (a) initial CT is (b) labeled row-wise, and (c) a test statistic $T$ is subsequently calculated after up to $k$ observations are simultaneously added to or removed from $u|v$. Further, the result is graphically represented for (d) each category of a third categorical variable (i.e., gender in this case) or using (e) surface and/or contour plots.

Figure 2 depicts the general workflow of the *e*GDT. The process begins with a two-way CT (Figure 2a) and the row-wise numbering of every single entry in the CT (Figure 2b). Further, for each of the $c\,!\,/\{(c-p)\,!\,p\,!\}$ total number of possible $p$-tuples ($p = 2, 3, \ldots$), we calculate $T_{rs}$ (Figure 2c). The last step is to produce a graphical representation of $T_{rs}$ or a function of it (i.e., the $p$-value) depending upon the value of $p$. When $p = 1$, the plots reduce to the GDT (Figure 2d); when $p = 2$, either or both a 3D or/and 2D plot can be produced (Figure 2e); and for $p = 3$, an alternative would be to produce similar plots for $T_{rs}$ given two entries of the CT and for every change in a third entry (see Example 3 for an illustration). A similar strategy can be applied to $p = 4$.

## 3. Illustrations

**1. Polygraph evaluation.** For illustration purposes, let us consider the polygraph evaluation data in Simonoff (2003, pp. 221) shown in Table 2. Here, a LRT of independence produces $G^2_M = 3.454$ and $p_M = 0.063$ from which we conclude, using a type I error probability of 5%, that the group individuals are classified by the polygraph is independent from the true group they actually come from. Observe that the number of cells in the CT is $c = 4$, and the number of pairs and triplets of cells that could be changed is $4!/\{(4-2)!\,2!\} = 6$ and $4!/\{(4-3)!\,3!\} = 3$, respectively. Following our notation, these combinations of entries would be given by 1|2, 1|3, 1|4, 2|3, 2|4 and 3|4 in the first case, and 1|2|3, 2|3|4 and 1|3|4 in the latter.

Table 2. Polygraph evaluation. Source: Simonoff (2003).

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

It was previously shown that, for this data, the model of independence hardly withstood in that 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 the initial conclusion Vélez et al. (2016). But, what if we simultaneously change the number of observations in pairs of cells? How would our previous conclusion change? Figure 3 shows the results of the *e*GDT applied to Table 2 when $p = 2$. Here, the combinations of changes in pairs of entries given by $1^{+1}\,|\,2^{-1}$, $1^{+1}\,|\,3^{-1}$, $1^{+1}\,|\,4^{-2}$, $2^{-1}\,|\,3^{+1}$, $2^{-1}\,|\,4^{+2}$ and $3^{-1}\,|\,4^{+1}$ would result in the rejection of the independence model initially tested. Overall, this shows how sensible our conclusions are to small changes in the CT.

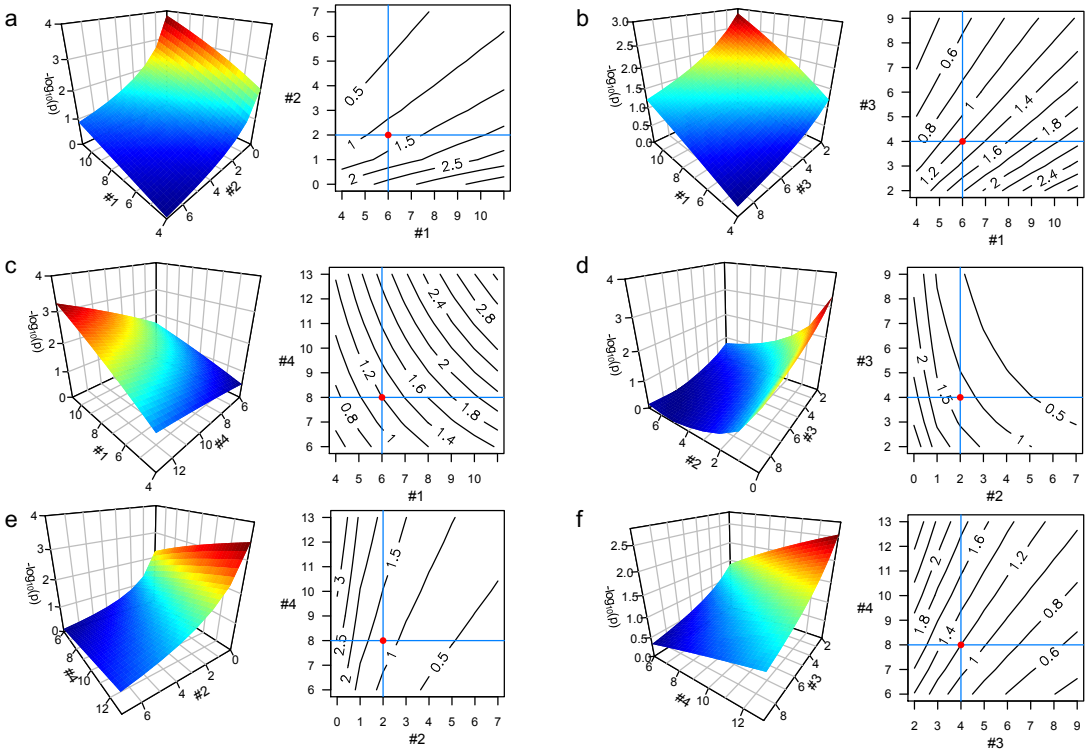


Figure 3. Extended GDT for the polygraph data in Simonoff (2003) when entries (a) 1|2, (b) 1|3, (c) 1|4, (d) 2|3, (e) 2|4 and (f) 3|4 are simultaneously changed. The blue lines correspond to the initial number of observations in that pair of entries, and the red dot to $-\log_{10}(p_M) = 1.2$. Values of $-\log_{10}(p) > 1.3$ indicate rejection of the independence model at 5% significance level.

**2. Postoperative delirium and statins.** In this example, we analyse the case of a stratified CT showing whether individuals undergoing cardiac surgery (and who reported previous use of statins) suffered postoperative delirium [Table 3]statins. We will determine whether the use of statins is associated with postoperative delirium using *e*GDT with the $G^2$ statistic. In contrast to the original study, here gender or other confounding variables are not controlled.

Our results indicate that the use of statins is associated with a reduction of postoperative delirium episodes in older ($G^2_M = 4.634, p_M = 0.031$), but not in younger individuals ($G^2_M = 0.318, p_M = 0.572$). As the authors discussed, this latter result is a consequence of

ageing as patients taking statins tend to be older and hence more likely to have delirium. In other words, the statins effect is attenuated in individuals with $< 60$ years of age.

When the GDT is applied to Table 3, several intriguing changes take place. In particular, adding three observations to cell $(1, 1)$ or seven observations in cell $(2, 2)$, or removing two or 25 observations from cell $(2, 1)$, would have resulted in no effect of statins use on the older group (Figure 4a and 4b, left). Furthermore, adding eight observations to cell $(2, 1)$, or removing five or nine observations from cells $(1, 1)$ and $(2, 1)$, respectively, would have resulted in an association between the use of statins and postoperative delirium in individuals with $< 60$ years of age (Figure 4a and 4b, right). This latter result, opposite to that in Katznelson et al. (2009), shows how fragile the conclusions are to small perturbations of the original data.

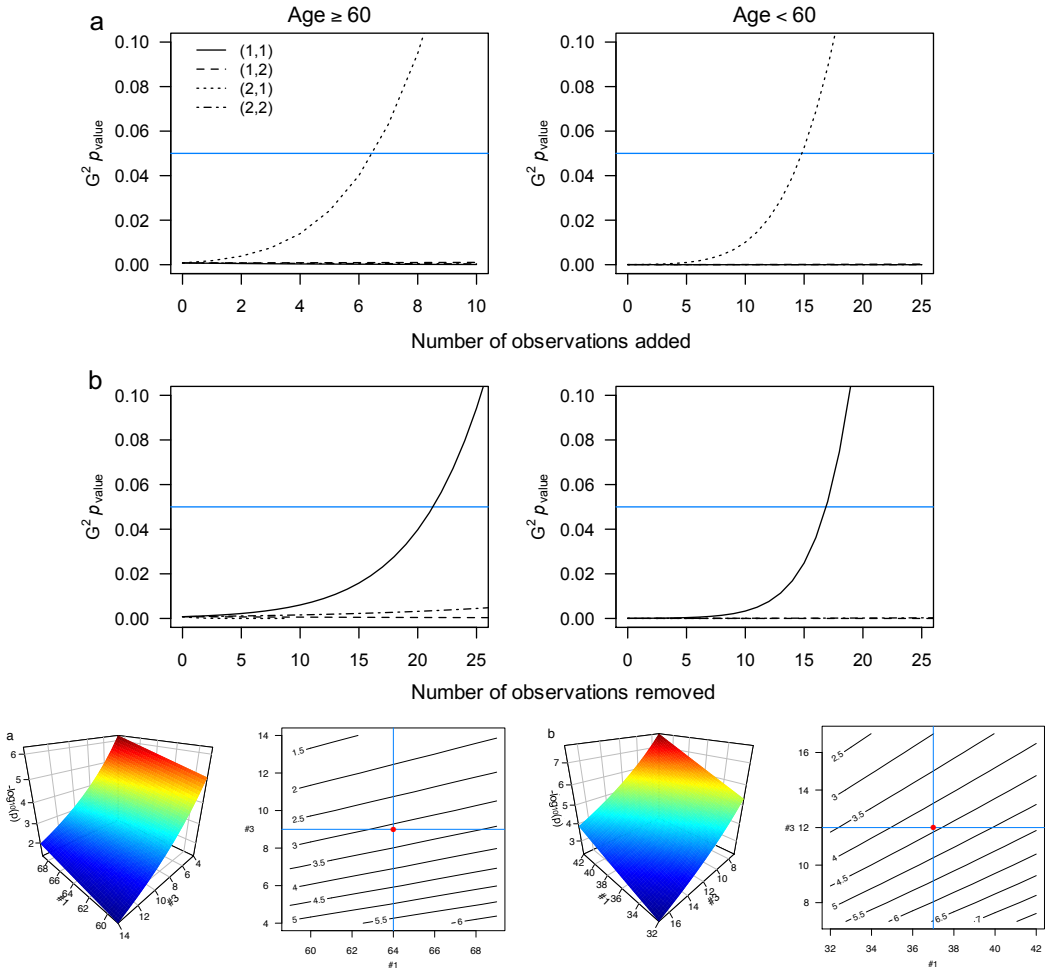


Figure 4. Extended GDT for the statin data in Katznelson et al. (2009). Observations are (a) added or (b) removed from individuals cells in Table 3. The 3D and contour representation of simultaneous changes in entries 1|3 of the resulting two-way CTs are shown for (c) age $\geq 60$ and (d) age $< 60$ years. The red dot is $-\log_{10}(p_M) = 1.51$ in (c) and $-\log_{10}(p_M) = 0.242$ in (d). Conventions as in Figures 1 and 3.

We also used the *e*GDT by changing entries 1|3 simultaneously. The results are presented in Figures 4c and 4d. In individuals aged $\geq 60$ years, changing $1^0|3^r$ with $r \leq -5$, and $1^r|3^0$ with $r > 3$, would have resulted in a completely different conclusion (that is, no effect of statin on postoperative delirium after cardiac surgery). Likewise, by changing $1^0|3^r$ with $r > 13$, and $1^r|3^0$ with $r < -6$, we would have concluded that statins reduce postoperative delirium in patients aged $< 60$ years. Once again, this result shows how small and simultaneous changes in two entries of the original CT could change our initial

conclusion entirely.

Table 3. Use of statins and presence of postoperative delirium by age. Source: Katznelson et al. (2009).

| Age | Statins | Delirium | |
|---|---|---|---|
| | | Yes | No |
| $\geq 60$ years | Yes | 64 | 424 |
| | No | 37 | 149 |
| $< 60$ years | Yes | 9 | 179 |
| | No | 12 | 185 |

**3. Smoking and Lung cancer.** Here we use the data presented in Table 1. Following Agresti (2002, pp. 46), the odds ratio is $(688 \times 59)/(650 \times 21) = 2.97$, the 95% confidence interval is $(1.78, 4.94)$ and the associated $p$-value is $p_M = 1.38 \times 10^{-5}$. Altogether, these results indicate that the odds of lung cancer for smokers is $\sim 3$ times that of nonsmokers, which in turn implies, as already known, that smoking is the main cause of lung cancer American Lung Association (2016).

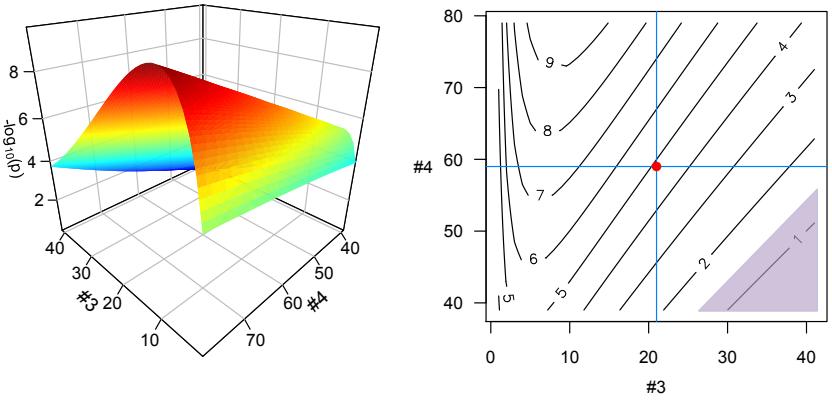


Figure 5. 3D and contour plots for the $e$GDT when entries 3|4 of Table 1 are simultaneously changed. The red dot corresponds to $-\log_{10}(p_M) = 4.86$. Combinations within the purple region on the right give $-\log_{10}(p) < 1.3$ and hence produce odds ratios that are not statistically significant at the 5% level.

Figure 5 shows the results of the $e$GDT applied to Table 1 when entries 3 (i.e., no smokers who developed lung cancer) and 4 (i.e., smokers who developed lung cancer) are simultaneously changed, and the odds ratio is used as the test statistic of interest to quantify the potential association between smoking habits and the development of lung cancer. Following previous recommendations Vélez et al. (2016), a total of $k = \{-20, -19, \ldots, 19, 20\}$ observations were independently added to these entries. Under this set up, a total of 128 combinations of entries 3|4 resulted in odds ratios that were not statistically significant (purple region, Figure 5); some of these combinations are $3^{+7}|4^{r}$ with $r = \{-19, -20\}$, and $3^{+20}|4^{r}$ with $-20 \leq r \leq -4$. Thus, whether including seven nonsmokers' cases and excluding 19 nonsmokers' controls, or adding 20 nonsmokers' cases and removing at least four nonsmokers controls would have led to a wrong conclusion that there is no association between smoking habits and the development of lung cancer.

## 4. Discussion

We have previously proposed a method that graphically represents the effects of modifying the number of observations in $I \times J$ CTs on the values of an association test diagtable. One of the drawbacks of such a method is that observations can be added or removed

one at a time in one cell while the other cells are kept constant. The current extension, *e*GDT, deals with that drawback by allowing the simultaneous modification of pairs of cells; that is, observations can be added or removed from two cells in tandem while the other cells are held constant. Furthermore, the *e*GDT can be applied to multiway CTs that are amenable to be broken down into several $I \times J$ CTs. In a couple of examples, it was illustrated how the *e*GDT can be used to model CT data and visualise several significant and non-significant scenarios given the data.

Our current implementation of the *e*GDT can be easily scaled to epidemiological studies to determine the effect of small-to-large changes in pair of cells on the difference between two proportions, the relative risk of individuals exposed and nonexposed to a particular event, or the odds ratio (as illustrated in Example 3). In order to take full advantage of the characteristics of the *e*GDT, the $\chi^2$ and $G^2$ tests, commonly used to establish whether two categorical variables $X$ and $Y$ are not independent McHugh (2013), can be substituted by other statistical measures of independence. In particular, measures such as the $\phi$ and $\gamma$ coefficients, the $C$ contingency coefficient, the Cramér's $V$ statistic, Yule's $Q$ measure of correlation, or the uncertainty coefficient (Agresti, 2002, Chapter 2), might be of interest. Similarly, the Fisher's exact and Barnad's tests could also be incorporated.

Another area of further development is the control for confounding variables within the *e*GDT. Confounding variables are well-known in masking the potential association between an outcome variable and an explanatory variable under the presence of a second explanatory variable that is not controlled (McDonald, 2014, pp. 21). Ultimately, this confounding effect may cause the researcher to incorrectly analyse the results and therefore make incorrect conclusions about their implications Skelly et al. (2012); Pourhoseingholi et al. (2012). By controlling confounding variables within the *e*GDT, these negative consequences can be ameliorated Pourhoseingholi et al. (2012). Furthermore, disciplines such as psychology, genetics and engineering, all of which make use of loglinear and generalised linear models (GLMs) Nelder and Wedderburn (1972); Agresti (2002), will instantly benefit from the inclusion of this feature in future implementations of the *e*GDT. It is worth mentioning that, in the context of multi- or two-way CTs, Logistic and Poisson regression models Nelder and Wedderburn (1972); Zeileis et al. (2008) are the best choice for studying the potential relationship between an outcome of interest and a set of predictors, and it would be desirable to have versions of the *e*GDT that include such models.

Missing values are common in psychology, data science, engineering, social sciences and genetic research. The effects of missing data and how they are dealt with in statistical analyses have been extensively discussed Rubin (1976); Little and Rubin (2002). Several easy-to-use implementations of some imputation methods that "spot" missing data are already available in `R` Harrell et al. (2016); van Buuren and Groothuis-Oudshoorn (2011); Su et al. (2011); Stekhoven (2013); Honaker et al. (2011). In the context of two-way CTs, Correa and Vélez (2014) determined the effect of partially missing data on the $\chi^2$ test of association. When the $\chi^2$ test is used and the missing information is accounted for in the statistical analysis (instead of dropping it as usual), the authors found that the $\chi^2$ test of association tends to reject the null hypothesis of independence more often than when the missing information is dropped. This result holds whenever $m$, the proportion of missing information in the two-way CT, is negligible compared to the sample size $n$ (Correa and Vélez, 2014, section 3.2). Considering that two-way CTs arise frequently in many research fields, thus it is key to include partially missing data in the analyses. *e*GDT enables the researcher to evaluate and graphically represent the effect of small-to-large changes in the entries of the CT; this feature constitutes a step forward in promoting the use of good practices for statistical modelling.

Finally, we want to focus on some potential applications of our previously proposed GDT Vélez et al. (2016) and the *e*GDT proposed herein. Consider an experiment to evaluate

whether a newly developed vaccine performs better than the one which is already available (i.e., the *gold* standard). The results of such an experiment are usually arranged in a two-way CT, which is often referred to as a confusion matrix; both vaccines are compared using performance measures such as sensitivity, specificity, and the positive and negative predictive values Parikh et al. (2008). These performance measures can be easily added to the repertoire of the *e*GDT so it is possible to compare two competitive diagnostic tools (as in the example above), or determine whether a relatively unknown statistical method is a plausible alternative Salazar et al. (2012); Vélez et al. (2014, 2015). Further improvements may include an enhanced version of the *e*GDT that incorporates Bayesian inference Gelman et al. (2004); Kerman and Gelman (2006); Congdon (2005) of these performance measures, as well as an interactive implementation of the *e*GDT to better understand potential associations between categorical variables in complex cases. For this purpose, the `R` packages such as `shiny` Chang et al. (2017) and `manipulate` Allaire (2014) constitute well-suited alternatives.

## Acknowledgements

We thank the Editor and two anonymous reviewers for their useful the comments and suggestions, and to Ms. Rosie Gronthos for proofreading the manuscript. JIV is in debt with Ms. Yolima Espinosa Jaramillo for her unconditional support.

**Authors' contributions.** JIV conceived the study; JIV and FMR analysed the data and wrote the paper.

**Computational details.** The `R` code to perform the *e*GDT based on the model of independence for two-way contingency tables is available from the authors by request.

**Conflict of interest.** None.

## References

Agresti, A., 2002. Categorical Data Analysis, 2nd Edition. Wiley, New Jersey.

Allaire, J., 2014. manipulate: Interactive Plots for RStudio. R package version 1.0.1. URL: https://CRAN.R-project.org/package=manipulate

American Lung Association, 2016. Lung Cancer Fact Sheet. Accessed 2016-07-05. URL: http://goo.gl/KLd6FC

Andersen, E.B., 1992. Diagnostics in categorical data analysis. Journal of the Royal Statistical Society, Series B 54, 781-791.

Belsey, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

Carlier, A., Ewing, G., 1992. Longitudinal analysis of contingency tables with application to demographic data. Computational Statistics 7, 329-353.

Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., 2017. shiny: Web Application Framework for R. R package version 1.0.0. URL: https://CRAN.R-project.org/package=shiny

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, 121-133.

Congdon, P., 2005. Bayesian Models for Categorical Data. Wiley, New York.

Correa, J.C., 2002. Un estudio de habilidad manual en niños 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 $\chi^2$. Comunicaciones en Estadística 7, 189-199.

Cung, C., 2013. Crime and Demographics: An Analysis of LAPD Crime Data. M.Sc. dissertation. Department of Statistics, UCLA, Los Angeles, California. URL: http://bit.ly/1jkR7O1

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, 1544-6115.

Doll, R., Hill, A.B., Sep 1950. Smoking and carcinoma of the lung; preliminary report. British Medical Journal 2, 739-748.

Friendly, M., 1994. Mosaic displays for multi-way contingency tables. Journal of the American Statistical Association 89, 190-200.

Friendly, M., 1995. Conceptual and visual models for categorical data. The American Statistician 49, 153-160.

Friendly, M., 1999. Extending mosaic displays: marginal, conditional, and partial views of categorical data. Journal of Computational and Graphical Statistics 8, 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, 395-398.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd Edition. Chapman & Hall/CRC, Washington.

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, 60-64.

Grizzle, J.E., Starmer, C.F., Koch, G.G., 1969. Analysis of categorical data by linear models. Biometrics 25, 489-504.

Harrell, F.E., 2001. Regression Modelling Strategies: With applications to linear models, logistic regression, and survival analysis. Springer, New York.

Harrell, F.E., 2016. Hmisc: Harrell Miscellaneous. R package version 3.17-4. URL: https://CRAN.R-project.org/package=Hmisc

Honaker, J., King, G., Blackwell, M., 2011. Amelia II: A program for missing data. Journal of Statistical Software 45, 1-47.

Hosmer, D., Lemeshow, S., 1989. Applied Logistic Regression. Wiley, 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, 2173-2184.

Kamish, H.J., 1988. Contingency table estimation of genetic parameters and disease risks. Statistics in Medicine 7, 591-600.

Katznelson, R., Djaiani, G.N., Borger, M.A., Friedman, Z., Abbey, S.E., Fedorko, L., Karski, J., Mitsakakis, N., Carroll, J., Beattie, W.S., 2009. Preoperative use of statins is associated with reduced early delirium rates after cardiac surgery. Anesthesiology 110, 67-73.

Kerman, J., Gelman, A., 2006. Bayesian data analysis using R. Rnews 6, 21-24.

Kleijnen, J., 1999. Validation of models: statistical techniques and data availability. Discussion Paper 1999-104, Center for Economic Research, Tilburg University. URL: http://ideas.repec.org/p/dgr/kubcen/1999104.html

Little, R., Rubin, D., 2002. Statistical Analysis with Missing Data, 2nd Edition. Wiley, New York.

Lustbader, E.D., Moolgavkar, S.H., 1985. A diagnostic statistic for the score test. Journal of the American Statistical Association 80, 375-379.

MacCullagh, P., 2002. What is a statistical model? The Annals of Statistics 30, 1225-1310.
Marcus, A.H., Elias, R.W., 1998. Some useful statistical methods for model validation. Environmental Health Perspectives 106, 1541-1550.
McDonald, J.H., 2014. Handbook of Biological Statistics, 3rd Edition. Sparky House Publishing, Baltimore, Maryland.
McHugh, M.L., 2013. The $\chi^2$ test of independence. Biochemia Medica 23, 143-149.
Nelder, J.A., Wedderburn, W.M., 1972. Generalized linear models. Journal of the Royal Statistical Society, Series A 135, 370-384.
Olivier, J., Bell, M.L., 2013. Effect sizes for $2 \times 2$ contingency tables. PLoS ONE 8(3), e58777. doi: 10.1371/journal.pone.0058777
Parikh, R., Mathai, A., Parikh, S., Sekhar, G.C., and Thomas, R., 2008. Understanding and using sensitivity, specificity and predictive values. Indian Journal of Ophthalmology 56, 45-50.
Pearson, K., 1904. Mathematical Contributions to the Theory of Evolution. Dulau and Co.
Pourhoseingholi, M.A., Baghestani, A.R., Vahedi, M., 2012. How to control confounding effects by statistical analysis. Gastroenterol Hepatol Bed Bench 5, 79-83.
R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/
Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581-592.
Salazar, D.A., Vélez, J.I., Salazar, J.C., 2012. Comparison between SVM and logistic regression: which one is better to discriminate? Revista Colombiana de Estadstica 35, 223-237.
Simonoff, J.S., 1988. Detecting outlying cells in two-way contingency tables via backwards-stepping. Technometrics 30, 339-345.
Simonoff, J.S., 2003. Analyzing Categorical Data. Springer, New York.
Skelly, A.C., Dettori, J.R., Brodt, E.D., (2012). Assessing bias: the importance of considering confounding. Evidence-Based Spine-Care Journal 3, 9-12.
Snee, R.D., 1977. Validation of regression models: methods and examples. Technometrics 19, 415-428.
Stekhoven, D.J., 2013. missForest: Nonparametric Missing Value Imputation using Random Forest. R package version 1.4. URL: https://cran.r-project.org/web/packages/missForest/index.html
Su, Y.-S., Gelman, A., Hill, J., Yajima, M., 2011. Multiple imputation with diagnostics (mi) in R: opening windows into the black box. Journal of Statistical Software 45, 1-31.
Tsujitani, M., Koch, G.G., 1991. Residual plots for log odds ratio regression models. Biometrics 47, 1135-1141.
van Buuren, S., Groothuis-Oudshoorn, K., 2011. mice: multivariate imputation by chained equations in R. Journal of Statistical Software 45, 1-67.
Vélez, J.I., Correa, J., Arcos-Burgos, M., 2014. A new method for detecting significant $p$-values with applications to genetic data. Revista Colombiana de Estadística 37, 67-76.
Vélez, J.I., Marmolejo-Ramos, F., Correa, J.C., 2016. A graphical diagnostic test for two-way contingency tables. Revista Colombiana de Estadística 39, 97-108.
Vélez, J.I, Jack, C.A., Chuah, A., Buckley, B., Correa, J.C., Easteal, S., Arcos-Burgos, M., 2015. Cross validation of pooling/resampling GWAS using the WTCCC data. Molecular Biology and Genetic Engineering 3, 1-10.
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, 6-10.

Wickens, T.D., 1969. Multiway Contingency Tables Analysis for the Social Sciences. Psychology Press, Lawrence Erlbaum Associates, Inc.
Zeileis, A., Kleiber, C., Jackman, S., 2008. Regression models for count data in R. Journal of Statistical Software 27, 1-25.

## Appendix A. Extended graphical diagnostic test

The `R` code to perform the *e*GDT based on the model of independence for contingency tables includes the following three functions:

(1) `runAssocTest(x, c1, c2, k, type)`.

(2) `plotSurface(M, ...)`.

(3) `plotContour(M, ...)`.

where

| | |
|---|---|
| `x` | $2 \times 2$ contingency table. |
| `c1` | Limits of the $y$-axis. By default it is the interval `c(0,1)`. |
| `c2` | 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` | Colour of the horizontal line. By default `lin.col = 2` (red). |
| ... | Additional arguments passed to `plot`. See `?plot` in R for more details. |

```
### ----------------------
###   Example
### ----------------------
## load the R code
source('https://www.dropbox.com/s/yknmfji0wkp7n81/egdt-src.R?dl=1')

## contingency table in Example 1
  x <- matrix(c(6, 2, 4, 8), ncol = 2, byrow = TRUE)

## values of k
  k <- -2:5

## run association test when entries 1 and 2 are changed, and the LRT is used
  out <- runAssocTest(x, 1, 2, k, type = 'lrt')
```

```
## 3D and contour plotting
  par(mfrow = c(1, 2), mar = c(4, 2, 1, 1))
  plotSurface(out, bty = 'b2', expand = 1, zlim = c(0, 4), theta = -135, phi = 25,
              resfac = 5)
  plotContour(out)
```