

Tarea 1

¿Cómo afecta la tendencia política a las votaciones?

Integrantes:	Antonia Arias Javiera Donoso Daniel Jara Carla Olivares
Profesor:	Daniel Schwartz Omar Perez
Auxiliar:	Catalina Chamy Francisca Contreras Karla Toledo Campos Nicole Troncoso Rienzi Roldan Santiago Del Río Sofía González Vicente Sule
Fecha de entrega:	20 de abril de 2023

1.

1.1. Elaboración de hipótesis

Existen varios estudios que sugieren una relación entre la diferencia de participación electoral cuando el voto pasa de ser voluntario a obligatorio. Un estudio de 1998 [1] comparó países con voto obligatorio y voto voluntario en donde se encontró que en el primer caso los partidos de izquierda obtienen más votos que los de derecha, mientras que para elecciones voluntarias esa diferencia se reduce significativamente. Esto podría indicar que la participación de los votantes de izquierda sí se ve afectada por la naturaleza de la elección.

Otro estudio analizó la introducción del voto obligatorio en Estados Unidos [2]. Encontró que aquellos votantes con preferencias de políticas progresistas tienen una mayor tasa de participación en comparación con los votantes conservadores. En particular, aumentó significativamente la participación entre aquellos con bajos niveles de educación y bajos ingresos, que tienden a favorecer partidos demócratas.

Basado en estos estudios, se postula la siguiente hipótesis: **La tendencia política tiene un efecto significativo en la diferencia de participación electoral cuando el voto pasa de ser voluntario a obligatorio.**

1.2. Construcción de la base de datos

Las 4 bases entregadas en la tarea se importaron a R con el objetivo de extraer las columnas específicas necesarias para realizar los siguientes análisis, utilizando la función `dplyr`.

Para crear la variable independiente de la segunda vuelta, se decidió trabajar con una base de datos distinta para cada candidato, debido a que es más sencillo trabajar con los datos por separado. De estas bases de datos, se extrajeron todos los votos obtenidos por José Antonio Kast y Gabriel Boric en la segunda vuelta presidencial. Luego, se combinaron ambas bases para crear la variable independiente en base al total de votos obtenidos por cada uno, dividido entre el total de votos obtenidos por ambos en cada comuna.

Un proceso similar se realizó en la base de datos de los votos de salida y entrada para encontrar la participación comunal. Debido a que nuestra variable dependiente utiliza tanto los votos de salida como de entrada, fue necesario obtener los datos del padrón electoral desde SERVEL.

Se seleccionaron cuatro variables CASEN además del factor de expansión comunal: la variable «edad», la variable de educación «educ», la variable «pobreza» y se creó una variable llamada «prevision», la cual entrega información acerca del sistema de salud al que pertenecen las personas de cada comuna, que puede ser ISAPRE o FONASA. Cabe destacar que

todos los grupos de FONASA fueron agrupados en uno solo y que dentro de ISAPRE se consideraron las FF.AA y de Orden para simplificar el estudio. Las variables se modificaron para que fueran concordantes con el factor de expansión comunal y se obtuvieron los promedios comunales correctos de las variables seleccionadas.

Es importante mencionar que la unidad escogida es la comuna, dadas las indicaciones de la tarea, lo que significa que cada observación de la base de datos es una comuna que tiene promedios de cada variable escogida.

2. Análisis exploratorio de datos

Para comenzar con el análisis .1, es importante declarar que cuando la diferencia de la tasa de participación varía, es debido a una sensibilidad frente al cambio de voto voluntario a obligatorio (que es lo que queremos analizar en este estudio) pero no se puede saber con total certeza el motivo de este cambio. Es decir, cuando vemos comunas en donde la diferencia de participación varía muy poco, no sabemos si es porque desde un comienzo la tasa de participación era alta, lo que implica que con el paso a voto obligatorio la tasa no varió, o si es porque a pesar de que la tasa de participación era baja, con el paso a voto obligatorio, el número continuó siendo bajo.

Con respecto a los valores atípicos u outliers, es importante mencionar que si bien existen comunas alejadas de la tendencia del gráfico, son un conjunto importante de comunas que siguen una tendencia diferente, y se considera importante mostrarlas dentro del estudio, ya que además en Chile son comunas que efectivamente tienen comportamientos distintos al promedio de las otras comunas.

Primero que todo, podemos ver en los gráficos la clara relación directamente proporcional entre las variables seleccionadas, es decir, que a mayor tendencia de derecha (porcentaje de votos Kast) mayor es la sensibilidad de la tasa de participación frente al cambio de voto voluntario a obligatorio.

Algunos comentarios importantes con respecto a los gráficos son los siguientes:

Al observar el gráfico de la riqueza podemos ver que la tendencia es que a mayor riqueza, menor es la diferencia de la tasa de participación, observando comunas con valores extremos como Vitacura, en donde de casi un 68 % de votos en el plebiscito de entrada, subieron a un 81 %, donde se explica la poca sensibilidad, debido a la gran participación que había en un sistema de votación voluntario.

En el gráfico de educación, podemos ver que existe una relación negativa. A mayor nivel de educación existe menor sensibilidad frente al cambio de sistema de votación.

Por último, con respecto a la zona, podemos ver que las zonas rurales, se caracterizan por ser más sensibles frente al cambio de sistema de voto, salvo por la Comuna Huara, en donde en el plebiscito de entrada votó aproximadamente un 16 % de las personas, versus un 34 % en el plebiscito de salida, donde se puede observar uno de los casos detallados al inicio, donde hay comunas que a pesar de que el voto se vuelva obligatorio, la tasa de votación sigue siendo baja.

La variable dependiente utilizada en los gráficos es:

$$\frac{V_s}{T_s} - \frac{V_e}{T_e} \quad (1)$$

V_s : cantidad de votos emitidos en el plebiscito de salida

V_e : cantidad de votos emitidos en el plebiscito de entrada

T_e : Total del padrón electoral para el plebiscito de entrada

T_s : Total del padrón electoral para el plebiscito de salida

La variable independiente utilizada en los gráficos es:

$$\frac{K}{K + B} = SV \quad (2)$$

K : cantidad de votos obtenidos por Kast en la segunda vuelta

B : cantidad de votos obtenidos por Boric en la segunda vuelta

Esta variable tendrá por abreviación SV dentro del modelo. Dentro del análisis exploratorio cabe destacar que se utiliza la ecuación (2) para los gráficos, sin embargo, para el estudio de regresión lineal utilizamos el complemento, es decir:

$$\frac{B}{K + B} = SV \quad (3)$$

Esto con el fin de poder encontrar conclusiones congruentes con el estudio previo.

2.1. Regresión lineal

En base a las observaciones realizadas en el análisis exploratorio de datos se considera que puede ser de significancia considerar a `casen_educ`, `casen_pobreza` y `casen_zona` como variables explicativas del modelo. En este caso, la variable `casen_prevision` muestra una correlación baja con respecto al resto de las variables, e incluso al graficar esta variable (revisar anexo figura .2) se puede apreciar que solamente en una fracción pequeña de las comunas toma valores mayores que 0.01, señalizando que la previsión FONASA es muy común y no

es un indicador relevante dentro del modelo.

Para la construcción de los modelos de regresión lineal se tomaron todas las combinaciones de las variables explicativas CASEN. En adición, se planteó un modelo que incluyera a todas las variables y la interacción entre dos de ellas. En el AED se muestra que la riqueza y la educación son las variables que menos correlación comparten, es por esto que se decidió analizar su interacción. A continuación en la figura .3 se muestran los resultados de las regresiones.

La variable *tasadp* representa el porcentaje de votos a favor de Gabriel Boric en la segunda vuelta presidencial y *SV* es la variable independiente descrita anteriormente. Entonces, por un aumento porcentual en los votos a favor de Boric, la diferencia de las tasas de participación entre el plebiscito de salida y entrada cambia en β_1 %. Como se asume que las elecciones presidenciales representan la tendencia política de la comuna, se puede decir que al tomar valores negativos, el coeficiente indica que los votantes de derecha son más sensibles ante el cambio de sistema de votación que aquellos votantes de izquierda.

Recordemos que *casen_pobreza* toma valores mayores cuando la comuna tiene mas riqueza, por lo tanto β_2 negativo indica que a mayores niveles de riqueza menos sensibilidad hay al cambio del tipo de elección. La variable *casen_zona* se acerca a 1 cuando la zona es rural, entonces β_3 positivo indica que mientras más rural sea la comuna más sensible son los votantes a los cambios de voto voluntario a obligatorio.

La interacción entre las variables pobreza y educación sí tiene significancia estadística, por lo que el efecto de la variable educación en la variable dependiente es diferente para las personas con diferentes ingresos. Por su cuenta los coeficientes de pobreza y educación son negativos, y se tiene que la relación entre la educación y la variable dependiente es más negativa para las comunas con más riqueza. Cabe notar que al agregar la interacción no se apreció cambio en el error estándar pero sí aumentó el coeficiente de interés en 0.002.

Al agregar variables explicativas el coeficiente que explica la relación de la variable dependiente con la tendencia política se acerca a cero, entre el modelo simple y aquel con todas las variables, el coeficiente cambió de -0.37 a -0.194, sin embargo este alcanza su máximo de -0.189 para cuando se toman en consideración las variables pobreza y zona, pero no educación. Para la elección de una regresión es importante considerar dos cosas, la primera es el nivel de significancia de los coeficientes, todos los modelos que se comparan cumplen con este punto. Lo segundo a considerar es el error estándar del coeficiente de interés. Este disminuye a medida que se agregan variables explicativas, un menor error estándar permite tener una mayor precisión. Como el propósito de este análisis es explicar el efecto de la tendencia

política en la diferencia en la tasa de participación, se valora el estimador de β_1 con menor error estándar por lo que se prefiere el modelo 9.

Cabe mencionar que el modelo debe cumplir con dos supuestos importantes, homocedasticidad y ausencia de multicolinealidad. Una forma de verificar esta última es con el cálculo del factor de inflación de la varianza. El VIF mide la magnitud en la cual la varianza de un coeficiente se infla debido a la multicolinealidad, en este caso incluso con la eliminación de variables explicativas, el FIV toma valores mayor a 1, lo que da cuenta de la alta correlación entre las variables. La homocedasticidad se da cuando la varianza de los errores (la varianza propia de la variable dependiente que no es explicada por las variables independiente) es constante e independiente. Puede ocurrir por ejemplo que para ciertos rango de la variable dependiente la educación tiene un efecto constante, pero para otros el efecto de la educación varía dependiendo de otros factores como por ejemplo el nivel de riqueza. En el modelo 9 se encontró que la variable de interacción sí tiene un efecto significativo en la variable dependiente y esto puede ser un indicador de heterocedasticidad, pues la educación afecta de manera diferente para niveles de riqueza mayores.

2.2. Discusión y Conclusión

Desde un comienzo la hipótesis postulada es que la tendencia política tiene un efecto significativo en la diferencia de participación electoral cuando el voto pasa de ser voluntario a obligatorio. Luego de un profundo análisis realizado a partir de la construcción de una base de datos se pueden concluir diferentes cosas al respecto.

Algunas conclusiones importantes obtenidas del estudio realizado tienen que ver con las variables CASEN. Primero que todo, es interesante analizar la importancia de la zona (rural y urbana) sobre la sensibilidad de la tasa de participación en las elecciones. se postula que esto se debe a la dificultad de movilización que puede desincentivar la votación para las elecciones voluntarias, generando una gran diferencia con respecto al caso de elección obligatoria. Por otra parte, es importante hablar del análisis con respecto a como influye la educación en la variación de la tasa de participación, ya que como se mencionó anteriormente esta se relaciona negativamente con la variación de la tasa, es decir, a mayor nivel educacional, menor es la variación en la tasa. Se propone que esto es debido a que personas con mayor educación, ya tenían una gran participación en el sistema de voto voluntario, lo que hace que su variación al momento de pasar a voto obligatorio sea pequeña. Análogo es el caso de la influencia de la variable pobreza (llamada también riqueza a lo largo del estudio). Esta analogía hace sentido también debido a la importante correlación que existe entre estas variables. Con respecto a las variables, también es importante señalar que a medida que se van añadiendo al modelo,

va disminuyendo la influencia de la variable de interés que es la tendencia política, lo que da a entender que la tasa de participación se ve afectada en gran medida por otras variables que no son la tendencia política.

También es relevante mencionar aspectos con respecto a la construcción de la base de datos y supuestos importantes que se hicieron al inicio de este estudio. Primero que todo, tenemos el hecho de usar la segunda vuelta de elecciones presidenciales, como medición de tendencia política, asumiendo así como tendencia política de derecha el haber votado por el candidato Kast, y tendencia política de izquierda, el haber votado por el candidato Boric. Esto podría haber generado invalidez del constructo, debido a que la medida utilizada para medir algo abstracto no es precisa, es decir, no necesariamente los votantes de Boric son de izquierda, ni los votantes de Kast, son exclusivamente de derecha. Por otra parte, se considera importante señalar el hecho de hacer el estudio en base a promedios comunales, debido a que podría existir una pérdida importante de información generando así un sesgo de medición.

Además, el haber utilizado los datos del plebiscito de entrada del año 2020, puede haber generado un problema de validez externa, debido a que en este plebiscito ocurrieron cosas muy específicas que no lo vuelve comparable con otras elecciones, partiendo por el hecho de que obtuvo la mayor participación electoral en la era del voto voluntario, y que votaron más personas jóvenes y con mayores niveles de pobreza, lo que nuevamente, no permite extrapolar estos resultados a otras votaciones en Chile o el mundo. Por otra parte, también está el sesgo de selección dado por la autoselección de los votantes, lo que puede generar que la población estudiada tenga características diferentes a la población total.

Finalmente, puede concluirse que la hipótesis planteada es correcta, sin embargo, dadas las falencias antes mencionadas del estudio, para obtener una evidencia más certera, deberían usarse otro tipo de mediciones y profundizar con datos más precisos, buscando al menos conseguir validez interna. A pesar de esto, dada la concordancia del trabajo con las bases de datos entregadas y el análisis posterior, para este estudio, la hipótesis sigue siendo válida. Vale mencionar además, lo importante de seleccionar y trabajar con buenas bases de datos para llegar a conclusiones certeras, que permitan obtener resultados útiles para otras investigaciones.

Referencias

- [1] y Agnieszka Dobrzynska, A. B., “Turnout in electoral democracies,” *European Journal of Political Research*, vol. 322, no. 10, pp. 239–261, 1998, doi:https://www.researchgate.net/publication/316674814_Turnout_in_Electoral_Democracies.
- [2] Leighley, J. E. y Nagler, J., “Who votes now? demographics, issues, inequality, and turnout in the united states,” *Princeton University Press*, vol. 322, no. 10, pp. 891–921, 2014, doi:<https://dx.doi.org/10.1002/andp.19053221004>.

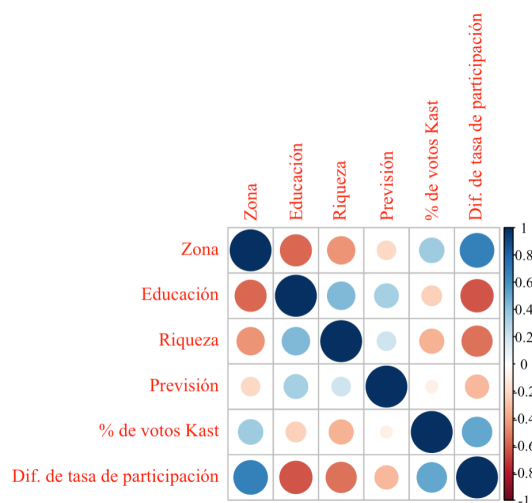


Tabla de correlaciones entre las variables

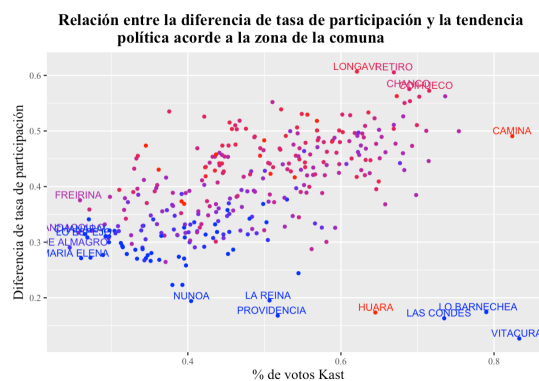
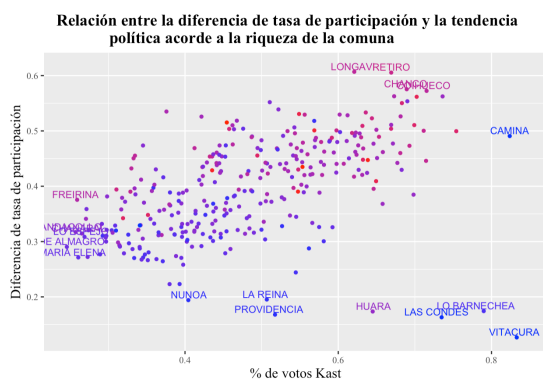
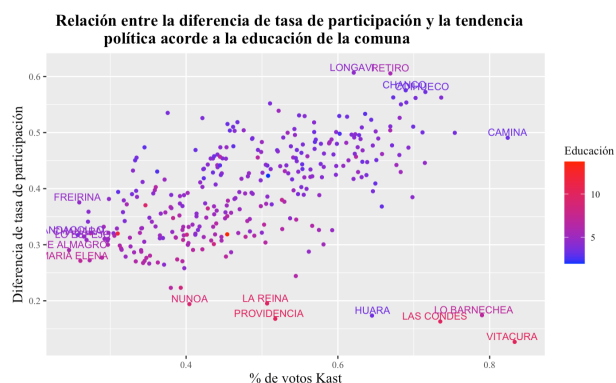


Figura .1: Gráficos del Análisis Exploratorio de Datos

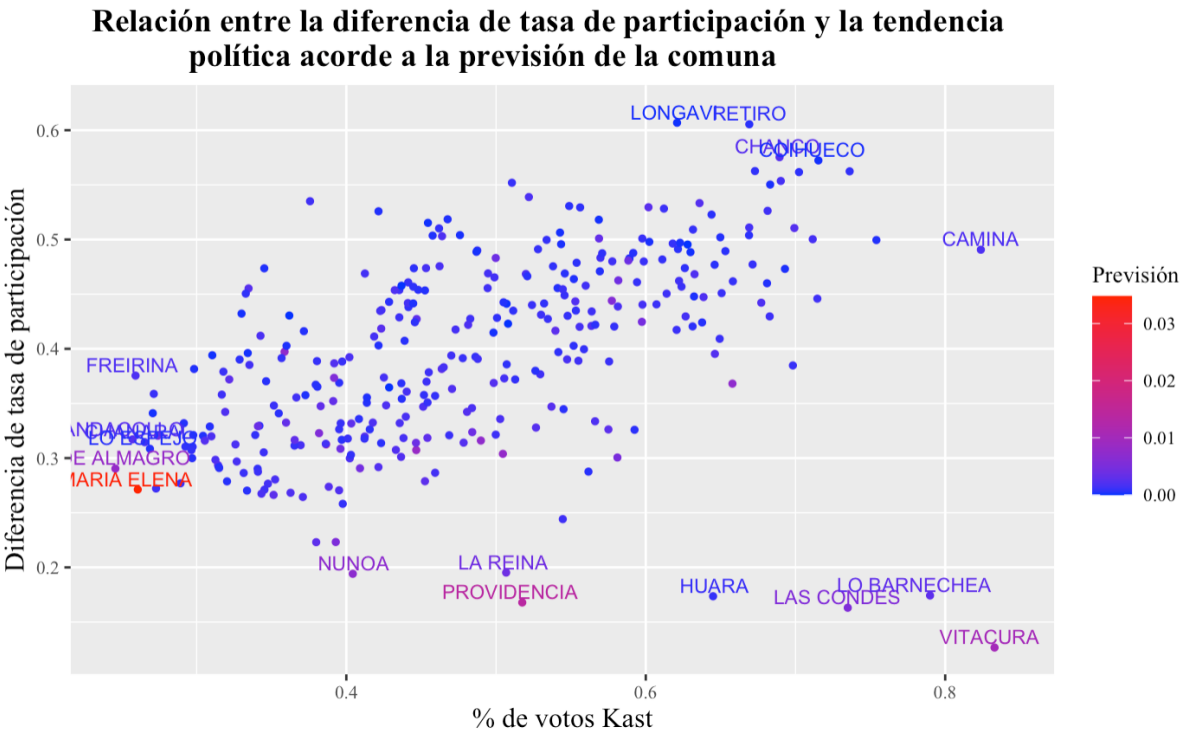


Figura .2: Gráfico de previsión

Dependent variable:									
	(1)	(2)	(3)	(4)	tasadp (5)	(6)	(7)	(8)	(9)
SV	-0.370*** (0.034)	-0.267*** (0.032)	-0.281*** (0.028)	-0.226*** (0.029)	-0.189*** (0.028)	-0.240*** (0.028)	-0.219*** (0.026)	-0.194*** (0.026)	-0.197*** (0.025)
casen_pobreza		-0.345*** (0.038)			-0.202*** (0.034)	-0.191*** (0.035)		-0.142*** (0.033)	0.349*** (0.098)
casen_educ			-0.032*** (0.002)			-0.026*** (0.002)	-0.020*** (0.003)	-0.017*** (0.003)	0.347*** (0.069)
casen_pobreza:casen_educ									-0.124*** (0.023)
casen_zona				0.171*** (0.012)	0.144*** (0.013)		0.113*** (0.013)	0.102*** (0.013)	0.106*** (0.013)
Constant	0.588*** (0.018)	1.521*** (0.104)	0.694*** (0.016)	0.450*** (0.017)	1.018*** (0.098)	1.194*** (0.094)	0.565*** (0.021)	0.950*** (0.092)	-0.493* (0.286)
Observations	324	324	324	324	324	324	324	324	324
R ²	0.269	0.419	0.535	0.544	0.588	0.573	0.619	0.640	0.669
Adjusted R ²	0.267	0.416	0.532	0.541	0.584	0.569	0.616	0.635	0.664
Residual Std. Error	0.074 (df = 322)	0.066 (df = 321)	0.059 (df = 321)	0.059 (df = 321)	0.056 (df = 320)	0.057 (df = 320)	0.054 (df = 320)	0.052 (df = 319)	0.050 (df = 318)
F Statistic	118.368*** (1; 322)	(df = 115.897*** 2; 321)	(df = 184.311*** 2; 321)	(df = 191.174*** 2; 321)	(df = 152.361*** 3; 320)	(df = 143.262*** 3; 320)	(df = 173.402*** 3; 320)	(df = 141.714*** 4; 319)	(df = 128.620*** 5; 318)

Note: *p<0.1; **p<0.05; ***p<0.01

Figura .3: Tabla de resultados de regresiones