Artículo Trabajo Final

**Introducción**

Durante las últimas elecciones en 2021, causó gran asombro cuando figuró el nombre Pedro Castillo Terrones, quien era partidario del partido político “Perú Libre”. Quien durante segunda vuelta fue victorioso contra Keiko Fujimori, su victoria fue inesperada, pues en Lima se esperaba que la segunda vuelta fuera entre Hernando de Soto y Rafael López Aliaga. Sin embargo, durante los siguientes se comentaba entre los limeños que “el Perú profundo se había levantado”, “Lima no era el Perú”, lo que resaltó bastante el hecho de ser el Perú un país centralista. El ex presidente había logrado ganarse el apoyo de la gran mayoría de personas que no residían en Lima ya que, se sentían identificados y representados por él, al ser considerada una persona “del pueblo”, creyendo que este los podría apoyar y mejorar el país dejando de ser centralista. Es por ello, que es relevante analizar las variables que provocaron que figure rápidamente entre los primeros lugares en primera vuelta porque para votar por un candidato no debe ser por las emociones o por el hecho de cómo se identifica con alguna persona sino por las propuestas que pueda presentar, su hoja de vida y su gabinete para representar al país. De esta manera, el trabajo presenta las variables de pobreza extrema, mujeres, jóvenes e IDH. Se considera la pobreza extrema porque el votante de Castillo formaba parte de los sectores C y D/E, es decir, incluía a una fracción de la población más afectada por la pobreza económica (Duárez, 2022). Asimismo, el hecho de ser mujer y que exista en su región un bajo IDH hace más probable que vote por Castillo porque tuvo mayor apoyo en la población rural de una zona marginalizada y al sentirse representados por la figura del maestro de escuela rural hizo que sea mayormente, votado destacando entre los otros candidatos en aquellas elecciones 2021 (Román, 2023). La variable central, dependiente, será el porcentaje de votos que obtuve Castillo; las variables independientes serán el porcentaje de mujeres que votaron por Castillo, el porcentaje de jóvenes y el IDH registrado de las 196 provincias del país; y la variable de control será el porcentaje de pobreza extrema de cada provincia. Por lo tanto, la pregunta del trabajo es: ¿Cuáles son las características sociales que tiene el votante de Castillo, que provocaron la mayoría de votos, por ende, la victoria de Castillo en las elecciones de segunda vuelta en 2021? Se presentará una hipótesis que el trabajo busca responder. La mayoría de votos conseguidos por Pedro Castillo fue que los electores tenían ciertas características como se encontraban en pobreza extrema, eran mujeres, jóvenes o tenían un IDH bajo en su provincia

**Metodología y data**

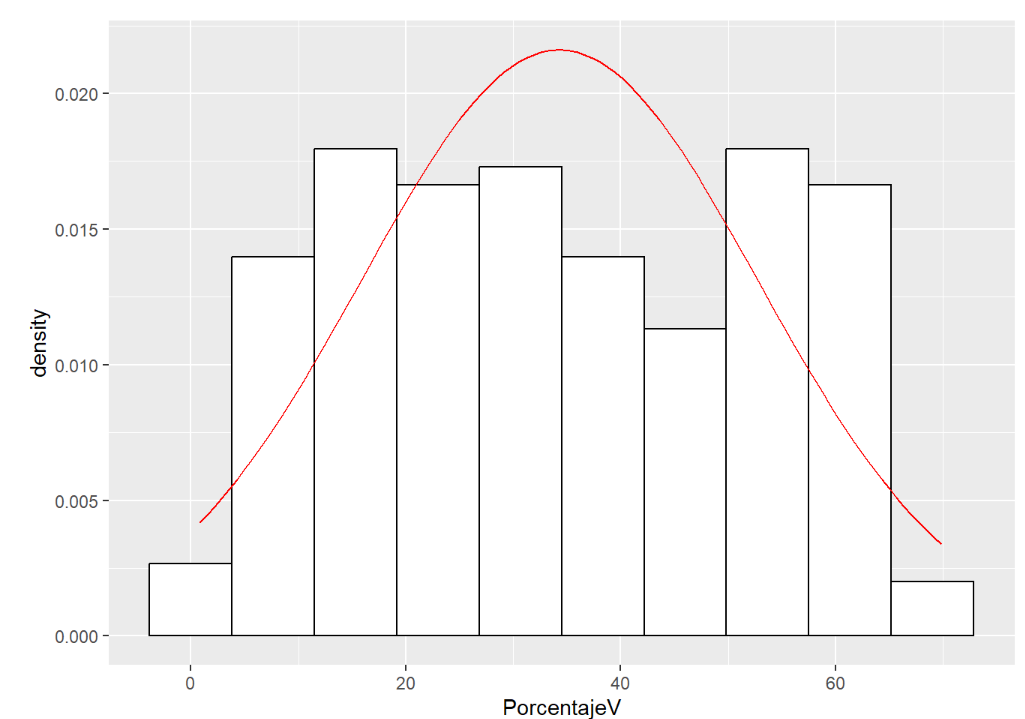
El proceso de recojo de fuentes fue mediante la página del gobierno, Infogob, de allí se logró conseguir las bases de datos sobre las elecciones 2021 y sobre el voto de mujeres y jóvenes. Al momento de limpiar, las bases estaban bastantes limpias y los nombres estaban sin tildes, dato importante. Luego, la tercera base de datos fue de Ceplan, donde se pudo recolectar las variables del IDH y de pobreza extrema. El hecho de juntar las dos primeras bases de datos, se realizó por *by* = Provincia y no hubo problemas. Sin mebargo, al juntar ambas bases de datos, algunos datos se perdieron ya que algunos nombres de provincia en la base de Ceplan tenían tilde. Así que, el *merge* lo reconocía como otra variable, por lo que fue un poco tedioso el hecho de cambiar las letras con tilde a sin tilde y cambiar algunos nombres porque estaban escritos de manera diferente. Sin embargo, se logró el cambio y la base de datos quedó limpia y sin datos perdidos. De allí en adelante, la *data* limpia ayudó a que los análisis fueran mejor y más exactas a la hora de usarlos en el programa Rstudio.

Lo que es la operacionalización de las variables dependiente, independientes y de control, lo que se busca lograr en el trabajo es que las variables independientes (votantes mujeres, jóvenes e IDH) expliquen la variable dependiente (votos por Castillo) mediante una variable de control (pobreza extrema). Por lo que, se mencionó anteriormente, la gran cantidad de votantes por el ex presidente Pedro Castillo pertenece al grupo más afectado en lo que se considera en el cuadro de pobreza económicamente, pero qué tanto puede significar que sea una mujer, o joven, en pobreza extrema y vote por Pedro Castillo porque según Román (2023), las personas se sentían identificadas y podrían creer que los ayudaría a salir de su condición, o al menos mejorarla, Además, el hecho de ser joven podría involucrar que sea más fácil de ser manipulado por grupos de izquierda, como lo era el gobierno de Castillo, para apoyarlo en estas elecciones, o el hecho de desconocer sobre las elecciones o propuestas de otros candidatos ya que si también existe un nivel de IDH bajo, implicaría la baja educación que pueden tener e implicaría desconocer diversos temas políticos.

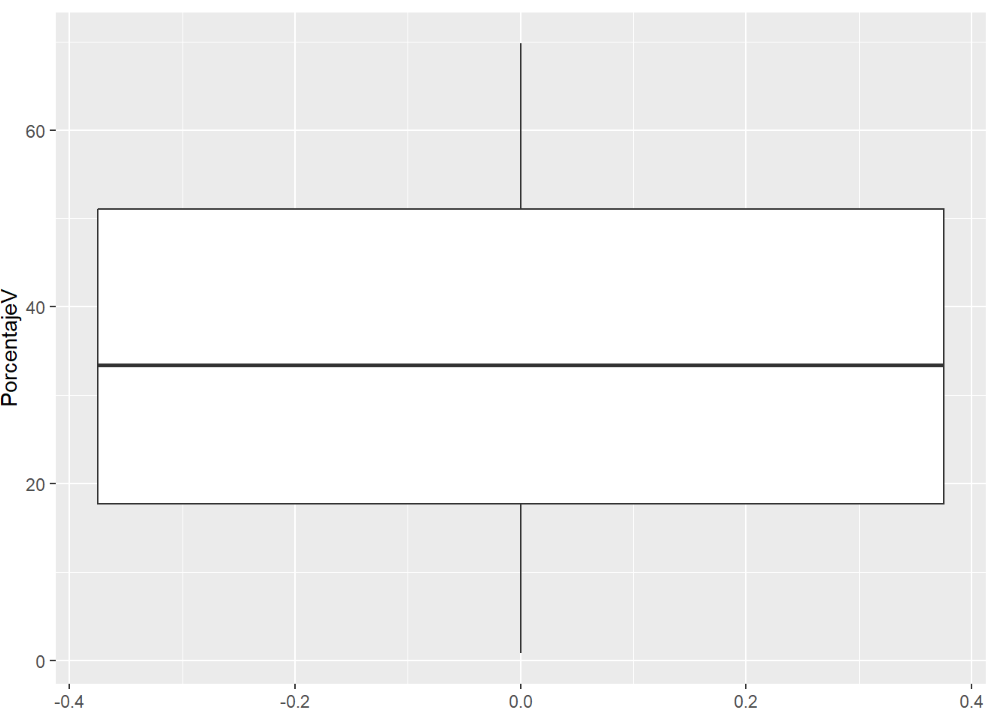
**Hallazgos**

***Variable central***

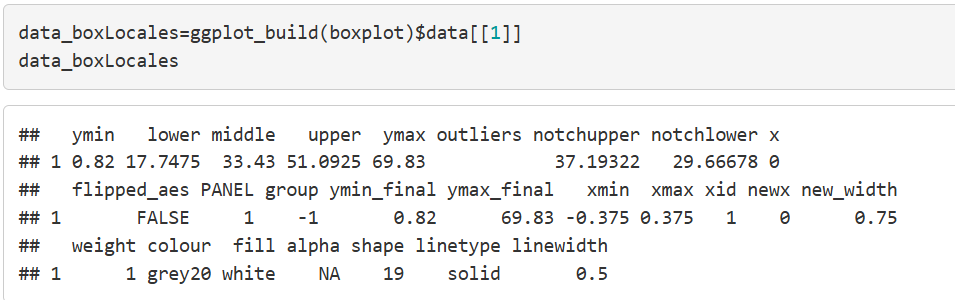
La variable central en la base de datos, como se mencionó, es *Porcentaje\_V,* la cual hace referencia al porcentaje de votos que recibió Pedro Castillo a nivel provincial. Por lo que, para analizarla se realizó un *summary* y se pudo observar que el mínimo valor es 0.8, el mediano es 33.4 y el máximo 69.8. Por lo que se realizó un histograma con los datos obtenidos.



Seguidamente, se realizó un bloxpot para verificar si habían *outliers.* Donde luego se pudo observar que no existían.

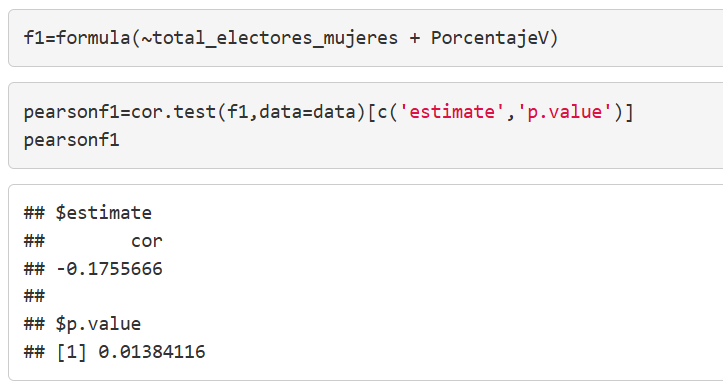


Y para verificar aún más, mediante un código de *ggplot\_bluid,* se verificó que, efectivamente, no había valores atípicos en la variable central.



***Bivariado***

Para los análisis bivariados se realizaron correlaciones Pearson, donde las figuraba que las variables independientes no tenían una fuerte correlación con la variable central.

* Primera correlación:

#total\_electores\_ mujeres

* Segunda correlación:

#porcentaje\_jovenes

* Tercera correlación:

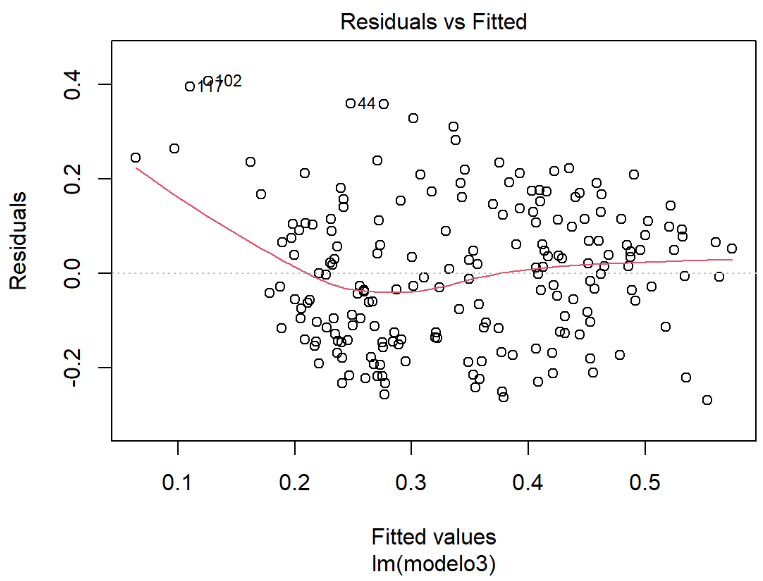


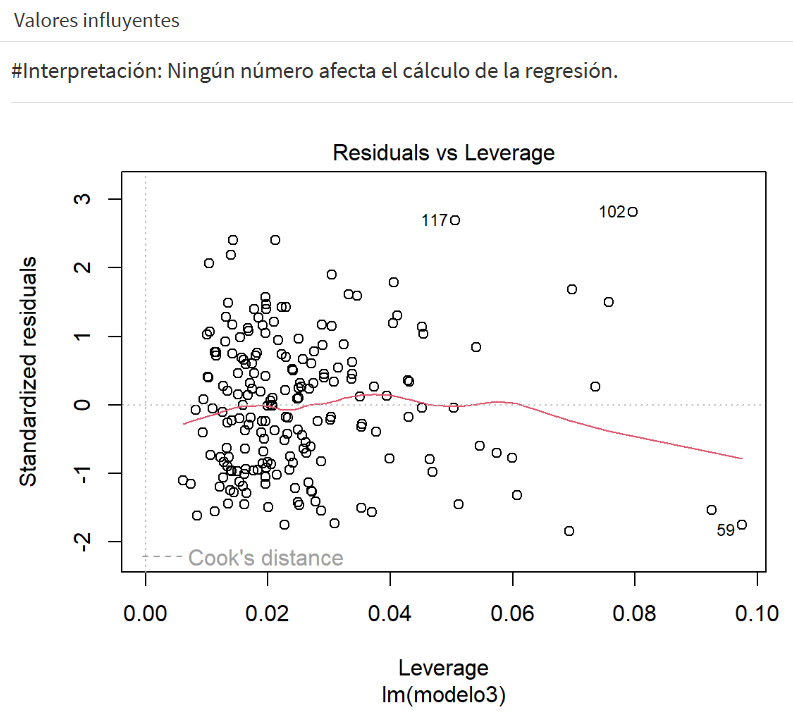
#IDH

En conclusión, en el análisis bivariado, el porcentaje de mujeres, jóvenes, IDH no estarían fuertemente, relacionadas con los votos por Castillo. Sin embargo, luego se realiza una regresión lineal para verificar si continúa siendo insignificante o tiene otro tipo de correlación.

***Análisis multivariado***

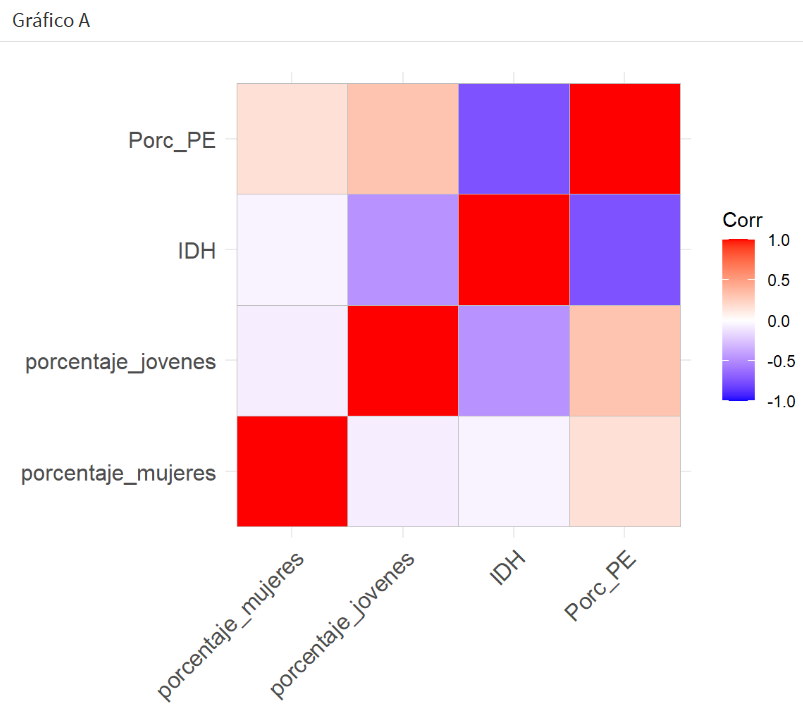
* Modelos de regresión

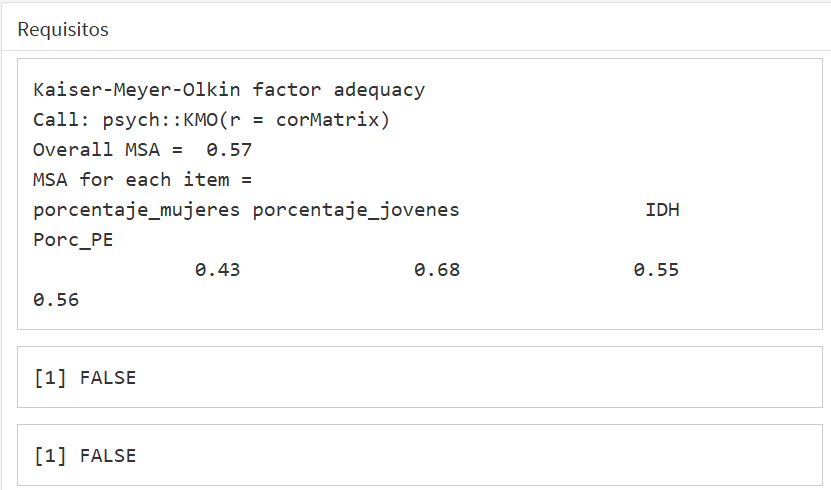
Se realizó un modelo de regresión lineal, en el cual en el primero modelo se observa que el porcentaje de mujeres votantes sí tiene efecto y es significativo, ya que tiene una relación directa y es controlada por la variable del porcentaje de pobreza extrema (0.020\*\*\*). Como segundo modelo, se añade la variable del porcentaje de jóvenes; es decir, está la variable porcentaje de mujeres + porcentaje de jóvenes + porcentaje de pobreza extrema y se puede apreciar que el porcentaje de jóvenes no tiene significancia en los votos de Castillo. Además, tiene un efecto inverso (-0.004). Por último, como tercer modelo, está la variable porcentaje de mujeres + porcentaje de jóvenes + IDH + porcentaje de pobreza extrema, donde el IDH tiene un efecto inverso pero significativo (-0.611\*\*\*). Y, se puede observar que la significancia en la variable de porcentaje de jóvenes ha cambiado (-0.008\*\*). Luego de obtener tres modelos, se utiliza la prueba Anova para compararlos y reconocer cuál es el mejor, el cual dio como resultado el tercer modelo con 0.0000760 a diferencia del modelo 2 que dio como resultado 0.1575174. Por lo tanto, el modelo 3 es el mejor. A partir de ello, se realizaron pruebas de Linealidad, lo que la falta de linearidad provocaría que el modelo no sriva para explicar las mismas variables con datos diferentes en otros estudios.

Asimismo, también se realizaron más pruebas como homocedasticidad, normalidad de los residuos, no multicolinealidad, y valores influyentes. Los dos últimos, se pueden apreciar a continuación:

(observar en [Dashboard](file:///C:\Users\soporte\Downloads\TrabajoFinal.html))

* Análisis factorial (EFA)

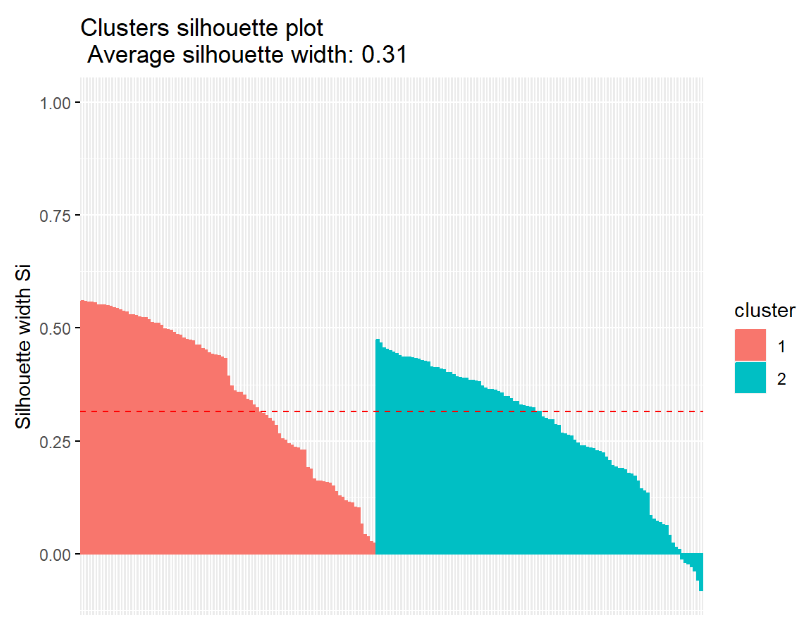
Por otro lado, se realizó un EFA, pero no se pudo concluir con el análisis ya que, a primera vista al graficar las correlaciones, se nota la poca correlación entre las variables. Y, efectivamente, al buscar cumplir con los requisitos: KMO, Bartlett y Singular.matrix, no cumple con todos los requisitos, por lo que se obtendría un resultado no esperado y continuar no sería lo mejor.



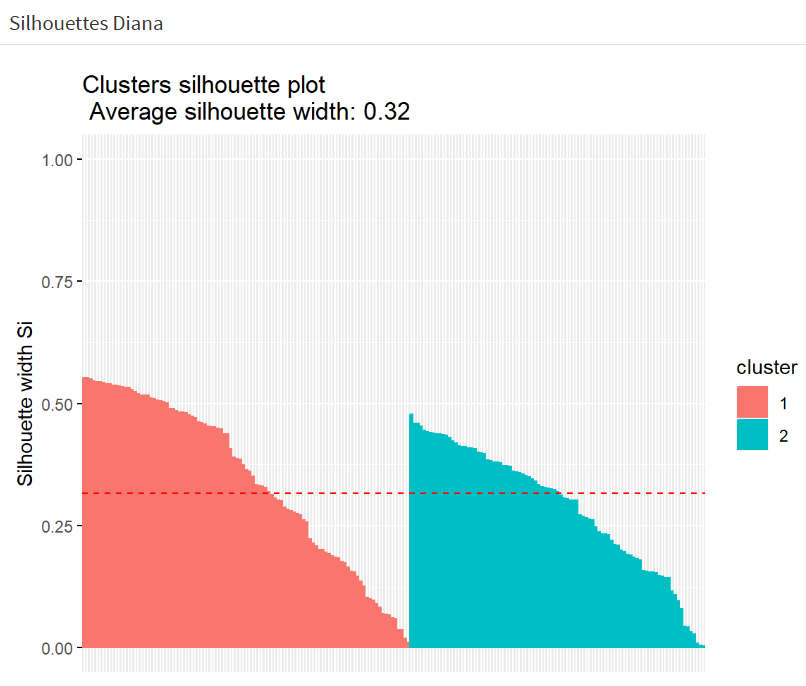
* Modelo de clusterización

Por último, se realizaron los modelos de clusterización: PAM, Agnes y Diana.

Para PAM, el resultado recomendó dos clusters, luego se realizó el gráfico de *silhouette* y se podía apreciar que habían provincias que estaban mal clusterizadas.



Mediante un código se observó que las provincias mal clusterizadas eran siete: Bongara, Cajatambo, Huamanga, Jaen, Lampa, Sandi y Utcubamba.

Luego, se quiso trabajar con el modelo de Agnes, sin embargo, el modelo recomendó un cluster, con el cual luego, no se pudo trabajar ya que el análisis no permitía un operador unitario. Así que, se prosiguió con el siguiente modelo: Diana. El modelo recomendó de igual manera dos clusters y al momento de realizar el gráfico y verificar las provincias mal clusterizadas, no había ninguna.

**Conclusiones**

Finalmente, se puede comprobar cierta parte de la hipótesis ya que las características de los votantes de Castillo es que sean mujeres en extrema pobreza con bajo IDH en su provincia son quienes lo apoyaron para las elecciones presidenciales de Perú 2021. Así que, los factores: porcentaje de mujeres e IDH tienen mayor significancia y efecto en los votos por Castillo. Por lo que se puede concluir que aquellas elecciones demostraron que el país está en pésimas condiciones de poder elegir un candidato más por sus propuestas que por sentirse identificados con él. De la misma manera, pasó con Toledo, quien decía ser “del pueblo” e igual no apoyó a las personas de zona rural o de zonas marginadas. Durante el gobierno de Castillo, no realizó grandes cambios en las zonas rurales y, es importante que el gobierno peruano, sea de izquierda o no, se sienta identificado con los provincianos o no, debe de estar presente en todos los departamentos del país porque “Lima no es Perú”.

Bibliografía

Duárez M., (2022). La deriva de la representación política en el Perú. El triunfo electoral de Pedro Castillo y la cuestión de la representatividad política. *Discursos del sur*, (10), 9-30. Epub 31 de diciembre de 2022. <https://dx.doi.org/10.15381/dds.n10.24404>

Román P., (2023). Politics as usual? El gobierno efímero de Pedro Castillo en Perú. Les Études du CERI, Amérique latine. L’Année politique 2022, n° 264-265, pp.30-34. <https://sciencespo.hal.science/hal-03968432/>