Self-betrayal voters: The Spaniards Case

Pérez-Asurmendi, Patrizia and de Andrés Calle, Rocío

Universidad Complutense de Madrid, Universidad de Salamanca

5 December 2023

Self-betrayal voters: The Spaniards Case★

P. Pérez-Asurmendi[a], R. de Andrés Calle[b,*]

[a] *PRESAD Research Group and IMUVA, Universidad Complutense de Madrid, E28223 Madrid, Spain*
[b] *BORDA and PRESAD Research Groups, Research Excellence Unit GECOS and Multidisciplinary Institute of Enterprise (IME), Universidad de Salamanca, E37007 Salamanca, Spain*

ARTICLE INFO

Keywords:
Cohen's Kappa statistic; Concordance; Categorical data; Voting behaviour

Abstract

This contribution deals with the measurement of the concordance between two characteristics reported by individuals at different points in time. Although intuition tell us that the declared opinion would be the same each time, the evidence does not always show it. In these cases, several measures for calculating agreement can be found in the literature, but by using them, much of the available information is lost, since they do not take disagreement into account. To overcome this drawback, we propose to use Cohen's Kappa statistic, which is an easy-to-interpret tool to measure the agreement between two categorical characteristics considering also the disagreement. To illustrate its applicability, we analyze the concordance between declared voting intention and voting decision, and between declared sympathy for political parties before and after the election with data from the 2015 General Election in Spain.

1. Introduction

In recent years we have confronted with a renewed interest in the ability of surveys to predict voters' decisions in several electoral processes. The disagreement between the declared and the voted could be promoted by several reasons of diverse nature. Among them, it can be highlighted the adopted methodology on the survey, potential incentives for voters to lie or possible changes in voters' opinions during the time between the survey and the election, among others.

This contribution tries to shed some light on the measurement of the difference between what voters declared before the electoral process and what they did on election day. The general open question we are trying to answer is the following. Presented with exactly the same situation, will an individual interpret data the same way and record exactly the same value for the variable each time these data are collected? Intuitively, it might seem that one person would behave the same way with respect to exactly the same phenomenon every time, but that is far from be real in several situations as in voting frameworks. In fact, researchers studying social behavior through surveys recognize differences between opinions and decisions surveys, although these include the same issues and people. That is an inter-rater agreement problem.

★De Andrés Calle is grateful to the Junta de Castilla y León and the European Regional Development Fund (Grant CLU-2019-03) for the financial support to the Research Unit of Excellence "Economic Management for Sustainability" (GECOS). De Andrés Calle would like to express her gratitude to the Spanish Agencia Estatal de Investigación for the financial support (Grant PID2022-139469NB-I00).

★★Pérez-Asurmendi is grateful to the Fundación Areces and the Spanish Agencia Estatal de Investigación for the financial support (Grants CISP18A6216 and PID2022-139469NB-I00, respectively).

*Corresponding author

ORCID(s): 0000-0002-2626-4456 (P. Pérez-Asurmendi); 0000-0002-8387-1164 (R.d. Andrés Calle)

		Voters' decisions		
		Party 1	Party 2	Party 3
Declared intention	Party 1	250	130	120
	Party 2	120	20	160
	Party 3	30	150	20

Table 1
An academic illustration of the problem.

This paper proposes a unified approach for measuring individuals' self-agreement in categorical data. Here, the quantities that reflect individuals' self-agreement are based on observed proportions from multidimensional contingency tables. To illustrate the problem, let us assume that we focus on the following two questions that might arise in an election: which are the voting intentions of the agents and which are their effective voting decisions. Table 1 provides an academic example with the voting intentions and decisions of 1,000 individuals in a three-party election. There, we can see that some individuals report the same voting intention and decision; specifically, 250 voters declare their voting intention and decision about Party 1, and 20 in the cases of Parties 2 and 3. Nevertheless, we can also find that other individuals report divergent voting intentions and decisions. For instance, if we look at the rest of the values in the first row of the table, we can find that 130 of the voters who declare their intention to vote for Party 1 ultimately decide to choose Party 2, and 120 decide to choose Party 3.

As it can be seen in the afore mentioned example our framework problem is a self cross-classified population with respect to two classifications and, obviously, questions about the degree of association existing between the different classifications often arise. To deal with them, we can consider statistical methods; among them, most traditional measures or indices of association are based on the standard chi-square statistic or on an underlying joint normality assumption. To calculate the magnitude of the concordance between classifications, we consider several alternative measures, almost all based upon a probabilistic model for activity to which the cross-classification may typically lead. In this context, Cohen's Kappa statistic emerges as the most suitable measure to our concern. It is important to know that only the case in which the population is completely known is considered; therefore, the question of sampling or measurement error does not arise in this scenario.

The paper is organized as follows. Section 2 is devoted to the notation and some starting hypothesis to introduce the different possible concordance measures in Section 3. Section 4 is dedicated to explain the Cohen's Kappa statistic, its characteristics and its interpretation in our context. The application to the Spaniards case is presented in Section 5 while Section 6 concludes the article with some comments and some insights about future research.

		D					
		1	2	3	···	α	**Total**
O	1	ρ_{11}	ρ_{12}	ρ_{13}	···	$\rho_{1\alpha}$	ρ_{1+}
	2	ρ_{21}	ρ_{22}	ρ_{23}	···	$\rho_{2\alpha}$	ρ_{2+}
	3	ρ_{31}	ρ_{32}	ρ_{33}	···	$\rho_{3\alpha}$	ρ_{3+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	α	$\rho_{\alpha 1}$	$\rho_{\alpha 2}$	$\rho_{\alpha 3}$	···	$\rho_{\alpha\alpha}$	$\rho_{\alpha +}$
	Total	ρ_{+1}	ρ_{+2}	ρ_{+3}	···	$\rho_{+\alpha}$	**1**

Table 2
Problem framework based on proportions.

2. Notation and starting hypothesis

This section introduces notation and several starting hypotheses necessary to define different concordance measures.

Let $\mathbf{N} = \{1, 2, ..., N\}$ be a set of agents than can be cross-classified regarding two issues, their opinion and their decision on someone or something. Let $\mathbf{X} = \{O\,(\text{opinion}), D\,(\text{decision})\}$ be the set of issues for which the agents are cross-classified (depending on the problem, the course of action may differ). Considering the both issues, agents are classified into α different unordered categories, $i = \{1, \ldots, \alpha\}$.

Let $\mathcal{P}$ be a *cross-clasiffication profile* of the set of agentes $\mathbf{N}$ on the set of issues $\mathbf{X}$:

$$\mathcal{P} = \begin{pmatrix} \rho_{11} & \cdots & \rho_{1\alpha} \\ \vdots & \ddots & \vdots \\ \rho_{\alpha 1} & \cdots & \rho_{\alpha\alpha} \end{pmatrix}_{\alpha \times \alpha}$$

where ρ_{ii} represents the proportion of the agents cross-classified as the i-th category for the both issues. Note the absence of category order implies equal consideration of all paired proportions. We write $\mathbf{P}_{\alpha\times\alpha}$ for the set of all cross-clasiffication profiles.

For one thing, define ρ_{i+} to be the proportion of the agents classified as i attending to the issue D. Therefore, ρ_{i+} can be computed like the sum of the elements of row i of the profile $\mathcal{P}$. For another, ρ_{+i} captures the proportion of the agents classified as i attending to the issue O. In the same way, ρ_{+i} can be computed like the sum of the elements of the column i of the profile $\mathcal{P}$.

An alternative formal modelization of the same information on the agents' cross-classification may be represented by a double-entry table. Table 2 gathers this information.

		Voters' decisions			
		Party 1	Party 2	Party 3	**Total**
Voters' intentions	Party 1	0.25	0.13	0.12	0.50
	Party 2	0.12	0.02	0.16	0.30
	Party 3	0.03	0.15	0.02	0.20
	Total	0.40	0.30	0.30	**1**

Table 3
Academic illustration of the problem.

In order to conduct this study, it is essential to consider the following starting assumptions. First, independence among the agents is assumed, i.e., neither the opinions nor the decisions of each individual determine the ones of the other agents. Second, it is assumed the issues are perceived as independent and there is not any cause-effect relationship between them. It is also assumed that the categories in each issue are independent, mutually exclusive, and exhaustive. Last, complete knowledge of the agents' categories' classifications is also taken for granted and there are no restrictions on how agents are distributed across categories for either issue.

Now and in order to improve understanding of the notation, the following illustrative example is introduced.

Example 1. *Let* $\mathbf{N} = \{1, 2, ..., N\}$ *be a set of voters than are cross-classified regarding their intention of vote and their decision of vote on three parties: Party 1, 2 and 3. The cross-clasiffication profile of the set of voters* $\mathbf{N}$ *on the set of issues* $\mathbf{X}$ *is:*

$$\mathcal{P} = \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{pmatrix}_{3\times3} = \begin{pmatrix} 0.25 & 0.13 & 0.12 \\ 0.12 & 0.02 & 0.16 \\ 0.03 & 0.15 & 0.02 \end{pmatrix}_{3\times3}$$

where ρ_{ii} *represents the proportion of the agents whose intentions of vote on the* i*-th party coincides with their decision of vote on the* i*-th party. Table 3 summarizes this information and includes the values of* ρ_{i+} *and* ρ_{+i}.

3. Concordance measurements

Categorizing agents based on multiple characteristics often leads to questions about the concordance between these classifications. Most traditional measures are derived from the standard chi-square statistic or assume underlying joint normality. This paper discusses the concordance between two variables, akin to the traditional statistical concept and a unified approach to measure the agreement for categorical data is proposed. Consistent with traditional literature, the concordance between the agents' cross-classification is quantified through observed proportions in contingency tables.

It is important to highlight that the specific measures of concordance described here are not presented as universally applicable measures. These suggestions should be seen as reasonable choices in certain scenarios, but it should be important to also consider and explore other measures. Then, the following definition is proposed to formalize the measurement of concordance between cross-classifications.

		D					
		1	2	3	$\cdots$	α	**Total**
O	1	ρ_{11}	ρ_{12}	ρ_{13}	$\cdots$	$\rho_{1\alpha}$	ρ_{1+}
	2	ρ_{21}	ρ_{22}	ρ_{23}	$\cdots$	$\rho_{2\alpha}$	ρ_{2+}
	3	ρ_{31}	ρ_{32}	ρ_{33}	$\cdots$	$\rho_{3\alpha}$	ρ_{3+}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	α	$\rho_{\alpha 1}$	$\rho_{\alpha 2}$	$\rho_{\alpha 3}$	$\cdots$	$\rho_{\alpha\alpha}$	$\rho_{\alpha +}$
	Total	ρ_{+1}	ρ_{+2}	ρ_{+3}	$\cdots$	$\rho_{+\alpha}$	**1**

Table 4
Row observed agreement, ROA.

Definition 1. *Let $\mathcal{P} \in \mathbf{P}_{\alpha\times\alpha}$ be a cross-clasiffication profile of the population* $\mathbf{N}$. *A cross-classification concordance measure is a mapping δ, that assigns a number $\delta(\mathcal{P})$ to each cross-clasiffication profile $\mathcal{P}$, with the property:*

i) $\delta(\mathcal{P}) = 1$ *when (and only when) there is perfect concordance between the classifications.*

Definition 1 rests on a low demanding condition and it is so general that can suit with diverse measures. In the literature, situations of the type herein considered have been variously handled.

Most of the research that has dealt with agreement among agents when categorical data are available is concerned with the derivation of some descriptive measures of such an agreement (see Gwet (2012)). The most primitive approach known as the *Raw Observed Agreement* (ROA) simply counts up the proportion of cases in which the classifications on the set of agents are agreed and its formal definition could be the following one.

Definition 2. *Let $\mathcal{P} \in \mathbf{P}_{\alpha\times\alpha}$ be a cross-clasiffication profile of the population* $\mathbf{N}$. *The Raw observed agreement (ROA) is a mapping $\delta_o : \mathbf{P}_{\alpha\times\alpha} \longrightarrow [0,1]$ that assigns a number $\delta_o(\mathcal{P})$ to each cross-clasiffication profile $\mathcal{P}$ as follows:*

$$\delta_o(\mathcal{P}) = \sum_{i=1}^{\alpha} \rho_{ii}$$

It is crucial to emphasize that the ROA ranges from 0 to 1. Thus, the former is got when there is minimum concordance between cross-classifications, i.e., when there is a maximum disagreement. And the latter denotes the maximum concordance, i.e., there is a perfect agreement between cross-classifications. Following the aforementioned alternative modelization represented by a double-entry table, in Table 4 are shown the values considered to compute $\delta_o(\mathcal{P})$.

Example 2. *Let's assume the data presented in Example 1. Table 5 shows the proportions of the voters' intentions and the voters' decisions included in the computation of $\delta_o(\mathcal{P})$. Formally,*

$$\delta_o(\mathcal{P}) = \sum_{i=1}^{3} \rho_{ii} = 0.25 + 0.02 + 0.02 = 0.29$$

		Voters' decisions			
		Party 1	Party 2	Party 3	**Total**
Voters' intentions	Party 1	0.25	0.13	0.12	0.50
	Party 2	0.12	0.02	0.16	0.30
	Party 3	0.03	0.15	0.02	0.20
	Total	0.40	0.30	0.30	**1**

Table 5
$\delta_o(\mathcal{P})$ for Example 1.

This result could be interpreted as follows:

1. *More than* 70% *of the voters show self-disagreement between their opinions and their decisions.*
2. *For instance,* 13% *and* 12% *of voters that declare their intention of vote for Party 1, finally decided to vote to Parties 2 and 3, respectively.*
3. *Furthermore, focusing on ρ_{+i} and ρ_{i+} proportions,* 50% *of the agents declare intention of vote on Party 1 although it only holds* 40% *of their voters' decision.*

As the academic example suggests, the $\delta_o(\mathcal{P})$ (ROA) measure may not be a suitable concordance measurement for this problem. It may be advantageous to use all available information in order to measure the agreement between the cross-classifications and capture the essence of such data. In other words, to consider not only the values relying on the agreement but also these relying on the disagreement. These kind of measures are defined as follows.

Definition 3. *Let $\mathcal{P} \in \mathbf{P}_{\alpha\times\alpha}$ be a cross-clasiffication profile of the population* **N**. *The Random corrected agreement (RCA) is a mapping $\delta_{RCA} : \mathbf{P}_{\alpha\times\alpha} \longrightarrow [-\infty, 1]$ that assigns a number $\delta_{RCA}(\mathcal{P})$ to each cross-clasiffication profile $\mathcal{P}$ as follows:*

$$\delta_{RCA}(\mathcal{P}) = \frac{\delta_o(\mathcal{P}) - p_e(\mathcal{P})}{1 - p_e(\mathcal{P})}$$

This could be directly interpretable as the proportion of joint classifications in which there is concordance, after chance concordance is excluded. It is based on two important quantities, the proportion of the population in which the classifications agreed, i.e., $\delta_o(\mathcal{P}) = \sum_{i=1}^{\alpha} \rho_{ii}$, and the proportion of the population for which agreement is fortuity expected, i.e., $p_e(\mathcal{P})$.

Figure 1 shows the RCA statistic interpretation when the values of the statistic are restricted to the unit interval. There, $\delta_o(\mathcal{P}) - p_e(\mathcal{P})$ is interpreted like the proportion of the real agreement cases, i.e., when those obtained by chance were ruled out, whereas $1 - p_e(\mathcal{P})$ represents the proportion of the cases for which the categories independence hypothesis would predict disagreement between the classifications.

It is worth noting that the chance concordance is not related to the idea of randomness. On the contrary, it is associated with the diverse reasons that could explain the divergence between cross-classifications. Take, for instance, the fact that some voters misrepresent their voting intention compared to their final decision, or that voters sometimes switch their vote after declaring their intention of vote.

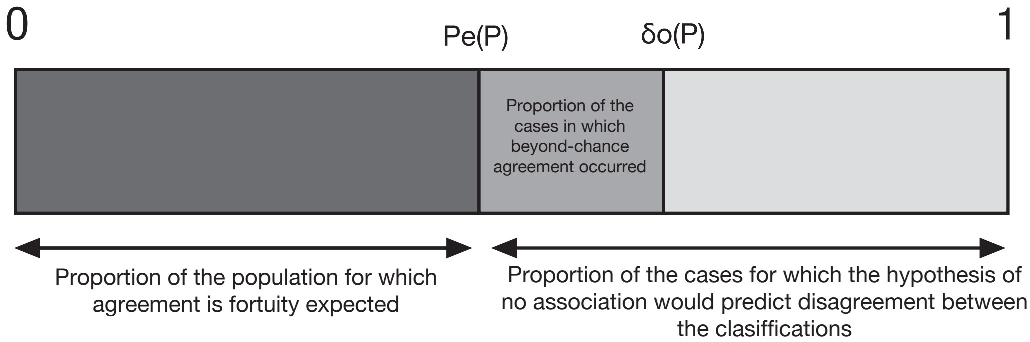

Figure 1: RCA statistic components, Abraira (2001).

The proportion of agreement by chance, $p_e(\mathcal{P})$, has been an object of research since the 1950s and it has been the subject of many classic studies. In Bennett, Alpert and Goldstein (1954), without the intention of being exhaustive, it is established as the inverse of the number of categories:

$$p_e(\mathcal{P}) = \frac{1}{\alpha}.$$

Therefore, the proportion is thought as a discrete uniform random variable.

Instead, in Scott (1955), $p_e(\mathcal{P})$ is associated with the sum of the squares of the averages of ρ_{i+} and ρ_{+i} for each category:

$$p_e(\mathcal{P}) = \sum_{i=1}^{\alpha} \left(\frac{\rho_{i+} + \rho_{+i}}{2} \right)^2,$$

behaving like a binomial random variable.

Going not so far Gwet (2008) considers that the proportion for agreement that occurs by chance has to do with the average of ρ_{i+} and ρ_{+i} in each category. To be more concrete:

$$p_e(\mathcal{P}) = \frac{1}{(\alpha - 1)} \sum_{i=1}^{\alpha} \pi_i \,(1 - \pi_i), \quad \text{where} \quad \pi_i = \frac{\rho_{i+} + \rho_{+i}}{2}, \quad \text{for} \quad i = 1, \ldots, \alpha$$

Other excellent definitions of these may be found in Tinsley and Weiss (1975), Landis and Koch (1977), Goodman and Kruskal (1979), Tinsley and Weiss (2000) and Gwet (2012), among others. The following section will introduce the Cohen's Kappa statistic, which is commonly used in empirical studies.

4. Cohen's Kappa statistic

The Cohen's Kappa statistic was introduced by Cohen in 1960 (Cohen, 1960). This method is utilized for assessing the agreement between two raters who assign ratings to a group of subjects using a nominal scale. Its main application

		D					
		1	2	3	⋯	α	**Total**
O	1	ρ_{11}	ρ_{12}	ρ_{13}	⋯	$\rho_{1\alpha}$	ρ_{1+}
	2	ρ_{21}	ρ_{22}	ρ_{23}	⋯	$\rho_{2\alpha}$	ρ_{2+}
	3	ρ_{31}	ρ_{32}	ρ_{33}	⋯	$\rho_{3\alpha}$	ρ_{3+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	α	$\rho_{\alpha 1}$	$\rho_{\alpha 2}$	$\rho_{\alpha 3}$	⋯	$\rho_{\alpha\alpha}$	$\rho_{\alpha+}$
	Total	ρ_{+1}	ρ_{+2}	ρ_{+3}	⋯	$\rho_{+\alpha}$	**1**

Table 6
Proportions of agreement due to chance in Cohen's statistic.

is in the fields of biostatistics and psychometrics research. Under Cohen's view, the proportion of agreement due to chance is understood as the sum of the product of ρ_{i+} and ρ_{+i} proportions for each category (see Table 6). In fact, Cohen pointed out that there is likely to be some level of agreement among data when population do not know the election results and they are merely guessing. Formally,

$$p_e(\mathcal{P}) = \sum_{i=1}^{\alpha} \rho_{i+}\ \rho_{+i}$$

Assuming the starting hypothesis about the independence of the agents, $\rho_{ii} = \rho_{i+}\ \rho_{+i}$ for $i = 1, \ldots, \alpha$. In broad terms, the observed agreement is equal to the agreement by chance when the ratings are completely unassociated. Then, each diagonal element of $\mathcal{P}$ is the product of the ρ_{i+} and ρ_{+i} proportions. These ideas determine the definition of the Cohen's RCA.

Definition 4. *Let $\mathcal{P} \in \mathbf{P}_{\alpha\times\alpha}$ be a cross-clasiffication profile of the population $\mathbf{N}$. The Cohen's Kappa statistic is a mapping $\delta^{\kappa}_{RCA} : \mathbf{P}_{\alpha\times\alpha} \longrightarrow [-1, 1]$ that assigns a number $\delta^{\kappa}_{RCA}(\mathcal{P})$ to each cross-clasiffication profile $\mathcal{P}$ as follows:*

$$\delta^{\kappa}_{RCA}(\mathcal{P}) = \frac{\delta_o(\mathcal{P}) - p_e(\mathcal{P})}{1 - p_e(\mathcal{P})} = \frac{\sum_{i=1}^{\alpha} \rho_{ii} - \sum_{i=1}^{\alpha} \rho_{i+}\, \rho_{+i}}{1 - \sum_{i=1}^{\alpha} \rho_{i+}\, \rho_{+i}}$$

This statistic is bounded by 1 in the case of maximum concondance and -1 in the opposite case. Going deeper into the meaning of the statistic is easy to see that:

- When observed agreement is greater than fortuitous agreement, it takes values above zero:

$$\delta^{\kappa}_{RCA}(\mathcal{P}) > 0 \longrightarrow \delta_o(\mathcal{P}) > p_e(\mathcal{P})$$

Kappa value	Level of agreement	% of trustworthy information
< 0	Less than chance agreement	-
$0 - 0.2$	None agreement	$0 - 4\%$
$0.2 - 0.4$	Minimal agreement	$15 - 35\%$
$0.4 - 0.6$	Weak agreement	$15 - 35\%$
$0.6 - 0.7$	Moderate agreement	$35 - 63\%$
$0.8 - 0.9$	Strong agreement	$64 - 81\%$
> 0.90	Almost perfect agreement	$82 - 100\%$

Table 7
Interpretation of Cohen's Kappa statistic.

- When observed agreement and fortuitous agreement are identical, its value becomes zero:

$$\delta^{\kappa}_{RCA}(\mathcal{P}) = 0 \longrightarrow \delta_o(\mathcal{P}) = p_e(\mathcal{P})$$

- When observed agreement is lower than fortuitous agreement, it takes values below zero:

$$\delta^{\kappa}_{RCA}(\mathcal{P}) < 0 \longrightarrow \delta_o(\mathcal{P}) < p_e(\mathcal{P})$$

Notice that in such a case, $\delta^{\kappa}_{RCA}(\mathcal{P})$ represents agreement worse than expected, or disagreement.

Also, the Cohen's Kappa is one of the most widely used because there is an interpretation of its different values (see Landis and Koch (1977) and McHugh (2012)). Table 7 collects its interpretation.

Example 3. *Based on the data provided in Example 1, the highlighted values used to calculate the Cohen's Kappa statistic are in Table 8.*

$$\delta_o(\mathcal{P}) = \sum_{i=1}^{3} \rho_{ii} = 0.25 + 0.02 + 0.02 = 0.29$$

$$p_e(\mathcal{P}) = \sum_{i=1}^{3} \rho_{i+}\ \rho_{+i} = 0.50 \cdot 0.40 + 0.30 \cdot 0.30 + 0.20 \cdot 0.30 = 0.35$$

$$\delta^{\kappa}_{RCA}(\mathcal{P}) = \frac{\delta_o(\mathcal{P}) - p_e(\mathcal{P})}{1 - p_e(\mathcal{P})} = \frac{0.29 - 0.35}{1 - 0.35} = -0.248$$

In this example, $\delta^{\kappa}_{RCA}(\mathcal{P})$ reaches the value of -0.248, a negative value that means that the observed agreement is lower than the expected by chance. To clarify, there is disagreement between cross-classifications.

One major issue in Cohen's Kappa statistic research concerned is the limits of the coefficient:

Self-betrayal voters: The Spaniards Case

		Voters' decisions			
		Party 1	Party 2	Party 3	**Total**
Voters' intentions	Party 1	0.25	0.13	0.12	0.50
	Party 2	0.12	0.02	0.16	0.30
	Party 3	0.03	0.15	0.02	0.20
	Total	0.40	0.30	0.30	**1**

Table 8
Cohen's Kappa statistic in our academic example.

		Voters' decisions		
		Party 1	Party 2	**Total**
Voters' intentions	Party 1	0	0.50	0.50
	Party 2	0.50	0	0.50
	Total	0.50	0.50	**1**

Table 9
The lower limit in two categories characteristics example.

- The upper limit: Perfect concordance o agreement.

  This happens when (and only when) there is complete agreement in classifying the characteristics, that is, when the off-diagonal (disagreement) cells are zero. Formallly,

$$\delta^{\kappa}_{RCA}(\mathcal{P}) = 1 \iff \delta_o(\mathcal{P}) = 1 \iff \rho_{i+} = \rho_{+i} = 0 \quad \text{for all} \quad i = 1, \dots, \alpha$$

  Notice that any $\delta^{\kappa}_{RCA}(\mathcal{P})$ less than the upper limit's value is a measure not only of agreement, but also of the reverse, i.e., the disagreement among classifications.

- The lower limit: Perfect disagreement.

  Keeping in mind the lower limit of the statistic, that is the value reflecting perfect disagreement, we found that as far as it depends on the marginal proportions, it is not always reachable. To illustrate that we can formulate several examples. In Table 9 we have that the observed agreement takes a value equal to 0 while the disagreement is maximum. In fact, there, voters decide just the contrary that they declare. Therefore, the value of Cohen's Kappa statistic reaches -1, the minimum posible value. When we face characteristics with three or more categories, there are more ways that one might disagree and it therefore quickly gets more complicated to disagree perfectly. In such cases, a situation should minimizes agreement in any combination. This would likely be a situation where there are no counts in some cells because it would be impossible to have perfect disagreement across all combinations at the same time.

		Voters' decisions			
		Party 1	Party 2	Party 3	**Total**
Voters' intentions	Party 1	0	0	0.12	0.12
	Party 2	0	0	0	0
	Party 3	0.88	0	0	0.88
	Total	0.88	0	0.12	**1**

		Voters' decisions			
		Party 1	Party 2	Party 3	**Total**
Voters' intentions	Party 1	0	0.40	0	0.40
	Party 2	0.60	0	0	0.60
	Party 3	0	0	0	0
	Total	0.60	0.40	0	**1**

The Cohen's Kappa statistic also has the benefit of computing an approximation to the standard error.

$$\sigma_{\delta^{\kappa}_{RCA}}(\mathcal{P}) = \sqrt{\frac{\delta_o(\mathcal{P})\left(1-\delta_o(\mathcal{P})\right)}{N\left(1-p_e(\mathcal{P})\right)^2}}$$

This formula is an approximation since treats $p_e(\mathcal{P})$ as a constant and treats $\delta_o(\mathcal{P})$ as if it were the population value. It should be adequate since ordinarily $p_e(\mathcal{P})$ will not vary greatly relative to α, particularly with N large (i.e., $\geqq 100$).

With N large, the sampling distribution of $\delta^{\kappa}_{RCA}(\mathcal{P})$ will approximate normality so that confidence limits can be set in the usal way:

$$95\% \quad \text{confidence limits} = \delta^{\kappa}_{RCA}(\mathcal{P}) \pm 1.96\, \sigma_{\delta^{\kappa}_{RCA}}(\mathcal{P})$$
$$99\% \quad \text{confidence limits} = \delta^{\kappa}_{RCA}(\mathcal{P}) \pm 2.58\, \sigma_{\delta^{\kappa}_{RCA}}(\mathcal{P})$$

5. A case of study: The Spaniards case

This section is devoted to present the results obtained for data on the panel study accomplished by the Spanish Sociological Research Center (CIS) for Spanish General Election in 2015 through two studies, specifically a pre-electoral one (Study 3117) and a post-electoral one (Study 3126). It should be noted that the Spanish democracy is a highly decentralized system in which we can find four different level of elections, namely, European, National or General, Regional and Municipal elections. In the General elections, voters elect the 350 seats in the Spanish Congress of Deputies and the majority of the members of the Spanish Senate (208 out of 266 senators). These two processes used to be held at the same time, but the electoral system for each is slightly different. In the first case, each voter casts a vote or does not for a party, while in the second one, she can directly elect the candidate to be a senator. Then, the d'Hont rule assigns the votes to the seats in both, the Congress and the Senate.

As mentioned above, data were collected before and after November 20th, 2015 election. The first phase took place from October 27th to November16th, 2015, whilst the second went from January 7th to March 19th, 2016. The pre-election survey interviewed a sample of 17,452 likely voters, while the post-election survey involved 6,242 likely voters.Therefore, we have complete information on 6,242 individuals. Nevertheless, we have restricted the sample to individuals whose characteristics have to do with parties competing throughout the national territory (hereinafter referred to as national parties).

With these data, we study two different situations. In the first one, we analyze the concordance between declared voting intention and declared vote. In the second one, we evaluate the concordance between declared sympathy for a party before and after the elections.

5.1. Concordance between declared voting intention and declared vote

In order to perform the analysis between declared voting intention and declared vote, we restrict the data to the 5,508 possible voters who did not declare their intention or vote for parties that did not compete in the whole Spanish territory (in the remainder, regional parties). We consider the following categories for both characteristics: national parties, i.e., PP, PSOE, IU, UPyD, C's and Podemos, and invalid ballots (IB), blank ballots (BB), abstention (ABS), don't know (DK) and decline to answer (DA).

Table 10 shows the frequencies of these data. Those relevant for the calculation of observed agreement are highlighted in yellow, while marginal proportion data are highlighted in red. At a first glance, we can observe several disparities between the observed agreement and the corresponding marginal proportions. And that happens with the big traditional Spanish parties, i.e., PP and PSOE, but also with the rest of them. In the case of PP, the winning party in that election, 15.3% of voters declared their intention and vote for it, while the total intention and declared vote reached 19.7% in the former, and 22.3% in the latter. Something similar happened in the case of the second most voted party, which was PSOE. If we look at the rest of the parties, the case of Podemos stands out. About 8.5% of voters declared their intention and vote in favor. However, the total intention declared by that party reached 10.5% and the voters who declared voting for it represented 16.1% of the individuals.

Tables 11 and 12 contain the proportions when filtering by gender. The differences between the observed agreement and the marginal proportions are slightly greater in the case of men than women considering four of the six parties, specifically in the cases of PP, UPyD, C's and Podemos.

Cohen's Kappa statistics, their corresponding confidence limits at 95% of significance and the values for the significance tests are collected in Table 13, considering the complete sample (General in the Table), filtering by gender and also by party. Notice that if we filter by party, we take into account both declared voting intention and declared vote. That means that a voter who declares intention to vote for a given party and another who declares her vote for it are considered voters in favor of that particular party. None of the statistics is negative and the only one statistically equal to zero corresponds to UPyD where the observed agreement is equal to that expected by chance. The highest value corresponds to the data on declared voting intention and vote for the PP. In this instance, less than 35% of voters who declared voting intention for the PP do not agree with their declared vote. The rest of the statistics reflect a moderate agreement between the two declarations. Taking into account the entire sample, at least half of the voters declared an intention different from their declared vote.

5.2. Concordance between sympathy for a party before and after the elections

We have also studied the change in declared sympathy by the parties after and before the elections. Somehow we can interpret the results in terms of the voters' satiety with the electoral campaign and the electoral process. We work with a sample of 5,532 possible voters who do not declare any sympathy, neither before nor after, for regional parties. The categories included in the two characteristics, namely, declared sympathy before and after the election are the following: PP, PSOE, IU, UPyD, C's, Podemos, None, Don't Know (DK) and Decline to Answer (DA). It is worth mentioning that before the elections, only 1,489 possible voters declared None, Don't Known or Decline to Answer whilst after the elections, the number increases to 3,059. It could indicate that some of those interviewed became very tired of the political competition.

We have also compare Cohen's Kappa statistics in pairs to find out if there are significant differences between them. To do this, we test the following hypotheses:

$$H_0 : \delta^{\kappa}_{RCA}\left(\mathcal{P}_i\right) = \delta^{\kappa}_{RCA}\left(\mathcal{P}_j\right)$$

against the inequality of them, considering all the possible comparisons between the six national parties. Table 14 shows that all of them can be considered unequal.

Tables 15, 16, and 17 contain the proportions for the 5,232 voters, the men and the women. In the three cases, the most striking result appears in the category None. As a sample, if all voters are taken into account, 16.3% declared that category before and after the election. If we look at the marginal proportions, we find 20.3% before and a surprising 52.9% after the election. This fact is even more pronounced in the case of women, where the marginal proportion for None after the election reached 54.1%.

These results agree with the Cohen's Kappa statistics obtained (see Table 18). More than 60% of voters answered differently when expressing their sympathies before and after the election. The two lowest statistics are those of UPyD and C's. The first one is not statistically significant given its high standard error, so it could be interpreted that there is no agreement. The second one reflects minimal agreement between before and after sympathies towards C's.

Finally, as can be seen in Table 19, we cannot reject the hypothesis of equality of Cohen's Kappa statistic in the case of UPyD and C's.

		Declared vote											
		PP	PSOE	IU	UPyD	C's	Podemos	IB	BB	ABS	DK	DA	**Total**
Declared intention	PP	0.15359	0.00527	0.00054	**0**	0.00690	0.00091	**0**	0.00127	0.01852	**0**	0.01053	0.19753
	PSOE	0.00763	0.12745	0.00236	**0**	0.00309	0.01235	0.00036	0.00145	0.01416	**0**	0.00944	0.17829
	IU	0.00054	0.00345	0.01761	**0**	0.00091	0.00999	0.00018	0.00018	0.00109	**0**	0.00290	0.03686
	UPyD	0.00018	0.00018	**0**	0.00036	0.00018	0.00036	**0**	0.00018	0.00018	**0**	0.00000	0.00163
	C's	0.01416	0.01307	0.00218	0.00073	0.07135	0.01107	0.00054	0.00109	0.00817	**0**	0.00708	0.12945
	Podemos	0.00073	0.00436	0.00327	**0**	0.00163	0.08551	0.00018	0.00000	0.00672	**0**	0.00345	0.10585
	IB	0.00036	0.00036	0.00054	**0**	0.00109	0.00054	0.00182	0.00036	0.00073	**0**	0.00036	0.00617
	BB	0.00418	0.00309	0.00054	0.00036	0.00363	0.00290	0.00073	0.00490	0.00744	**0**	0.00290	0.03068
	ABS	0.00781	0.00563	0.00127	0.00018	0.00490	0.00617	0.00036	0.00127	0.04793	**0**	0.00708	0.08261
	DK	0.02760	0.03086	0.00617	0.00091	0.02015	0.02633	0.00145	0.00309	0.02505	0	0.03958	0.18119
	DA	0.00635	0.00708	0.00109	**0**	0.00272	0.00508	0.00018	0.00054	0.00508	**0**	0.02160	0.04975
	Total	0.22313	0.20080	0.03558	0.00254	0.11656	0.16122	0.00581	0.01434	0.13508	0	0.10494	1.00000

IB: Invalid Ballot BB: Blank Ballot ABS: Abstention DK: Don't Know DA: Decline to Answer

Table 10
Declared voting intention and declared vote.

Men		Declared vote											
		PP	PSOE	IU	UPyD	C's	Podemos	IB	BB	ABS	DK	DA	**Total**
Declared intention	PP	0.15920	0.00753	0.00038	0.00000	0.00979	0.00151	0.00000	0.00075	0.01995	0.00000	0.01091	0.21001
	PSOE	0.00640	0.11216	0.00226	0.00000	0.00376	0.01204	0.00038	0.00038	0.00979	0.00000	0.00677	0.15393
	IU	0.00000	0.00263	0.01844	0.00000	0.00000	0.01242	0.00000	0.00000	0.00113	0.00000	0.00339	0.03801
	UPyD	0.00038	0.00038	0.00000	0.00075	0.00038	0.00038	0.00000	0.00038	0.00038	0.00000	0.00000	0.00301
	C's	0.01581	0.01242	0.00188	0.00000	0.08054	0.01393	0.00075	0.00113	0.00979	0.00000	0.00828	0.14452
	Podemos	0.00113	0.00527	0.00301	0.00000	0.00113	0.10425	0.00000	0.00000	0.00715	0.00000	0.00414	0.12608
	IB	0.00075	0.00075	0.00038	0.00000	0.00075	0.00075	0.00188	0.00038	0.00038	0.00000	0.00038	0.00640
	BB	0.00414	0.00339	0.00038	0.00038	0.00376	0.00452	0.00075	0.00452	0.00715	0.00000	0.00151	0.03049
	ABS	0.00941	0.00527	0.00151	0.00000	0.00489	0.00979	0.00038	0.00188	0.04893	0.00000	0.00640	0.08845
	DK	0.02559	0.02936	0.00602	0.00075	0.01505	0.02409	0.00113	0.00151	0.01694	0.00000	0.02898	0.14942
	DA	0.00715	0.00565	0.00038	0.00000	0.00339	0.00489	0.00038	0.00075	0.00602	0.00000	0.02108	0.04968
	Total	0.22996	0.18479	0.03463	0.00188	0.12345	0.18856	0.00565	0.01167	0.12759	0.00000	0.09183	1.00000

IB: Invalid Ballot BB: Blank Ballot ABS: Abstention DK: Don't Know DA: Decline to Answer

Table 11
Declared voting intention and declared vote by men.

Women		Declared vote											
		PP	PSOE	IU	UPyD	C's	Podemos	IB	BB	ABS	DK	DA	**Total**
Declared intention	PP	0.14837	0.00316	0.00070	**0**	0.00421	0.00035	**0**	0.00175	0.01719	**0**	0.01017	0.18590
	PSOE	0.00877	0.14170	0.00246	**0**	0.00246	0.01263	0.00035	0.00246	0.01824	0.00000	0.01193	0.20098
	IU	0.00105	0.00421	0.01684	**0**	0.00175	0.00772	0.00035	0.00035	0.00105	**0**	0.00246	0.03578
	UPyD	**0**	**0**	**0**	0	**0**	0.00035	**0**	**0**	**0**	**0**	**0**	0.00035
	C's	0.01263	0.01368	0.00246	0.00140	0.06278	0.00842	0.00035	0.00105	0.00666	**0**	0.00596	0.11540
	Podemos	0.00035	0.00351	0.00351	**0**	0.00210	0.06805	0.00035	**0**	0.00631	**0**	0.00281	0.08699
	IB	0.00075	0.00075	0.00038	**0**	0.00075	0.00075	0.00188	0.00038	0.00038	**0**	0.00038	0.00640
	BB	0.00421	0.00281	0.00070	0.00035	0.00351	0.00140	0.00070	0.00526	0.00772	**0**	0.00421	0.03087
	ABS	0.00631	0.00596	0.00105	0.00035	0.00491	0.00281	0.00035	0.00070	0.04700	**0**	0.00772	0.07717
	DK	0.02946	0.03227	0.00631	0.00105	0.02490	0.02841	0.00175	0.00456	0.03262	0	0.04946	0.21080
	DA	0.00561	0.00842	0.00175	**0**	0.00210	0.00526	**0**	0.00035	0.00421	**0**	0.02210	0.04981
	Total	0.21677	0.21571	0.03648	0.00316	0.11014	0.13574	0.00596	0.01684	0.14206	0	0.11715	1.00000

IB: Invalid Ballot BB: Blank Ballot ABS: Abstention DK: Don't Know DA: Decline to Answer

Table 12
Declared voting intention and declared vote by women.

			Confidence limits			
	$\delta^{\kappa}_{RCA}(\mathcal{P})$	$\sigma_{\delta^{\kappa}_{RCA}}(\mathcal{P})$	Lw	Up	Tailed Test	p^*
General	0.46210	0.00585	0.45064	0.47356	101.911	0
Men	0.48124	0.00834	0.46490	0.49758	72.384	0
Women	0.44324	0.00817	0.42722	0.45926	71.118	0
PP	0.65841	0.03584	0.58816	0.72866	18.369	0
PSOE	0.59580	0.03613	0.52498	0.66662	16.490	0
IU	0.46691	0.05632	0.35651	0.57730	8.289	0
UPyD	0.17218	0.21004	-0.23949	0.58386	0.819	0.412
C's	0.52118	0.03858	0.44557	0.59678	13.511	0
Podemos	0.58496	0.03796	0.51055	0.65937	15.409	0

Table 13
Cohen's Kappa statistics for declared voting intention and declared vote.

	Tailed Test	p^*
Men-Women	170.643	0
PP-PSOE	57.642	0
PP-IU	65.663	0
PP-UPyD	11.100	$1.248e^{-28}$
PP-C's	106.749	0
PP-Podemos	59.299	0
PSOE-IU	44.013	0
PSOE-UPyD	9.671	$4.005e^{-22}$
PSOE-C's	56.844	0
PSOE-Podemos	8.558	$1.147e^{-17}$
IU-UPyD	6.715	$1.873e^{-11}$
IU-C's	-18.040	$9.288e^{-73}$
IU-Podemos	-39.503	0
UpyD-C's	-7.966	$1.634e^{-15}$
UPyD-Podemos	-9.422	$4.401e^{-21}$
C's-Podemos	-44.247	0

Table 14
Tests of equality of Cohen's Kappa statistics for declared voting intention and declared vote.

		After election									
		PP	PSOE	IU	UPyD	C's	Podemos	None	DK	DA	**Total**
Before election	PP	0.11226	0.00253	**0**	**0**	0.00416	0.00072	0.08134	**0**	0.00416	0.20517
	PSOE	0.00362	0.10665	0.00343	0.00018	0.00217	0.00687	0.10033	0	0.00434	0.22758
	IU	0.00036	0.00235	0.02223	**0**	0.00054	0.00741	0.01880	**0**	0.00108	0.05278
	UPyD	**0**	0.00018	**0**	0.00127	0.00054	0.00054	0.00398	**0**	**0**	0.00651
	C's	0.01103	0.00687	0.00108	0.00036	0.02965	0.00416	0.07664	**0**	0.00199	0.13178
	Podemos	0.00054	0.00524	0.00380	**0**	0.00018	0.05405	0.04140	**0**	0.00181	0.10701
	None	0.01085	0.01067	0.00307	0.00018	0.00343	0.00850	0.16341	0.00018	0.00343	0.20372
	DK	0.00145	0.00199	0.00108	**0**	0.00072	0.00127	0.02513	**0**	0.00217	0.03380
	DA	0.00362	0.00163	0.00054	**0**	0.00090	0.00217	0.01826	**0**	0.00452	0.03163
	Total	0.14371	0.13811	0.03525	0.00199	0.04230	0.08568	0.52928	0.00018	0.02350	1.00000

DK: Don't Know DA: Decline to Answer

Table 15
Declared sympathy before and after the elections.

Men		After election									
		PP	PSOE	IU	UPyD	C's	Podemos	None	DK	DA	**Total**
Before election	PP	0.11465	0.00337	**0**	**0**	0.00412	0.00112	0.08805	**0**	0.00337	0.21469
	PSOE	0.00112	0.09554	0.00300	**0**	0.00262	0.00749	0.08767	**0**	0.00300	0.20045
	IU	**0**	0.00300	0.02286	**0**	0.00075	0.00824	0.02248	**0**	0.00112	0.05845
	UPyD	**0**	**0**	**0**	0.00187	0.00037	0.00037	0.00487	**0**	**0**	0.00749
	C's	0.01103	0.00687	0.00108	0.00036	0.02965	0.00416	0.07664	**0**	0.00199	0.13178
	Podemos	0.00112	0.00637	0.00525	**0**	0.00037	0.06557	0.04758	**0**	0.00262	0.12889
	None	0.01349	0.01161	0.00412	**0**	0.00262	0.00937	0.14912	**0**	0.00300	0.19333
	DK	0.00187	0.00075	0.00112	**0**	0.00075	0.00187	0.01986	**0**	0.00150	0.02773
	DA	0.00412	0.00112	0.00037	**0**	0.00075	0.00225	0.01499	**0**	0.00450	0.02810
	Total	0.15024	0.12739	0.03784	0.00225	0.04346	0.10191	0.51592	**0**	0.02098	1.00000

Table 16
Declared sympathy before and after the elections by men.

Women		After election									
		PP	PSOE	IU	UPyD	C's	Podemos	None	DK	DA	**Total**
Before election	PP	0.11002	0.00175	**0**	**0**	0.00419	0.00035	0.07510	**0**	0.00489	0.19630
	PSOE	0.00594	0.11701	0.00384	0.00035	0.00175	0.00629	0.11212	**0**	0.00559	0.25288
	IU	0.00070	0.00175	0.02166	**0**	0.00035	0.00664	0.01537	**0**	0.00105	0.04750
	UPyD	**0**	0.00035	**0**	0.00070	0.00070	0.00070	0.00314	**0**	**0**	0.00559
	C's	0.00838	0.00803	0.00105	0.00035	0.02829	0.00279	0.07230	**0**	0.00210	0.12330
	Podemos	**0**	0.00419	0.00244	**0**	**0**	0.04331	0.03563	**0**	0.00105	0.08662
	None	0.00838	0.00978	0.00210	0.00035	0.00419	0.00768	0.17674	0.00035	0.00384	0.21341
	DK	0.00105	0.00314	0.00105	**0**	0.00070	0.00070	0.03004	**0**	0.00279	0.03947
	DA	0.00314	0.00210	0.00070	**0**	0.00105	0.00210	0.02131	**0**	0.00454	0.03493
	Total	0.13762	0.14810	0.03283	0.00175	0.04122	0.07056	0.54174	0.00035	0.02585	1.00000

DK: Don't Know DA: Decline to Answer

Table 17
Declared sympathy before and after the elections by women.

			Confidence limits			
	$\delta^{\kappa}_{RCA}(\mathcal{P})$	$\sigma_{\delta^{\kappa}_{RCA}}(\mathcal{P})$	Lw	Up	Tailed Test	p^*
General	0.37834	0.00547	0.36761	0.38906	72.30236	0
Men	0.37249	0.00794	0.35694	0.38805	50.13231	0
Women	0.38265	0.00547	0.36761	0.39742	51.80298	0
PP	0.56820	0.03637	0.49692	0.63948	15.62324	0
PSOE	0.49005	0.03618	0.41914	0.56096	13.54467	0
IU	0.48235	0.05219	0.38006	0.58464	9.24262	0
UPyD	0.29488	0.14803	0.00474	0.58501	1.99201	0.046
C's	0.27774	0.04182	0.19578	0.35970	6.64174	0
Podemos	0.51417	0.04071	0.43437	0.59396	12.62923	0

Table 18
Cohen's Kappa statistics for declared sympathy before and after the elections.

	Tailed Test	p^*
Men-Women	-48.759	0
PP-PSOE	67.703	0
PP-IU	34.265	$2.628e^{-257}$
PP-UPyD	12.649	$1.132e^{-36}$
PP-C's	183.661	0
PP-Podemos	36.101	$2.162e^{-285}$
PSOE-IU	3.083	0.002
PSOE-UPyD	9.0327	$1.674e^{-19}$
PSOE-C's	135.285	0
PSOE-Podemos	-16.251	$2.168e^{-59}$
IU-UPyD	8.630	6.098^{e-18}
IU-C's	75.174	0
IU-Podemos	-11.901	$1.162e^{-32}$
UpyD-C's	0.792	0.428
UPyD-Podemos	-10.139	$3.709e^{-24}$
C's-Podemos	-128.777	0

Table 19
Tests of equality of Cohen's Kappa statistics for declared sympathy before and after the elections.

6. Conclusions and future research

In this paper a well-known concept in the area of medical diagnosis is introduced to analyze voters' self-agreement considering their declarations about their voting intention and voting decision, and their sympathies for the parties before and after the elections. Cohen's Kappa statistic allows us to provide a measure of concordance considering categorical characteristics like ours. Its beauty lies in its ability to capture the proportion of the cases in which beyond-chance agreement occurred, representing the fortuitous agreement as a function of the marginal proportions of each category. In other words, Cohen's Kappa statistic measures the concordance between two categorical characteristics taking into account the observed agreement, but also the disagreement.

To carry out this study, the data provided by the panel study accomplished by the Spanish Sociological Research Center (CIS) for Spanish General Elections in 2015 was taken. The results show us that in the case of declared voting intention and declared vote more than half of the considered voters did not agree in their answers. That proportion decreases to just over a third for the voters self-declared as PP voters and to forty per cent for the PSOE ones. The greatest disagreement is found in the case of UPyD. Considering the results for declared sympathy before and after the elections, almost two thirds of the voters show self disagreement between both declarations. Such disagreement is less in the cases of PP and Podemos, and greater for UPyD and C's.

In spite of the preliminar nature of this contribution, it provides a new framework for developing new research in the area of the concordance in individuals decisions. It leaves several interesting open questions. For instance, in this first approach, we have considered that both the categories and the characteristics are independent. Such simplification seems unrealistic in several cases and to overcome that drawback we are considering how to collect possibles relationships between categorical characteristics in Cohen's Kappa statistic. Moreover, the concordance between more than two characteristics remains open. Also the application of this interesting tool to other examples far from voting situations, but related to decision making.

References

Abraira, V., 2001. El índice kappa. Medicina de Familia. SEMERGEN 27, 247–249.

Bennett, E.M., Alpert, R., Goldstein, A.C., 1954. Communications Through Limited-Response Questioning. Public Opinion Quarterly 18, 303–308.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement , 37–46.

Goodman, L., Kruskal, W., 1979. Measures of association for cross classifications, in: Measures of Association for Cross Classifications. Springer New York. Springer Series in Statistics, pp. 2–34.

Gwet, K., 2008. Computing inter-rater reliability and its variance in the presence of high agreement. The British Journal of Mathematical and Statistical Psychology , 29–48.

Gwet, K., 2012. Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

McHugh, M., 2012. Interrater reliability: the kappa statistic. Biochemia Medica 22, 276–282.

Scott, W.A., 1955. Reliability of content analysis: The case of nominal scale coding. The Public Opinion Quarterly 19, 321–325.

Tinsley, H.E., Weiss, D.J., 1975. Interrater reliability and agreement of subjective judgments. Journal of Counseling Psychology 22, 358–376.

Tinsley, H.E., Weiss, D.J., 2000. Interrater reliability and agreement, in: Handbook of Applied Multivariate Statistics and Mathematical Modeling. Academic Press, San Diego, pp. 95–124.