



FACULTAD DE CIENCIAS SOCIALES

IMPUTANDO LA NO RESPUESTA EN LA PREFERENCIA ELECTORAL DE LAS
ENCUESTAS CEP DE SEPTIEMBRE-OCTUBRE DEL 2013 Y 2017

Un acercamiento a los indecisos

por

TOMÁS IGNACIO SOTO RAMÍREZ

Tesis presentada al Instituto de Sociología de la Pontificia Universidad Católica de Chile,
para optar al título de Magister en Diseño y Análisis de Encuestas Sociales

Profesor guía:
Matías Bargsted Valdés.

Septiembre, 2021
Santiago, Chile

© 2021, Tomás Ignacio Soto Ramírez

© 2021, Tomás Ignacio Soto Ramírez

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica que acredita el trabajo y a su autor.

DEDICATORIA

A mis amigos por transmitirme el amor por el conocimiento y la ciencia.

TABLA DE CONTENIDO

DEDICATORIA	ii
ÍNDICE DE TABLAS	v
ÍNDICE DE ILUSTRACIONES	vi
RESUMEN	vii
I. INTRODUCCIÓN	1
II. MARCO CONCEPTUAL	4
2.1. Preferencia Electoral	4
2.1.1. Teoría socio estructural del voto.....	4
2.1.2. Teoría de la identificación partidaria	5
2.1.3. Enfoque de la elección racional	6
2.1.4. Otras variables de interés.....	7
2.2. La no respuesta en las encuestas electorales	8
2.2.1. La no respuesta al ítem	9
2.2.2. El caso de las encuestas electorales	10
2.3. Modelo del votante probable	11
2.4. Métodos de imputación	12
2.4.1. Imputación múltiple.....	13
2.5. Métodos de imputación en la estimación de la preferencia electoral	14
2.5.1. España	14
2.5.2. Estados Unidos	15
2.5.3. Chile	15
III. DATOS Y METODOLOGÍA	17
3.1. Tratamiento de las variables	17
3.1.1. Predictores de preferencia electoral	17
3.1.2. Variables Dependientes	18
3.2. Realización de la Imputación	20
3.2.1. Creación de modelos de regresión	20
3.2.2. Creación de bases de datos para imputación.....	22
3.2.3. Imputación.....	23
IV. RESULTADOS	26
4.1. Predictores seleccionados para los modelos de regresión	26
4.2. Modelos de regresión multinomial para explicar la preferencia electoral	26
4.3. Imputaciones	32
4.3.1. Comparación entre escenarios	32

V. CONCLUSIONES	37
BIBLIOGRAFÍA	38

ÍNDICE DE TABLAS

<i>Tabla 3.1. Creación de preferencia electoral con votante probable elección 2013</i>	24
<i>Tabla 3.2. Creación de preferencia electoral con votante probable elección 2017</i>	24
<i>Tabla 3.3. Datos perdidos a imputar en la preferencia electoral de las elecciones 2013 y 2017</i>	25
<i>Tabla 4.1. Regresión 1 y 2</i>	27
<i>Tabla 4.2. Regresión 3 y 4</i>	30
<i>Tabla 4.3. Imputación 1</i>	33
<i>Tabla 4.4. Imputación 2</i>	35

ÍNDICE DE ILUSTRACIONES

<i>Figura 3-1 Preferencia electoral para la encuesta CEP de septiembre-octubre 2013 y 2017</i>	<i>20</i>
------------------------------------------------------------------------------------------------------	-----------

RESUMEN

La presente tesis se encarga de evaluar si es posible mejorar la capacidad predictiva de las encuestas preelectorales en Chile por medio de técnicas de imputación de datos. Para dar respuesta a esta interrogante, los objetivos se centraron en 1) la utilización del modelo de votante probable de González y Mackenna, para tener más precisión sobre quiénes realmente votan, 2) estimar un modelo de regresión logística multinomial para identificar los predictores que indican la preferencia electoral y 3) estimar un modelo de imputación múltiple de ecuaciones encadenadas de la preferencia electoral.

Para llevar a cabo los modelos de regresión multinomial y posterior imputación, se utilizaron 9 variables independientes y 1 dependiente de las encuestas CEP de septiembre-octubre de 2013 y 2017. En términos analíticos se decidió comparar la preferencia electoral en un escenario con y sin votante probable para observar sus diferencias. De lo anterior se concluye que: 1) no existe mayor diferencia entre la imputación con y sin votante probable. 2) pese a que la imputación se adapta a la distribución de la preferencia electoral en ambas mediciones, tiende a estimar de mejor manera a los candidatos de derecha por ser los que presentan una mayor fuerza predictiva en los modelos de regresión. Adicionalmente, la imputación en los candidatos con mayor favoritismo provoca que las estimaciones se alejen de los valores reales de la elección. 3) la imputación por sí sola no mejora la predicción electoral, debido a que se necesita que el instrumento logre medir la preferencia electoral de buena manera y esté lo mejor representado posible.

I. INTRODUCCIÓN

El desarrollo de las encuestas electorales ha tenido una presencia intermitente en la historia de Chile. En 1957, con las denominadas “Encuestas Hamuy” comenzaron a realizarse mediciones regulares hasta 1973, observándose una baja pronunciada durante los primeros años de la dictadura militar, para volver a reactivarse en la década de los 80 con un número relevante de estudios a medida que se acercaba el plebiscito de 1989 (Cordero y Tapia, 2007).

Desde 1990 hasta el 2006, la industria de las encuestas electorales experimentó un incremento importante de trabajos debido a la incorporación de empresas, fundaciones y Think Thnaks, llevando a una profesionalización paulatina del rubro, donde los temas éticos como la transparencia, procesado e información metodológica pública se convirtieron en la propuesta de valor de las diferentes encuestadoras (Cordero y Tapia, 2007).

Simultáneamente, los medios de comunicación se presentaron no solo como difusores, sino también como productores de investigaciones, donde se buscaba la credibilidad de las audiencias y la industria a través de la realización de sondeos, demostrando que las encuestas servían para crear contenido y articular el debate público.

En este contexto, las encuestas de opinión pública y, por sobre todo las electorales, fueron haciéndose conocidas por la ciudadanía, manteniendo algún grado de confianza por los resultados que se obtenían al medir la intencionalidad del voto o la preferencia de los electores, por la experiencia acumulada desde las “Encuestas Hamuy”.

En el 2012, con la implementación de la inscripción automática y el voto voluntario, las instituciones encuestadoras comenzaron a tener problemas para pronosticar los resultados electorales porque no se sabía quiénes realmente participaban, generando una sobreestimación de algunos candidatos y subestimación de otros (Corvalán, Cox y Hernández, 2015), trayendo como consecuencia desconfianza por parte de la mayoría de segmentos de la sociedad y un problema a nivel de industria con la salida de GFK-Adimark del mercado de las encuestas electorales.

Así, la pregunta sobre la intencionalidad de los ciudadanos respecto al candidato que votará se convierte en una de las variables más importantes dentro de una encuesta electoral, una pregunta considerada de carácter sensible, que posee, por un lado, un componente importante de deseabilidad social (Corvalán, Cox y Hernández, 2015; Braña, Rial, Varela y Vázquez, 1998), provocando que algunos encuestados tengan mayor propensión a mentir por el candidato de su preferencia.

Y por otro, presenta una alta presencia de no sabe y no responde que, según Cruz (1990) formaría parte de uno de los principales problemas de esta variable, porque indirectamente evaluaría si los sujetos están informados sobre los candidatos y si se identifican con alguno, lo que llevaría a un proceso cognitivo más elevado con respecto a otras preguntas, provocando una mayor probabilidad de rechazo (Braña, et al., 1998).

El año 2012, por lo tanto, marcó un antes y un después en el análisis del comportamiento electoral en Chile, no solo para dar cuenta de la abstención y volatilidad electoral de los ciudadanos chilenos, sino también para establecer y generar nuevas investigaciones en torno a las metodologías de las encuestas electorales para mejorar la precisión y disminuir los errores de estimación que trajo consigo la reforma.

Una de las soluciones alternativas para mejorar la estimación de los resultados electorales en un contexto de mayor incertidumbre, según Perry (1973), Dimock, Keeter, Shulman y Miller (2001), Traugott (2014), González y Mackenna (2017), Marshall (2018), y Rentsch, Schaffner y Gross (2019), consistía en la creación de un modelo de votante probable, el cual permite discriminar entre aquellos encuestados que participarán en una elección, quienes optarán por abstenerse y quienes se encuentran indecisos.

Para el caso de la preferencia electoral, se han desarrollado los enfoques de carácter estadístico metodológico que evalúan distintos modelos de imputación en esta variable (Braña et al., 1998; King, Honaker, Joseph y Scheve, 2001; Bernhagen y Marsh, 2007; Rivas, Martínez y Galindo, 2010; Markham, Falk, McGrath y Schneider, 2011; Escobar y Jaime, 2013; Liu, 2014; Bautista 2016; Cabrera y Escobar 2019; Bernhagen y Marsh 2010) cuyos propósitos se centran en el hecho que, al completar los valores faltantes, estos se acerquen o alejen de los resultados reales de una elección.

En términos generales, la imputación es considerada como un conjunto de métodos estadísticos que pretenden completar la información faltante de una pregunta o un conjunto de variables (Useche y Mesa, 2006). La literatura en torno al tema ha determinado que existen una variedad importante de tipos (tradicionales, simples y múltiples), los cuales se pueden adaptar de mejor manera dependiendo de la distribución de los datos faltantes o la cantidad de información disponible en la base de datos (INE, 2019).

En resumen, la estimación de la preferencia electoral se ha transformado en uno de los problemas más relevantes que poseen las encuestas electorales, porque desde que se instauró en Chile la inscripción automática y el voto voluntario el 2012, solo algunas instituciones encuestadoras se han acercado a los valores reales de las elecciones (encuesta opina-El Mercurio 2013 y la encuesta UDD en el 2017).

Haciéndose necesario acudir, como alternativa, al modelo de votante probable, que calcula la participación electoral (si la persona votará en las elecciones) y a los métodos de imputación estadísticos que tratan de ajustar y solucionar el problema de la no respuesta en la pregunta de preferencia electoral (por cuál candidato votarán).

En la presente investigación, se utilizó el método de imputación múltiple de ecuaciones encadenadas porque ha sido el más probado en la literatura en las encuestas electorales (Markham et al., 2011; Liu, 2014; Rivas, Martínez y Galindo, 2010; Cabrera y Escobar, 2019; Bernhagen y Marsh 2010), y en las de tipo poblacional como la Encuesta Nacional de Presupuestos Familiares (2018).

Para llevar a cabo la imputación, se utilizaron los datos del Centro de Estudios Públicos (CEP), correspondientes a septiembre-octubre del 2013 y septiembre-octubre del 2017,

porque en ambos instrumentos se observan todos los candidatos que participaron en sus respectivas primeras vueltas, además que el Centro de Estudios Públicos fue una de las instituciones que presentó mayores problemas en términos de estimación de preferencia electoral, sobre todo en la elección 2013.

Finalmente, estas dos mediciones permiten comprobar la validez y replicabilidad de los modelos utilizados.

En consecuencia, de todo lo anterior, cabe preguntarse si ¿Es posible mejorar la capacidad predictiva de las encuestas preelectorales en Chile por medio de técnicas de imputación de datos?

Para responder esta pregunta, 1) se calculó el modelo de votante probable de González y Mackenna (2017) con los datos de la Encuesta CEP septiembre-octubre del 2013 y 2017, 2) identificaron las variables predictoras que afectan las preferencias electorales en Chile y 3) estimó el modelo de imputación múltiple presentado anteriormente.

Hay que considerar que es importante llevar a cabo esta investigación, en primer lugar, porque teóricamente se buscará crear un modelo parsimonioso que explique la participación electoral, para ser replicable en diferentes encuestas políticas.

En segundo lugar, posee una pertinencia metodológica porque el modelo en sí mismo permitirá identificar no solo los predictores sino también, ayudará a mejorar la medición de la encuesta, debido a que al ser más precisa la predicción, baja el error de procesamiento y por ende el error total de la encuesta.

Finalmente, cabe destacar que no se han desarrollado investigaciones en Chile que utilicen métodos de imputación múltiples en encuestas electorales, por lo que este trabajo es pionero en lo que respecta al análisis de la no respuesta en el ámbito de la opinión pública.

II. MARCO CONCEPTUAL

El sustento teórico que abraza esta investigación se centra en 5 puntos: 1) preferencia electoral; 2) la no respuesta en las encuestas políticas; 3) modelo de votante probable; 4) métodos de imputación; y 5) los métodos de imputación en la preferencia electoral.

2.1. Preferencia Electoral

El comportamiento electoral de las personas que no responden a la pregunta de preferencia electoral se presenta de manera heterogénea y la inclinación de su voto ha sido a diferentes candidatos o partidos políticos a lo largo del tiempo (Escobar y Jaime, 2013)

Para dar solución a este escenario de incertidumbre se utilizan los métodos de imputación, los que se basan en utilizar la información completa de los sujetos que declararon abiertamente por cuales candidatos votarán para predecir a aquellos que no respondieron. En este sentido, se hace necesario identificar los predictores que explican la preferencia electoral en Chile.

Dentro de la vasta literatura enfocada en la preferencia electoral, se trabajó con la teoría socio estructural del voto, la teoría de la identificación partidaria y el enfoque del elector racional o voto económico.

2.1.1. Teoría socio estructural del voto

En términos generales, esta teoría plantea que, el voto está definido por la posición del votante en la estructura social o los grupos con características similares, los que tienen un comportamiento electoral homogéneo (Escobar y Jaime, 2013), por lo mismo, en el escenario político, los intereses colectivos de cada grupo se ven representados por los partidos políticos.

El origen de la diferencia de cada grupo viene dado por los clivajes sociales, quienes responden a quiebres profundos en la historia de las sociedades contemporáneas (Torres, 2016). Así, estos permiten que los individuos tomen posturas con respecto a temas de carácter controversial, los que se hacen más evidentes en el ámbito de las políticas públicas.

Lipset y Rokkan (1967) como fue citado en Torres (2016), definieron 4 tipos de clivajes: 1) cultura; 2) religión; 3) etnia; y 4) clase.

El primero toma como referencia el conflicto centro-periferia o identitario, el que surge en el momento mismo de la formación de los Estados Nación. Se trata del conflicto entre la cultura de construcción nacional del centro y la resistencia de las poblaciones en las periferias o entre las distintas etnias o religiones en un mismo territorio (Torres, 2016). Por lo tanto, el conflicto viene dado por una oposición identitaria como en caso de los mapuches en Chile o del País Vasco en España.

El segundo viene dado por el conflicto de secularización, donde se busca despojar a la Iglesia del control económico a favor de los Estados modernos (Torres, 2016). En este sentido, generalmente el mecanismo de secularización se ha dado en áreas como la educación, sin embargo, en las últimas décadas, el conflicto se ha desarrollado en torno a la salud como el aborto o la eutanasia.

El tercero hace referencia al conflicto campo vs ciudad o tierra vs industria. Un conflicto que nace entre la elite terrateniente que domina el campo y la burguesía. Por lo tanto, está contextualizado posterior a la revolución industrial con la creación de los estados-nación y la distribución del poder político y económico en la nueva sociedad (Torres, 2016). Un caso de este tipo de clivaje se presenta en Argentina, específicamente en la confrontación entre Buenos Aires y el resto de las provincias del Interior.

El cuarto se centra en el conflicto entre empresarios versus trabajadores, un clivaje que nació con la consolidación de la sociedad industrial. Según Torres (2016), se establece que el conflicto ha tendido a desaparecer en Latinoamérica, sin embargo, el debate sigue estando presente porque se ha expandido más allá de los Estados para crear un conflicto a nivel internacional consecuencia de la globalización.

En las últimas décadas se ha observado que las fisuras que determinan el origen de los partidos políticos se han diluido casi por completo, consecuencia de los cambios socioeconómicos de la era posindustrial, resolviendo que la clase social es un factor cada vez menos importante para predecir la participación electoral (Escobar y Jaime, 2013). Sin embargo, el hecho que los clivajes pierdan cada vez más importancia en la teoría del comportamiento electoral, no quiere decir que no estén presentes en algunos sujetos de la sociedad actual.

Por lo tanto, las variables a utilizar bajo esta perspectiva teórica serán: 1) asistencia a servicios religiosos que está presente en el clivaje religioso, 2) grupo socioeconómico, donde se espera que aquellos votantes que pertenezcan a un grupo socioeconómico más elevado tiendan a votar por candidatos de derecha en comparación a votantes de segmentos socioeconómicos más bajos (Altman, 2004; Morales, 2010; Dodds, 2017), y 3) nivel educacional, una variable que está fuertemente relacionada con el grupo socioeconómico.

2.1.2. Teoría de la identificación partidaria

Este enfoque teórico tiene como idea central que el comportamiento de los votantes es explicado en base a su identificación con los diferentes partidos políticos (Escobar y Jaime 2013), siendo estructurado a través de los procesos de socialización política, desarrollándose un sistema actitudinal de afectos o aversiones hacia los distintos elementos del sistema político.

De esta forma, la identificación hacia un partido se adquiere en una primera instancia por la familia, para luego ser reforzada por la pertenencia a ciertos grupos secundarios y, manifestada en el voto en un partido determinado en el tiempo (Roche, 2008),

Así, este comportamiento se ve condicionado por la socialización política y, en algunas ocasiones, puede sufrir modificaciones por la percepción de los candidatos o las relaciones mantenidas con los partidos o sus representantes.

Harrup y Miller (1987) citado en Escobar y Jaime (2013) resumen esta teoría en 5 dimensiones: 1) la mayoría de los electores sienten una afinidad hacia un partido, la cual es heredada por la familia; 2) la identificación partidista influye en la determinación del voto, pero también afecta la percepción de las políticas, candidatos y los vínculos entre grupos sociales; 3) la identificación con un partido se correlaciona con la edad del votante, puesto que esta se refuerza con el paso del tiempo. Los cambios en la identidad partidaria son personales, ya que reflejan el cambio de grupos de pertenencia, a menudo con la movilidad geográfica o social 4) los electores, aunque estén influenciados por una campaña electoral contra el partido que se identifican, mantendrán su identidad y volverán a votar por él; 5) la distribución de las identidades partidistas entre un electorado provee la base para calcular el pronóstico electoral, si no intervinieran variables de corto plazo a favor de ningún partido.

En síntesis, esta perspectiva da énfasis a las identidades partidistas, la imagen de los políticos y la influencia de las campañas electorales en los medios de comunicación. En este contexto, el voto es considerado como un mecanismo de afirmación, pero no de pertenencia a una clase social (como la teoría estructural del voto), sino que es un proceso de largo plazo que ha llevado al individuo a identificarse con un partido en particular.

En consecuencia, la variable que estará presente en los modelos para realizar la imputación será identificación por partido político.

2.1.3. Enfoque de la elección racional

Esta perspectiva teórica es una aplicación de la economía a la política. En este contexto, el votante se transforma en un ser racional, donde evalúa los costos y beneficios de los procesos que envuelven el contexto del voto. Por lo tanto, bajo esta perspectiva, un votante es un sujeto que maximiza sus utilidades con motivaciones egoístas, concibiendo la política como un mercado, en donde los partidos políticos compiten por votos como un medio para obtener recompensas una vez logran llegar al poder (Escobar y Jaime 2013).

El creador de esta teoría es Downs (1957), el cual se basaba en 4 premisas: 1) los votantes calculan la utilidad que han recibido del gobierno como suma de todos los beneficios obtenidos; 2) los votantes asumen que recibirán en el futuro la misma utilidad que en el pasado si ese gobierno se mantiene en el poder, es decir, los votantes no discriminan en torno a las promesas, sino más bien sobre el desempeño de los gobiernos; 3) los votantes calculan la utilidad que recibirían si el partido en la oposición accediera al gobierno, lo cual se presenta como un desafío puesto que no pueden juzgar su desempeño real, pero es necesario para comparar entre los partidos; 4) los electores calculan la diferencia entre sus alternativas a través de sus utilidades y votarán por el partido que maximiza sus beneficios. En este sentido, Fiorina (1977), argumentó que la evaluación que los ciudadanos hacen de los partidos, no se limita al régimen actual, sino también, en las experiencias anteriores del gobierno.

Por lo tanto, considerando que la hipótesis central de este modelo se basa en la responsabilidad (Escobar y Jaime, 2013; Montecinos, 2007), los ciudadanos recompensarán con su voto a los partidos que tengan buenos resultados económicos y castigarán electoralmente a aquellos gobernantes que produzcan una recesión, siendo complementada con la perspectiva de Navia y Del Pozo (2012), donde observaron que, existe una menor predisposición a votar entre las personas que consideraban que la situación económica actual del país era mala o muy mala, al igual que la percepción de la situación económica futura (Navia y del Pozo, 2012).

No obstante, una de las principales críticas al modelo económico es que la información no es ilimitada y se distribuye de manera desigual, por lo tanto, los ciudadanos no toman decisiones exclusivamente racionales. En este sentido, la aprobación de un gobierno es una mezcla de acciones racionales: una evaluación de beneficios más una fuente subjetiva (experiencias previas, expectativas) más niveles de simpatía político-partidista.

Por lo tanto, cuando un elector está identificado con un partido político tanto a nivel cognitivo como emocional, se identifica con su plataforma ideológico-política, generando una comunidad de intereses y afectividad (Montecinos, 2007).

En este contexto, la variable a utilizar será la percepción de la situación económica del país.

2.1.4. Otras variables de interés

Edad

Se considera relevante utilizar la variable edad, porque es uno de los predictores más significativos tanto en la participación electoral (Navia, 2004; Toro, 2008; Bargsted, Somma y Muñoz, 2019; PNUD, 2019) como en la preferencia electoral (Morales, 2010; Escobar y Jaime, 2013).

En términos de la participación electoral, la edad podría ser un indicador proxy de indicadores de maduración en la perspectiva del ciclo de vida político (Highton y Wolfinger 2001, como se citó en Bargsted, Somma y Muñoz, 2019), es decir, la transición de la juventud a una vida adulta como el optar por estudios superiores, búsqueda de una pareja e ingresar al mercado laboral, podría llevar a los jóvenes a restarse del proceso electoral (Bargsted, Somma y Muñoz, 2019).

Una vez adulto esto tiende a cambiar debido a que, al poseer más recursos, formar familia y el poseer vínculos con la comunidad, aumenta el interés en la política y esta, a su vez, aumenta la participación (Bargsted, Somma y Muñoz, 2019).

Para el caso de la preferencia electoral, la edad está presente tanto en la teoría socio-estructural del voto debido a que permite diferenciar a un grupo social de otro, como en la teoría de la identificación partidaria donde se refuerza la filiación a medida que aumenta la edad del votante (Escobar y Jaime, 2013).

Ideología

La postura ideológica, al igual que la edad es considerada una variable fundamental en la preferencia electoral (Escobar y Jaime, 2013; Sulmont, 2015; Torrico y Solis, 2020).

Una de las principales conclusiones de Sulmont (2015) en su estudio comparado del voto ideológico y sistema de partidos en Brasil, Chile, México y Perú, es que la dimensión izquierda derecha tiene un peso en la decisión electoral. Sin embargo, está condicionado por la forma en que los partidos políticos logran diferenciarse entre sí (Sulmont, 2015).

En aquellos países donde ocurre esa diferenciación, los electores son capaces de utilizar la dimensión izquierda/derecha para identificarse e identificar a los partidos. En este sentido, Chile sería un ejemplo donde se manifestaría este comportamiento.

Aprobación presidencial

La literatura especializada ha trabajado la aprobación presidencial desde tres perspectivas: 1) la identidad partidaria, 2) la percepción de personalidad de los líderes y 3) la teoría del voto económico.

La primera se enfoca en que aquellos sujetos que se sientan identificados con el partido de gobierno tenderán a aprobar de manera positiva al presidente (Campbell, Converse, Miller y Donald, 1960; Fiorina, 1981, como se citó en Ortiz y García, 2014).

En el caso de la segunda, se establece un vínculo entre las percepciones de la ciudadanía con respecto a la personalidad del presidente (Kinder, 1981, como se citó en Ortiz y García, 2014), es decir, si un sujeto percibe que quien está gobernando tiene las cualidades necesarias para liderar, aprobará su labor. Entre las características más importantes del gobernante destacan: el carisma, simpatía, conocimiento, liderazgo, etc. (Ortiz y García, 2014).

Finalmente, desde la perspectiva del votante racional, un ciudadano aprobará o desaprobará al presidente dependiendo de la percepción de su economía y la situación económica del país.

Sexo

Sexo es una variable que está considerada en la literatura de comportamiento electoral como poco explicativa o ambigua (PNUD, 2017). No obstante, se integrará al análisis por ser una característica sociodemográfica básica en la identificación de un ciudadano.

2.2. La no respuesta en las encuestas electorales

El fenómeno de la no respuesta ha sido estudiado por diferentes disciplinas, como la sociología, psicología, ciencia política, economía y todas aquellas ciencias que deben utilizar cuestionarios como métodos de recolección de datos.

En términos metodológicos, la no respuesta se puede manifestar de dos formas: la primera - no respuesta a la unidad- ocurre cuando no logramos recolectar datos a partir de una unidad

seleccionada en la muestra, y la segunda cuando la falta de información se restringe a algunos ítems del cuestionario, denominándola no respuesta al ítem (Groves, et al, 2009). Ambos tipos de no respuesta tienen un doble impacto en la calidad de los datos que se recolectan, provocando un potencial sesgo de no respuesta, y reducción del tamaño efectivo de muestra, generando una disminución de la precisión de las estimaciones.

2.2.1. La no respuesta al ítem

La falta de respuesta en algún ítem de un cuestionario es común en muchas disciplinas, y pueden ser generados por el diseño del estudio, errores o fallas en el proceso de producción de datos decisiones durante el proceso de revisión, edición o validación de los datos recolectados y por decisión (voluntaria o no) de los entrevistados (Groves, et al., 2009).

2.2.1.1. *Diseño:*

Ítems que, por diseño, se aplican solo a parte de los entrevistados, por ejemplo, solo a los mayores de 60 años (filtros de subpoblación), o solo a una muestra aleatoria de la población (submuestra).

2.2.1.2. *Fallas:*

Pérdida involuntaria de cuestionarios -como robo de material, incendio, etc., o datos -pérdida de datos electrónicos, errores de procesamiento-

2.2.1.3. *Procesamiento de los datos:*

El objetivo final del proceso de edición y validación de datos es producir una base de datos consistente y con la menor cantidad de errores posible, sin embargo, no es posible evaluar la validez de todos los reportes entregados por los entrevistados. En vez de eso, se busca lograr que los reportes sean consistentes pese a que nunca es posible eliminar todos esos errores.

Los resultados del proceso de edición/validación incluyen: dejar los datos tal y como están; modificar los datos observados con información auxiliar, o reemplazar los datos observados por datos perdidos.

2.2.1.4. *Por decisiones del entrevistado:*

Según Beatty y Herrmann (2002), el fenómeno de la no respuesta al ítem está asociado a tres mecanismos: cuanto saben las personas acerca del tema, sus juicios acerca de la precisión de su conocimiento y lo que la pregunta parece requerir; y su voluntad de reportar la respuesta.

Los primeros dos factores son cognitivos o informativos, mientras que el tercero es motivacional. En este contexto, el modelo aplicado a preguntas de ingreso, por ejemplo, predice que los entrevistados no van a responder cuando, por ejemplo: No sepan el ingreso familiar; tengan una idea vaga del monto, pero asuman que la pregunta exige una cantidad exacta, o sepan la respuesta, pero no estén dispuestos a revelarla al encuestador o a la institución a cargo del estudio (Beatty y Herrmann, 2002).

2.2.2. El caso de las encuestas electorales

En el caso específico de las encuestas políticas, Cruz (1990), establece que la tasa de no respuesta a las cuestiones políticas depende de predisposiciones sociológicas (sexo, nivel de estudios y la residencia urbana-rural), del nivel de implicación política y de la orientación ideológica. Estas características, además de estar muy relacionadas entre sí, ejercen cada una de ellas una influencia, de manera independiente.

En este sentido, la autora observa que la no respuesta se encuentra más frecuentemente entre los mayores de treinta años, casados y de un nivel medio de ingresos; entrevistados de mediana edad, de un nivel bajo de estudios, así como en el colectivo de personas solteras y otras personas aisladas (Cruz, 1990). Además, la tasa de no respuesta es siempre más elevada entre los que tienen sólo estudios primarios o no tienen estudios que entre aquellos que han adquirido un nivel de educación más elevado, y entre los habitantes de zonas rurales (comunidades de menos de 2.000 habitantes) que entre los que residen en zonas urbanas.

Siguiendo esta misma línea, en el trabajo de Cruz (1990), se observa un listado con las temáticas que causaron más problemas en un estudio llevado a cabo por el CIS de España, en marzo de 1989 sobre recabar la opinión de los trabajadores respecto a distintos aspectos de su trabajo, donde destacan: la desconfianza del anonimato (68%); longitud del cuestionario (66%); ignorancia -no sabe- (57%); miedo (35%); preguntas improcedentes (28%); no poder encuadrarse dentro de las categorías (17%); falta de tiempo (11%); falta de interés por el tema.

En el caso específico de las preguntas que más problema se tuvo, destacan: intención de voto (88%); política (75%); preguntas personales o datos de identificación (47%); equipamiento del hogar (46%); sobre la vida íntima (35%) y el número de teléfono.

Se hace necesario destacar esta información porque la pregunta de intención de voto es de carácter sensible y por lo mismo requiere de una mayor destreza para el encuestador o un proceso cognitivo mayor para el entrevistado (Medway y Tourangeau, 2015).

En síntesis, las causas de la no respuesta se pueden encontrar, en diferentes momentos de la puesta en marcha de una encuesta y, por tanto, en los diversos instrumentos y técnicas utilizados, por ejemplo:

- En la elaboración de un cuestionario (diseño).
- Durante la realización de la entrevista, ya sea debido a que el entrevistador posee una escasa capacitación, como al entrevistado que no esté informado, se niegue a participar.
- Características sociodemográficas que favorecen la no respuesta (edad o género)
- Seleccionar una mala fecha para llevar a cabo el cuestionario.
- Nivel de involucramiento político (interés por la temática en cuestión).
- Ambigüedad Ideológica.

2.3. Modelo del votante probable.

Uno de los principales problemas que trajo el paso del sistema electoral de inscripción voluntaria con voto obligatorio a uno de inscripción automática con voto voluntario, para las encuestas de opinión pública, y, por sobre todo a las encuestas electorales, fue la incertidumbre para predecir correctamente no solo la intencionalidad del voto, sino también la probabilidad de participación. En este contexto, diferentes autores como Marshall (2018), Traugott (2014), González y Mackenna (2017), entre otros, están estudiando implementar el modelo de votante probable en diferentes países que tengan un sistema de voto voluntario para disminuir la volatilidad de no conocer los votantes que realmente participarán en una elección.

Para la presente investigación se utilizará el modelo de González y Mackenna (2017), el cual utiliza 4 variables, (intención de participación, edad, interés en la elección y voto en elecciones pasadas) para identificar a los votantes probables.

Para llegar al votante probable se requiere seguir 5 pasos:

- Se calificará como votante a la persona que haya respondido “Sí, con toda seguridad irá a votar” a la pregunta “Y en su caso, ¿irá Ud. a votar en las próximas elecciones presidenciales de 2017?”.
- Se calificará como no votante a la persona que haya respondido “Sí, con toda seguridad irá a votar” a la pregunta “Y en su caso, ¿irá Ud. a votar en las próximas elecciones presidenciales de 2017?”, pero que no haya votado en la elección presidencial pasada.
- Los jóvenes que no tenían edad para votar en la elección presidencial pasada se calificarán como votantes si respondieron: “Sí, con toda seguridad irá a votar” a la pregunta “Y en su caso, ¿Irá Ud. a votar en las próximas elecciones presidenciales de 2017? y han declarado estar “muy” o “bastante” interesados en la elección presidencial de 2017.
- Se clasificará como votante a la persona que haya respondido “Probablemente sí” a la pregunta “Y en su caso ¿irá Ud. a votar en las próximas elecciones presidenciales de 2017?, haya votado en la elección presidencial pasada y haya declarado estar “muy” o “bastante” interesado en la elección presidencial de 2017.
- Se calificará como votante a la persona que no sepa o no haya respondido a la pregunta, “Y en su caso, ¿Irá a votar en las próximas elecciones presidenciales de 2017?, haya votado en la elección presidencial pasada y haya declarado estar “muy” o “bastante” interesado en la elección presidencial de 2017.

En términos de resultado, los autores estipularon que:

“Hemos mostrado que el modelo produce estimaciones cercanas a los niveles de participación observados en las elecciones municipales de 2012 y 2016 y la primera vuelta

presidencial de 2013, utilizando la información de las encuestas preelectorales del CEP, y se aproxima con razonable precisión a la participación electoral por sexo y edad” (González y Mackenna, 2017: p.8).

Se hace necesario calcularlo en esta investigación porque al mostrar que no existen tantas diferencias entre los resultados de las elecciones y el modelo, la estimación de la participación electoral se simplifica de gran manera, obteniendo valores más precisos.

2.4. Métodos de imputación

La no respuesta puede aparecer en las encuestas por varias razones: un entrevistador puede no hacer la pregunta, un encuestado puede negarse a responder, un empleado que ingresa los datos puede omitir un valor, etc. A veces, los elementos con respuestas se cambian a faltantes cuando el conjunto de datos se limpia o existe un valor que no cumple con los criterios lógicos de la encuesta, es decir, se debe contar como perdido si un niño de 3 años votó en la última elección presidencial (Lohr, 2009).

La imputación se usa comúnmente para asignar valores a los elementos faltantes. Un valor de reemplazo, a menudo, de otra persona que es similar al elemento que no responde en otras variables, imputando o completando el valor faltante con estas características. Cuando se utiliza la imputación, se debe crear una variable adicional para el conjunto de datos, la cual indica si la respuesta se midió o se imputó (Lohr, 2009).

En términos históricos, los métodos de imputación son relativamente nuevos, debido a que comenzaron a desarrollarse en la década de los 60 con la utilización del procedimiento de Hot Deck en todas sus variantes, para reemplazar los datos faltantes de encuestas, censos y estudios (CEPAL, 2007).

En 1976, Rubin desarrolló un marco conceptual para solucionar el problema de la no respuesta a través de técnicas de inferencia estadística, para que después, con la llegada del algoritmo Expectation Maximization (EM) permitiera generar estimadores robustos a partir de la aplicación del método de máxima verosimilitud (CEPAL, 2007). Llegando, finalmente a la concepción de imputación múltiple, un modelo estadístico sofisticado que tiene como premisa que cada valor debe ser reemplazado a partir de $m > 1$ simulaciones (Rubin, 1976).

Finalmente, llegando a los 90, los avances se enfocaron en sofisticar y adaptar los métodos clásicos, como el procedimiento de reponderación, utilizado en la imputación de encuestas, se modificó para ser aplicado en términos de regresión, además de desarrollarse el campo de la bioestadística con el propósito de imputar cuando las características dependían de características poblacionales.

Un punto previo que se debe tener en cuenta para llevar a cabo cualquier tipo de imputación es observar el patrón de cómo se encuentran distribuidos los valores perdidos. Según la literatura existente, esto se denomina “noción de mecanismo de no respuesta” (Rubin, 1976), a partir de la cual se distinguen tres tipos de datos faltantes, según el patrón de aleatoriedad que éstos sigan.

En primer lugar, se encuentran los datos faltantes tipo MCAR (missing completely at random), o lo que Groves y Peytecheva (2008) denominan “modelo de causas separadas”, donde tanto la propensión a responder como la variable de interés no depende de los datos observados, es decir, los datos completos son equivalentes a los datos incompletos.

En segundo lugar, MAR (missing at random) o denominada por Groves y Cooper (1998) como “modelo de causa común”, se observa cuando la propensión a responder y la variable de interés pueden ser explicadas por un mismo conjunto de variables, estableciendo que, pese a que los sujetos con datos completos son diferentes a los incompletos, estos últimos pueden ser predichos a partir de los primeros. En este escenario se encuentran la mayoría de las encuestas a hogares.

Finalmente, se encuentra el modelo MNAR (missing not at random), donde la variable de interés determina la propensión a responder, en este sentido, la no respuesta adquiere un carácter no aleatorio (INE, 2019). En estos casos, el mismo fenómeno que se pretende medir influye en la propensión de las personas a responder.

En la vasta literatura especializada de los métodos de imputación, existe una variedad importante de procedimientos que se pueden trabajar en ciencias sociales (Braña et al, 1998; King, et al, 2001; Bernhagen y Marsh 2007; Rivas, Martínez y Galindo 2010; Markham et al., 2011; Escobar y Jaime 2013; Liu, 2014; Bautista 2016; Cabrera y Escobar 2019; Bernhagen y Marsh 2010), en este sentido, como se ha ido observando en la presente investigación, el método que más se ha probado y han tendido resultados interesantes en términos de la participación electoral, ha sido, la imputación múltiple con un método encadenado (King, et al., 2001; Bernhagen y Marsh 2007; Rivas, Martínez y Galindo 2010; Markham, et al., 2011; Liu, 2014; Bautista, 2016; Cabrera y Escobar, 2019; Bernhagen y Marsh, 2010).

2.4.1. Imputación múltiple.

El método de imputación múltiple se basa en imputar más de un valor para cada caso ausente. En este sentido, la imputación consiste en generar $m > 1$ valores aleatorios para cada valor perdido por no respuesta, de manera que se disponga de m conjuntos de datos completos que permitan realizar los análisis estadísticos usuales (media, regresión, etc.) de los m conjuntos de datos generados para las m estimaciones, con la finalidad de producir una estimación con buenas propiedades estadísticas y con la posibilidad de estimar la varianza de las estimaciones.

En términos metodológicos, Román (2017), citando a Rubin (1976) estipula que este modelo de imputación posee 3 etapas: 1) Imputación; 2) Análisis y 3) Combinación.

Por lo general, en una imputación múltiple, se desarrolla bajo el procedimiento de Markov Chain Monte Carlo (Román, 2017). El MCMC utiliza simulación paramétrica generando muestras aleatorias a partir de métodos bayesianos y, en el método de imputación múltiple,

se aplica para generar las m selecciones independientes de valores faltantes, las cuales se utilizan en la etapa de inferencia o análisis.

Como método de extracción de los valores perdidos, se decidió utilizar ecuaciones encadenadas, las cuales se presentan como un algoritmo cuya base se presenta en el hecho de calcular una ecuación de regresión para cada variable con datos faltantes.

Para comenzar con el algoritmo, en primer lugar, se deben ordenar las variables de menor a mayor en términos de sus valores perdidos. En segundo lugar, el algoritmo predice los valores faltantes de la primera variable hasta rellenar sus celdas vacías. Una vez que la variable está completa, se utilizará en el modelo para imputar la siguiente variable y así sucesivamente hasta completar la base de datos.

Cabe destacar que se debe definir el modelo de regresión adecuado a las variables imputadas, por ejemplo, si son categóricas nominales, multinomiales u ordinales, lineales, etc.

2.5. Métodos de imputación en la estimación de la preferencia electoral.

En el presente apartado se revisarán diferentes estudios respecto a la estimación de encuestas electorales en países donde se trabaja en torno a la no respuesta y a los modelos de imputación utilizados para la solución de este problema, en España, Estados Unidos y Chile.

2.5.1. España

En este país se han encontrado variadas investigaciones respecto a la predicción de encuestas electorales e imputación, destacando:

- Métodos de imputación múltiple para predecir resultados electorales (Escobar y Jaime, 2013).
- La imputación múltiple como alternativa al análisis de la no respuesta en la variable de voto (Rivas, Martínez y Galindo, 2010).
- Especial encuestas: errores, cocina y predicción (Penadés, 2015).
- Estimación de la respuesta de los "no sabe/no contesta" en los estudios de intención de voto (Braña et al., 1998).
- Del no sabe al no contesta: un lugar de encuentro para diversas respuestas (Cruz, 1990).

Estos textos evalúan los diferentes modelos de imputación que utilizaron para poder corregir la no respuesta en la pregunta de preferencia electoral. En el caso de Escobar y Jaime (2013), los autores usaron dos tipos de imputación: uni univariado (imputación por regresión) y uno múltiple de carácter encadenado, resultando el primero más cercano a los votos reales obtenidos en la elección, en comparación al segundo.

Para el caso de Rivas, Martínez y Galindo (2010) se observa un resultado similar al estudio anterior, en el sentido que, los autores estipulan que, el método no mejora sustantivamente los resultados de las elecciones generales de España en el 2000. Ahora bien, en el caso de Penadés (2015), se establece que la forma más antigua y básica para llegar a la preferencia electoral es asignar a los indecisos en función de su partido más cercano y corregir el resultado por el recuerdo del voto.

Las dos investigaciones siguientes conceptualizan el fenómeno de la no respuesta tanto a la unidad como al ítem. En el caso específico de Cruz (1990), la investigación esboza los rasgos más importantes de la no respuesta en las encuestas de opinión pública, donde se hace hincapié en las diferentes concepciones del “no sabe”, “no contesta”, sus causas y los diferentes modos de abordar el fenómeno.

En el caso de Braña et al, (1998), desarrollan una investigación propia donde se define y comparan 4 métodos de imputación: Listwise, asignación proporcional, análisis discriminante y Hot-Deck. Los resultados indicaron que el Hot-Deck combinado con el análisis discriminante fueron los métodos más eficaces para estimar el voto probable de los sujetos que no se pronunciaron sobre la intención de voto en la elección de 1996.

2.5.2. Estados Unidos

Este país es uno de los más importantes en términos del desarrollo de teorías relacionadas a las encuestas electorales y a la predicción de voto, no obstante, para objeto de este apartado, destacaré una investigación de Traugott (2014), presentada en la revista “Estudios Públicos N°137” del Centro de Estudios Públicos (CEP), denominada: “métodos alternativos para la estimación de resultados electorales”.

En términos generales, este ensayo abarca la dificultad de la predicción de una encuesta política en el paso de un sistema de voto obligatorio a uno voluntario en Chile, entregando como solución, la creación de un votante probable. Para reforzar su teoría, el autor extrae información de las distintas elecciones en Estados Unidos. Asimismo, se encuentran una serie de investigaciones (King et al., 2001; Bernhagen y Marsh, 2007; Markham et al., 2011; Liu, 2014; Bernhagen y Marsh, 2010) donde se aplica la imputación múltiple en los estudios de preferencia electoral, mejorando levemente las estimaciones, no obstante, actualmente se encuentra la discusión de que esta ausencia de precisión se deba a que se asume MAR (Missing at Random) siendo que la naturaleza del fenómeno de la no respuesta en la mayoría de la ciencias sociales, tiene un patrón Not missing at Random (Liu, 2014; CEPAL 2007; Restrepo y Marín, 2012). En este contexto, se presenta una alternativa, presentando variables auxiliares.

2.5.3. Chile

Finalmente, la predicción de encuestas electorales a través de métodos de imputación no está muy trabajada en Chile, ya que las investigaciones que más se asocian al tema, se han

enfocado en determinar al votante probable y a un informe de evaluación al Centro de Estudios Públicos:

- Análisis Metodológico de la Encuesta del Centro de Estudios Públicos (Marshall, 2018).
- ¿Quiénes votan?... O por qué es tan difícil predecirlo (Carvacho, et al., 2018).
- Un modelo de votante probable para la encuesta CEP (González & Mackenna, 2017).

El primer texto hace referencia a un informe metodológico realizado por Pablo Marshall (2018), donde analizó cada una de las etapas de la encuesta, identificando variables contextuales (estado del arte de las instituciones encuestadoras en Chile) y metodológicas (votante probable, muestra, pregunta con no sabe o no responde, ponderadores, trabajo de campo y efecto encuestadores). Para terminar con sugerencias relacionadas a mejorar el diseño muestral de la encuesta, generar un modelo estandarizado de votante probable (modelo Perry – Gallup) y modificar los ponderadores.

El segundo estudio revisado hace referencia a una investigación realizada por un grupo de científicos sociales de la Pontificia Universidad Católica de Chile. En ella se revisan las variables que ayudan a construir la conducta de voto como las características sociodemográficas del votante, la historia de participación del votante, las características de la elección y de los candidatos, actitudes y conocimiento político del votante, y las condiciones asociadas a la conducta del voto.

En este contexto, los autores concluyen que las encuestas podrían enriquecer su capacidad de predecir los resultados electorales si incorporaran modelos más complejos de la identificación de votantes probables, que incorporen aspectos psicosociales, tales como las normas sociales (Carvacho, et al., 2018).

Finalmente, el estudio de (González y Mackenna, 2017), tiene como objetivo crear un modelo de votante probable, el cual se compone de 4 variables: intención de participación, edad, interés en elección, y voto en elecciones pasadas, y fue testeado en las elecciones municipales de 2012 y 2016; y la primera vuelta presidencial de 2013, cuyo resultado fue cercano a los datos reales de las votaciones.

III. DATOS Y METODOLOGÍA

Los datos que se utilizaron en la presente investigación se obtuvieron de dos encuestas: 1) CEP 70, correspondiente a septiembre y octubre del 2013 y 2) CEP 81, realizada entre septiembre y octubre del 2017 porque fueron las últimas mediciones previo a la elección presidencial de esos años.

En términos metodológicos, la encuesta CEP tiene una modalidad de entrevista presencial, con un muestro probabilístico por conglomerados trietápico, un error muestral aproximadamente de 3% y un nivel de confianza de un 95%. La muestra para la primera encuesta fue de 1.437 sujetos y la segunda, 1.424.

Las variables que se utilizaron fueron aquellas que según la literatura presentan un efecto en la variable dependiente y las que permiten calcular el modelo de votante probable.

3.1. Tratamiento de las variables.

3.1.1. Predictores de preferencia electoral

Las variables trabajadas en esta investigación fueron: 1) ideología; 2) sexo; 3) edad; 4) identificación con partido político; 5) percepción de la situación económica del país; 6) aprobación presidencial 7) nivel educacional; 8) grupo socioeconómico y 9) asistencia a servicios religiosos.

Los predictores ordinales fueron tratados como variables de razón a diferencia de las nominales que fueron transformadas en dummies para ser integradas a los modelos.

Ideología se presenta como una variable nominal con 5 categorías: Derecha, Centro, Izquierda, Independiente y Ninguno, dejando como categoría de referencia Derecha (para ambos años). Sexo es una variable nominal binaria con 1 Hombre y 2 con Mujer, siendo esta última la categoría de referencia. Edad se trabaja como una variable de razón.

Identificación con algún partido político para el año 2017 poseía 10 categorías: Renovación Nacional (RN), Partido Demócrata Cristiano (DC), Partido por la Democracia (PPD), Unión Demócrata Independiente (UDI), Partido Socialista (PS), Partido Radical Social Demócrata (PRSD), Partido Comunista (PC), Partido Evolución Política (EVOPOLI), Otros Partidos y Ningún Partido.

El tratamiento de esta variable consistió en agrupar los partidos de la UDI y RN por ser los más importantes de la derecha chilena y, en transformar las categorías restantes a dummy. Dejando, finalmente al partido Demócrata Cristiano (DC) como referencia y, la agrupación anterior con “Ningún Partido” como las categorías a incluir en el modelo, debido a que permitieron no solo discriminar entre los candidatos, sino también, porque presentó una significación cercana a 0 con respecto a la variable dependiente.

Para la medición del 2013, la encuesta CEP consideró 11 partidos: Partido Demócrata Cristiano (DC), Unión Demócrata Independiente (UDI), Partido Comunista (PC), Renovación Nacional (RN), Partido Socialista (PS), Partido Radical Social Demócrata

(PRSD), Partido Humanista (PH), Partido Regionalista Independiente (PRI), Partido Progresista (PRO), Otro Partido Político y Ningún Partido.

El tratamiento de esta variable, al igual que el año 2017, fue agrupar RN y UDI y, dejar como indicador la categoría “DC” y Ningún Partido debido a que permitió diferenciar a los candidatos, y porque la agrupación de RN y UDI presentó valores muy altos en los efectos de la regresión multinomial.

Percepción de la situación económica del país es una variable de carácter ordinal donde se pregunta a los sujetos cómo calificarían la actual situación económica del país, cuyas categorías son: Muy Mala, Mala, Regular, Buena y Muy Buena. Como se mencionó anteriormente, esta variable es ordinal y fue trabajada como razón, es decir, sus categorías de respuesta fueron analizadas como valores numéricos del 1 al 5.

La aprobación presidencial se expresa como “independiente de su posición política: ¿usted aprueba o desaprueba la forma como Michelle Bachelet//Sebastián Piñera está conduciendo su gobierno?” es una variable nominal cuyas categorías de respuesta fueron: Aprueba=1; Desaprueba=2; No aprueba ni Desaprueba=3 (dejando como referencia la primera categoría para ambos años de medición).

El nivel educacional se presentó como una variable con 9 categorías para cada año: 0) No estudió, 1) Básica Incompleta, 2) Básica Completa, 3) Media Incompleta, 4) Media Completa, 5) Técnica Superior Incompleta, 6) Técnica Superior Completa, 7) Universitaria Incompleta, 8) Universitaria Completa y, 9) Postgrado. Esta variable al ser ordinal se analizó de manera numérica, es decir, del 0 al 9.

Grupo socioeconómico es una variable ordinal que se presenta con 5 categorías en ambas mediciones: ABC1=1, C2=2, C3=3; D=4 y E=5. Esta variable se trabajó con las categorías de 1 al 5, es decir, como una variable numérica.

Frecuencia para asistir a servicios religiosos es ordinal con 9 categorías en ambas mediciones (Varias veces a la semana; una vez a la semana; 2 o 3 veces al mes; Una vez al mes; Varias veces al año; Una vez al año; Menos frecuente que una vez al año; Nunca y Rechaza).

El tratamiento de esta variable se centró, por un lado, en considerar la categoría “Rechaza” como perdido y, por otro, al ser ordinal se trabajó con las categorías del 1 a 8.

3.1.2. Variables Dependientes

La variable dependiente en esta investigación fue la preferencia electoral o por cuál candidato votará en las próximas elecciones presidenciales. Como ya se ha planteado, uno de los mayores problemas que trajo el cambio de sistema electoral al campo de la medición de la opinión pública fue que no solo se tiene incerteza por los candidatos a escoger, sino también que no se conoce quienes realmente participarán en la elección.

Una solución para enfrentar el último problema fue con el modelo de votante probable realizado por González y Mackenna (2017), el cual, permitió observar no solo cuan precisa

fue la pregunta de preferencia electoral de la encuesta, sino también porque se tiene un número más realista de aquellos sujetos a los cuales se les pretende imputar la no respuesta.

En consecuencia, votante probable, se manifiesta como una variable nominal de 3 categorías: Votante Probable=1, Votante No Probable=2 e Indecisos=3), siendo un símil de propensión a votar.

Preferencia electoral con votante probable (simil de preferencia electoral), para la encuesta CEP 70, se compone de 9 categorías: Evelyn Matthei, Michelle Bachelet, Marco Enríquez Ominami, Franco Parisi, Marcel Claude, Roxana Miranda, Alfredo Sfeir, Ricardo Israel y No Votaría. En el caso de la CEP 81, se compuso de 8 categorías: Marco Enriquez Ominami, Carolina Goic, Alejandro Guillier, José Antonio Kast, Alejandro Navarro, Sebastián Piñera, Beatriz Sánchez y No votaría.

En términos del tratamiento de estas variables, se decidió, en primer lugar:

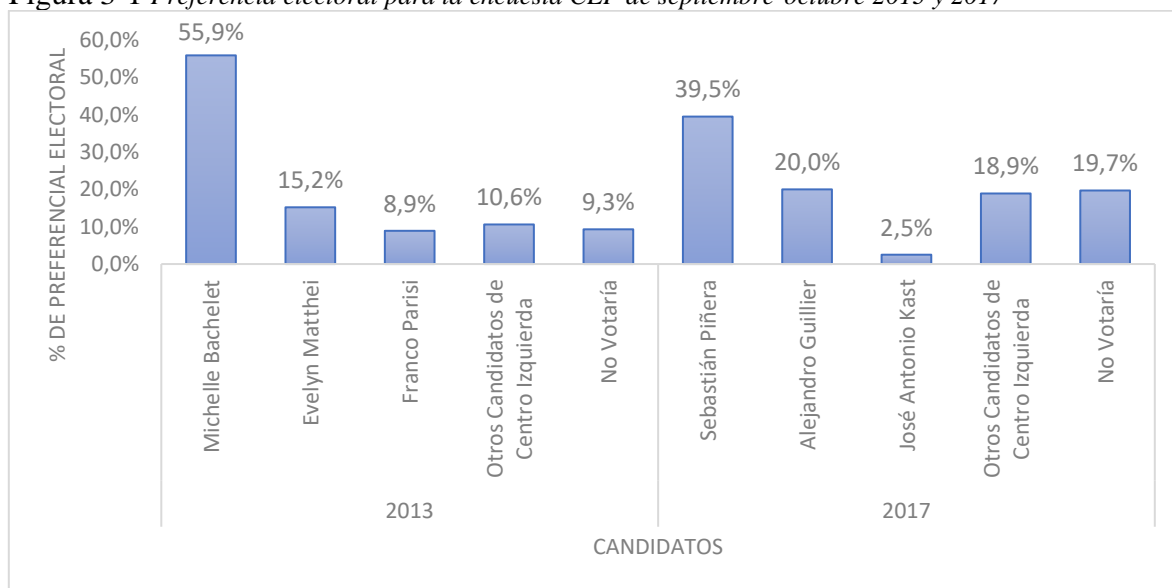
- Eliminar del análisis a Ricardo Israel en la encuesta 2013 porque contaba con una sola preferencia y presentaba valores perdidos en las variables predictoras, dificultando el proceso de imputación.
- Eliminar del análisis al candidato Eduardo Artés en la encuesta 2017 porque, al igual que el caso anterior, contaba con una sola preferencia y tenía valores perdidos en alguna de las variables predictoras. Asimismo, según el criterio de González y Mackenna (2017), el sujeto que prefería al candidato era considerado como votante no probable.

En segundo lugar, se agruparon aquellos aspirantes con mayor cantidad de preferencias y a los candidatos de derecha, dejando a los demás como “Otros candidatos de Centro e Izquierda”. Por lo tanto, para el caso de la encuesta CEP 70, la variable dependiente se configura con 5 categorías: Evelyn Matthei=1; Michelle Bachelet=2; Franco Parisi=3; Otros Candidatos de Centro Izquierda (Marco Enrique Ominami, Marcel Claude, Roxana Miranda, Alfredo Sfeir)=4; y No Votaría=5. Siendo la categoría de referencia “Otros candidatos de centro e izquierda”.

El proceso de la encuesta CEP 81 fue similar. Los candidatos sometidos a evaluación y posterior imputación fueron: Sebastián Piñera, Alejandro Guillier y José Antonio Kast; el resto de los candidatos se consideran de Centro e Izquierda (Carolina Goic, Marco Enrique Ominami, Beatriz Sánchez y Alejandro Navarro) y No Votaría .

La categoría No Votaría para ambas mediciones se eliminó del análisis debido a que el objetivo principal es imputar los valores perdidos en base a los datos de los candidatos. Por lo tanto, la muestra se compondrá exclusivamente de los casos que correspondan a los candidatos, más los valores perdidos.

Figura 3-1 Preferencia electoral para la encuesta CEP de septiembre-octubre 2013 y 2017



Fuente: Elaboración propia a partir de la encuesta CEP de septiembre-octubre 2013 y 2017

3.2. Realización de la Imputación.

Los pasos para realizar una imputación son 4:

- Realización de un modelo de regresión (multinomial) con los predictores que predicen la variable dependiente a imputar.
- Construcción de la base de datos con los predictores y la variable a imputar.
- Imputación.
 - Ordenar la base de datos de menor a mayor cantidad de valores perdidos.
 - Imputar los valores perdidos a través de la técnica de “imputación múltiple con ecuaciones encadenadas”.
- Comparación de los valores imputados con los resultados de las elecciones.

3.2.1. Creación de modelos de regresión

La creación del votante probable provocó que la preferencia electoral se pudiera analizar en un contexto normal (como viene dada en la encuesta) y bajo la idea de votante probable. Por lo mismo, se realizaron 2 modelos de regresión multinomial para cada medición (2013 y 2017): el primero en un contexto sin votante probable y uno con votante probable, cuyo detalle se expresa de la siguiente manera:

Modelo multinomial 1 y 2: Preferencia electoral de las candidatas con mayor favoritismo más Franco Parisi, en un contexto *con y sin votante probable* para la primera vuelta presidencial del 2013.

$$\left(\ln \frac{P(\text{Michelle Bachelet} = 1)}{P(\text{Otros Candidatos de Centro Izquierda} = 0)} \right)$$

$$= \beta_{10} + \beta_{11}(\text{Ideología (Centro)}) + \beta_{12}(\text{Ideología (Izquierda)})$$

$$+ \beta_{13}(\text{Ideología (Independiente)}) + \beta_{14}(\text{Ideología (Ninguno)}) + \beta_{15}(\text{Sexo (Hombre)})$$

$$+ \beta_{16}(\text{Edad}) + \beta_{17}(\text{Identificación con Partido Político (Ninguno)})$$

$$+ \beta_{18}(\text{Identificación con Partido Político (DC)}) + \beta_{19}(\text{Situación económica del país})$$

$$+ \beta_{110}(\text{Desempeño del Gobierno (Desaprueba)})$$

$$+ \beta_{111}(\text{Desempeño del Gobierno (No Aprueba ni Desaprueba)}) + \beta_{112}(\text{Nivel Educativo})$$

$$+ \beta_{113}(\text{Grupo Socioeconómico}) + \beta_{124}(\text{Asistencia a Servicios Religiosos})$$

$$\left(\ln \frac{P(\text{Evelyn Matthei} = 2)}{P(\text{Otros Candidatos de Centro Izquierda} = 0)} \right)$$

$$= \beta_{20} + \beta_{21}(\text{Ideología (Centro)}) + \beta_{22}(\text{Ideología (Izquierda)})$$

$$+ \beta_{23}(\text{Ideología (Independiente)}) + \beta_{24}(\text{Ideología (Ninguno)}) + \beta_{25}(\text{Sexo (Hombre)})$$

$$+ \beta_{26}(\text{Edad}) + \beta_{27}(\text{Identificación con Partido Político (Ninguno)})$$

$$+ \beta_{28}(\text{Identificación con Partido Político (DC)}) + \beta_{29}(\text{Situación económica del país})$$

$$+ \beta_{210}(\text{Desempeño del Gobierno (Desaprueba)})$$

$$+ \beta_{211}(\text{Desempeño del Gobierno (No Aprueba ni Desaprueba)}) + \beta_{212}(\text{Nivel Educativo})$$

$$+ \beta_{213}(\text{Grupo Socioeconómico}) + \beta_{214}(\text{Asistencia a Servicios Religiosos})$$

$$\left(\ln \frac{P(\text{Franco Parisi} = 3)}{P(\text{Otros Candidatos de Centro Izquierda} = 0)} \right)$$

$$= \beta_{30} + \beta_{31}(\text{Ideología (Centro)}) + \beta_{32}(\text{Ideología (Izquierda)})$$

$$+ \beta_{33}(\text{Ideología (Independiente)}) + \beta_{34}(\text{Ideología (Ninguno)}) + \beta_{35}(\text{Sexo (Hombre)})$$

$$+ \beta_{36}(\text{Edad}) + \beta_{37}(\text{Identificación con Partido Político (Ninguno)})$$

$$+ \beta_{38}(\text{Identificación con Partido Político (DC)}) + \beta_{39}(\text{Situación económica del país})$$

$$+ \beta_{310}(\text{Desempeño del Gobierno (Desaprueba)})$$

$$+ \beta_{311}(\text{Desempeño del Gobierno (No Aprueba ni Desaprueba)}) + \beta_{312}(\text{Nivel Educativo})$$

$$+ \beta_{313}(\text{Grupo Socioeconómico}) + \beta_{314}(\text{Asistencia a Servicios Religiosos})$$

Modelo multinomial 3 y 4: Preferencia electoral de los candidatos con mayor favoritismo más José Antonio Kast, en un contexto *con y sin votante probable* para la primera vuelta presidencial del 2017.

$$\left(\ln \frac{P(\text{Alejandro Guillier} = 1)}{P(\text{Otros Candidatos de Centro Izquierda} = 0)} \right)$$

$$= \beta_{10} + \beta_{11}(\text{Ideología (Centro)}) + \beta_{12}(\text{Ideología (Izquierda)})$$

$$+ \beta_{13}(\text{Ideología (Independiente)}) + \beta_{14}(\text{Ideología (Ninguno)}) + \beta_{15}(\text{Sexo (Hombre)})$$

$$+ \beta_{16}(\text{Edad}) + \beta_{17}(\text{Identificación con Partido Político (Ninguno)})$$

$$+ \beta_{18}(\text{Identificación con Partido Político (DC)}) + \beta_{19}(\text{Situación económica del país})$$

$$+ \beta_{110}(\text{Desempeño del Gobierno (Desaprueba)})$$

$$+ \beta_{111}(\text{Desempeño del Gobierno (No Aprueba ni Desaprueba)}) + \beta_{112}(\text{Nivel Educativo})$$

$$+ \beta_{113}(\text{Grupo Socioeconómico}) + \beta_{124}(\text{Asistencia a Servicios Religiosos})$$

$$\left(\ln \frac{P(\text{Sebastián Piñera} = 2)}{P(\text{Otros Candidatos de Centro Izquierda} = 0)} \right)$$

$$= \beta_{20} + \beta_{21}(\text{Ideología (Centro)}) + \beta_{22}(\text{Ideología (Izquierda)})$$

$$+ \beta_{23}(\text{Ideología (Independiente)}) + \beta_{24}(\text{Ideología (Ninguno)}) + \beta_{25}(\text{Sexo (Hombre)})$$

$$+ \beta_{26}(\text{Edad}) + \beta_{27}(\text{Identificación con Partido Político (Ninguno)})$$

$$+ \beta_{28}(\text{Identificación con Partido Político (DC)}) + \beta_{29}(\text{Situación económica del país})$$

$$+ \beta_{210}(\text{Desempeño del Gobierno (Desaprueba)})$$

$$+ \beta_{211}(\text{Desempeño del Gobierno (No Aprueba ni Desaprueba)}) + \beta_{212}(\text{Nivel Educativo})$$

$$+ \beta_{213}(\text{Grupo Socioeconómico}) + \beta_{214}(\text{Asistencia a Servicios Religiosos})$$

$$\left(\ln \frac{P(\text{José Antonio Kast} = 3)}{P(\text{Otros Candidatos de Centro Izquierda} = 0)} \right)$$

$$= \beta_{30} + \beta_{31}(\text{Ideología (Centro)}) + \beta_{32}(\text{Ideología (Izquierda)})$$

$$+ \beta_{33}(\text{Ideología (Independiente)}) + \beta_{34}(\text{Ideología (Ninguno)}) + \beta_{35}(\text{Sexo (Hombre)})$$

$$+ \beta_{36}(\text{Edad}) + \beta_{37}(\text{Identificación con Partido Político (Ninguno)})$$

$$+ \beta_{38}(\text{Identificación con Partido Político (DC)}) + \beta_{39}(\text{Situación económica del país})$$

$$+ \beta_{310}(\text{Desempeño del Gobierno (Desaprueba)})$$

$$+ \beta_{311}(\text{Desempeño del Gobierno (No Aprueba ni Desaprueba)}) + \beta_{312}(\text{Nivel Educativo})$$

$$+ \beta_{313}(\text{Grupo Socioeconómico}) + \beta_{314}(\text{Asistencia a Servicios Religiosos})$$

El Pseudo R cuadrado escogido para comparar los modelos fue Nagelkerke porque permite una mejor interpretación que Cox y Snell al tener su escala de 0 a 1. Asimismo, los criterios para determinar qué variables conforman los modelos explicativos recaen en la parsimonia y en el aporte predictivo del modelo en su conjunto.

3.2.2. Creación de bases de datos para imputación.

Un paso previo para realizar la imputación fue la creación de las bases de datos tanto para el escenario normal como para el votante probable.

Para cada una de las mediciones, se definieron 2 bases de datos: 1) Preferencia electoral sin votante probable sin “No Votaría”; y, 2) preferencia electoral con votante probable sin “No votaría”.

En términos concretos, la/las bases de datos deben estar ordenadas de menor a mayor en términos de los valores perdidos.

Para el año 2013, preferencia electoral “sin votante probable”, se ordenó con: sexo, edad, grupo socioeconómico, nivel educativo, situación económica del país, aprobación del gobierno, identificación con partido político, ideología, asistencia a servicios religiosos y preferencia electoral.

En el caso con votante probable, las variables fueron: sexo, edad, grupo socioeconómico, nivel educativo, situación económica del país, ideología, identificación con partido político, desempeño del gobierno, asistencia a servicios religiosos y preferencia electoral.

La creación de la base para imputar la preferencia electoral sin votante probable en el año 2017 las variables fueron: sexo, edad, grupo socioeconómico, situación económica del país,

nivel educacional, asistencia a servicios religiosos, aprobación del gobierno, ideología, identificación con partido político y preferencia electoral.

Finalmente, la base para imputar la preferencia electoral con votante probable se compone por: sexo, edad, grupo socioeconómico, nivel educacional, situación económica del país, asistencia a servicios religiosos, desempeño del gobierno, ideología, identificación con partido político y preferencia electoral.

3.2.3. Imputación

Las imputaciones que se llevaron a cabo para responder al objetivo general de esta investigación se centraron en un método de imputación múltiple.

El método de imputación múltiple consiste en generar $m > 1$ valores aleatorios para cada valor perdido por no respuesta, de manera que se disponga de m conjuntos de datos completos que permitan realizar los análisis estadísticos usuales de los m conjuntos de datos generados para las m estimaciones con la finalidad de producir una estimación con buenas propiedades estadísticas y con la posibilidad de estimar la varianza de las estimaciones.

Como método de extracción de los valores perdidos, se decidió utilizar ecuaciones encadenadas, las cuales se presentan como un algoritmo cuya base se presenta en el hecho de calcular una ecuación de regresión para cada variable con datos faltantes, es decir, se debe ordenar la base de datos de menor a mayores valores perdidos, para que el algoritmo prediga los valores faltantes de la primera variable hasta rellenar sus celdas vacías. Una vez que la variable está completa, se utilizará en el modelo para imputar la siguiente variable y así sucesivamente hasta completar la base de datos.

La librería utilizada para llevar a cabo la imputación fue “*mice*”, albergada en el software estadístico R, un paquete cuyos parámetros se centraron en: la base de datos con valores perdidos, los tipos de regresión a utilizar para las diferentes variables a imputar, la semilla de partida del algoritmo, la cual fue 132

(criterio aleatorio) y la cantidad de bases de datos a generar, las que en este caso fueron 20 porque según Rivero (2011), aumenta la precisión a medida que aumenta el número de bases de datos.

El criterio para fusionar las diferentes bases de datos se basó en el principio de combinación elaborado por Rubin (1987), el cual se llevó a cabo por el comando “complete” de MICE. De acuerdo con lo expuesto anteriormente, se define que los datos a imputar en la presente investigación son los que se observan en la tabla 3.3.

Donde la cantidad de valores perdidos en la encuesta normal (sin votante probable) es mayor en comparación a la preferencia electoral cuando se utiliza un modelo de votante probable, porque este último es un subconjunto de la muestra total.

En otras palabras, los valores perdidos en un contexto sin votante probable (como viene dado en la encuesta), tenderán a ser mayores que en un modelo con votante probable porque en el

primero se tienen todos los valores perdidos de la pregunta (no sabe y no responde), en comparación al segundo que, al ser un subconjunto de la muestra, los valores perdidos serán igual o menores a los de la muestra original. Este fenómeno se puede observar en las tablas 3.1 y 3.2.

Tabla 3.1. Creación de preferencia electoral con votante probable elección 2013

Categorías de Respuesta	Intención de Participación Encuesta	%	Preferencia Electoral Encuesta	%	Votante Probable	%	Preferencia Electoral con Votante Probable	%
Vota	1.077	74,95	1.136	79,05	719	50,03	650	90,40
No Vota	314	21,85	117	8,14	447	31,11	0	0
Casos Perdidos	46	3,20	184	12,80	271	18,86	69	9,60
Total	1.437	100,00	1.437	100,00	1.437	100	719	100,00

Fuente: Elaboración propia a partir de la encuesta CEP de septiembre-octubre 2013 y 2017

La diferencia entre los votantes del modelo con votante probable con la nueva variable de preferencia electoral responde a que, si bien se observaron 719 votantes propensos a participar en la elección, 69 de ellos declararon en la nueva pregunta de preferencia electoral que no votarían (13), no saben (32) o no respondieron (24), por lo mismo, todos esos valores fueron trabajados como casos perdidos para ser posteriormente imputados.

Tabla 3.2. Creación de preferencia electoral con votante probable elección 2017

Categorías de Respuesta	Intención de Participación Encuesta	%	Preferencia Electoral Encuesta	%	Votante Probable	%	Preferencia Electoral con Votante Probable	%
Vota	1.003	70,44	970	68,12	703	49,37	600	85,35
No Vota	371	25,05	237	16,64	437	30,69	0	0
Casos Perdidos	50	3,51	217	15,24	284	19,94	103	14,65
Total	1.424	100,00	1.424	100,00	1.424	100,00	703	100,00

Fuente: Elaboración propia a partir de la encuesta CEP de septiembre-octubre 2013 y 2017

En la encuesta CEP 81 (elección 2017), se observa una diferencia de 103 casos (703 menos 600) entre los votantes probables y la preferencia electoral, siendo considerados como casos perdidos para ser imputados, porque se observaron votantes probables que en la pregunta de preferencia electoral respondieron que no votarían (25), no saben (34) o no respondieron (44).

En consecuencia, los valores perdidos a imputar en la presente investigación serán los observados en la tabla 3, donde:

- Fue eliminada la categoría “No votaría” en el escenario de votante probable para ambas encuestas.
- Se sumó un valor perdido en ambos escenarios de la medición 2013 porque el candidato Ricardo Israel presentó problemas en la imputación al presentarse un valor perdido en una de las variables predictoras.
- Se sumó un valor perdido en la medición 2017 en el escenario sin votante probable (como viene dado en la encuesta), porque en el modelo de votante probable, la persona que indicó a Eduardo Artés como preferencia, fue clasificado como votante no probable.

Tabla 3.3. Datos perdidos a imputar en la preferencia electoral de las elecciones 2013 y 2017

Año	Sin votante probable	Con votante probable
2013	185	57
2017	218	78

Fuente: Elaboración propia a partir de las encuestas CEP septiembre-octubre del 2013 y 2017

Adicionalmente, como la preferencia electoral sin votante probable posee más cantidad de valores perdidos, es esperable que los modelos de regresión muestren un mayor número de muestra en comparación a un escenario con votante probable, donde la cantidad de sujetos que indicaron que participarán y que votarán por algún candidato son menores que en el primer escenario.

IV. RESULTADOS

Previo a la presentación de los resultados, es necesario aclarar algunos criterios metodológicos: en primer lugar, se decide no considerar los votos nulos y blancos en la imputación porque la encuesta CEP los agrupa en una sola categoría junto a “No Votaría”. En segundo lugar, se decide excluir del análisis a la categoría No Votaría porque el objetivo principal de la imputación es predecir los valores perdidos a través de los observados, es decir, los datos imputados utilizarán como base los valores de los candidatos en la preferencia electoral.

En tercer lugar, Ricardo Israel fue considerado como valor perdido en la encuesta CEP 70, porque al ser solo un caso y al presentarse un NA en una de las variables explicativas (identificación con partido político), el modelo de ecuaciones encadenadas no logró predecir el valor de la casilla; en tercer lugar, Eduardo Artés presentó una situación similar a Ricardo Israel, siendo considerado como NA, siendo descartado del análisis.

Finalmente, las comparaciones de las mediciones con los resultados reales de las elecciones fueron realizadas con los resultados brutos, es decir, no se aplicó el ponderador diseñado por el Centro de Estudios Públicos.

4.1. Predictores seleccionados para los modelos de regresión.

Como se mencionó en el apartado anterior, la selección de los predictores dependía de dos criterios: 1) que estuvieran justificados de manera teórica y 2) aportaran al modelo, pero manteniendo un criterio de parsimonia.

Los predictores que cumplieron con ese criterio fueron todos, por lo que se incluirán en el modelo las variables independientes descritas en el apartado anterior.

4.2. Modelos de regresión multinomial para explicar la preferencia electoral.

En esta sección se interpretarán de manera general los modelos obtenidos con las variables seleccionadas, porque el objetivo final es la capacidad predictiva del modelo en comparación al efecto de cada variable independiente sobre la dependiente.

El modelo de regresión 1 muestra los candidatos con la mayor cantidad de preferencias y Franco Parisi en un contexto sin votante probable para la elección 2013. La mayoría de las variables observadas se muestran significativas a excepción de la variable dummy de género y situación económica del país.

Tabla 4.1. Regresión 1 y 2

Modelos logísticos multinomiales para explicar la preferencia electoral de las candidatas con mayor favoritismo más Franco Parisi, en un contexto con y sin votante probable para la primera vuelta presidencial del 2013

Variables	Modelo 1 (Sin Votante Probable)			Modelo 2 (Con Votante Probable)		
	Evelyn Matthei (1)	Michelle Bachelet (2)	Franco Parisi (3)	Evelyn Matthei (1)	Michelle Bachelet (2)	Franco Parisi (3)
Ideología (Centro)	-3.013*** (0.697)	-1.088 (0.693)	-1.488** (0.725)	-2.500*** (0.929)	0.375 (0.924)	-0.348 (0.996)
Ideología (Izquierda)	-4.633*** (0.874)	-0.435 (0.681)	-1.928*** (0.737)	-5.079*** (1.304)	0.436 (0.884)	-1.609 (1.016)
Ideología (Independiente)	-2.650*** (0.969)	-0.957 (0.874)	-1.351 (0.983)	-3.208*** (1.237)	-1.254 (1.079)	-2.086 (1.452)
Ideología (Ninguna)	-2.780*** (0.688)	-0.982 (0.682)	-1.535** (0.720)	-2.596*** (0.897)	0.145 (0.893)	-1.238 (0.993)
Sexo (Hombre)	-0.234 (0.289)	-0.117 (0.219)	0.339 (0.284)	-0.125 (0.421)	-0.408 (0.318)	-0.280 (0.444)
Edad	0.041*** (0.009)	0.025*** (0.007)	0.001 (0.010)	0.056*** (0.015)	0.041*** (0.011)	-0.006 (0.016)
Identificación Partido Político (Ninguno)	-0.469 (0.417)	-0.107 (0.320)	0.052 (0.411)	-0.584 (0.645)	-0.502 (0.483)	-0.323 (0.652)
Identificación Partido Político (DC)	-1.879* (1.013)	1.052* (0.596)	0.783 (0.727)	-2.198* (1.177)	0.612 (0.723)	0.249 (0.919)
Situación Económica del País	0.184 (0.188)	-0.066 (0.136)	0.188 (0.184)	0.240 (0.278)	0.097 (0.198)	0.112 (0.283)

Desempeño del Gobierno (Desaprueba)	-1.951*** (0.375)	0.581** (0.276)	-0.837** (0.339)	-2.295*** (0.562)	0.619 (0.426)	-0.567 (0.557)
Desempeño del Gobierno (No Aprueba ni Desaprueba)	-1.422*** (0.423)	0.206 (0.331)	-0.655 (0.407)	-2.305*** (0.643)	-0.144 (0.485)	-0.555 (0.626)
Nivel Educativo	-0.147* (0.088)	-0.312*** (0.068)	-0.014 (0.087)	-0.097 (0.126)	-0.283*** (0.096)	-0.196 (0.137)
Grupo Socioeconómico	0.107 (0.223)	0.473*** (0.172)	0.248 (0.220)	0.189 (0.321)	0.358 (0.237)	-0.101 (0.332)
Asistencia a Servicios Religiosos	-0.101 (0.064)	-0.103** (0.051)	0.017 (0.067)	-0.063 (0.093)	-0.084 (0.071)	-0.027 (0.101)
Constante	2.514* (1.480)	1.436 (1.220)	0.066 (1.513)	0.974 (2.092)	-0.245 (1.667)	2.573 (2.206)
N	1,060	1,060	1,060	609	609	609
AIC	1,717.220	1,717.220	1,717.220	870.016	870.016	870.016
BIC	1,668.188	1,668.188	1,668.188	807.254	807.254	807.254
Pseudo R2 Nagelkerke	0.5181	0.5181	0.5181	0.6022	0.6022	0.6022

Nota:

*p<0.1; **p<0.05; ***p<0.01

Evelyn Matthei es la candidata que presenta mayor cantidad de efectos significativos con respecto a los demás, permitiendo teóricamente una mejor predicción. Para el caso de Michelle Bachelet, las variables que mostraron mayor efecto fueron, edad, identificación con el partido político de la Democracia Cristiana, desaprobación del desempeño del gobierno grupo socioeconómico y nivel educacional. Finalmente, las variables que explican la preferencia electoral para el caso de Franco Parisi fueron Ideología de Centro, Ideología de Izquierda, Ninguna Ideología y desaprobación de la gestión del gobierno. El pseudo R cuadrado de este modelo se presenta como 0,52.

El segundo modelo de regresión responde a los candidatos con mayor cantidad de preferencias en la elección presidencial del 2013 en un escenario de votante probable.

Como se observa en el modelo 2, las variables más significativas para la candidata Evelyn Matthei se centraron en la ideología, no aprueba ni desaprueba la gestión del gobierno, y desaprobación del desempeño del gobierno. Michelle Bachelet logró significación estadística en edad y nivel educacional. Para el caso de Franco Parisi no se observaron variables significativas. El Pseudo R cuadrado de Nagelkerke de modelo corresponde a 0,60.

Tabla 4.2. Regresión 3 y 4

Modelos logísticos multinomiales para explicar la preferencia electoral de los candidatos con mayor favoritismo más José Antonio Kast, en un contexto con y sin votante probable para la primera vuelta presidencial del 2017

Variables	Modelo 3 (Sin Votante Probable)			Modelo 4 (Con Votante Probable)		
	Alejandro Guillier (1)	Sebastián Piñera (2)	José Antonio Kast (3)	Alejandro Guillier (1)	Sebastián Piñera (2)	José Antonio Kast (3)
Ideología (Centro)	-0.295 (0.538)	-1.761*** (0.430)	-1.287 (0.832)	0.578 (0.714)	-1.241** (0.578)	-1.232 (1.013)
Ideología (Izquierda)	0.601 (0.496)	-2.391*** (0.437)	-1.295 (0.860)	0.918 (0.644)	-2.754*** (0.577)	-1.459 (1.022)
Ideología (Independiente)	0.417 (0.658)	-2.790*** (0.574)	-1.285 (0.995)	0.615 (1.002)	-1.028 (0.893)	0.107 (1.221)
Ideología (Ninguna)	-0.408 (0.496)	-2.012*** (0.391)	-1.688** (0.742)	-0.092 (0.657)	-1.976*** (0.503)	-1.738** (0.844)
Sexo (Hombre)	0.138 (0.216)	0.021 (0.211)	-0.181 (0.494)	-0.071 (0.289)	0.087 (0.297)	-0.083 (0.601)
Edad	0.034*** (0.007)	0.015** (0.007)	-0.009 (0.015)	0.032*** (0.010)	0.007 (0.010)	-0.023 (0.020)
Identificación Partido Político (UDI y RN)	-0.995 (0.874)	2.542*** (0.542)	1.868* (0.958)	-13.067*** (0.00001)	4.231*** (1.111)	3.077** (1.512)
Identificación Partido Político (Ninguno)	0.090 (0.255)	1.652*** (0.318)	1.313* (0.733)	-0.031 (0.319)	2.012*** (0.426)	2.020** (0.906)
Situación Económica del País	-0.101 (0.137)	0.003 (0.135)	0.295 (0.303)	-0.124 (0.197)	-0.068 (0.199)	0.305 (0.375)

Desempeño del Gobierno (Desaprueba)	-0.122 (0.251)	1.835*** (0.262)	1.997** (0.813)	-0.389 (0.339)	1.880*** (0.362)	1.810** (0.889)
Desempeño del Gobierno (No Aprueba ni Desaprueba)	-0.009 (0.274)	0.493 (0.310)	-0.524 (1.271)	-0.340 (0.350)	0.171 (0.414)	-0.797 (1.305)
Nivel Educativo	0.081 (0.062)	-0.101 (0.064)	-0.061 (0.145)	0.147* (0.082)	-0.089 (0.091)	-0.045 (0.176)
Grupo Socioeconómico	0.082 (0.160)	-0.180 (0.155)	-0.765** (0.353)	0.350* (0.209)	0.032 (0.213)	-0.609 (0.418)
Asistencia a Servicios Religiosos	0.001 (0.047)	-0.061 (0.045)	-0.261*** (0.100)	0.004 (0.060)	-0.090 (0.063)	-0.271** (0.121)
Constante	-1.970* (1.144)	0.407 (1.079)	0.190 (2.442)	-3.037** (1.547)	0.115 (1.505)	0.255 (2.837)
N	878	878	878	556	556	556
AIC	1,570.489	1,570.489	1,570.489	926.053	926.053	926.053
BIC	1,521.698	1,521.698	1,521.698	874.484	874.484	874.484
Pseudo R2 Nagelkerke	0.4803	0.4803	0.4803	0.5762	0.5762	0.5762

Nota:

*p<0.1; **p<0.05; ***p<0.01

Para el caso de la CEP 81, el modelo 3 se enfocó en explicar a los candidatos con mayor preferencia electoral y José Antonio Kast en un contexto sin votante probable.

La variable estadísticamente significativa para el candidato Alejandro Guillier es edad. Ideología, edad, identificación con partido político y desempeño del gobierno son las variables significativas para el caso de Sebastián Piñera. Finalmente, las variables estadísticamente significativas para José Antonio Kast son ideología, identificación con algún partido político, desempeño del gobierno, grupo socioeconómico y asistencia a servicios religiosos.

Dicho lo anterior, el Pseudo R cuadrado del modelo es de 0,48.

El modelo 4 muestra que Alejandro Guillier posee 4 variables significativas: edad, identificación con partido político, nivel educacional y grupo socioeconómico. El candidato Sebastián Piñera posee 6 factores significativos encontrando ideología, identificación con partido político y desempeño del gobierno. Para el caso de José Antonio Kast, las variables que explican su preferencia son: ideología, identificación con partido político, desempeño del gobierno y asistencia a servicios religiosos. El pseudo R2 de Nagelkerke para el presente modelo es de 0,58.

Adicionalmente, se observa que, la muestra en un escenario sin votante probable es levemente superior que, en una medición con votante probable, lo cual se debe que se trabaja solo con el 80,9% de la muestra (el 19,1% corresponde a No Votaría), en comparación al año 2013 que la preferencia electoral abarcó el 90,8% de la muestra (ver figura 1).

4.3. Imputaciones

En este apartado se observarán las principales imputaciones para cada encuesta a través de las comparaciones entre la preferencia electoral con votante probable y sin votante probable.

4.3.1. Comparación entre escenarios.

2013.

Para el caso de la encuesta CEP 70 (2013), no se observan mayores diferencias en términos de imputación en el escenario de estar o no con votante probable¹.

¹ La suma del modelo de votante probable sin imputar da 649 y no 650 porque Ricardo Israel fue considerado como missing, perdiendo un valor en candidatos de centro e izquierda, al igual que en la medición original sin imputación.

Tabla 4.3. Imputación 1

Comparación entre modelos con y sin votante probable sin "No vota" para la elección 2013

Elecciones			Encuesta CEP			Modelo sin votante probable				Modelo de votante probable sin imputar				Modelo con votante probable			
Candidatos	Votos	%	CEP	%	DIF (%)	IMP	%	DIF ELE (%)	DIF CEP	MVP	%	DIF ELE (%)	DIF CEP	IMP	%	DIF ELE (%)	DIF CEP
Michelle Bachelet	3,075,839	46.7	701	56.0	9	795	60.2	14	4	419	64.6	18	9	447	63.3	17	7
Evelyn Matthei	1,648,481	25.0	191	15.3	-10	228	17.3	-8	2	127	19.6	-5	4	143	20.3	-5	5
Franco Parisi	666,015	10.1	111	8.9	-1	134	10.2	0	1	41	6.3	-4	-3	45	6.4	-4	-2
Otros Candidatos de Centro Izquierda	1,195,473	18.2	132	10.5	-8	163	12.3	-6	2	62	9.6	-9	-1	71	10.1	-8	0
Votos Válidamente Emitidos	6,585,808		1.135	90.7	91	1.320	100			649	100			706	100		
Nulos y Blancos	113,203	0.8	0	0	-1	0	0	-1	0	0	0	-1	0	0	0	-1	
No Votaron	6,874,132	50.6	117	9.3	-41	0	0	-51	0	0	0	-51	-9	0	0	-51	0
Total (PEV 2013)	13,573,143	100	1.252	100		1.320	100		-9	649	100			706	100		-9.3

Fuente: Elaboración propia en base a la encuesta CEP de septiembre-octubre 2013

Según lo observado en la tabla, la encuesta CEP sobrestimó la preferencia electoral de Michelle Bachelet en un 9%. Cuando se imputa la pregunta de la encuesta (sin votante probable), la estimación se aleja en un 14%, observando la misma tendencia en un escenario de votante probable (alejándose de los datos reales en un 17%).

Evelyn Matthei fue subestimada por la encuesta CEP en 10%. En un escenario sin votante probable, la imputación de la preferencia electoral mejora en 2 puntos porcentuales (17,3%). En la misma línea, el modelo de votante probable mejora la estimación de la candidata en la Encuesta en 5% (20,3%).

Para el caso de Franco Parisi, la encuesta CEP subestimó al candidato en 1,2%. La imputación sin votante probable mantuvo los mismos porcentajes, en comparación al escenario con votante probable que, en vez de mejorar la precisión, se alejó de los resultados de la elección en un 2%.

La categoría “otros candidatos de izquierda” fue subestimado por la encuesta CEP (10,6%). Observando una leve mejora en la imputación sin votante probable (12,3%). Sin embargo, en un modelo con votante probable la estimación es similar a lo observado a la medición de la CEP sin someterse a ningún tipo de imputación.

De acuerdo a lo observado, ambas imputaciones (con y sin votante probable) se adaptan a la distribución de los candidatos, no obstante, cuando se utiliza el modelo de votante probable, existe una leve mejora para el caso de la candidata Evelyn Matthei (5 puntos porcentuales) y, un alejamiento para el caso de la candidata Michelle Bachelet en 14 y 17 puntos.

2017.

Al igual que en el año 2013, en términos generales, no existe gran diferencia entre un escenario sin votante probable y con votante probable en términos de los candidatos.

Tabla 4.4. Imputación 2

Comparación entre modelos con y sin votante probable sin "No vota" para la elección 2017

Elecciones			Encuesta CEP			Modelo sin votante probable				Modelo de votante probable sin imputar				Modelo con votante probable			
Candidatos	Votos	%	CEP	%	DIF (%)	IMP	%	DIF ELE (%)	DIF CEP	MVP	%	DIF ELE (%)	DIF CEP	IMP	%	DIF ELE (%)	DIF CEP
Alejandro Guillier	1,496,560	22.7	241	20.0	-3	293	24.7	2	5	153	26	3	6	179	26.4	4	6
José Antonio Kast	522,946	7.9	24	2.0	-6	29	2.4	-5	0	18	3	-5	1	19	2,8	-5	1
Sebastián Piñera	2,416,054	36.6	476	39.5	3	578	48.7	12	9	309	52	15	12	338	49.9	13	10
Otros Candidatos de Centro Izquierda	2,158,462	32.7	228	18.9	-14	287	24.2	-9	5	120	20	-13	1	142	20.9	-12	2
Válidamente Emitidos	6,594,022		969	80.3	0	1.187	100			600	100			678	100		
Nulos y Blancos	105,605	0.7	0	0,0	-1	0	0	-1	0	0	0	-1	0	0	0	-0,7	0,0
No Votaron	7,608,524	53.2	237	19.7	-34	0	0	-53	-20	0	0	-53	-20	0	0	-53.2	-19.7
Total (PEV 2013)	14,308,151	100	1.206	100		1.187	100			600	100			678	100		

Fuente: Elaboración propia en base a la encuesta CEP de septiembre-octubre 2017

La tabla muestra los resultados de la preferencia electoral para la imputación sin votante probable (como viene dada en la encuesta) y con votante probable.

Para el caso de Alejandro Guillier, la encuesta CEP lo sobre estimó en 3 puntos porcentuales. Cuando se realizó la imputación sin votante probable, la preferencia electoral del candidato subió de 20,0% a 24,7% alejándose levemente del valor real de la elección (22,7%), situación similar se observó al utilizar el modelo de votante probable porque de 22,7% subió a 26,4%.

José Antonio Kast fue subestimado por la encuesta CEP en 6%. Al imputar la pregunta en un escenario sin votante probable la preferencia electoral baja en un punto porcentual, al igual que en un modelo con votante probable.

El candidato ganador de las primarias de Chile Vamos “Sebastián Piñera”, fue sobre estimado en un 3% por la Encuesta CEP 81. Si se aplica una imputación a la pregunta de preferencia electoral, la medición empeora de 39,5% a 48,7%, un comportamiento similar al utilizar un modelo de votante probable porque se genera una diferencia de 13% con respecto a los resultados de la elección.

Para el caso de “otros candidatos de centro izquierda”, la encuesta CEP los subestimó en un 14% (18,9%). Cuando se realiza la imputación a la pregunta sin votante probable, tiende a mejorar el resultado (24,2%). Para el caso de la imputación de la preferencia electoral en un contexto con votante probable, si bien mejora la estimación, es menor que en un escenario sin votante probable.

De acuerdo con lo observado en la tabla, no existe mayor diferencia para los candidatos en un escenario con o sin votante probable, sin embargo, pese a que la distribución de la imputación se adapta a la de los datos, la estimación hacia el candidato de derecha tiende a empeorar la estimación en vez de mejorarla.

V. CONCLUSIONES

Los métodos de imputación han sido diseñados para estimar los valores perdidos en una determinada pregunta de una encuesta, sin embargo, requieren de otros materiales para desempeñar de mejor manera su labor, como el diseño de una buena pregunta, habilidad de los encuestadores para abordar los temas sensibles y una muestra representativa que abarque la mayor cantidad de la población a encuestar, entre otros.

En términos del trabajo realizado se puede establecer que los escenarios de intencionalidad de voto con y sin votante probable se presentan de manera similar respecto a la distribución de la pregunta imputada, manteniendo la tendencia de la medición original, es decir, la imputación por si sola no mejora los pronósticos electorales, lo cual es una situación que se puede dar y está respaldada por la literatura especializada (Rivas, Martínez y Galindo 2010; Escobar y Jaime 2013; Liu 2014; Cabrera y Escobar 2019), no obstante, se hace necesario establecer criterios para mejorar la imputación, como poseer mayor información en las variables clave.

En términos teóricos, se demostró que los candidatos que lograron un mayor número de votos en la estimación de la preferencia electoral fueron los de derecha con mayor preferencia en cada encuesta (Sebastián Piñera y Evelyn Matthei) por la fuerza de la significación en los modelos de regresión multinomial. Logrando que, en la medición del 2013, la imputación se acercara a los resultados electorales y en la del 2017, se alejara.

El hecho que exista un número importante de candidatos de centro izquierda, genera un desafío no solo para los ciudadanos que los tienen que ubicar en una escala ideológica (si se quiere utilizar la teoría espacial del voto), sino también para los investigadores que están interesados en mejorar las estimaciones, porque metodológicamente se requiere de nuevas variables para poder diferenciarlos y de una muestra más grande para que las imputaciones se lleven a cabo de mejor manera.

Finalmente, se toma como recomendación práctica que se construyan las bases de datos (para imputar la no respuesta) a partir de variables binarias (dummies) porque así se logra incrementar de manera importante los pseudo R cuadrado de los modelos de regresión multinomial, porque se comprobó que esta era una forma más eficiente que aquella que contenía todas las categorías de las variables.

Queda para futuros trabajos el seguir investigando sobre los predictores que mejoran la predicción de la imputación, como el estudiar la importancia de las categorías de respuesta en la variable ideología. ¿es lo mismo que ideología sea una variable escalar de 1 a 10 o que tenga categorías de respuesta?, ¿influye el agregar más categorías de respuesta a la variable ideología?

BIBLIOGRAFÍA

Altman, D. (2004). “Redibujando el mapa electoral chileno: incidencia de factores socioeconómicos y género en las urnas”. *Revista de Ciencia Política*, 24 (2): 49-66.

Bargsted, M., Somma, N., Muñoz, B. (2019). “Participación electoral en Chile. Una aproximación de edad, periodo y cohorte”. *Revista de ciencia política*, 39 (1): 75-98.

Bautista, R. (2016). “Unit Nonresponse in a Presidential Election Day Survey: An Initial Multilevel Exploration Using Fully Conditioned Imputed Data”. *Research Methodologist at NORC at the University of Chicago, Working Paper: N°1*.

Beatty, P., y Herrmann, D. (2002). “To answer or not to answer: Decision processes related to survey item nonresponse”. En R. Groves, D. Dillman, J. Eltinge y R. Little (Eds.), *Survey nonresponse* (pp. 71–69). John Wiley y Sons.

Bernhagen, P., y Marsh, M. (2007). “The partisan effects of low turnout: Analyzing vote abstention as a missing data problem”. *Electoral Studies*, 26 (3): 548-560.

Bernhagen, P., y Marsh, M. (2010). “Missing Voters, Missing Data: Using Multiple Imputation to Estimate the Effects of Low Turnout”. *Journal of Elections, Public Opinion and Parties*, 20(4): 447-472.

Braña, T., García, A., Rial, A., Varela, J., Vazquez, X. (1998). “Estimación de la respuesta de los no sabe/no contesta en los estudios de intención de voto”. *REIS: Revista Española de Investigaciones Sociológicas*, 83 (1): 269-287.

Cabrera, P., y Escobar, M. (2019). “The Effect of Weighting and Multiple Imputation on Bias in Spanish Election Polls”. *Revista Española de Investigaciones Sociológicas*, 165 (1): 45-64.

Campbell, A., Converse, P., Miller, W., y Donald, S. (1960). *The American Voter*. John Wiley and Sons.

Carvacho, H., Miranda, D., Pinto, P., Mackenna, B., Matulic, C., Sepúlveda, F. (2018). ¿Quiénes votan? ¿O por qué es tan difícil predecirlo? Mide UC. Recuperado de <https://www.mideuc.cl/sociedad/encuesta-foco-ciudadano/quienes-votan-en-chile-o-por-que-es-tan-dificil-predecirlo/>.

CEPAL. (2007). *Imputación de datos: teoría y práctica* (54). Recuperado de https://repositorio.cepal.org/bitstream/handle/11362/4755/1/S0700590_es.pdf

Cordero, R y Tapia, G. (2007). *Antecedentes históricos y Desarrollos Metodológicos de la Industria de la Opinión Pública de Chile*. 1er Congreso Latinoamericano WAPOR "Opinión pública, conflicto social y orden político”.

Corvalán, A., Cox, P., y Hernández, C. (2015). Evaluando los determinantes de la participación electoral en Chile: sobre el uso de datos individuales y el sobre-reporte en encuestas. En *Condicionantes de la participación electoral en Chile*. Santiago de Chile: Programa de las Naciones Unidas para el Desarrollo.

Cruz, P. (1990). “Del no sabe al no contesta: un lugar de encuentro para diversas respuestas”. *REIS: Revista Española de Investigaciones Sociológicas*, 52 (1): 139-158.

Dimock, M., Keeter, S., Schulman, M., Miller, C. (2001). “A Voter Validation Experiment: Screening for Likely Voters in Pre-Election. Paper prepared for presentation at the 56th Annual American Association for Public Opinion Research Conference. May 17-20, 2001 Montreal, Quebec, Canada.

Dodds, T. (2017). Efecto del postmaterialismo y nivel socioeconómico en el comportamiento de voto chileno. *Ciencia Política*, 12(24), 205-235.

Downs, A. (1957). *An economic theory of democracy*. New York. Harper.

Escobar, M., Jaime, Antonio., (2013). “Métodos de imputación múltiple para predecir resultados electorales”. En Alfonso Mendoza Velázquez (Eds.), *Aplicaciones en Economía y Ciencias Sociales con Stata*, StataCorp LP.

Fiorina, M. (1977). An Outline for a Model of Party Choice. *American Journal of Political Science*, 21 (3): 601-625.

Fiorina, M. (1981). *Retrospective Voting in American National Elections*. Yale University Press.

González, R., Mackenna, B. (2017). “Un modelo de votante probable para encuesta CEP”. *Puntos de Referencia*, 466 (1): 1-10.

Groves, R. y Cooper, M. (1998). *Nonresponse in Household Interview Surveys*. French Edition.

Groves, R. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, 72(1): 167-189.

Groves, R. (2009). *Survey Methodology*. 2nd Edition. John Wiley y Sons.

Harrup, M., y Miller, W. (1987). *Elections and voters: a comparative introduction*. Hampshire.

Highton, B. y Wolfinger, R. (2001). “The First Seven Years of the Political Life Cycle”. *American Journal of Political Science* 45 (1): 202-209.

INE. (2018). Informe de principales resultados VIII Encuesta de presupuestos familiares (EPF). Recuperado en: [https://www.ine.cl/docs/default-source/encuesta-de-presupuestos-familiares/publicaciones-y-anuarios/viii-epf---\(julio-2016---junio-2017\)/informe-de-principales-resultados-viii-epf.pdf?sfvrsn=d5bd824f_2](https://www.ine.cl/docs/default-source/encuesta-de-presupuestos-familiares/publicaciones-y-anuarios/viii-epf---(julio-2016---junio-2017)/informe-de-principales-resultados-viii-epf.pdf?sfvrsn=d5bd824f_2)

INE. (2019). Métodos de imputación VIII EPF: Gastos diarios e ingresos de la actividad laboral y jubilaciones. Working Paper: N°7.

Kinder, David R. 1981. "Presidents, Prosperity, and Public Opinion". *The Public Opinion Quarterly* 45 (1): 1-20.

King, G., Honaker, J., Joseph, A., Scheve, K. (2001). Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review*, 95 (1), 49-69.

Lipset, S., y Rokkan, S. (1967). *Cleavage Structures, Party Systems and Voter Alignments: An Introduction*. Free Press.

Lohr, S. (2009). *Sampling: Design and Analysis*. 2nd Edition. CRC Press.

Markham, T., Falk, E., McGrath, D., y Schneider, J. (2011). *Analysis of Multiple Imputation Techniques for the Survey of Local Election Officials*.

Marshall, P. (2018). Análisis de la Encuesta del Centro de Estudios Públicos. *Estudios Públicos*. Rescatado en <https://www.cepchile.cl/cep/otros-documentos/analisis-metodologico-de-la-encuesta-del-centro-de-estudios-publicos>.

Medway, R., y Tourangeau, R. (2015). "Response Quality in Telephone Survey. Do prepaid cash incentives make a difference?". *Public Opinion Quarterly*, 79(2): 524-543.

Montecinos, E. (2007) Análisis del comportamiento electoral: de la elección racional a la teoría de redes. *Revista de Ciencias Sociales*, 13 (1): 9-22.

Morales, M. (2010). ¿Quiénes son, dónde están, qué quieren?: las bases electorales de los candidatos presidenciales 2009. En C. Fuentes (Coord.), *Chile 2009: Percepciones y actitudes sociales*. Santiago de Chile: Universidad Diego Portales.

Navia, P., y Del Pozo, B. (2012). Los efectos de la voluntariedad del voto y de la inscripción automática en Chile. *Estudios Públicos* (127). Recuperado en <https://www.cepchile.cl/cep/estudios-publicos/n-121-a-la-150/estudios-publicos-n-127-2012/los-efectos-de-la-voluntariedad-del-voto-y-de-la-inscripcion-automatica>.

Ortiz, A., García, M., (2014). Porque te quiero te apoyo: Estilo de gobierno y aprobación presidencial en América Latina. *Revista de Ciencia Política*, 34 (2): 373-398.

Liu, F. (2013). "Using Multiple Imputation for Vote Choice Data: A Comparison across Multiple Imputation Tools". *Open Journal of Political Science*, 4(02): 39-46.

Penades, A. (2015). "Especial encuestas: errores, cocina y predicción". Recuperado en <https://www.fundacionalternativas.org/las-publicaciones/informes/especial-encuestas-errores-cocina-y-prediccion>

Perry, P. (1973). A Comparison of the Preferences of Likely Voters and Likely Nonvoters. *The Public Opinion Quarterly*, 7(1). Oxford University Press.

PNUD (2017). Diagnósticos sobre la participación electoral en Chile. Programa de Gobernabilidad Democrática. Recuperado en: https://www.cl.undp.org/content/chile/es/home/library/democratic_governance/diagnostico-sobre-la-participacion-electoral-en-chile.html.

Rentsch, A., Schaffner, B., y Gross, J. (2019). The Elusive Likely Voter: Improving Electoral Predictions With More Informed Vote-Propensity Models. *Public Opinion Quarterly*, 83 (4): 782-804.

Rivas, C., Martínez, R., Galindo, P. (2010). La imputación múltiple como alternativa al análisis de la no respuesta en la variable intención de voto. *Revista Española de Ciencia Política*. 22(22): 99-118.

Roche, M. (2008). Modelos de comportamiento electoral y político. Tesis Doctoral Universidad Autónoma de Nuevo León.

Román, G. (2017) Los valores perdidos en el muestreo de poblaciones finitas. Técnicas de imputación. *Actas XIII Congreso Galego de Estatística e Investigación de Operacións*. 293-298

Rubin (1976). *Multiple imputation for nonresponse in surveys*. 1st Edition. Wiley.

Sánchez, N. (2015). Imputación de datos faltantes en Encuestas de Intención de Voto. Tesis de grado Universidad Carlos III.

Sulmont, D. (2015). Voto ideológico y sistema de partidos en América Latina: El Peso de la Dimensión Izquierda-Derecha en el Comportamiento Electoral en Brasil, Chile, México y Perú. Working Paper: N°12.

Traugott, M. (2014). Métodos alternativos para la estimación de resultados electorales. *Estudios Públicos*. Recuperado en: <https://www.cepchile.cl/cep/encuestas-cep/metodologia/metodologia-general/1-metodos-alternativos-para-la-estimacion-de-resultados-electorales>.

Torres, R. (2016). Perspectiva de la teoría de clivajes para el caso latinoamericano. *Cuadernos Americanos: Nueva Epoca*, 1 (155): 97-115.

Torrico, M., Solís, D. (2020). Voto ideológico, ¿por qué los latinoamericanos votan por la izquierda o la derecha? *Foro Internacional*, 60 (1): 175-226.

Toro, S. (2008). De lo épico a lo cotidiano: Jóvenes y generaciones políticas en Chile. *Revista de Ciencia Política*, 28 (2): 143-160.

Useche, L., Mesa, D., (2006). Una introducción a la imputación de valores perdidos. *Terra Nueva Etapa*, 22 (13): 127-155.