TEC | Tecnológico de Costa Rica

Ingeniería en Computación

IC-6200 Inteligencia Artificial

Juan Manuel Esquivel Rodriguez

Proyecto Corto 1 - Simulador Votantes

Fabio Mora Cubillo - 2013012801

Sergio Moya Valerin - 2013015682

Gabriel Venegas Castro - 2013115967

I Semestre, 2018

Naturaleza del proyecto

El proyecto consiste en generar datos sintéticos que sigan la distribución teórica de los datos de las elecciones nacionales de Costa Rica del año 2018. Consta de dos funciones principales:

- generar_muestra_pais(n): retornará una muestra simulada con n votantes con etiquetas y atributos proporcionales a la distribución de resultados nacionales.

- generar_muestra_provincia(n, nombre_provincia): retornará una muestra simulada siguiendo la distribución de datos para una provincia específica (donde nombre_provincia es uno de "SAN JOSE", "ALAJUELA", "CARTAGO", "HEREDIA", "GUANACASTE", "PUNTARENAS", "LIMON")

Datos a utilizar

Los datos a utilizar en el proyecto son archivos csv generados a partir de los siguientes:

- Actas de escrutinio: Votos por junta receptora: http://www.tse.go.cr/elecciones2018/actas_escrutinio.htm
- Mapeo juntas a cantones: Información sobre mapeo juntas-cantones: http://www.tse.go.cr/pdf/nacional2018/JRV.pdf
- Indicadores cantonales: Atributos e indicadores cantonales: https://www.estadonacion.or.cr/files/biblioteca_virtual/otras_publicaciones/Indicadorescantonales_Censos2000y2011.xlsx

Decisiones técnicas sobre los datos

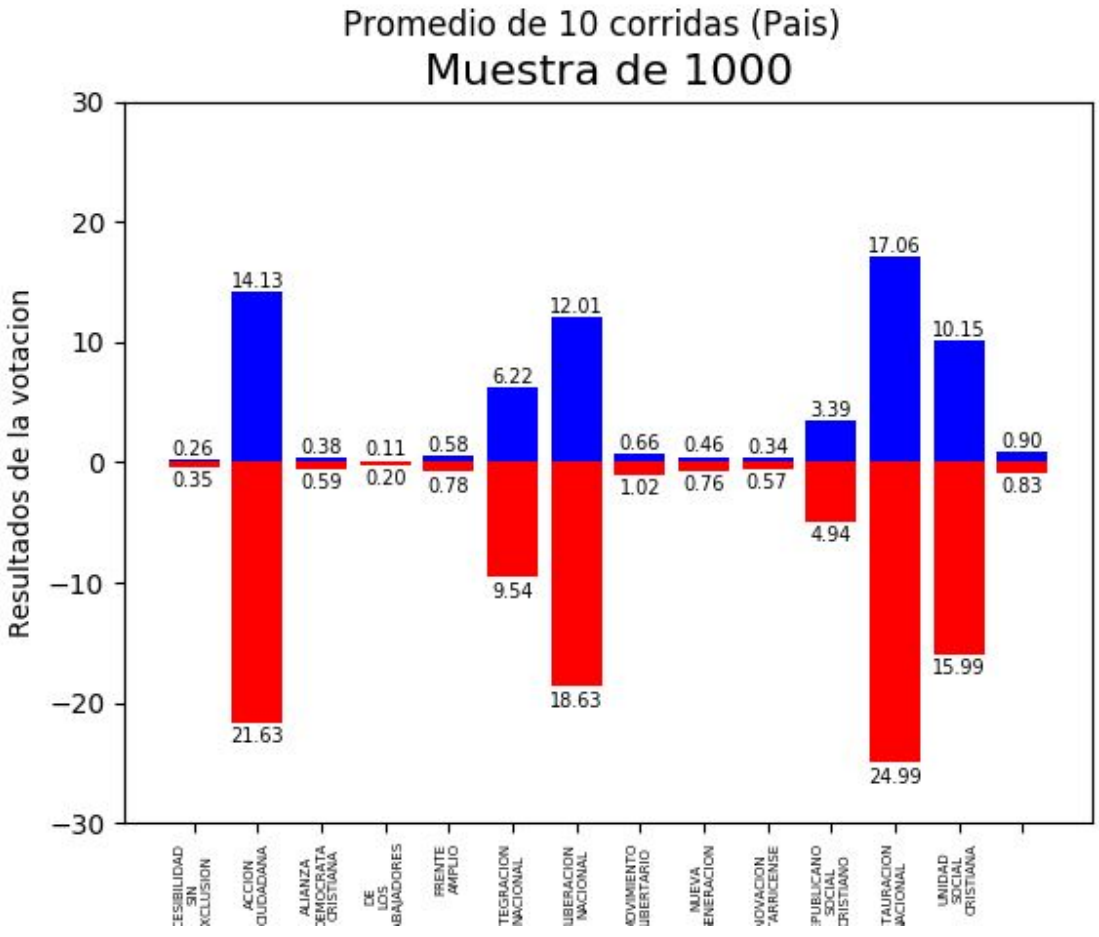
1. Se toman en cuenta los indicadores de ruralidad/urbanidad.
2. La relación de cantidad entre hombres y mujeres se modifica mediante software para ser usado como porcentaje. (Hombres*100 / hombres+100 = Porcentaje de hombres)
3. El dato “relación de dependencia” se toma como sólo las personas mayores de 65, dejando fuera a las menores de 15, ya que no forman parte del padrón electoral.
4. Se incluyen en la generación de la población todos los posibles indicadores.
5. Se agrega un indicador “PROVINCIA” en la penúltima posición de la lista de indicadores
6. Se agrega un indicador “VOTO” en la última posición de la lista de indicadores
7. El tamaño de la lista interna que retornan las funciones (la que contiene todos los indicadores) es [numero_atributos+2], ya que se añaden PROVINCIA y VOTO.

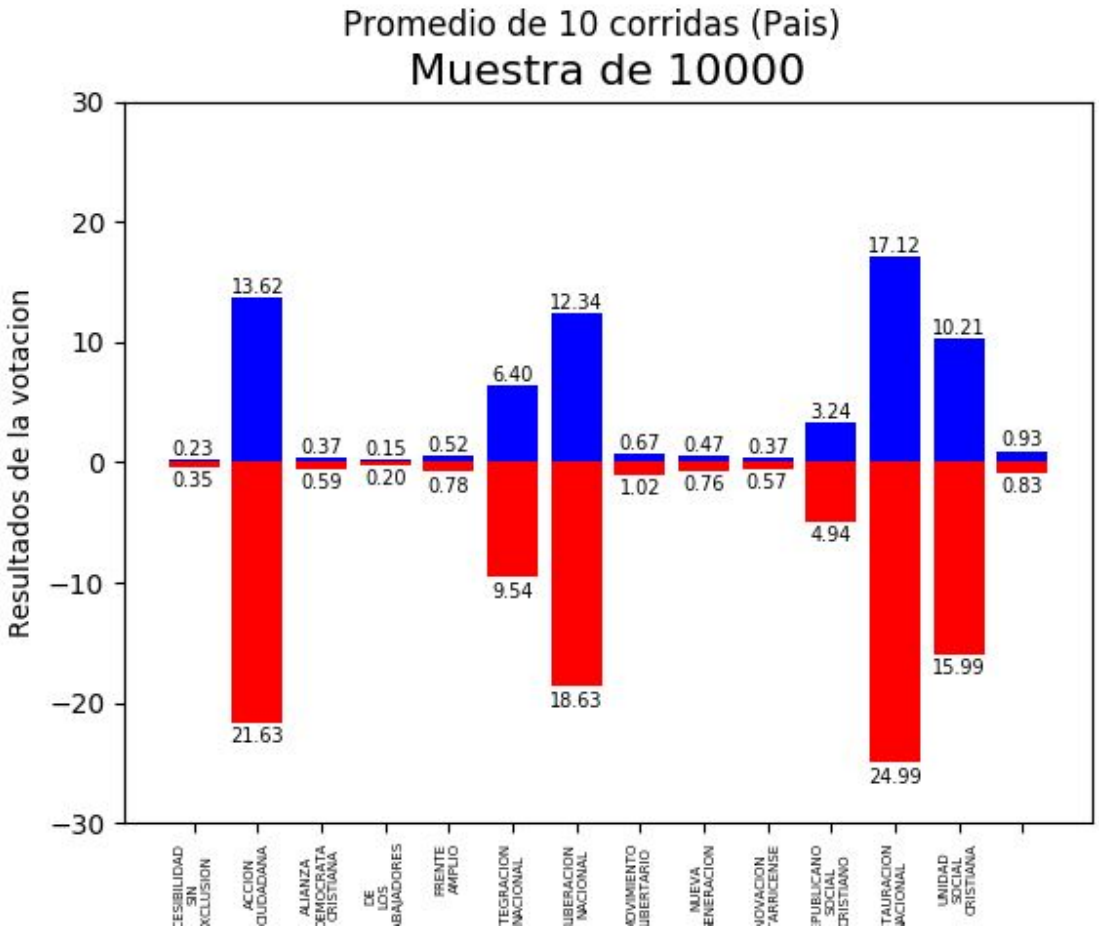
Otras consideraciones

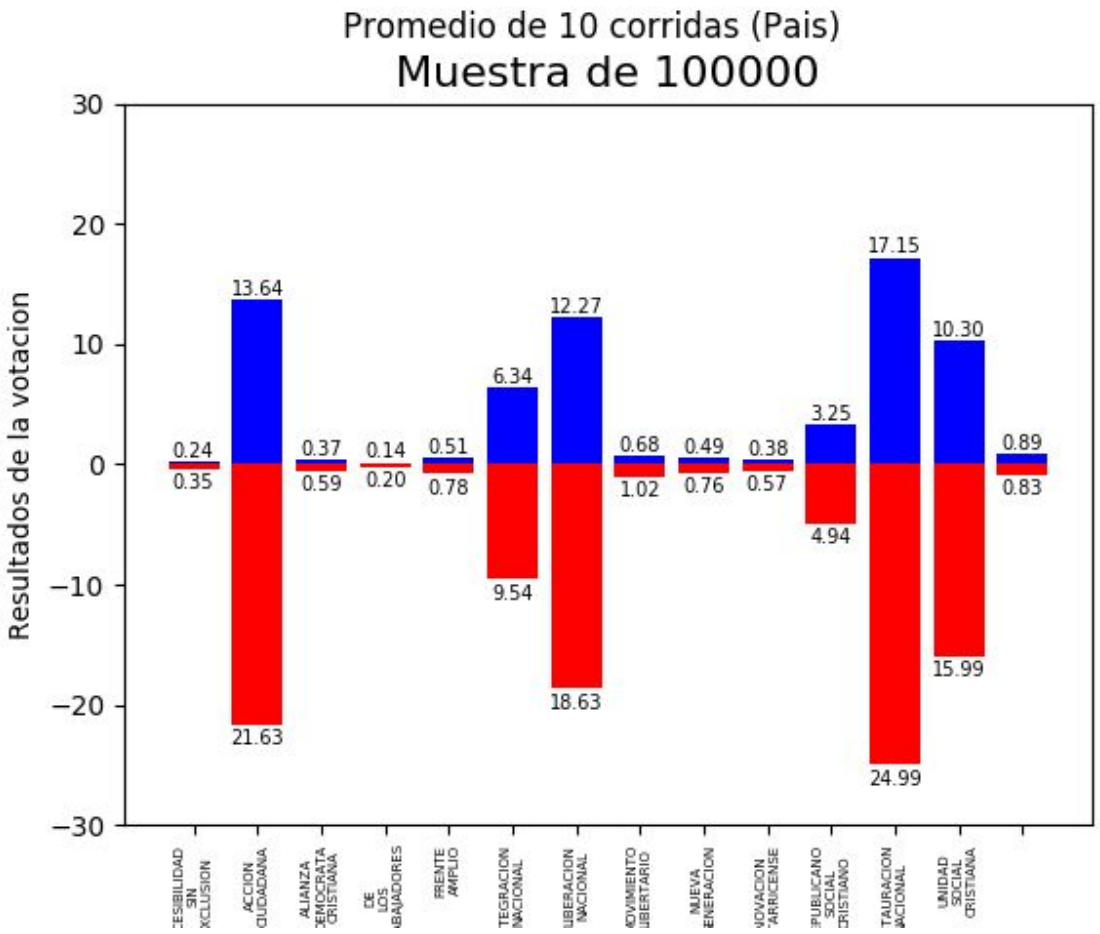
1. Para las pruebas pytest, se toman los porcentajes como un valor flotante entre 0.0 y 1.0
2. Para las pruebas pytest, se hacen pruebas de aleatoriedad para verificar que el generador de números aleatorios tenga el comportamiento esperado. Se puede especificar el tamaño de la muestra y el porcentaje de tolerancia con respecto al porcentaje fijo establecido.
3. Para muestras pequeñas, no tiene sentido mantener un valor de tolerancia bajo; para muestras medianas y grandes, la distribución converge, permitiendo tener valores de tolerancia más bajos.

Análisis de resultados

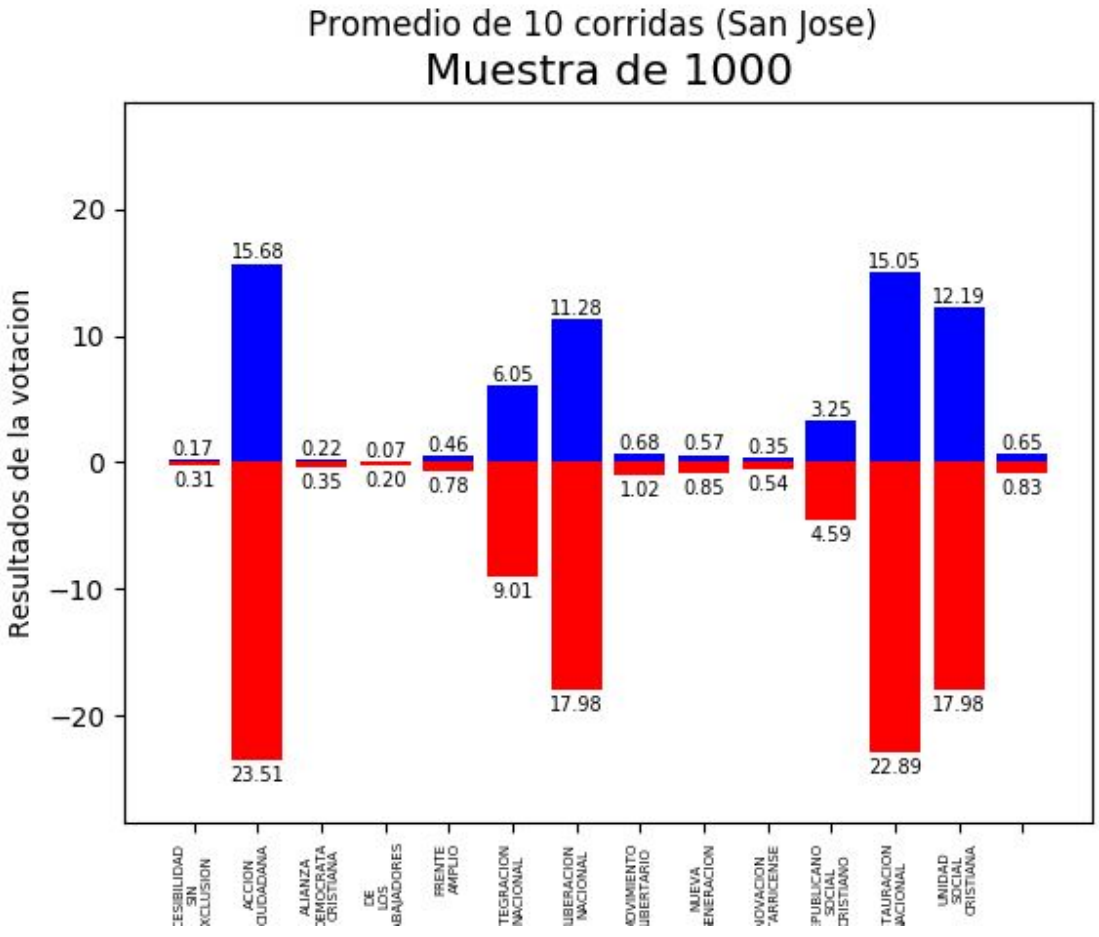
Para el análisis de los resultados se determinó un estándar que consiste en promediar 10 corridas de cada caso. Los casos son por país y por provincias, con 1000 y 10000 muestras. Se mostrarán una serie de gráficos y su descripción relativa. En dichos gráficos las barras azules equivalen al valor promediado de las 10 corridas y el rojo al valor real provisto por el TSE, además el último valor a la derecha son los votos vacíos y blancos.
Se utilizó matplotlib para automatizar la generación de los gráficos.

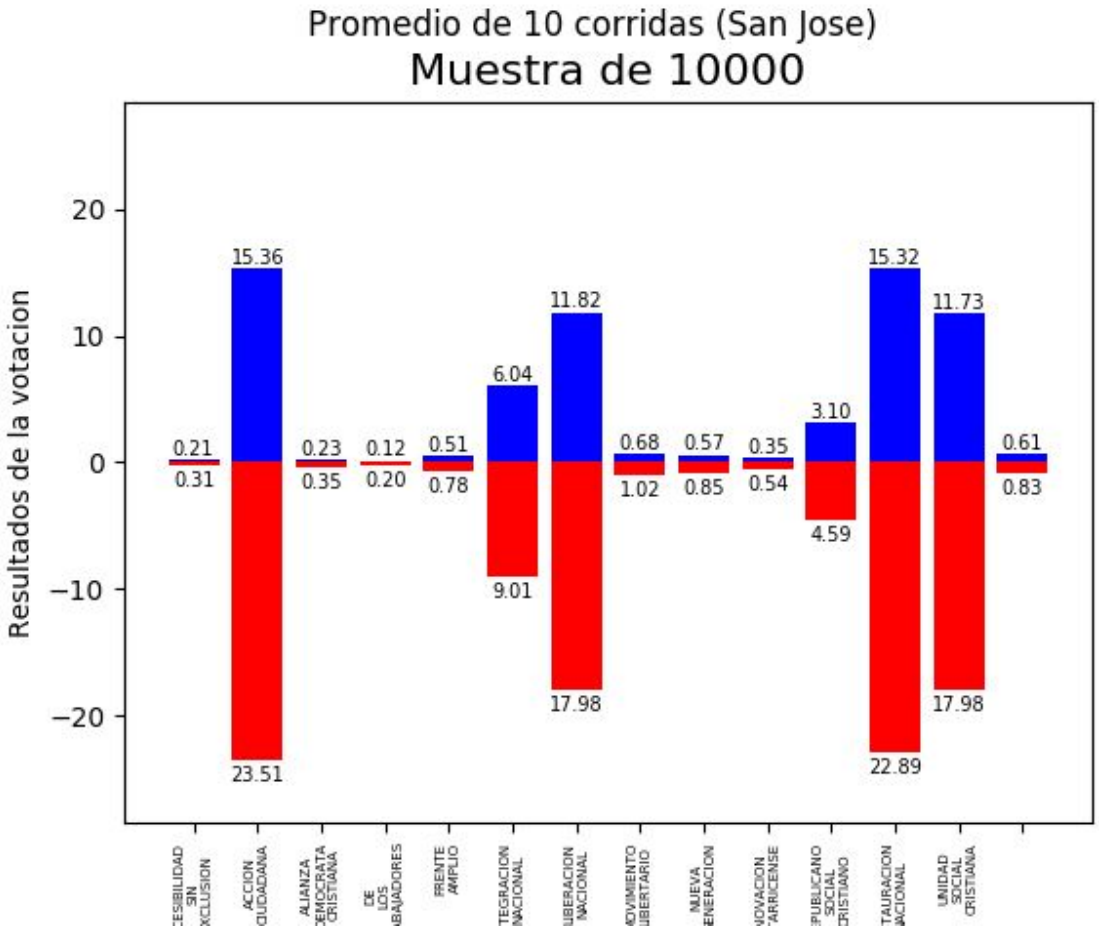

Promedio de 10 corridas (Pais)
Muestra de 1000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.26
0.35
14.13
21.63
0.38
0.59
0.11
0.20
0.58
0.78
6.22
9.54
12.01
18.63
0.66
1.02
0.46
0.76
0.34
0.57
3.39
4.94
17.06
24.99
10.15
15.99
0.90
0.83
ACCION CIUDADANA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

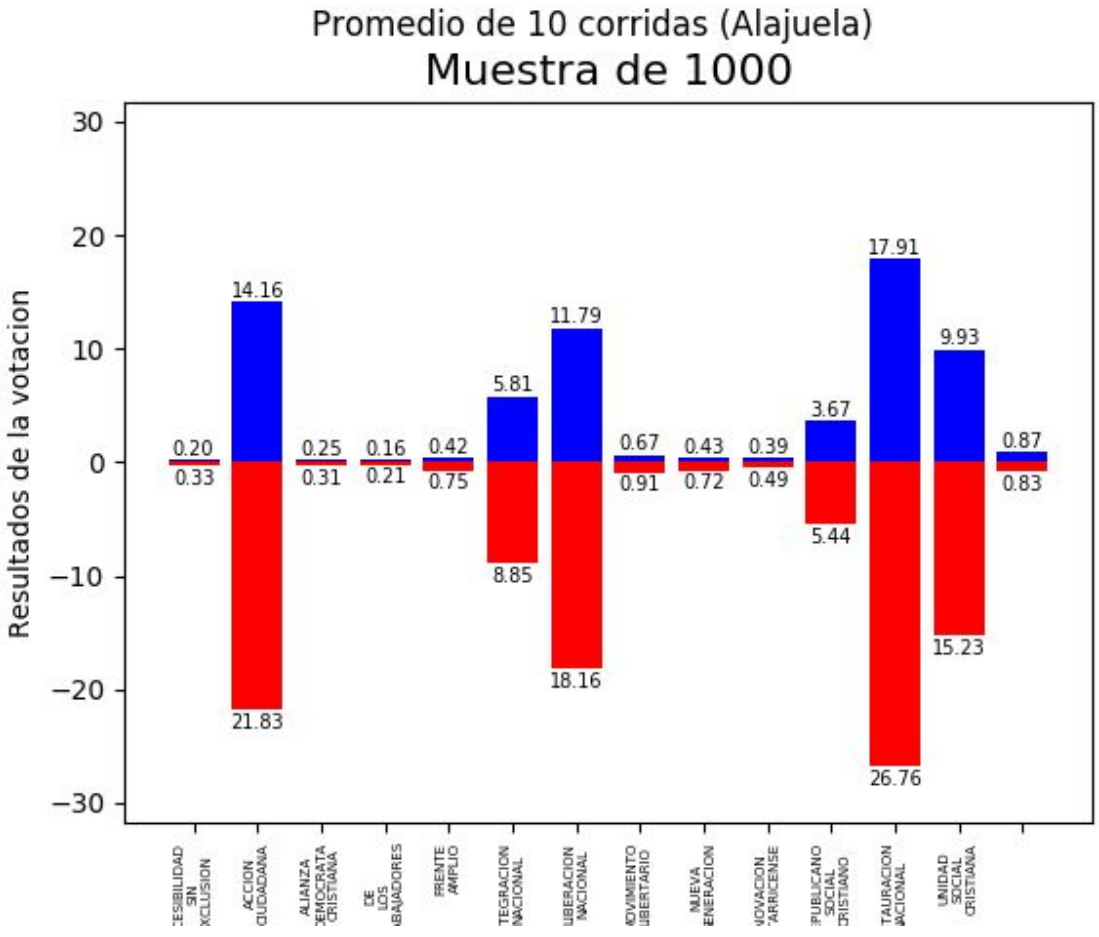

Promedio de 10 corridas (Pais)
Muestra de 10000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.23
0.35
13.62
21.63
0.37
0.59
0.15
0.20
0.52
0.78
6.40
9.54
12.34
18.63
0.67
1.02
0.47
0.76
0.37
0.57
3.24
4.94
17.12
24.99
10.21
15.99
0.93
0.83
ACCION CIUDADANA
ALIANZA DEMOCRATA CRISTIANA
DE LOS TRABAJADORES
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

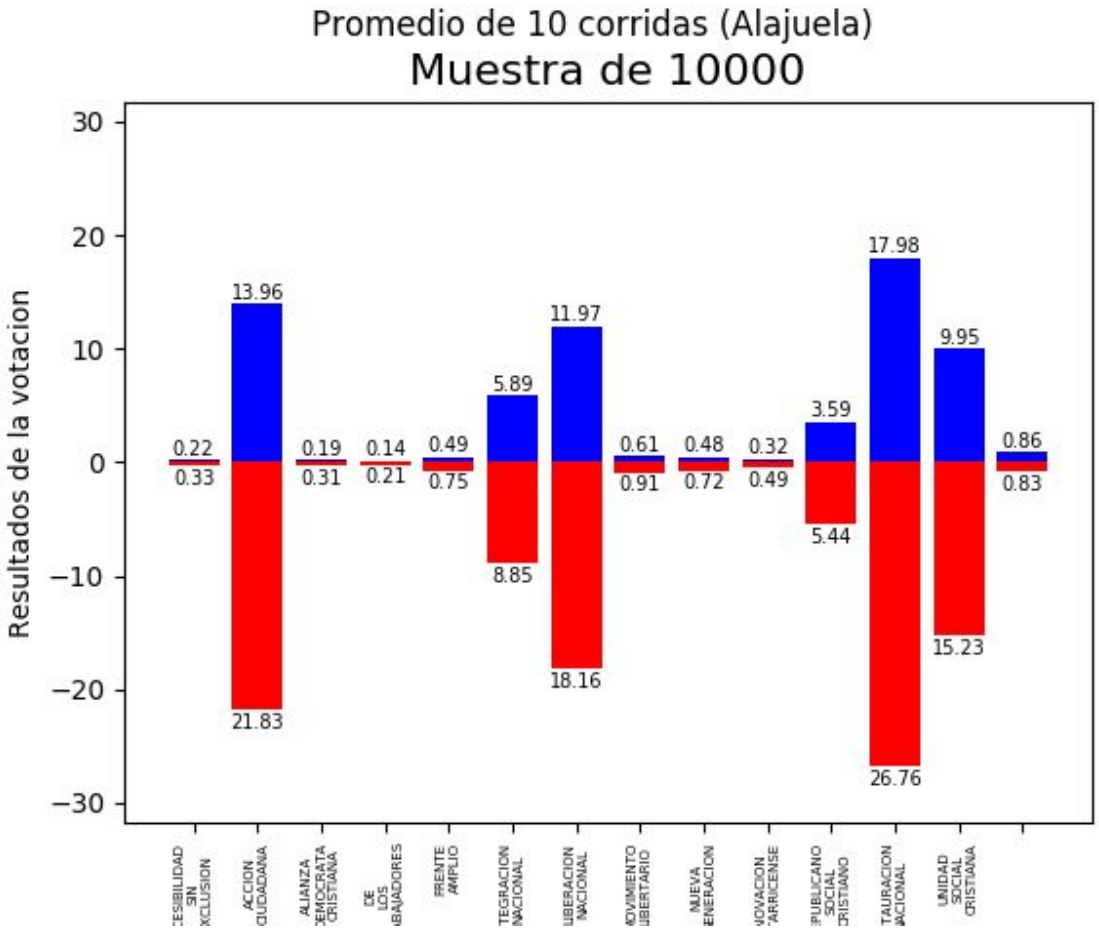


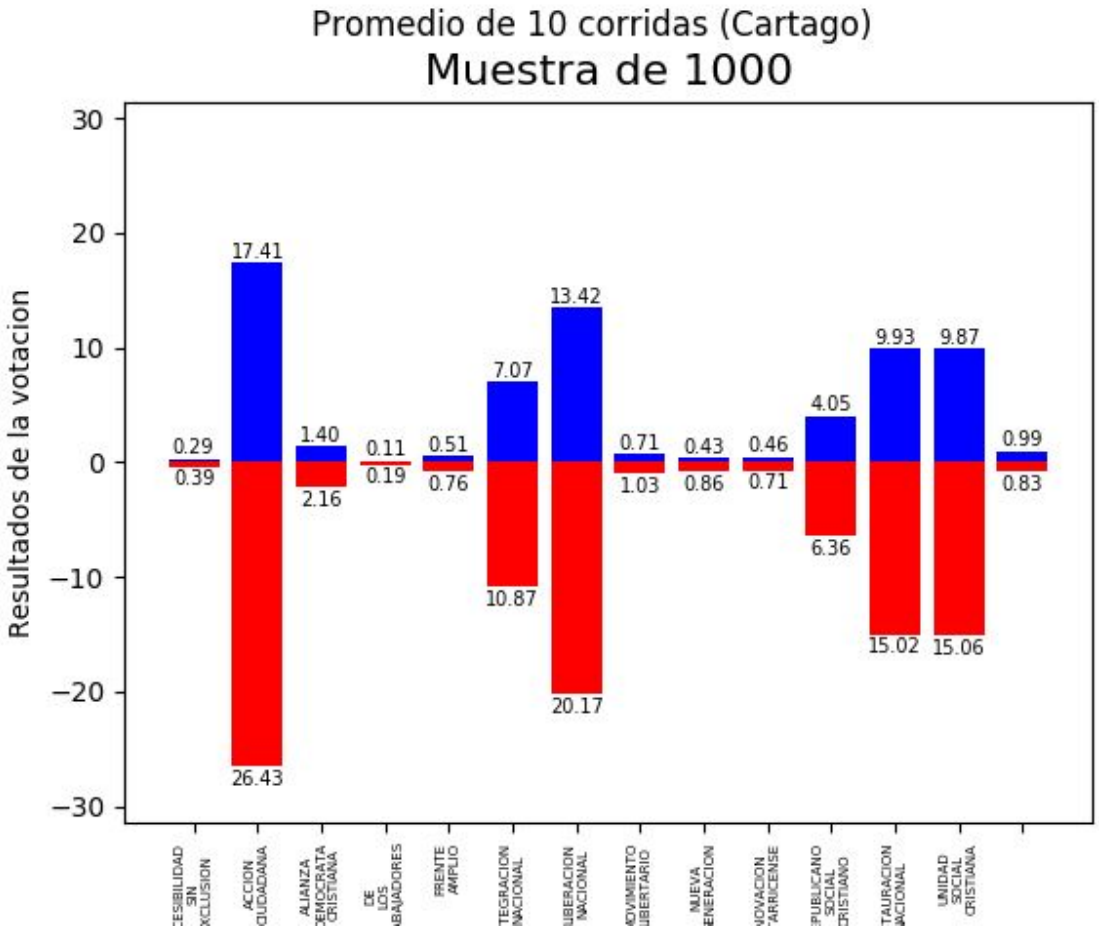
Al generar la muestra de cien mil para el país se determinó que la generación de muestras es lo suficientemente sólida e independientemente del tamaño de la muestra, mientras esta sea lo suficientemente grande, siempre lanzará resultados casi iguales.

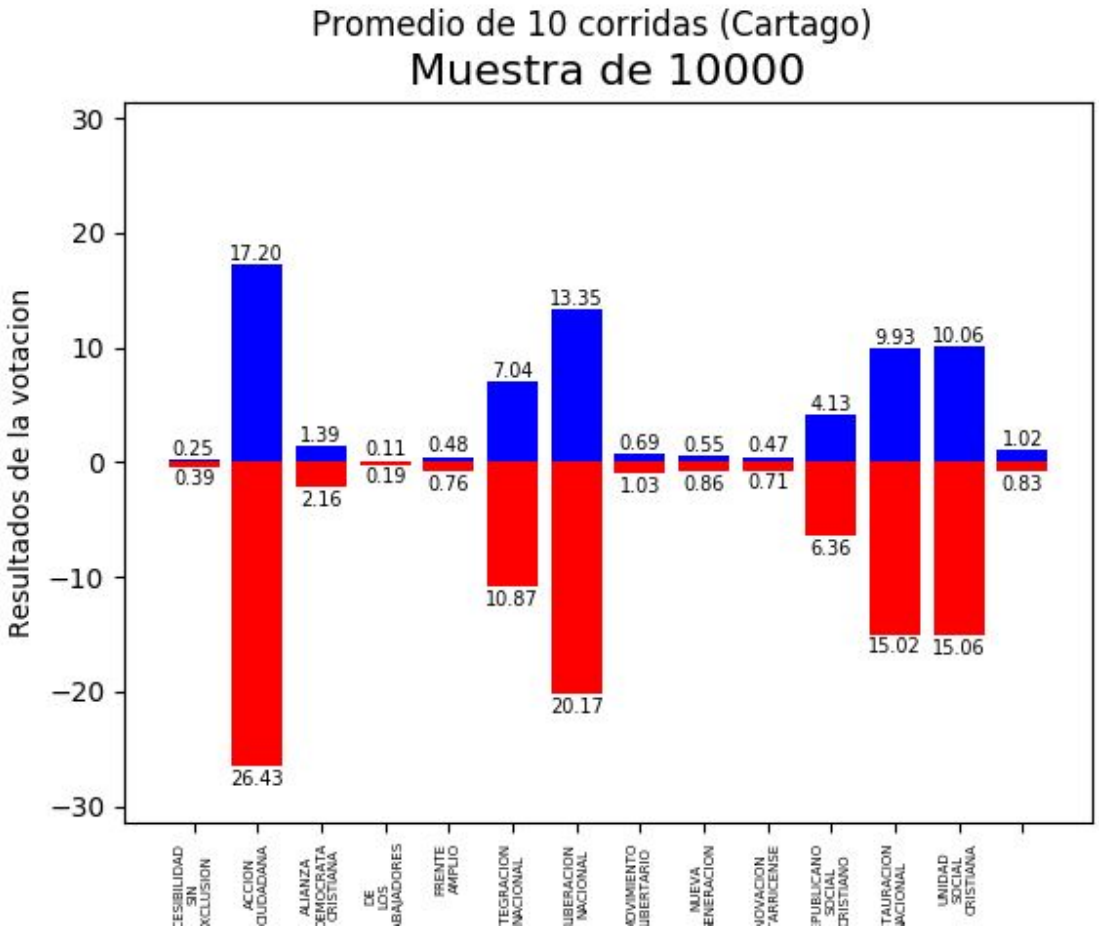

Promedio de 10 corridas (San Jose)
Muestra de 1000
Resultados de la votacion
20
10
0
−10
−20
0.17
0.31
15.68
23.51
0.22
0.35
0.07
0.20
0.46
0.78
6.05
9.01
11.28
17.98
0.68
1.02
0.57
0.85
0.35
0.54
3.25
4.59
15.05
22.89
12.19
17.98
0.65
0.83
ACCION CIUDADANA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

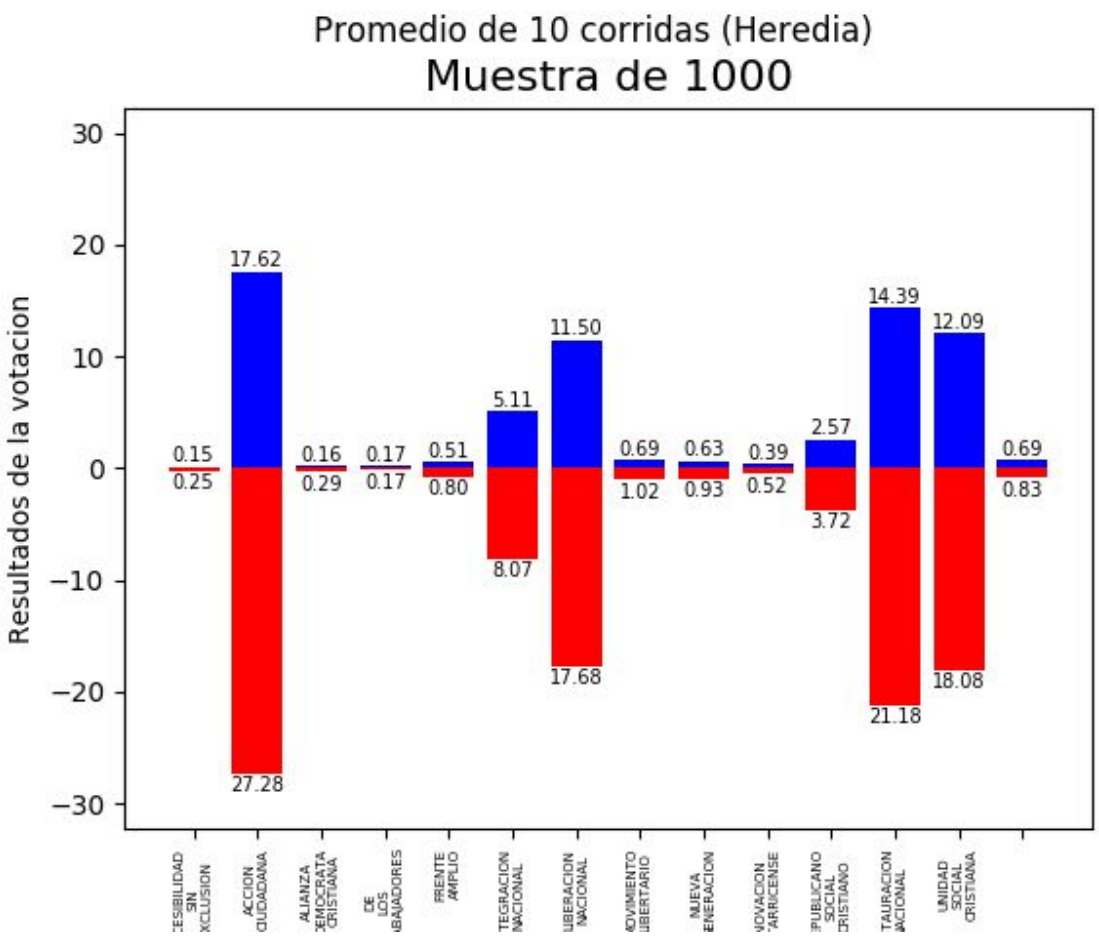

Promedio de 10 corridas (San Jose)
Muestra de 10000
Resultados de la votacion
20
10
0
−10
−20
0.21
0.31
15.36
23.51
0.23
0.35
0.12
0.20
0.51
0.78
6.04
9.01
11.82
17.98
0.68
1.02
0.57
0.85
0.35
0.54
3.10
4.59
15.32
22.89
11.73
17.98
0.61
0.83
ACCION CIUDADANA
ALIANZA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

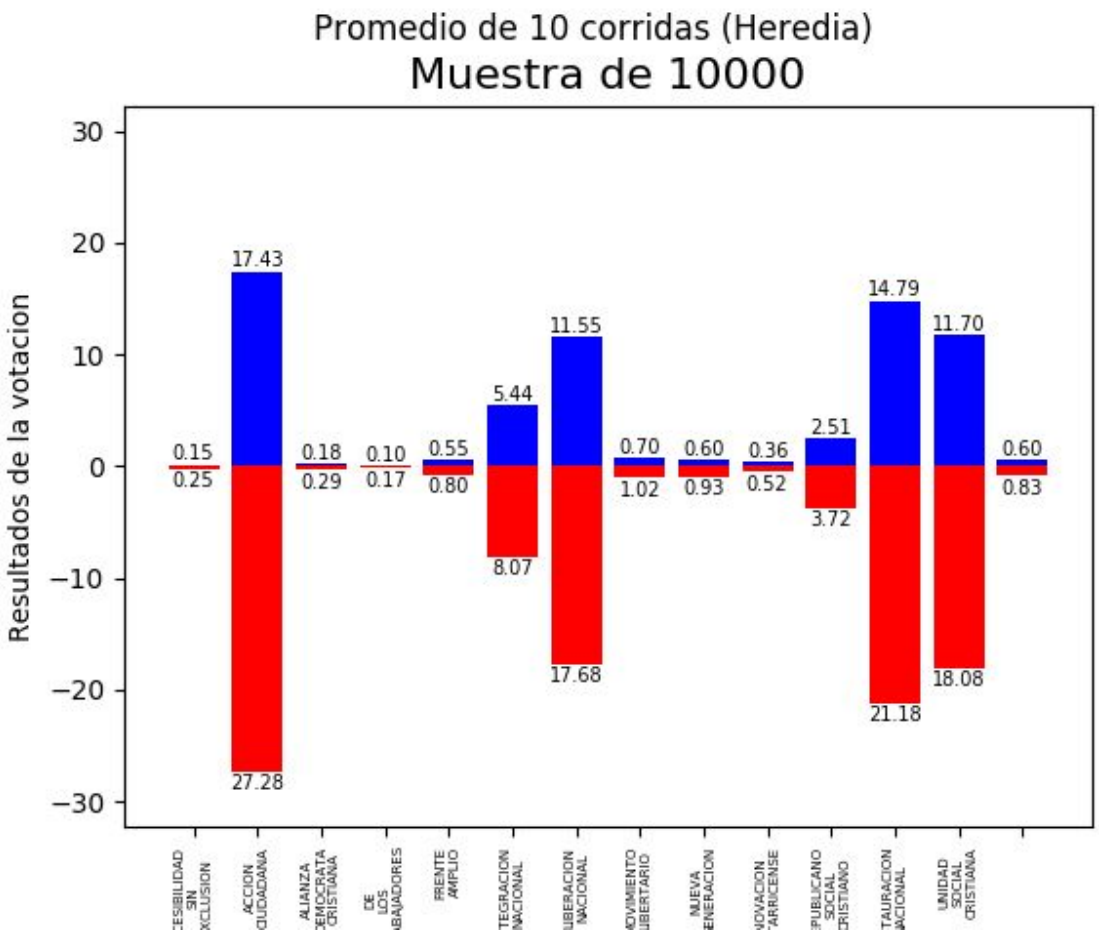

Promedio de 10 corridas (Alajuela)
Muestra de 1000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.20
0.33
14.16
21.83
0.25
0.31
0.16
0.21
0.42
0.75
5.81
8.85
11.79
18.16
0.67
0.91
0.43
0.72
0.39
0.49
3.67
5.44
17.91
26.76
9.93
15.23
0.87
0.83
ACCION CIUDADANA
ALIANZA DEMOCRATA CRISTIANA
DE LOS TRABAJADORES
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

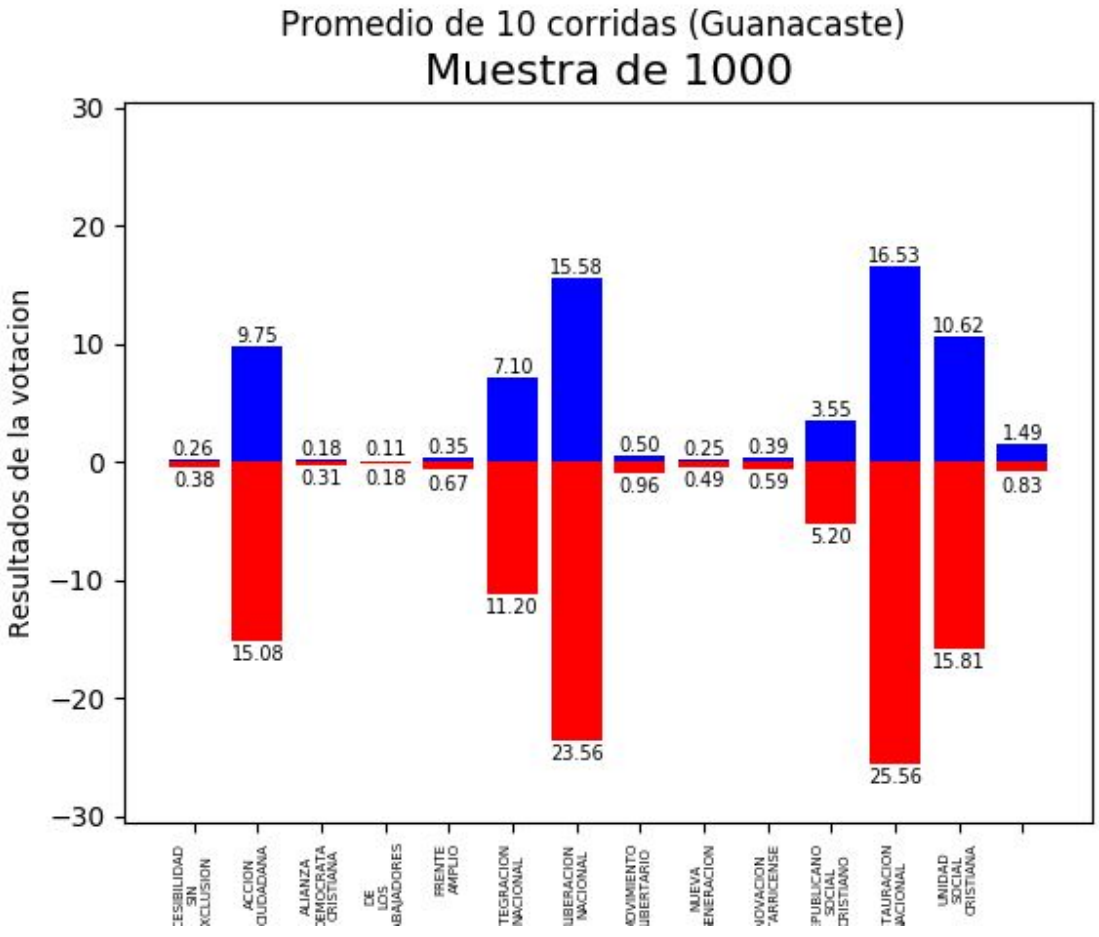

Promedio de 10 corridas (Alajuela)
Muestra de 10000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.22
0.33
13.96
21.83
0.19
0.31
0.14
0.21
0.49
0.75
5.89
8.85
11.97
18.16
0.61
0.91
0.48
0.72
0.32
0.49
3.59
5.44
17.98
26.76
9.95
15.23
0.86
0.83
ACCION CIUDADANA
ALIANZA DEMOCRATA CRISTIANA
DE LOS TRABAJADORES
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

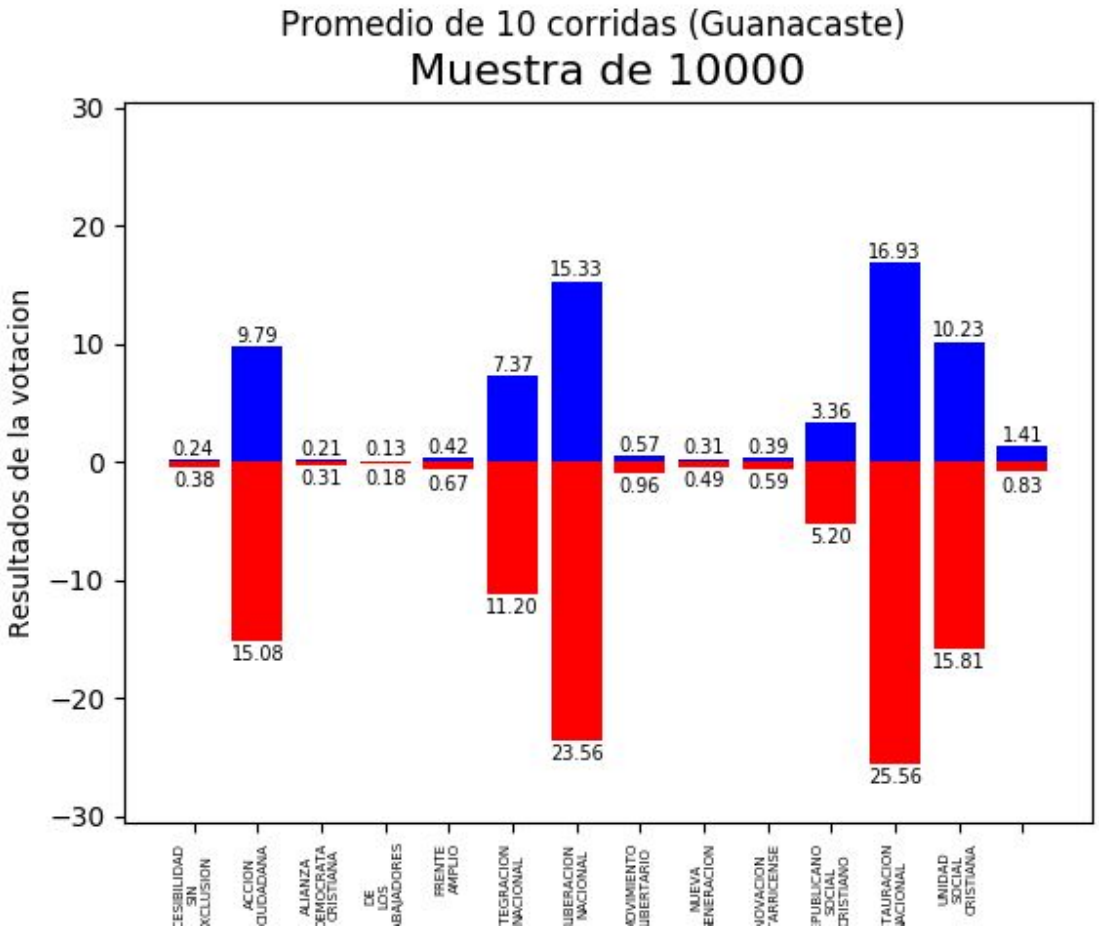

Promedio de 10 corridas (Cartago)
Muestra de 1000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.29
0.39
17.41
26.43
1.40
2.16
0.11
0.19
0.51
0.76
7.07
10.87
13.42
20.17
0.71
1.03
0.43
0.86
0.46
0.71
4.05
6.36
9.93
15.02
9.87
15.06
0.99
0.83
ACCION CIUDADANA
ALIANZA
DE LOS
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

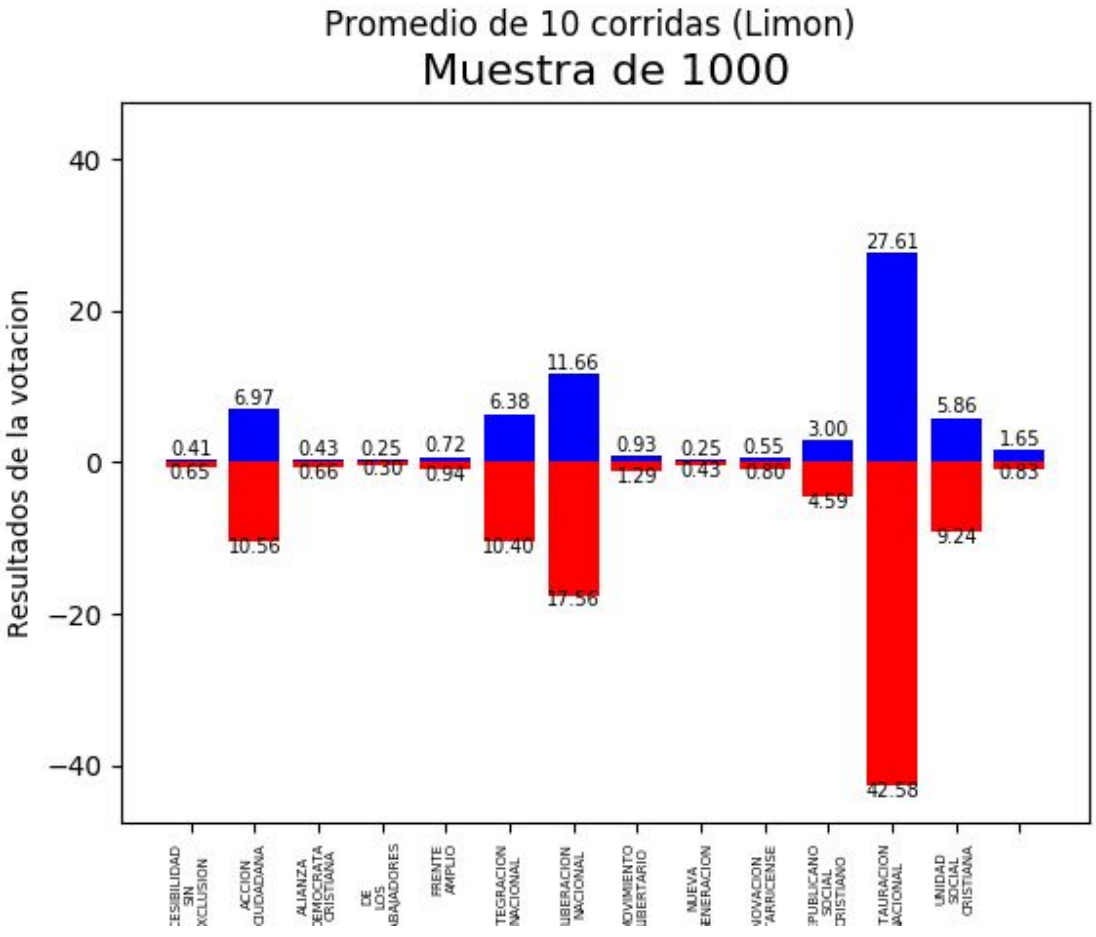

Promedio de 10 corridas (Cartago)
Muestra de 10000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.25
0.39
17.20
26.43
1.39
2.16
0.11
0.19
0.48
0.76
7.04
10.87
13.35
20.17
0.69
1.03
0.55
0.86
0.47
0.71
4.13
6.36
9.93
15.02
10.06
15.06
1.02
0.83
ACCION CIUDADANA
ALIANZA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

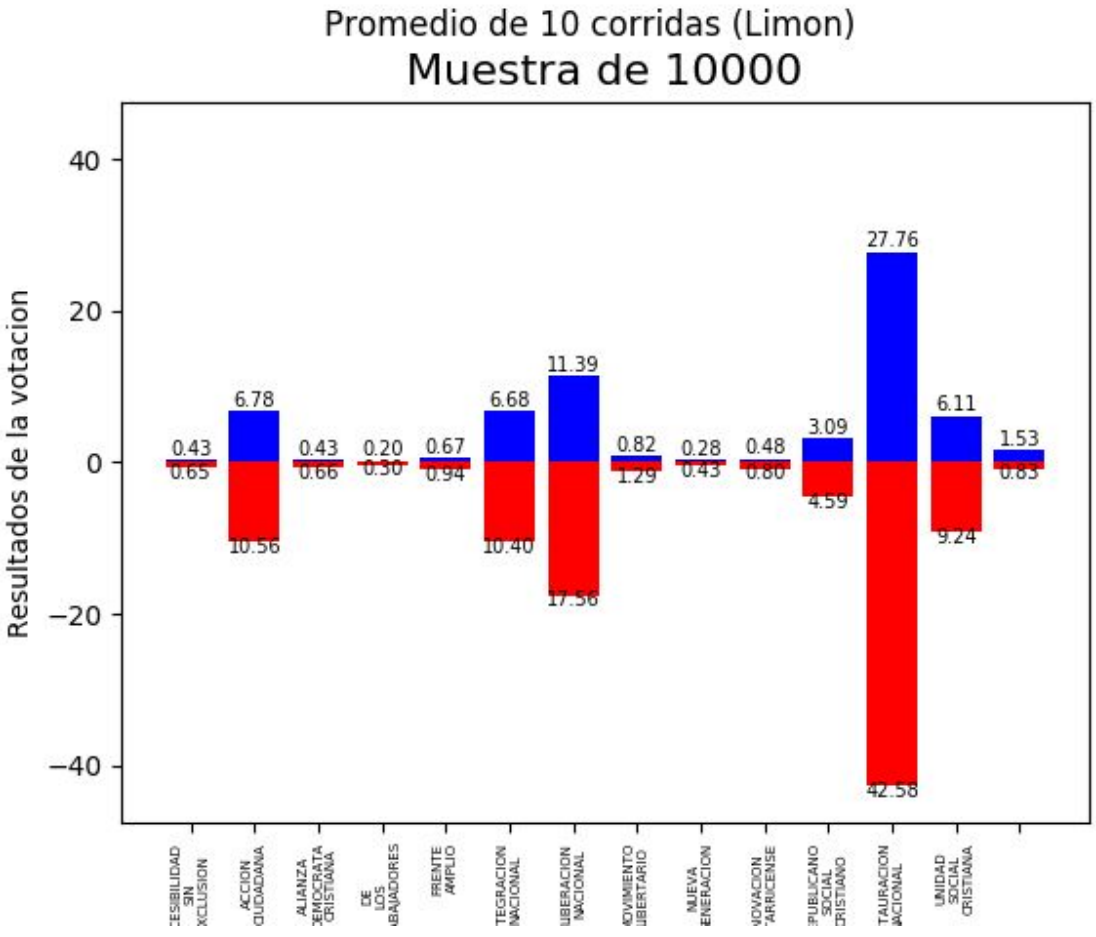

Promedio de 10 corridas (Heredia)
Muestra de 1000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.15
0.25
17.62
27.28
0.16
0.29
0.17
0.17
0.51
0.80
5.11
8.07
11.50
17.68
0.69
1.02
0.63
0.93
0.39
0.52
2.57
3.72
14.39
21.18
12.09
18.08
0.69
0.83
ACCION CIUDADANA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

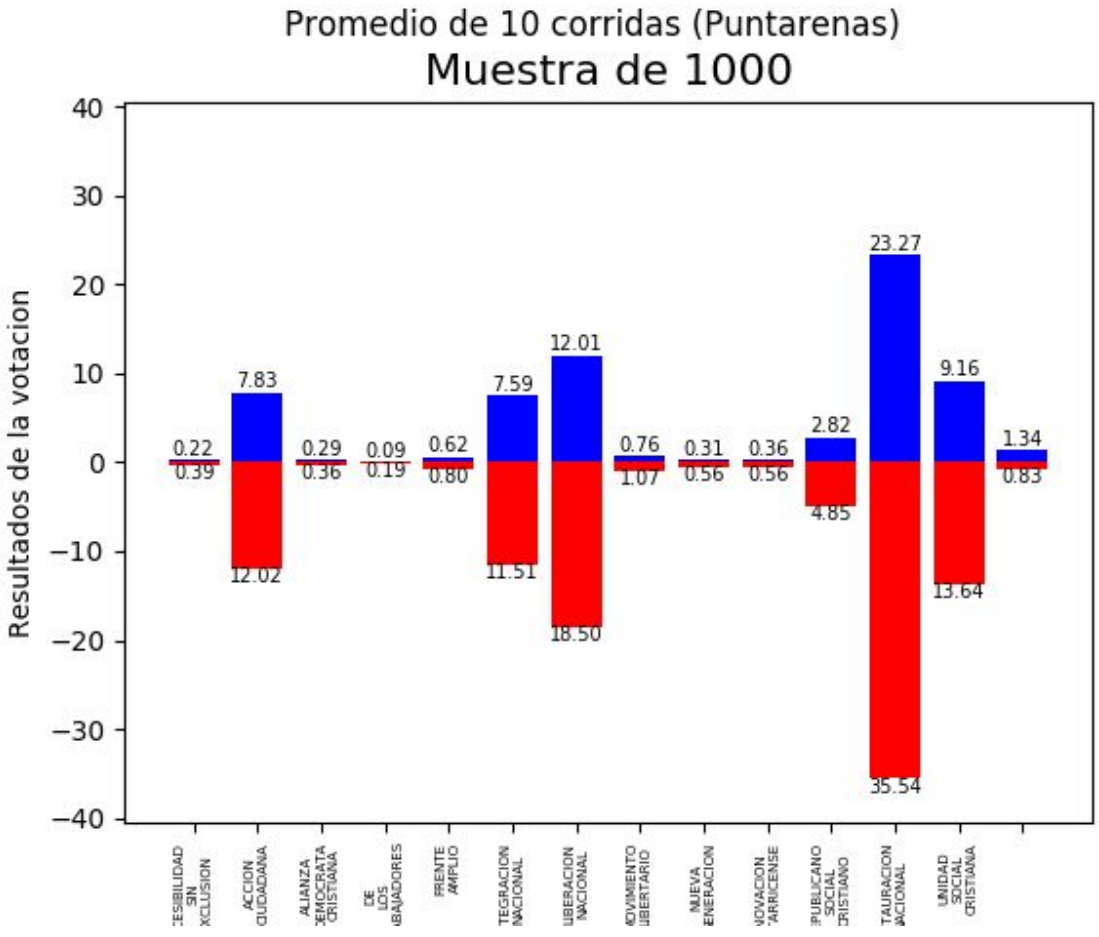

Promedio de 10 corridas (Heredia)
Muestra de 10000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.15
0.25
17.43
27.28
0.18
0.29
0.10
0.17
0.55
0.80
5.44
8.07
11.55
17.68
0.70
1.02
0.60
0.93
0.36
0.52
2.51
3.72
14.79
21.18
11.70
18.08
0.60
0.83
ACCION CIUDADANA
ALIANZA
DE LOS
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA

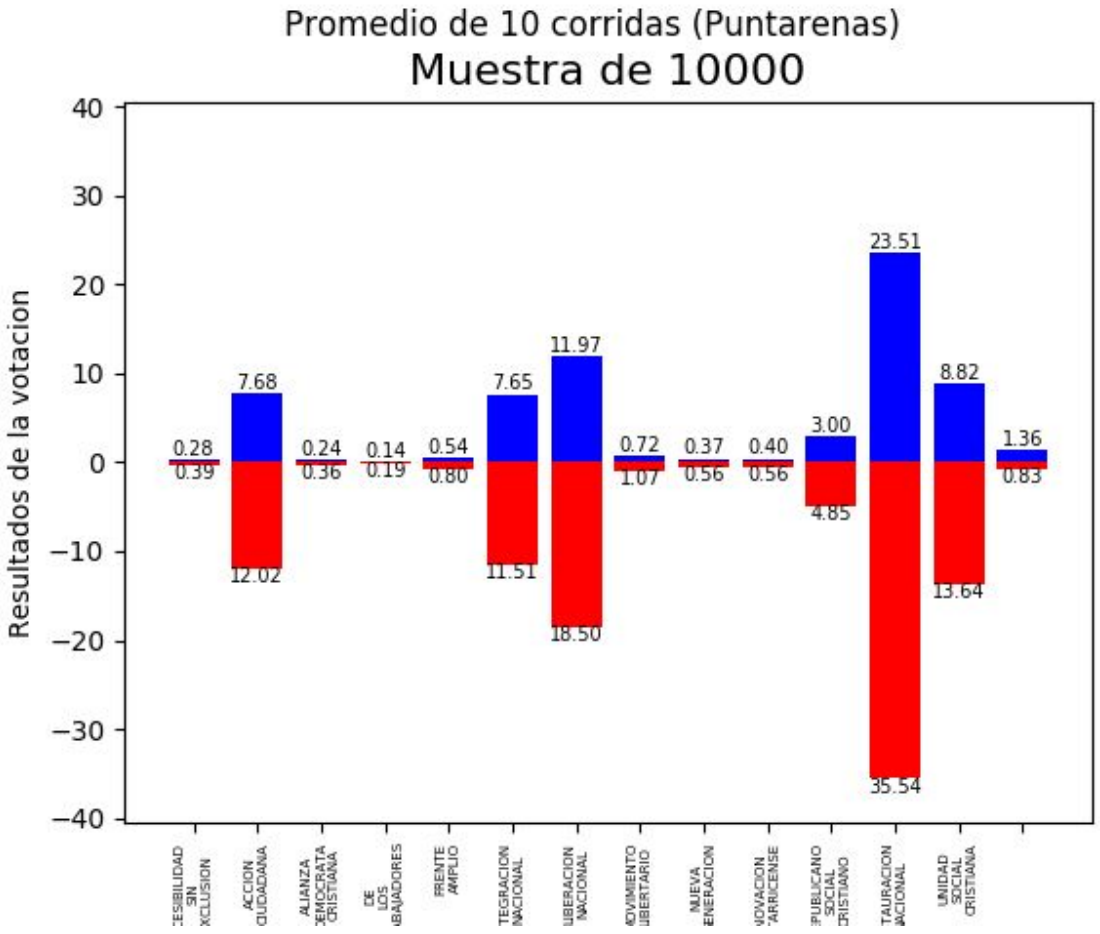

Promedio de 10 corridas (Guanacaste)
Muestra de 1000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.26
0.38
9.75
15.08
0.18
0.31
0.11
0.18
0.35
0.67
7.10
11.20
15.58
23.56
0.50
0.96
0.25
0.49
0.39
0.59
3.55
5.20
16.53
25.56
10.62
15.81
1.49
0.83
ACCION CIUDADANA
ALIANZA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA


Promedio de 10 corridas (Guanacaste)
Muestra de 10000
Resultados de la votacion
30
20
10
0
−10
−20
−30
0.24
0.38
9.79
15.08
0.21
0.31
0.13
0.18
0.42
0.67
7.37
11.20
15.33
23.56
0.57
0.96
0.31
0.49
0.39
0.59
3.36
5.20
16.93
25.56
10.23
15.81
1.41
0.83
ACCION CIUDADANA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA


Promedio de 10 corridas (Limon)
Muestra de 1000
Resultados de la votacion
40
20
0
−20
−40
0.41
0.65
6.97
10.56
0.43
0.66
0.25
0.30
0.72
0.94
6.38
10.40
11.66
17.56
0.93
1.29
0.25
0.43
0.55
0.80
3.00
4.59
27.61
42.58
5.86
9.24
1.65
0.83
ACCION CIUDADANA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA


Promedio de 10 corridas (Limon)
Muestra de 10000
Resultados de la votacion
40
20
0
−20
−40
0.43
0.65
6.78
10.56
0.43
0.66
0.20
0.30
0.67
0.94
6.68
10.40
11.39
17.56
0.82
1.29
0.28
0.43
0.48
0.80
3.09
4.59
27.76
42.58
6.11
9.24
1.53
0.83
ACCION CIUDADANA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA


Promedio de 10 corridas (Puntarenas)
Muestra de 1000
Resultados de la votacion
40
30
20
10
0
−10
−20
−30
−40
0.22
0.39
7.83
12.02
0.29
0.36
0.09
0.19
0.62
0.80
7.59
11.51
12.01
18.50
0.76
1.07
0.31
0.56
0.36
0.56
2.82
4.85
23.27
35.54
9.16
13.64
1.34
0.83
ACCION CIUDADANA
FRENTE AMPLIO
UNIDAD SOCIAL CRISTIANA



Finalmente se tiene que, pese a que el porcentaje de votos para cada partido baja con respecto al porcentaje real, es esperable ya que el valor correspondiente a los porcentajes de la muestra es un promedio. Lo importante es visualizar que la distribución de votos se mantiene y el partido ganador de cada provincia en los datos reales también gana en los datos de la muestra, por lo que se determina que el generador de población es válido para representar la población de votantes en el contexto dado.