A statistical approach to assess referendum results: The venezuelan recall referendum 2004

Maria M. Febres Cordero and Bernardo Márquez

Independent Consultants
CCS 6854, 11010 NW 30TH Street, STE 104, Miami, FL 33172-5032, USA
E-mail: fecodelco@yahoo.com, bb_marquezz@yahoo.es

Summary

This article presents a statistical approach to assess the coherence of official results of referendum processes. The statistical analysis described is divided in four phases, according to the methodology used and the corresponding results:
(1) Initial Study, (2) Quantification of irregular certificates of election, (3) Identification of irregular voting centers and (4) Estimation of recall referendum results.
The technique of cluster analysis is applied to address the issue of heterogeneity of the parishes with respect to their political preferences.

The Venezuelan recall referendum 2004 is the case study we used to apply the proposed methodology, based on the data published by the "Consejo Nacional Electoral" (CNE-National Electoral Council). Finally, we present the conclusions of the study which we summarize as follows: The percentage of irregular certificates of election is between 22,2% and 26,5% of the total; 18% of the voting centers show an irregular voting pattern in their certificates of election, the votes corresponding to this irregularity are around 2.550.000; The result estimate, using the unbiased votes as representative of the population for the percentage of YES votes against President Chávez is 56,4% as opposed to the official result of 41%.

Key words: Cluster analysis, confidence interval, hypothesis testing, Recall referendum

1. Background

On August 15th, 2004, Venezuelan citizens went to the polls to vote in the Recall Referendum on the presidency of Hugo Chávez.

The "Consejo Nacional Electoral" (CNE-National Electoral Council), the government institution in charge of regulating and executing any/all electoral processes, officially declared that 59% of the voters had voted in favor of President Chávez (NO votes) and 41% of voters had voted against (YES votes).

However, the veracity of these results has been seriously questioned by members of the Opposition. Within this context, the purpose of this study is to present a statistical approach designed to assess the coherence of the official results of the presidential recall referendum process.

2. Scope of data

The data used was the official result for the recall referendum, published by the CNE and quoted in Table 1. The table describes the distribution (i.e., yes, no or null) of total votes followed by a breakdown per municipalities, parishes, voting centers, and certificates of election.

Table 1
Official Data, provided by the CNE

Total Votes	**9,815,631**
NO	5,800,629
YES	3,989,008
NULL	25,994
Municipalities	**397 (Includes 62 outside Venezuela)**
Average Votes by Municipality	24,725
Average Parishes by Municipality	3.1
Parishes	**1,228 (Includes 96 outside Venezuela)**
Average Votes by Parish	7,993
Average Voting Centers by Parish	6.8
Voting Centers	**8,335**
Average Votes by Voting Center	1,178
Average Certificates of Election by Center	2.8
Certificates of Election	**23,681**
Average Votes by Certificate of Election	414

Since it was allowed to vote overseas there were 62 virtual municipalities and 96 virtual parishes, usually associated with a Venezuelan embassy or consular office in different countries.

A mixed voting system (computerized and manual) was used in this referendum process. Of the total votes, 87% were cast through voting machines and the remaining 13% through manual procedures. It is interesting to mention that when the voting process was officially closed in each voting machine, the results were transmitted first to the tabulation center (CNE), and then, all the pertaining counting documents (certificates of election) were printed in the voting centers. After that, the manual certificates were sent to the CNE with the printed documents.

3. Statistical Methodology

The statistical analysis described in this study is divided in four phases, according to the methodology used and the corresponding results, as follows: 3.1. Initial Study; 3.2. Quantification of irregular certificates of election; 3.3. Identification of irregular voting centers and 3.4. Estimation of recall referendum results.

3.1. Initial Study

This phase of analysis is based on the testing, for each voting center, of the following hypothesis: at the certificate of election level, there are no significant differences in % of YES votes amongst the various certificates of election in a center.

Since each voter is assigned at random to a certificate of election (computerized or manual), it can be stated that each certificate of election is a random sample of the voting center population. Therefore, significant differences are not expected among certificates of election within a center and the severity of the suggested test is very high. However, there are few certificates of election (samples) per center, which makes irregular voting patterns difficult to identify with high levels of confidence. In this context, if inconsistencies show up in the official data published by the CNE, the above hypothesis can be rejected.

The statistical analysis to be performed (Mood, 1974) can determine if the data prove or disprove this hypothesis. Firstly, estimated %YES/NO votes were calculated for the binomial referendum process using the normal approximation (Feller, 1973) for the voting population at the voting center level (8,335). We may use this normal approximation to the binomial distribution when the estimated value of %YES/NO votes is not extreme and the estimated mean is not too small. Secondly, confidence intervals were estimated for different levels of significance. Thirdly, certificates of election with vote patterns not contained in the confidence intervals were rejected as irregular.

The following example illustrates the methodology: if there are 450 votes in a certificate of election, belonging to a voting center where the YES obtained 50% of the votes, it is expected with 95% confidence that there are between 204 and 246 YES votes.

If the results from the certificate of election was 190 YES, it can be stated that the certificate of election is irregular, since the result is below the calculated minimum indicating that this certificate of election could be sub valuated; if the results from the certificate of election was 260 YES, it can be stated that the certificate of election is irregular, since the result is above the calculated maximum indicating that other certificates of election could be sub valuated; and if the result from the certificate of election was 225 YES, the certificate of election is accepted.

Results

Initial Study

At 99% confidence level, the results for the hypothesis tested are that the Percentage of Rejected Certificates of Election is (1%) and the Number of Rejected Certificates of election is (263).

These results do not indicate major inconsistencies in the official data at the level of voting center. For this reason, further analysis at the Parish level will be

performed, after the heterogeneity of the parishes with respect to the political preferences is disscused in tables 2 and 3.

Table 2 describes the parishes according to two dimensions: first the Heterogeneity defined as the Standard Deviations of the % of YES votes amongst the various Voting Centers within a Parish, and secondly, the number of voting centers per parish.

Table 2

Heterogeneity - Average and Standard Deviation

Heterogeneity (%)	Average (%)	Parishes (%)
Does not apply	Does not apply	19
Up to 9	5.8	41
More than 9	13.1	40
Average heterogeneity: 9.4%, Standard Deviation: 4.8%		

Number of Voting Centers - Average and Standard Deviation

Number of Voting Centers	Average	Parishes(%)
1	1	19
Between 2 and 5	3.4	40
More than 5	12.6	41
Average Voting Centers per Parish: 6.8, Standard Deviation: 7.9		

The calculated parameters induce the description of the 1,228 parishes in the 3x3 contingency table 3.

Table 3

Parish Description - Heterogeneity and Number of Voting Centers

Voting Centers per Parish	Homogeneous	Quasi-Homogeneous Up to 9%	Heterogeneous More than 9%	TOTALS
More than 5 *(Mayor parish)*	0 Votes: 0%	208 Votes: 28%	304 Votes: 55%	**512 Votes: 83%**
Between 2 and 5 *(Minor parish)*	0 Votes: 0%	296 Votes: 10%	191 Votes: 5%	**487 Votes: 15%**
1 *(Minimum parish)*	229 Votes: 2%	0 Votes: 0%	0 Votes: 0%	**229 Votes: 2%**
TOTALS	**229 Votes: 2%**	**504 Votes: 38%**	**495 Votes: 60%**	**1,228**

The information included in Tables 2 and 3 imply that 60% of the votes were found to be associated to parishes with an average heterogeneity greater than

9%. In this context, we decided a reasonable approach would be to subdivide the parishes by using the Minimum Heterogeneity distance for clustering analysis.

3.2. Quantification of irregular certificates of election

The Minimum Heterogeneity distance (MINH) is defined as the estimated standard deviation of the % of YES votes at the level of certificates of election within a parish. It guarantees that no certificate of election is rejected under the assumption that all the certificates have the same expected % of YES votes.

Accordingly, MINH is calculated by MINH = Square root (P*(1-P)/N), where P is the % of YES votes in the parish and N is the greatest number of votes per certificate of election in a parish.

Then, MINH Average = 2.1 %, MINH Std Deviation = 0.7 %
MINH Maximum = 3.2 % (few votes, % very similar),
MINH Minimum = 1.3 % (many votes, % very different)

The second phase of the analysis is used to quantify irregular certificates of election:

1. Parishes are subdivided, according to their heterogeneity in **clusters** as discussed by (Sokal, 1973) and (Press, 1982), until the heterogeneity of each cluster is equal or less than the MINH of the cluster. (Homogeneous and Quasi-homogeneous)

2. For each of the clusters, using its complete set of certificates of election, the probability that the % of YES votes in the certificates have different expected values is calculated using Chi-square (Choi, 1978). This probability (P_i) is assigned to each cluster.

3. Given that the parishes were subdivided to the maximum possible in step 1, then independence is assumed for Chi-squared random variables from step 2. Furthermore, performing Bernoulli trials with variable probability (Feller, 1973):

   R_i = 1 with probability P_i ; R_i= 0 with probability (1-P_i) where
   R_i: is the random variable cluster (i) is rejected
   N_i: is the number of certificates of election in cluster (i)
   Taking $R = \Sigma N_i * R_i$, number of rejected certificates of election

   we have:
   Expected (R) = $\Sigma N_i * P_i$
   Standard deviation (R) = Square root($\Sigma N_i^2 * P_i * (1-P_i)$)

For an illustration of the methodology, table 4 shows the result of step 1 for Parish La Candelaria and table 5 the resultant clusters for all parishes.

Table 4
Methodology - Step 1
Parish La Candelaria

Center	%NO	%YES			
150	61.5	38.5	Cluster A		MINH: 2.2%
160 170	55.9 54.2	44.1 45.8	Cluster B	%NO: 55.0 %YES: 45.0	SDEV: 1.2% MINH: 2.2%
152	50.6	49.4	Cluster C		MINH: 2.7%
120	46.8	53.2	Cluster D		MINH: 2.3%
190 200 210 221	33.7 36.2 36.6 37.3	66.3 63.8 63.4 62.7	Cluster E	%NO: 36.3 %YES: 63.7	SDEV: 1.6% MINH: 2.1%
180 220 230	31.5 30.6 31.4	68.5 69.4 68.6	Cluster F	%NO: 31.0 %YES: 69.0	SDEV: 0.5% MINH: 2.1%
201 231 241	26.3 24.4 27.9	73.7 75.6 72.1	Cluster G	%NO: 26.5 %YES: 73.5	SDEV: 1.7% MINH: 1.9%
Weighted Average	**36.9**	**63.1**	**SDEV: 11.8%**		

Table 5
Methodology – Step 1
Resultant clusters

Voting Centers per Cluster	Homogeneous	Quasi-Homogeneous Up to 3.7%	TOTALS
More than 5 *(Mayor cluster)*	0 Votes: 0%	156 Votes: 19%	**156** **Votes: 19%**
Between 2 and 5 *(Minor cluster)*	0 Votes: 0%	1728 Votes: 57%	**1728** **Votes: 57%**
1 *(Minimum cluster)*	2395 Votes: 24%	0 Votes: 0%	**2395** **Votes: 24%**
TOTALS	**2395** **Votes: 24%**	**1884** **Votes: 76%**	**4279**

Note: *Reduction in Heterogeneity: 87%*
Average Heterogeneity: 1.2%
Average Centers x Cluster: 2.0

Steps 2 and 3 of the methodology are illustrated in table 6 for parish La Candelaria and Figure 1. summarizes the frequency distribution of rejection probabilities for the resultant clusters.

Table 6
Methodology – Steps 2 and 3
Parish La Candelaria

Cluster	Chi-Squared	Degrees of Freedom	Rejection Probabilities
A	5.22	4	73.4
B	1.19	3	24.4
C	0.00	1	0.4
D	1.01	1	68.6
E	9.94	25	0.3
F	4.79	18	0.1
G	3.90	10	4.8

Note: Degrees of Freedom = Number of Certificates of Election – 1

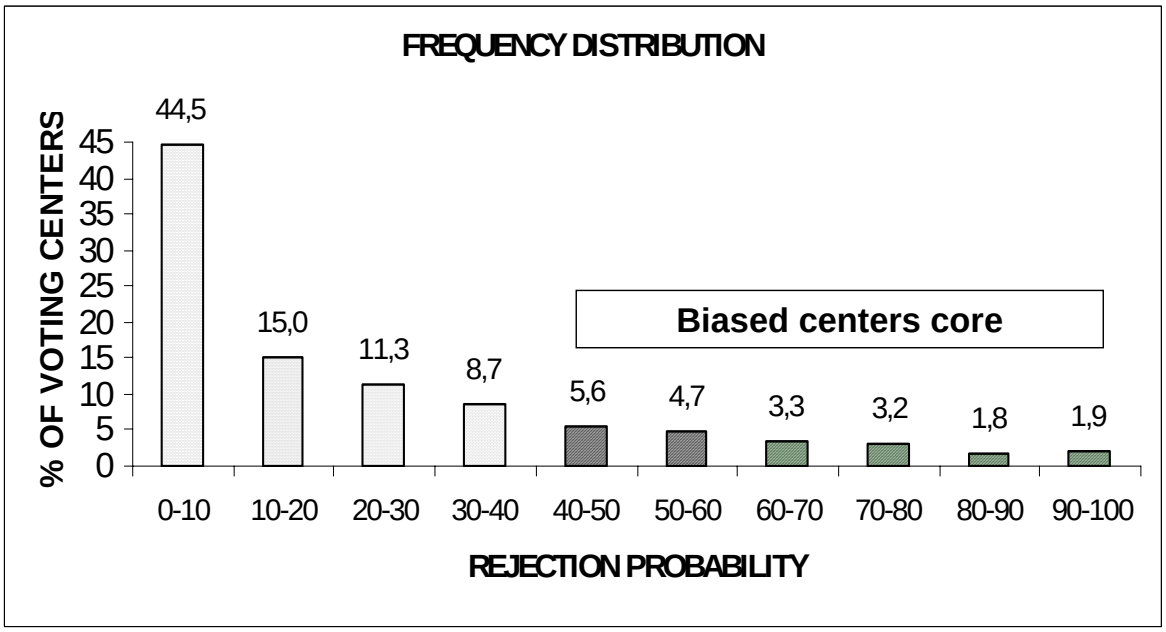

Figure 1. Frequency distribution of rejection probabilities

Results

Using the random variable R, defined in the step 3 of the methodology, as the number of Rejected certificates of election, with the maximum subdivision of parishes, we have:

- Expected (Rejected certificates of election) = 5,761 (24.3%)
- Standard Deviation (Rejected certificates of election) = 198

With the extreme binding implied by the subdivision applied, the number of rejected certificates of election is contained within the interval (MIN: 5,251, MAX: 6,271). The probability of falling outside of this interval is 1%.

This result rigorously narrows the process of subdividing the parishes in clusters (Homogeneous and Quasi-homogeneous) for the next phase.

3.3. Identification of irregular voting centers

The purpose of this section is to identify irregular voting centers using the following methodology:

1. Parishes are subdivided into analysis clusters (Press, 1982) such that the resultant average heterogeneity is contained within the minimum heterogeneity interval determined in section 3.2.
2. The hypothesis “The certificates of election of the resultant clusters (Homogeneous, Quasi-Homogeneous) have the same expected % of YES votes” is tested.
3. Clusters (subsets of centers within a parish), where the hypothesis is not accepted, are rejected.

For an illustration of the methodology, table 7 shows the result of step 1 for Parish La Candelaria and table 8 the resultant clusters for all parishes.

Table 7
Methodology – Step 1
Parish La Candelaria

Center	%NO	%YES			
150	61.5	38.5	Cluster A		
160	55.9	44.1	Cluster B	%NO: 51.9	%YES: 48.1 SDEV: 4.0%
170	54.2	45.8			
120	46.8	53.2			
152	50.6	49.4			
190	33.7	66.3	Cluster C	%NO: 36.3	%YES: 63.7 SDEV: 1.6%
200	36.2	63.8			
210	36.6	63.4			
221	37.3	62.7			
180	31.5	68.5	Cluster D	%NO: 29.4	%YES: 70.6 SDEV: 0.5%
201	26.3	73.7			
220	30.6	69.4			
230	31.4	68.6			
231	24.4	75.6			
241	27.9	72.1			
Weighted Average	**36.9**	**63.1**	**SDEV: 11.8%**		

Table 8
Methodology – Resulting Clusters

Voting Centers per Cluster	Homogeneous	Quasi-Homogeneous Up to 7%	TOTALS
More than 5 *(Mayor cluster)*	0 Votes: 0%	316 Votes: 40%	**316 Votes: 40%**
Between 2 and 5 *(Minor cluster)*	0 Votes: 0%	1462 Votes: 48%	**1462 Votes: 48%**
1 *(Minimum cluster)*	1347 Votes: 12%	0 Votes: 0%	**1347 Votes: 12%**
TOTALS	**1347 Votes: 12%**	**1778 Votes: 88%**	**3125**

Note: Reduction in Heterogeneity: 76%
Average Heterogeneity: 2.3%
Average Centers by Cluster: 2.7

The example presented in Table 9 shows the result of steps 2 and 3 for Parish La Candelaria.

Table 9
Methodology – Steps 2 and 3
Parish La Candelaria, Cluster A: Unidad Educativa José Marti (Center 150)

Certificate of Election	NO Votes	YES Votes	Votes	%NO	%YES
1	297	161	458	64.8%	35.2%
2	247	186	433	57.0%	43.0%
3	278	159	437	63.6%	36.4%
4	296	174	470	63.0%	37.0%
5	247	174	421	58.7%	41.3%
	1365	854	2219	61.5%	38.5%

Even though significant differences exist between the certificates of election in the voting center, the center can only be rejected with a confidence level of 73.4%. This is mainly caused by the existence of very few samples (certificates of election) by cluster.

Results

The consolidated results for hypothesis testing “The certificates of election of the resultant clusters (Homogeneous, Quasi-Homogeneous) have the same expected % of YES votes”, using the resultant clusters of table 8 are shown in table 10 both for computerized and manual certificates of election. Table 11

shows the results for the computerized subset of certificates and table 12 for the manual subset.

Table 10
Computerized and Manual Certificates of Election

Confidence Level	85%	90%	95%
Rejected Voting Centers	21%	18%	15%
Rejected Certificates of Election	28%	25%	21%
Rejected Votes	29%	26%	22%
Number of Rejected Voting Centers	1.781	1.470	1.216
Number of Rejected Certificates of Election	6.558	5.805	4.915
Number of Rejected Votes	2.841.281	2.534.537	2.146.176
Rejected NO Votes	68%	68%	68%
Rejected YES Votes	32%	32%	32%

Table 11
Computerized Certificates of Election

Confidence Level	85%	90%	95%
Rejected Voting Centers	30%	26%	22%
Rejected Certificates of Election	31%	28%	24%
Rejected Votes	31%	28%	24%
Number of Rejected Voting Centers	1.357	1.200	1.014
Number of Rejected Certificates of Election	5.982	5.419	4.627
Number of Rejected Votes	2.656.813	2.406.873	2.051.102
Rejected NO Votes	67%	67%	67%
Rejected YES Votes	33%	33%	33%

Table 12
Manual Certificates of Election

Confidence Level	85%	90%	95%
Rejected Voting Centers	11%	7%	5%
Rejected Certificates of Election	13%	9%	6%
Rejected Votes	14%	10%	7%
Number of Rejected Voting Centers	424	270	202
Number of Rejected Certificates of Election	576	386	288
Number of Rejected Votes	184.468	127.664	95.074
Rejected NO Votes	74%	74%	74%
Rejected YES Votes	26%	26%	26%

3.4. Estimation of recall referendum results

In this section the following data is used to estimate Referendum results**:**

1) Official data provided by the CNE
   Total Votes: 9,815,631
   - Manual Votes: 1,309,764 (13.3%)
   - Computerized Votes: 8,505,867 (86.7%)

2) Unbiased Votes: votes in clusters with probability of rejection less than or equal to 10%, according to results of section 3.3.
   Unbiased Votes: 1,948,574
   - Manual Votes: 294,030 (15.1%)
   - Computerized Votes: 1,654,544 (84.9%)

In this phase methodology is applied as follows:

1. Referendum results were estimated using unbiased votes.

2. Result estimates were analyzed under two premises:

   a) Representative Population
   a.1 The proportion of votes by state given by the CNE is representative of the population.
   a.2 Unbiased votes are representative of the population.

b) Results by type of certificate of election

b.1 Unbiased manual and computerized certificates of election are considered.

b.2 Unbiased computerized certificates of election are considered.

Results

Table 13

Result Estimates - %YES Votes

Representative Population	b.1 Manual and Computerized	b.2 Computerized
a.1 Proportion given by the CNE is representative of the population	52,2%	54.5%
a.2 Unbiased votes are representative of the population	56,4%	59,7%

To illustrate the strength of the methodology applied in sections 3.3 and 3.4, we present the relationship between the % of YES votes and the rejection probabilities of the resultant clusters in Figure 2. The dimension % of votes is added to provide the size of the population at each point.

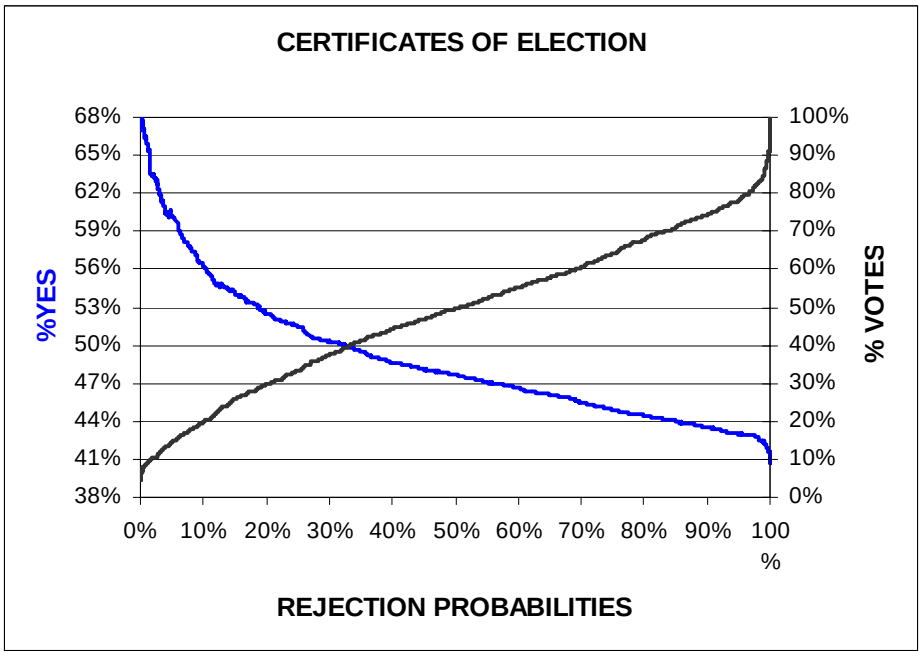

Figure 2. % YES votes vs. % Rejection Probabilities

In this context, using votes in clusters with probability of rejection less than or equal to 10%, the estimate % of YES votes equals 56,4% with the unbiased 20% of the population.

As observed from Figure 2. the % of YES votes continuously decreases as the rejection probability increases, therefore it can be stated that there is a very strong inverse correlation between the two variables. This is a crucial statistical finding, which means that the % NO votes have higher rejection probabilities than the % of YES votes in all the intervals (0%-100%).

4. Conclusions

The statistical approach introduced in the preceding sections to assess the coherence of official results of a recall referendum, can be applied in practice to any binomial electoral process (p,1-p). For the Venezuelan case study under analysis, we now summarize our findings:

4.1. Initial Study

When comparing the voting patterns of different certificates of election within a voting center, only 1% shows irregularities at this level.

4.2. Quantification of irregular certificates of election

The expected value and standard deviation of the random variable R (number of rejected certificates of election) were calculated using the maximum subdivision of parishes in clusters (subsets of centers within a parish). The resultant interval of irregular certificates of election falls between 5,251 and 6,271 with probability 99%, this represents between 22,2% and 26,5% of the total of certificates of election.

4.3. Identification of irregular voting centers

When comparing the voting patterns of certificates of election within clusters, it can be stated that 18% of the voting centers show an irregular voting pattern in their certificates of election. The votes corresponding to this irregularity are around 2,550,000 in number, and are mainly (68%) related to the NO option of the presidential referendum.

4.4. Estimation of recall referendum results

The result estimates for the %YES votes in the referendum fall between 52.2% and 59.7% regarding the representative population and the results by type of the electoral certificate.

Within this context, it is relevant to mention that the proposed statistical approach was powerful enough to identify irregular voting centers because bias was not homogeneous within parishes. The annexes of this paper provide the

lists of Rejected Centers, Unbiased Centers and Other Centers according to the computed rejection probabilities.

Finally, these findings lead us to conclude that the Venezuelan opposition has statistical evidence to reject the official results given by the CNE. The irregularities detected were observed consistently in numerous voting centers and the magnitude of the irregularities imply that the official results do not reflect the intention of voters with statistical confidence.

Note: The authors can provide the annexes List of Rejected Centers, List of Unbiased Centers and List of Other Centers for further research studies on request.

Résumé

Cet article présente une approche statistique pour dégager une cohérence des résultats officiels des processus du référendum. L'analyse statistique décrite se divise en quatre phases, selon la méthodologie utilisée et les résultats qui correspondent:
(1) Étude initiale, (2) La quantification des certificats d'élection irréguliers, (3) L'identification des centres de suffrage irréguliers et (4) Estimation des résultats d´appel de referendum.
La technique d'analyse de grappes est utilisée pour traiter le problème d'hétérogénéité des paroisses en relation avec leurs préférences politiques.

Le cas d'étude choisi pour appliquer la méthodologie proposée est le référendum vénézuélien de l'année 2004. Les données utilisées comme point de référence sont celles publiées par le CNE (Conseil National Electoral). Les conclusions de cette étude se résument comme suit :
Le pourcentage des certificats d'élection irréguliers est entre 22,2% et 26,5% du total ; 18% des bureaux de vote montrent un patron irrégulier de vote dans les certificats d'élection; les votes irréguliers sont environs 2.550.000; le résultat estimé, en utilisant les votes impartiaux comme représentatifs de la population dans le pourcentage de votes OUI contre le président Chavez est de 56,4% en comparaison avec le résultat officiel de 41%.

Mots clés: analyse de grappes, intervalle de confiance, essai d'hypothèse, Rappeler le référendum

References

Choi, S.C. (1978). Introductory applied statistics in Science, Prentice Hall Inc., New Jersey

Feller, W. (1973). Introducción a la teoría de probabilidades y sus aplicaciones, Vol. 1, Limusa-Wiley S.A., México, DF

Mood, A.M., Graybill, F.A., Boas, D.C. (1974). Introduction to the theory of statistics, Mc Graw-Hill Inc., New York

Press, S.J. (1982). Applied Multivariate Analysis, R. E. Krieger Publishing Company Inc., Malobar, Florida

Sokal, R.R.,Sneath, P.H.A. (1973). Numerical Taxonomy: The principles and Practice of Numerical Classification, W.H. Freeman & Co., San Francisco