

Interrogación 3

EYP2417 - Muestreo

Camila Echeverría

Francisca Vilca

Análisis de la elección presidencial del año 2021

En el siguiente informe, se darán a conocer algunos resultados interesantes, con respecto a las elecciones presidenciales del año 2021, se analizará al votante de 1ra Vuelta con respecto a sus características como el Sexo, Edad y GSE, además se intentarán proyectar los resultados de 2da Vuelta, para ellos se usarán diferentes técnicas de muestreo y de balanceo para que los resultados sean lo más verosímiles a la realidad.

Diseño Muestral

En este caso se realizará un muestreo de conglomerados en dos etapas, estratificadas por zona (considera solo viviendas urbanas – censo 2017). Primeramente, se obtiene una muestra aleatoria simple de manzanas al interior de cada estrato y de cada manzana se obtiene una muestra aleatoria de viviendas. Al interior de cada vivienda se selecciona un elector según la técnica del “cumpleaños”. No se realiza actualización del número de viviendas de la manzana seleccionada, de aquí son seleccionadas 1007 personas de un total de 1.520084×10^7 .

Probabilidades de Selección y Factores de Expansión según Zona, Sexo y cat-Edad de acuerdo al diseño Muestral

Por un lado si se usa la información del censo realizado el 2017, es posible obtener la información de las manzanas y las viviendas totales para calcular las probabilidades de selección. Por otro lado, si se usan las estimaciones y proyecciones de la población que se tienen para el 2021, es posible conocer la cantidad de personas por sexo y edad que habrá en Chile y usando la primera tabla de las referencias, se puede conocer los GSE que existen al 2021.

Con toda esta información se tiene que las probabilidades de selección y factores de Expansión serán:

zona	sexo	cat.edad	Prob de Selección	Factores de Expansión
1	1	1	0.000133	7537.188
1	1	2	0.000086	11693.560
1	1	3	0.000076	13141.294
1	1	4	0.000045	22064.875
1	2	1	0.000091	11017.476
1	2	2	0.000094	10617.741
1	2	3	0.000080	12523.444
1	2	4	0.000043	23180.333
2	1	1	0.000086	11684.565
2	1	2	0.000037	27304.125
2	1	3	0.000051	19460.613
2	1	4	0.000058	17209.000
2	2	1	0.000095	10549.260
2	2	2	0.000084	11966.875
2	2	3	0.000055	18050.229
2	2	4	0.000062	16015.884
3	1	1	0.000197	5063.837
3	1	2	0.000065	15467.056
3	1	3	0.000062	16182.062
3	1	4	0.000058	17291.462
3	2	1	0.000069	14558.000
3	2	2	0.000074	13562.286
3	2	3	0.000078	12845.600
3	2	4	0.000053	18807.214
4	1	1	0.000070	14200.684
4	1	2	0.000037	27337.026
4	1	3	0.000044	22727.969
4	1	4	0.000045	22166.192
4	2	1	0.000068	14601.566
4	2	2	0.000076	13192.851
4	2	3	0.000072	13852.982
4	2	4	0.000048	20705.676

Cruce de votos 2da Vuelta vs 1ra Vuelta de la muestra sin ponderar

Para este caso, sólo es de interés ver la redistribución de los votos tal y como es lo expresado por los ciudadanos de la muestra. Los resultados pueden ser apreciados en la siguiente tabla:

	Boric	Kast	Provoste	Sichel	Ártes	Enriquez	Parisi	No vote	Blanco /Nulo	No sabe
Boric	92.683	2.454	70.213	20.000	78.571	71.111	50.943	31.098	25	20.833
Kast	1.951	93.252	4.255	63.636	0.000	6.667	16.038	20.732	5	12.500
Nulo	0.976	0.613	8.511	1.818	7.143	2.222	12.264	8.232	25	8.333
No	0.000	0.613	2.128	3.636	7.143	4.444	4.717	4.573	0	0.000
Votaré										
No	4.390	3.067	14.894	10.909	7.143	15.556	16.038	35.366	45	58.333
sabe										
Total	205.000	163.000	47.000	55.000	14.000	45.000	106.000	328.000	20	24.000

Cruce de votos 2da Vuelta vs 1ra Vuelta de la muestra con ponderación

En la misma línea del ítem anterior, ahora se hará el cruce de votos pero haciendo una ponderación según la zona en la que habita la persona, si es hombre o mujer y en qué categoría se encuentra según su edad. Para poder lograr esto, debemos calcular las probabilidades de selección para cada uno de estos casos, y así podremos obtener los factores de expansión w_{ij} , cuyo procedimiento y detalle podemos encontrar en el script. De esta manera, la distribución de los votos es la apreciada en la siguiente tabla:

	Boric	Kast	Provoste	Sichel	Ártes	Enriquez	Parisi	No vote	Blanco /Nulo	No sabe
Boric	93.143	2.084	72.590	17.364	82.815	70.316	51.984	29.143	19.613	19.463
Kast	1.790	93.662	3.666	68.336	0.000	6.503	18.285	21.053	4.383	10.181
Nulo	0.912	0.866	7.436	1.560	5.324	2.005	9.779	8.628	24.704	5.365
No	0.000	0.573	2.927	2.277	5.859	6.366	4.421	5.012	0.000	0.000
Votaré										
No	4.154	2.814	13.381	10.463	6.001	14.810	15.531	36.164	51.299	64.990
sabe										
Total	205.000	163.000	47.000	55.000	14.000	45.000	106.000	328.000	20.000	24.000

Con los cruces de votos de ambas tablas, es posible observar algunas tendencias. Por ejemplo, para las personas que votaron por Boric o por Kast se mantuvieron relativamente constantes en primera y segunda vuelta, mientras que las personas que se inclinaron por Provoste, Ártes o MEO en primera vuelta decidieron ir por Boric en la segunda instancia, lo cual tiene mucho sentido al ver los discursos políticos de estos candidatos. Sin embargo, uno de los personajes más destacados fue Parisi, ya que al ser el tercer candidato con más votos, la inclinación que tomen sus simpatizantes resultará clave para determinar a un ganador. Los resultados de nuestra muestra, con ponderación y sin ella, indican que prácticamente el 50% de los votantes de Parisi apoyarían a Boric, aunque este análisis podría verse opacado por la gran masa de quienes no votaron, cuya distribución es en general equitativa entre los dos candidatos.

Cruce de votos 2da Vuelta vs 1ra Vuelta de la muestra con g-weights

Una de las particularidades que tiene el método de g-weights, es que calcula los pesos de una manera que resulta proporcional entre la población muestreada y la población total. Luego de realizar este procedimiento para las categorías solicitadas, notemos que, de manera muy similar al caso anterior, la siguiente tabla nos muestra que la mitad de los votantes de Parisi votarían por Boric en la segunda vuelta. Por otro lado, resulta interesante recalcar que quienes no sabían por quién votar en la primera vuelta, en general tampoco saben para esta segunda vuelta. Algo similar ocurre con la gente que votó nulo o blanco, pues la mayoría de esas personas tampoco sabe por quién votar. Las tendencias antes mencionadas no presentan mayores cambios en este caso.

	Boric	Kast	Provoste	Sichel	Ártes	Enriquez	Parisi	No vote	Blanco /Nulo	No sabe
Boric	91.810	2.736	68.027	22.308	69.444	71.429	50.171	31.579	22.807	16.901
Kast	1.524	92.289	5.442	63.846	0.000	5.357	13.652	19.361	7.018	11.268
Nulo	1.143	0.995	7.483	2.308	11.111	2.679	12.287	8.741	14.035	7.042
No	0.000	0.498	2.041	2.308	8.333	6.250	5.461	4.229	0.000	0.000
Votaré										
No sabe	5.524	3.483	17.007	9.231	11.111	14.286	18.430	36.090	56.140	64.789
Total	205.000	163.000	47.000	55.000	14.000	45.000	106.000	328.000	20.000	24.000

Cruce de votos 2da Vuelta vs 1ra Vuelta usando raking según Zona, Sexo, cat.edad y GSE

En general es común encontrar diferencias entre la población y la muestra que se selecciona, esto se debe principalmente a las variables demográficas, por lo que una vez que se tiene seleccionada la muestra es bueno, que antes de hacer las estimaciones hacer un ajuste en las diferencias usando ponderaciones para que los totales marginales concuerden, en este caso se usará el método de raking, ya que se sabe que es más flexible cuando se usa más de una variable, que en este caso serán Zona, Sexo, cat.edad y GSE, de esta manera quedará mejor calibrado y mejorará la eficiencia de nuestras estimaciones. Lo cual se puede observar en la Tabla 5, donde se puede notar que el porcentaje de los votos de Boric y Kast aumentaron en aquellos votantes que desde la primera vuelta estaban decididos por aquel candidato. Por otro lado y punto clave de análisis, al agregar el análisis con raking, es claro notar que los votos que originalmente eran Parisi, disminuyeron en la segunda vuelta al dar el apoyo a Boric y se sumó al grupo de no saber por quien votar. Esto resulta interesante, ya que eso es de ayuda para ambos candidatos para saber a que público dirigir su discurso con el fin de conseguir los sufragios de aquellos ciudadanos.

	Boric	Kast	Provoste	Sichel	Ártes	Enriquez	Parisi	No vote	Blanco /Nulo	No sabe
Boric	92.228	2.924	68.620	16.742	66.936	71.385	48.410	30.641	15.046	15.373
Kast	1.699	92.433	6.982	70.172	0.000	5.149	13.679	18.462	11.280	9.387
Nulo	1.243	1.508	6.567	2.072	13.094	2.677	11.270	9.315	14.949	4.009
No	0.000	0.317	2.572	2.034	5.041	6.061	5.569	4.002	0.000	0.000
Votaré										
No sabe	4.829	2.818	15.258	8.980	14.929	14.727	21.072	37.580	58.725	71.231
Total	205.000	163.000	47.000	55.000	14.000	45.000	106.000	328.000	20.000	24.000

Cruce de votos 2da Vuelta vs 1ra Vuelta usando raking según Zona, Sexo, cat.edad y voto en 1ra vuelta

Para este último caso, a diferencia del anterior, se le da importancia al ciudadano que no sabe por quien votar o que simplemente no vota, por lo que en la Tabla 6, al hacer el raking con estas nuevas variables, es posible notar, nuevamente que el cambio significativo no esta en los candidatos si no en los porcentajes de quienes no votaron, donde se ve que se mantiene la tendecia de que en porcentajes exista una alta cantidad que no vote ni en primera ni en segunda vuelta. Por lo que los candidatos, en general tienen que ser capaces de captar a ese ciudadano a través de sus propuestas para que poder declararse ganador de la elección presidencial. Esta tendencia en general no cambia la dinamica que se traía de los otros análisis solo le da mayor fuerza y validez a las ideas originales.

	Boric	Kast	Provoste	Sichel	Ártes	Enriquez	Parisi	No vote	Blanco /Nulo	No sabe
Boric	92.730	2.173	70.542	16.482	81.456	70.477	49.865	29.775	19.602	19.208
Kast	2.058	93.939	3.924	67.088	0.000	6.386	17.859	20.903	5.761	11.106
Nulo	1.012	0.636	8.948	2.031	5.824	2.552	10.770	8.530	26.408	6.362
No	0.000	0.430	3.267	2.644	5.521	4.231	4.660	4.706	0.000	0.000
Votaré										
No	4.200	2.822	13.318	11.756	7.199	16.353	16.846	36.086	48.229	63.324
sabe										
Total	205.000	163.000	47.000	55.000	14.000	45.000	106.000	328.000	20.000	24.000

Predicciones para el Total Poblacional

En general, podemos ver que para todos los ponderadores el caso es similar:

Boric sale elegido como nuevo presidente.

Sin embargo, algunas de las diferencias que se pueden apreciar en cada uno de los casos se dan por la metodología utilizada. Por ejemplo, el panorama que se veía más estrecho en cuanto al resultado es el de la d), sobre todo porque toma en cuenta la abstención al momento de hacer el cálculo de los pesos y posteriores predicciones. En cambio, una situación contraria se puede observar en el caso del ponderador de b), donde la diferencia es notoria. Como en este caso se calculan los pesos en base a las proporciones muestrales versus las reales, sin considerar la abstención, muestra más o menos qué ocurriría si todas estas personas concurrieran a votar.

Ponderador de a)

	Cantidad de Votos	Porcentaje
Boric	6636812	43.66
Kast	4441365	29.22
Nulo	803478	5.29
No Votaré	420248	2.76
No sabe	2898934	19.07
Total	15200840	100.00

Ponderador de b)

	Cantidad de Votos	Porcentaje
Boric	6899547	45.39
Kast	4302992	28.31
Nulo	888595	5.85
No Votaré	400933	2.64
No sabe	2708770	17.82
Total	15200840	100.00

Ponderador de c)

	Cantidad de Votos	Porcentaje
Boric	6413205	42.19
Kast	3985401	26.22
Nulo	939090	6.18
No Votaré	396845	2.61
No sabe	3466297	22.80
Total	15200840	100.00

Ponderador de d)

	Cantidad de Votos	Porcentaje
Boric	5705098	37.53
Kast	4289800	28.22
Nulo	985504	6.48
No Votaré	498565	3.28
No sabe	3721871	24.48
Total	15200840	100.00

Finalmente, para finalizar todo este análisis exhaustivo sobre las elecciones presidenciales, es posible concluir que para poder obtener buenos resultados y/o estimaciones en este caso para conocer quien

podría ser el próximo presidente de la República de Chile, el muestreo juega un rol clave, ya que si no se es consciente de las características de la población o no se utilizan correctamente los supuestos, se obtienen malos resultados y por ello puede que algún candidato se vea perjudicado, lo que afectaría al funcionamiento de la Democracia que según diversos historiadores, esta solo funcionará cuando todos tengan la misma información.