



FACULTAD DE CIENCIAS ECONÓMICAS Y  
EMPRESARIALES.

**ANÁLISIS SOCIOECONÓMICO DE LA  
VACUNACIÓN CONTRA EL COVID-19 ENTRE  
LOS JÓVENES EN ESPAÑA.**

**¿En qué medida afectan, a lo largo de la pandemia, a la  
reticencia hacia la vacuna la ideología política, el nivel  
de estudios y el miedo a contagiarse del Covid-19?**

Autor: Manuel María Jiménez-Poyato Narváez

Director: Antonio Rúa Vieites

Madrid

Junio 2022

## Resumen

*El objetivo de este trabajo de investigación es conocer la evolución de la actuación de los jóvenes en España cuya edad oscila entre los 18 y los 26 años en relación con su reticencia hacia la vacuna contra el COVID-19 o no. En concreto, se va a analizar si existen patrones de comportamiento relacionados con su ideología política, su nivel de estudios y su grado de miedo a contagiarse del Coronavirus.*

*Para poder realizar dicha síntesis, primero se ha reflejado el contexto de la pandemia en España, reflejando la evolución de las cifras de contagios, fallecimientos y porcentaje de vacunación. Más adelante, se han explicado las tres encuestas llevadas a cabo por la Fundación Española para la Ciencia y Tecnología (FECYT), de las cuales se han extraído los datos a utilizar en esta investigación. A continuación, se ha expuesto, a modo explicativo, la importancia de la aplicación práctica del Machine Learning en casos como el presente, en concreto de las técnicas de clustering y regresión lineal múltiple en RStudio. Una vez preparado el escenario, se ha indicado, paso a paso, cómo se han implementado dichas técnicas de análisis de datos, encuesta a encuesta, buscando patrones de comportamiento a medida que transcurre la pandemia. Por último, a modo de conclusión, se han reflejado los patrones de comportamiento descubiertos y se les ha dado una posible explicación obtenida de los acontecimientos sanitarios, sociales y políticos a lo largo de la pandemia expuestos en la introducción de la investigación.*

**Palabras clave:** COVID-19, vacunación, reticencia, jóvenes, patrones de comportamiento, Machine Learning.

## Abstract

*The aim of this research work is to know the evolution of the behavior of young people in Spain between 18 and 26 years of age in relation to their reluctance to be vaccinated against COVID-19 or not. Specifically, we will analyze whether there are behavioral patterns related to their political ideology, their level of education and their degree of fear of contracting the Coronavirus.*

*In order to carry out this synthesis, first the context of the pandemic in Spain has been reflected, showing the evolution of the numbers of infections, deaths and vaccination rates. Later, the three surveys carried out by the Spanish Foundation for Science and Technology (FECYT), from which the data to be used in this research were extracted, were explained. Next, the importance of the practical application of Machine Learning in cases such as the present one was explained, specifically the clustering techniques and multiple linear regression in RStudio. Once the scenario has been prepared, it has been indicated, step by step, how these data analysis techniques have been implemented, survey by survey, looking for patterns of behavior as the pandemic progresses. Finally, in conclusion, the behavioral patterns discovered have been reflected and given a possible explanation drawn from the health, social and political events throughout the pandemic that were outlined in the introduction to this research.*

**Keywords:** COVID-19, vaccination, reluctance, young people, behavioral patterns, Machine Learning.

## Índice

|   |       |
|---|-------|
| <b>1. INTRODUCCIÓN</b>  | p. 5  |
| <b>1.1 COVID-19. Cifras generales</b>   | p. 5  |
| <b>1.2 Cronología de la pandemia en España.</b>                                       | p. 5  |
| <b>1.3 Encuestas llevadas a cabo por la FECYT.</b>                                    | p. 20 |
| <b>1.4 Justificación de la investigación</b>  | p. 13 |
| <br>  |       |
| <b>2. MACHINE LEARNING. APLICACIÓN PRÁCTICA</b>                                       | p. 14 |
| <b>2.1 Introducción</b>   | p. 14 |
| <b>2.2 Aprendizaje no supervisado. <i>Clustering</i></b>                              | p. 15 |
| <b>2.3 Aplicación práctica</b>  | p. 16 |
| <b>2.4 Regresión lineal: múltiple</b>   | p. 17 |
| <b>2.5 Análisis de las variables escogidas</b>  | p. 18 |
| <br>  |       |
| <b>3. CLUSTERING Y REGRESIÓN LINEAL MÚLTIPLE EN RSTUDIO</b>                           | p. 21 |
| <b>3.1 Preparar la base de datos</b>  | p. 21 |
| <b>3.2 Encuesta julio 2020</b>  | p. 21 |
| <b>3.2.1 P13_1 y D4. Reticencia hacia la vacuna e ideología política</b>              | p. 22 |
| <b>3.2.2. P13_1 y D5. Reticencia hacia la vacuna y nivel de estudios</b>              | p. 23 |
| <b>3.2.3 P13_1 y P7 Reticencia hacia la vacuna y miedo a contagiarse del COVID-19</b> | p. 23 |
| <b>3.2.4 Regresión lineal múltiple</b>  | p. 25 |
| <b>3.2.5 Conclusión</b>   | p. 28 |
| <b>3.3 Encuesta enero 2021</b>  | p. 28 |
| <b>3.3.1 P13_1 y D4. Reticencia hacia la vacuna e ideología política</b>              | p. 30 |
| <b>3.3.2 P13_1 y D5. Reticencia hacia la vacuna y nivel de estudios</b>               | p. 31 |
| <b>3.3.3 P13_1 y P7 Reticencia hacia la vacuna y miedo a contagiarse del COVID-19</b> | p. 32 |
| <b>3.3.4 Regresión lineal múltiple</b>  | p. 32 |
| <b>3.3.5 Conclusión</b>   | p. 36 |
| <b>3.4 Encuesta julio 2020</b>  | p. 36 |
| <b>3.4.1 P13_1 y D4. Reticencia hacia la vacuna e ideología política</b>              | p. 37 |

|   |       |
|---|-------|
| <b>3.4.2</b> P13_1 y D5. Reticencia hacia la vacuna y nivel de estudios               | p. 38 |
| <b>3.4.3</b> P13_1 y P7 Reticencia hacia la vacuna y miedo a contagiarse del COVID-19 | p. 39 |
| <b>3.4.4</b> Regresión lineal múltiple  | p. 40 |
| <b>3.4.5</b> Conclusión   | p. 43 |
| <b>4. RELACIÓN ENTRE LOS DATOS OBTENIDOS CON LA REALIDAD</b>                          | p. 43 |
| <b>4.1</b> <i>Clustering</i>  | p. 43 |
| <b>4.1.1</b> Ideología política y la reticencia hacia la vacunación                   | p. 43 |
| <b>4.1.2</b> Nivel de estudios y reticencia hacia la vacunación                       | p. 45 |
| <b>4.1.3</b> Miedo a contagiarse del COVID-19 y la reticencia hacia la vacunación     | p. 48 |
| <b>4.2</b> Regresión lineal múltiple  | p.49  |
| <b>5. CONCLUSIONES</b>  | p. 50 |
| <b>6. BIBLIOGRAFÍA</b>  | p. 51 |
| <b>ANEXO</b>  | p. 57 |

## 1. INTRODUCCIÓN

### 1.1 COVID-19. Cifras generales

En el mes de marzo de 2020 comenzó a nivel mundial una pandemia debida a un virus denominado "SARS-CoV-2", más comúnmente conocido como "COVID-19" o "Coronavirus". Dicho virus ha causado un antes y un después en la forma de actuar del ser humano debido a la alta propagación y la mortandad que ha dejado a su paso (Matús-Abásolo, Nemeth-Kohanszky e Inostroza-Tapia, 2020).

El COVID-19 es un virus originario de la ciudad china Wuhan y se caracteriza por tener síntomas similares a la gripe. Los efectos secundarios más comunes son tos, fiebre y fatiga. Además, es curioso que es común entre las personas que contraen el Coronavirus la pérdida del gusto y del olfato, siendo en un considerable número de ocasiones irreversible. El problema tiene lugar cuando el COVID-19 deriva en una neumonía, en problemas respiratorios severos o en aquellos relacionados con la coagulación de la sangre y trombosis (Pérez Abreu, Gómez Tejada y Dieguez Guach, 2020, p.2). Una de las principales características del Coronavirus es que está en constante evolución, adaptándose a cada escenario que se le presenta. Es decir, a medida que se producen cambios en su código genético, dicho virus muta. Se han registrado sucesivas variantes del COVID-19 que han convivido entre ellas, destacando así la variante “Alfa” desde el principio de la pandemia, “Delta” desde junio de 2021 en adelante de 2021 y “Ómicron” de noviembre 2021 en adelante (Orellana-Centeno y Guerrero Sotelo, 2022, p. 28).

Haciendo referencia a las principales cifras que el Coronavirus ha traído a España hasta el mes de abril de 2022, estas son una cantidad de casi 11,6 millones de casos y alrededor de 103.000 fallecidos registrados. A nivel global, el COVID-19 ha provocado la muerte de más de 6 millones de personas ("COVID-19 en España", 2022; "COVID-19 en Todo el mundo", 2022).

### 1.2 Cronología de la pandemia en España.

La actitud de la población española y las decisiones gubernamentales ante el Coronavirus ha ido variando a lo largo de la pandemia. Por ejemplo, al principio, ante tal situación de incertidumbre, se optó por el confinamiento domiciliario, considerándose

esta la medida más severa que se ha implementado. No obstante, a medida que los meses han transcurrido se ha ido optando por diferentes "estrategias" para erradicarlo. Dichas estrategias son, por ejemplo, el uso de la mascarilla, el cierre perimetral de algunas Comunidades Autónomas, la imposición de un "toque de queda" domiciliario a partir de una determinada hora de la noche, la limitación a un número concreto de personas a la hora de reunirse socialmente, o, la más determinante, la opción de inyectarte la vacuna contra el COVID-19 o no (Hernández y de la Rosa, 2021, p.64). Esta última medida ha sido la más decisiva en cuanto al freno de la propagación del virus o la reducción de los síntomas del que lo padece. Es interesante cómo España ha sido un ejemplo como país a la hora de llevar a cabo una vacunación en masa a prácticamente toda la población (85,7% en abril de 2022) con la intención de obtener una inmunidad total ("COVID-19 en España", 2022). Esto no debe de sorprender, ya que no es la primera vez que, gracias a una campaña a favor de una vacuna, se erradica una enfermedad en España. Esto ocurrió entre las décadas de 1950 y 1960 con la poliomielitis, la cual quedó prácticamente erradicada gracias a una inmunización por vacunación en masa de la población (Nájera, 2021).

En este apartado de la Introducción se va a llevar a cabo un análisis cronológico de la evolución del Coronavirus en España desde marzo 2020 hasta finales de febrero 2022. En dicho análisis se analizarán las sucesivas olas de la pandemia a la vez que el desarrollo de los contagios y defunciones con el objetivo de contextualizar la investigación.

La primera ola en España tuvo lugar durante los meses de marzo a junio de 2020, en los cuales el país llevó a cabo un confinamiento domiciliario. El Instituto Nacional de Estadística (INE) registró un total de 45.648 muertes causadas por el COVID-19. En comparación con el resto de los países de la Unión Europea, cabe destacar que España e Italia fueron los más afectados por el COVID-19 durante la primera ola (Mateos y González, 2021).

Después de los meses en los que tuvo lugar el confinamiento, comenzó la segunda ola del Coronavirus en nuestro país el 20 de julio de 2020, transcurriendo hasta el 20 de diciembre de 2020. El principal factor que caracterizó a esta ola fue la vuelta a la normalidad de la población en los meses de verano. Cabe destacar que esta fue la última ola en la que la población combatió el virus sin la ayuda de la vacuna, ya que esta no llegaría hasta finales de diciembre de 2020. Durante esta segunda ola, el Ministerio de

Sanidad anunció la cantidad total de 48.926 defunciones por Coronavirus en España (Arroyo, 2021). Durante estos meses, las dos principales estrategias que se llevaron a cabo para frenar la curva de contagios fueron el uso de la mascarilla permanente y la implantación de un toque de queda en horario nocturno en octubre de 2020 (Hernández y de la Rosa, 2021, p.64).

Tras estas dos primeras olas, llegó el primer cargamento de vacunas *Pfizer BioNTech* contra el COVID-19. No obstante, el crecimiento inicial de la puesta en marcha de la vacunación en masa fue lento, ya que se administró la vacuna primero al personal sanitario y a los ciudadanos considerados más vulnerables. Más adelante, llegaron nuevas vacunas contra el COVID-19, entre las cuales caben destacar *AstraZeneca*, *Moderna* y *Janssen* (González, 2021).

Sin embargo, las vacaciones de Navidad y las reuniones sociales no ayudaron a la hora de frenar el número de contagios en nuestro país, de forma que desde el 15 de enero a mediados de febrero de 2021 tuvo lugar la tercera ola. La principal característica de esta tercera ola fue el aumento de los ingresos en hospitales. No obstante, el hecho de que la vacuna se comenzó a administrar supuso una reducción considerable a la hora de tratar el número de muertos por el virus entre la población más vulnerable, como, por ejemplo, en las residencias de personas de la tercera edad (Sanitaria, 2021).

Más adelante tuvo lugar la cuarta ola durante los meses de marzo y abril de 2021. Esta se caracteriza por ser mucho más estable y controlada en comparación con la ola anterior. No obstante, se siguieron sufriendo las consecuencias de las hospitalizaciones que se llevaron a cabo durante los meses de invierno. No obstante, el efecto de dicha ola no tiene comparación con los de las olas anteriores. De hecho, en el mes de abril de 2021 se suministraron tantas dosis como se habían suministrado en total en el periodo enero-marzo, lo que indica un crecimiento exponencial de la vacunación (Tovar, 2021).

Sin embargo, una nueva variante del Coronavirus provocó la llegada de la quinta ola a España durante el verano de 2021, comenzando en el mes de junio y terminando en septiembre. El origen detonante de dicha ola fue el rebrote masivo en Mallorca, llegando así la variante “Delta” a España provocando un gran número de contagios entre los jóvenes. Dicha variante, originaria de la India, supuso una mayor transmisibilidad a la hora de expandirse entre la población. Cabe destacar que la vacunación en masa supuso que el número de muertos por Coronavirus durante dichos meses descendiera. En

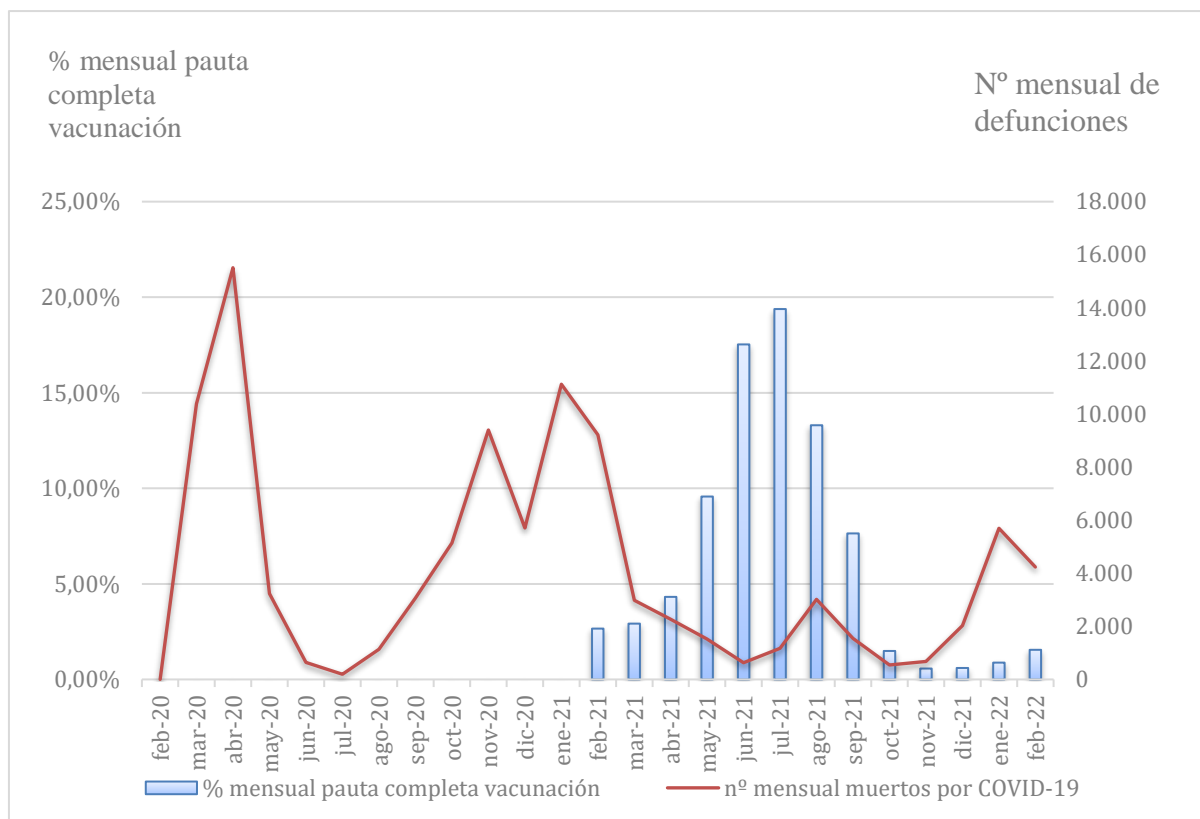


septiembre de 2021 el 70% de la población española tenía la pauta completa de vacunación. Es cierto que influye que el Gobierno de España, a la hora de administrar las vacunas, consideró que el hecho de haber contraído el Coronavirus suponía la carga equivalente de una dosis, de forma que esas personas que ya lo habían contraído en un intervalo de 6 meses, con solo una dosis, tenían la pauta completa de vacunación (Martos, Martín, Zambrana, Bianchi, 2021).

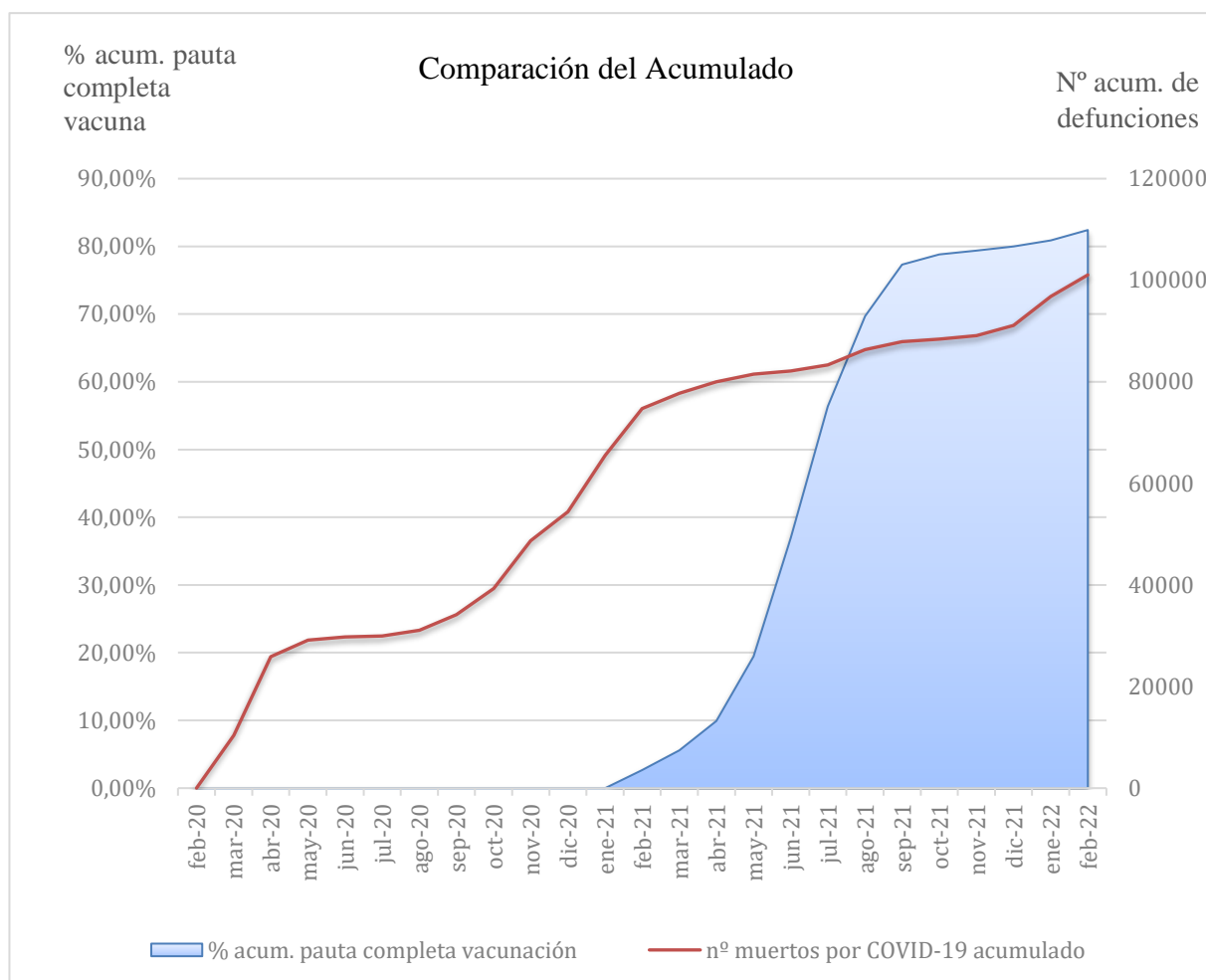
No obstante, aunque el final parecía que estaba presente después de la quinta ola, una nueva variante denominada “Ómicron” volvió a azotar las vidas de la población en el mes de diciembre de 2022. Debido a un altísimo porcentaje de la población con la pauta completa (84,86%), e incluso con una tercera dosis de refuerzo al sector de la población más vulnerable frente al Coronavirus, el número de muertos durante la sexta ha sido ínfimo. En cambio, dicha variante es una de las más contagiosas, provocando la recaída de una gran parte de la población que ya contrajo dicho virus durante los meses pasados, e incluso la infección en personas que ya estaban vacunadas con la pauta completa. Es curioso que la Organización Mundial de la Salud (OMS) ya empezaba a decir que la variante “Ómicron” podía ser el principio del fin del COVID-19 en Europa (BBC News Mundo, 2022). Lo que se ha evidenciado, una vez más como con la poliomielitis, es que si no hubiera sido por la vacunación en masa en España no habría sido posible frenar la curva de contagios, muertes y hospitalizaciones. Por otra parte, cabe destacar que en otros países europeos en los que no se ha querido optar por la vacunación en masa, se han sufrido en la sexta ola severas consecuencias. Por ejemplo, Austria, cuyo porcentaje de personas sin ninguna dosis es 30%, ha registrado en enero de 2022 máximos históricos de número de contagios diarios, siendo 17.000 el récord (Punz, 2022). Un dato representativo de España es que el 100% de la población mayor de 80 en España tiene en marzo de 2022, al menos, la pauta completa de vacunación (Datadista, 2022).

A continuación, se puede observar un cuadro de elaboración personal a partir de los datos recabados de las plataformas Expansión (2022) y Datadista (2022) en el cual se estudia la comparación entre el número de muertos y de vacunados a lo largo de la pandemia, tanto mensual como acumulado.

**Figura 1. Comparación del número de fallecidos con el porcentaje de vacunación de la población española en cifras mensuales a lo largo de la pandemia (Expansión, 2022; Datadista, 2022).**



**Figura 2. Comparación del número de fallecidos con el porcentaje de vacunación de la población española en cifras acumuladas a lo largo de la pandemia (Expansión, 2022; Datadista, 2022).**



En ambos gráficos se muestra cómo el incremento de la vacunación de la población es inversamente proporcional al número de defunciones. En concreto, se puede observar cómo en la primera mitad de 2021 (cuando tuvo lugar el inicio de la vacunación en masa) se neutralizó el número total de defunciones. Sin embargo, la propensión a la vacunación no siempre ha seguido un patrón constante.

### 1.3 Encuestas llevadas a cabo por la FECYT:

La Fundación Española para la Ciencia y la Tecnología ("FECYT") ha llevado a cabo tres encuestas a lo largo de la pandemia con la intención de estudiar y mostrar la percepción que la población española tiene del COVID-19 y de la vacuna para combatirlo.

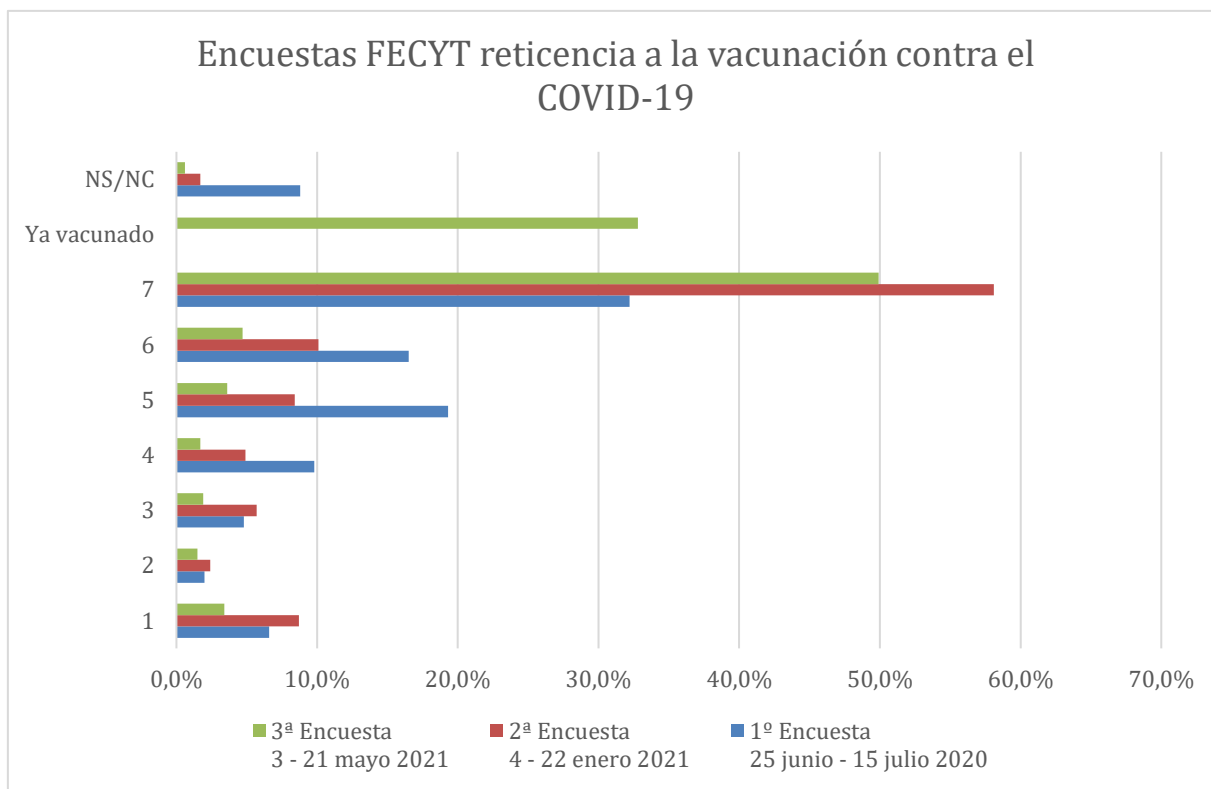
En concreto, las encuestas estudian en qué medida la población está dispuesta a vacunarse o no y cuáles son los factores sociológicos relacionados. A lo largo de estas tres encuestas se puede observar cómo el porcentaje de reticencia hacia la vacunación no permanece estable. Se debe recordar que todos los individuos que participaron son personas con nacionalidad española, mayores de edad y que llevaran 5 años o más residiendo en España.

En la tabla que se muestra a continuación, elaborada a partir de los datos provenientes de las encuestas del FECYT, se pueden observar los resultados obtenidos a la pregunta principal que se hizo a los participantes. En concreto, esta fue: “Si mañana mismo estuviese disponible una vacuna para prevenir la COVID-19, ¿Se la pondría? Por favor, responda en una escala en la que 1 es “No se la pondría en absoluto” y 7 es “Totalmente seguro que se la pondría” (FECYT, 2021).

**Figura 3. Evolución de las respuestas globales en las tres encuestas a la pregunta acerca de la reticencia hacia la vacunación (FECYT, 2021).**

| ESCALA         | 1º Encuesta<br>25 junio - 15<br>julio 2020 | 2ª Encuesta<br>4 - 22 enero<br>2021 | 3ª Encuesta<br>3 - 21 mayo<br>2021 |
|----------------|--|-------------------------------------|------------------------------------|
| 1              | 6,6%                                       | 8,7%                                | 3,4%                               |
| 2              | 2,0%                                       | 2,4%                                | 1,5%                               |
| 3              | 4,8%                                       | 5,7%                                | 1,9%                               |
| 4              | 9,8%                                       | 4,9%                                | 1,7%                               |
| 5              | 19,3%                                      | 8,4%                                | 3,6%                               |
| 6              | 16,5%                                      | 10,1%                               | 4,7%                               |
| 7              | 32,2%                                      | 58,1%                               | 49,9%                              |
| Ya<br>vacunado | 0,0%                                       | 0,0%                                | 32,8%                              |
| NS/NC          | 8,8%                                       | 1,7%                                | 0,6%                               |
| Muestra (N)    | 2.100                                      | 2.108                               | 2.100                              |

**Figura 4. Gráfico con las respuestas globales en las tres encuestas a la pregunta acerca de la reticencia hacia la vacunación (FECYT, 2021).**



Analizando el gráfico superior, se puede concluir que, en rasgos generales, la reticencia hacia la vacunación decrece con el transcurso de la pandemia. Sin embargo, hay ciertos matices a destacar, como, por ejemplo, que en la segunda encuesta hay más personas que "en absoluto" se vacunarían respecto con la primera. Por otra parte, el tanto por ciento de personas que "no saben/no contestan" decrece. Eso puede significar que la población cada vez ha estado más informada y comprometida a la hora de interesarse e informarse acerca de la vacuna contra el Coronavirus. Aunque, a simple vista, hay más "Totalmente" dispuestos a vacunarse en la segunda en comparación con la tercera encuesta, es capcioso ya que se deberían sumar a los números 7 el colectivo "Ya vacunados", resultando ser un total de 82,7%.

Sin embargo, en estas tres encuestas también se preguntó por muchos otros aspectos sociales relacionados con la vacuna y el COVID-19, como, por ejemplo, la confianza de la población en el gobierno, su predisposición a usar mascarilla o a practicar

el distanciamiento social, la edad, ideología política o nivel de estudios. Más adelante se comentará este aspecto (FECYT, 2021).

#### **1.4. Justificación de la investigación**

La situación generada por el estallido del COVID-19 ha supuesto la creación de un escenario sin precedentes en la historia reciente (Guterres, 2020). Es decir, la población ha visto cómo hábitos cotidianos que eran impensablemente modificables, han sido trastocados. Es decir, a cambio de conseguir seguridad, se ha restringido en parte su libertad (Rueda, 2020).

Atendiendo a Phillips (2020), en la historia contemporánea se ha podido comprobar que, ante todo evento trascendente y con repercusión social, la sección juvenil de la población cada vez ha ido cobrando más protagonismo a la hora de reaccionar a ello. A pesar de que muchos autores confirman que fue la revolución de mayo de 1968 el “antes y después” en este aspecto, otros autores lo retrasan a la participación de los jóvenes en ambas guerras mundiales y en el aumento de protagonismo de los grupos de jóvenes en las grandes ciudades. Es decir, el "joven" se ha consagrado a lo largo de los últimos años como participante "ruidoso" de la vida social (Phillips, 2020).

Por lo tanto, esta restricción de seguridad debido a la actual pandemia supone un potencial escenario reaccionario para los jóvenes. Es decir, esta es una oportunidad única para estudiar el comportamiento de los jóvenes ante una situación socialmente trascendente y sin precedentes recientes en la historia (Guterres, 2020). Esta es la principal causa por la que se ha decidido, en este trabajo de investigación, estudiar el comportamiento de la sección juvenil de la población en España.

A modo de resumen, el panorama ante el que se encuentra el "joven medio" español se podría definir como una crisis sanitaria cuyo número de fallecimientos en España solo encuentra semejanza reciente a la Guerra Civil española (TeleSUR, 2021). Las diversas formas que existen de combatir y sobrevivir a esta crisis son a través de medidas de precaución (como el uso de mascarilla o el distanciamiento social). No obstante, una posible solución definitiva a recuperar su vida anterior al Coronavirus es por medio de la vacuna. Sin embargo, el tiempo medio de desarrollo de una vacuna es de 4 años, y la vacuna contra el COVID-19 se ha desarrollado en menos de un año

(Asociación Española de Vacunología, 2020). Además, el principal administrador de dicha vacuna es el gobierno vigente de España, el cual es de ideología progresista (RTVE, 2021). Por lo tanto, el abanico de posibilidades de actuación del "joven medio español" es muy amplio.

Gracias a las encuestas llevadas a cabo por la FECYT, en las cuales muchas preguntas ayudan a conocer diferentes aspectos de la conducta de los jóvenes, se permite llevar a cabo un estudio de la evolución, a lo largo de la pandemia, de los patrones de comportamiento de los jóvenes de cara a adoptar la vacuna o no como solución a la crisis del COVID-19. La decisión de ponerse la vacuna o no puede venir afectada por infinitos aspectos condicionantes que sean influyentes sobre los jóvenes. Sin embargo, para acotar la investigación, se ha decidido que los tres aspectos principales a estudiar son el nivel de estudios, la ideología política y el grado de miedo a contagiarse del Coronavirus. Aquí radica el objetivo de la investigación. Es decir, el propósito es analizar esas tres variables que influyen a los jóvenes a vacunarse contra el COVID-19 o no, en qué medida influyen y a raíz de eso qué patrones de comportamiento juvenil en este contexto se pueden obtener de dicho análisis.

Las tres variables mencionadas anteriormente se recogen en la encuesta bajo la forma de tres preguntas. Más adelante, se analizarán las variables escogidas con su pregunta pertinente de la encuesta.

## **2. MACHINE LEARNING. APLICACIÓN PRÁCTICA**

### **2.1 Introducción.**

"Machine Learning" es como se denomina una rama de la Inteligencia Artificial la cual, por medio de ciertas técnicas, se consigue el aprendizaje automático utilizando grandes volúmenes de datos (Russo et al., 2016). Es decir, Machine Learning se podría definir como una "ciencia que trata de hacer que los ordenadores aprendan y actúen como seres humanos inteligentes, alimentándolos con datos e información en forma de observaciones e interacciones del mundo real" (Faggella, 2020). Dicha forma de actuar se basa en algoritmos, los cuales nos llevan a diferenciar tres tipos de aprendizaje diferente:

- Aprendizaje supervisado: el algoritmo asocia a las variables de entrada un valor de salida adecuado (su "etiqueta" correspondiente), de forma que, con una gran cantidad de datos, se puede llegar a predecir ese valor de salida a nuevos valores de entrada (Recuero de los Santos, 2021). Por ejemplo, se consideraría aprendizaje supervisado una encuesta de sintomatología del COVID-19 que, según se termine, te confirma si padeces de Coronavirus o no. Esa es la etiqueta que se asigna a la nueva variable.
- Aprendizaje no supervisado: en este ámbito no se trabaja con datos a los que se les pueda asignar un valor "etiqueta" de salida. Por lo tanto, la única forma de trabajar con ellos es buscando estructuras que organicen las variables de entrada de tal forma que se clasifiquen los datos de entrada con similitudes encontradas (Recuero de los Santos, 2021). Por ejemplo, se consideraría aprendizaje no supervisado si se analiza el total de clientela de Zara en España y se agruparan en diferentes grupos en función de su edad, sexo y el dinero gastado. Este es el tipo de aprendizaje que concierne a este trabajo de investigación.
- Aprendizaje reforzado: es aquel ámbito en el que las variables carecen de etiqueta (no se conoce el valor salida asociado) pero tampoco tienen como objetivo estructurarse para clasificar por grupos buscando similitudes y patrones en común (Aprende Machine Learning, 2020).

## **2.2 Aprendizaje no supervisado. *Clustering***

Antes de comenzar a desarrollar este apartado, se debe aclarar en qué consiste la técnica de "clusterización" o *clustering*. Atendiendo a Villardón (2007), "el análisis de *clusters* (o análisis de conglomerados) es una técnica de Análisis Exploratorio de Datos para resolver problemas de clasificación. Su objeto consiste en ordenar objetos (personas, cosas, animales, plantas, variables, etc ...) en grupos (conglomerados o *clusters*) de forma que el grado de asociación/similitud entre miembros del mismo *cluster* sea más fuerte que el grado de asociación/similitud entre miembros de diferentes *clusters*. Cada *cluster* se describe como la clase a la que sus miembros pertenecen" (Villardón, 2007). Es decir, que la homogeneidad entre los integrantes de un mismo *cluster* sea elevada, y que la heterogeneidad entre integrantes de diferentes *clusters* lo sea también (Larrañaga, Inza y Moujahid, 2012). A raíz de la distancia entre los valores observados, se permite analizar



la similitud entre ellos, agrupándolos en *clusters* homogéneos (Vilá-Baños, Rubio-Hurtado, Berlanga-Silvente y Torrado-Fonseca, 2014).

Dentro de la técnica *clustering*, se pueden encontrar dos clases de métodos: el jerárquico y el no jerárquico. El método jerárquico consiste en “agrupar *clusters* para formar uno nuevo o bien separar alguno ya existente para dar origen a otros dos, de tal forma que, si sucesivamente se va efectuando este proceso de aglomeración o división, se minimice alguna distancia o bien se maximice alguna medida de similitud” (Gallardo). El resultado que se obtiene es una estructura con forma de árbol en la que existe una jerarquía. Al contrario, el método no jerárquico genera el número de *clusters* que haya sido predeterminado. El método principal no jerárquico es a través del algoritmo K-Means ("K-Medios").

### **2.3. Aplicación práctica.**

Como se ha explicado con anterioridad, el Coronavirus ha llevado a la población a una situación muy particular y sin precedentes en la historia reciente (Guterres, 2020). Ante la vacuna como posible solución ante la crisis sanitaria, no toda la población, como se ha estudiado, ha reaccionado igual.

En este trabajo de investigación se va a analizar la evolución de la disposición de la población joven (entre los 18 y los 26 años), a lo largo de las 3 encuestas realizadas por la FECYT, a vacunarse contra el COVID-19 o no. Además, se han escogido tres variables para poder estudiar su grado de afectación sobre la disposición a vacunarse o no. Estas tres variables son la ideología política, el nivel de estudios de la población y el grado de miedo que tiene la población joven a la hora de contagiarse o no del COVID-19. Se analizarán a continuación.

El objetivo es, a través de la técnica *clustering* por medio del programa RStudio, buscar perfiles de comportamiento entre la población joven en la primera encuesta que se llevó a cabo en julio de 2020 y, replicando el mismo proceso con los datos de las dos encuestas posteriores (enero 2021 y mayo 2021), analizar cómo han evolucionado los patrones de comportamiento juvenil a lo largo de la pandemia. La importancia de la investigación radica en la explicación de la evolución de dicho comportamiento. Esa explicación beberá de dos fuentes relacionadas: primero, de los macro-datos sociales

expuestos en la introducción de este trabajo de investigación (nuevas cepas, evolución del número de contagiados y fallecidos o porcentaje de pautas completas administradas) y segundo, beberá de las conclusiones obtenidas de los análisis de *clustering* practicados a continuación en RStudio.

Para el desarrollo de este trabajo de investigación se ha preferido la implementación del método no jerarquizado. Esto se debe a que, en RStudio, se han ido probando en cada *clustering* cantidades diferentes de centroides, de cara a elegir cuál es el que mejor representa y explica los diversos patrones de comportamiento.

## 2.4 Regresión Lineal: Múltiple.

La regresión lineal es una técnica estadística por medio de la cual se calcula el grado de influencia que una variable (X) tiene sobre otra (Y). A la variable Y se le denomina "variable respuesta" o "variable dependiente" y a la X se le denomina "predictor" o "variable independiente" (Silva, Jadue, Crivelli, y Buccioni, 2017).

En el ámbito de la regresión lineal, se pueden encontrar dos tipos: simple (cuando solo hay una variable independiente) o múltiple (cuando hay más de una) (Carrasquilla-Batista et al., 2016). En este trabajo de investigación, una vez se hayan realizado los *clusterings* correspondientes a cada encuesta, se va a practicar una regresión lineal múltiple. En ella, será la variable dependiente aquella que hace referencia a la reticencia hacia la vacunación, y las otras tres variables restantes serán los predictores. Atendiendo a Amat (2016), la ecuación que se obtendría a la hora de practicar una regresión de este tipo tendría la siguiente estructura:

$$Y = (B0 + B1 * X1 + B2 * X2 + B3 * X3) + e$$

Los componentes de esta ecuación son:

- $Y$  = variable que recoge la reticencia de la población joven a la vacunación.
- $X_i$  ( $i=1,2,3$ ) = variables "predictoras", de las cuales una recoge el nivel de estudios de la población joven, otra su ideología política y otra su miedo a contraer el COVID-19 o no.
- $B0$  = valor de la ordenada en el origen, es decir, es el valor de la variable dependiente cuando las variables independientes valen 0.

- $B_i$  ( $i=1,2,3$ ) = coeficientes de regresión, los cuales miden el efecto que tiene cada una de las variables predictoras sobre la variable respuesta.
- $e$  = es el error que recoge la diferencia entre el valor que se ha observado y el que ha sido estimado por el modelo.

(Amat, 2016).

El objetivo que se quiere alcanzar es obtener los coeficientes de la ecuación de regresión a través de RStudio. Dichos coeficientes indican en qué medida cada variable predictora está influyendo sobre la variable respuesta. Una vez se hayan analizado las tres encuestas y se hayan practicado la regresión lineal múltiple correspondiente a cada una, se podrá estudiar la evolución de dichos coeficientes. De esta forma, se expondrá cómo ha cambiado la influencia que cada una de las variables predictoras tienen sobre la variable de reticencia hacia la vacunación.

## **2.5 Análisis de las variables escogidas.**

Como se ha mencionado en el apartado 1.4 de esta investigación, las tres variables que se han escogido a la hora de estudiar la reticencia juvenil hacia la vacunación son la ideología política, el nivel de estudios y el grado de miedo a contagiarse del Coronavirus. Cada una de esas tres variables se recoge en cada encuesta hecha por la FECYT en forma de pregunta. A continuación, se desarrollará cada pregunta pertinente.

La primera variable (con nombre "P13\_1" en las dos primeras encuestas y "P12\_1" en la tercera) es la que refleja la reticencia hacia la vacunación. Dicha variable se mide, como se ha explicado con anterioridad, con la siguiente pregunta: "Si mañana mismo estuviese disponible una vacuna para prevenir la COVID-19, ¿Se la pondría? Por favor, responda en una escala en la que 1 es "No se la pondría en absoluto" y 7 es "Totalmente seguro que se la pondría". Los valores "No sabe" o "No contesta", representados en la base de datos con valores 98 y 99, se asignarán como valores perdidos "NA" sin tenerse en cuenta. Por lo tanto, las respuestas posibles serán del 1 al 7 (FECYT, 2021). Esta variable ha sido escogida ya que recoge la información esencial de este trabajo de investigación, es decir, la predisposición o no de la población joven hacia la vacuna. En concreto, la investigación a realizar es cómo afectan a esta variable las tres variables restantes.

La segunda variable (con nombre "D4") es aquella que mide la ideología de la población participante de la encuesta. En concreto, la pregunta que se llevó a cabo fue: "Cuando se habla de política se utilizan normalmente las expresiones izquierda y derecha. En una escala de 1 a 7, en la que 1 significa 'extrema izquierda' y 7 'extrema derecha', ¿en qué punto se colocaría Ud.?". Se debe recordar que, a lo largo de toda la pandemia del COVID-19, el Gobierno vigente ha sido una coalición de dos partidos políticos de ideología progresista. Relacionado con esta variable, serán las personas que respondan con cifras bajas las que sean más afines al Gobierno. Esta variable fue escogida para poder estudiar la ideología política de los jóvenes que han mostrado más o menos reticencia hacia la vacuna. Se interpreta a "izquierda" como progresista y a "derecha" como conservador. Partiendo de que el Gobierno vigente a lo largo de toda la pandemia ha sido de corte "progresista" (más afín a la respuesta 1 que a la 7 en esta variable), y que el Gobierno ha sido el "gestor" general de la pandemia a nivel nacional y el administrador principal de la vacuna contra el Coronavirus, va a ser interesante analizar la evolución a lo largo de la pandemia de la predisposición a vacunarse de los jóvenes afines y no afines a la ideología del Gobierno.

La tercera variable (con nombre "D5"), la cual recoge el nivel de estudios de la población, se estudió a través de la siguiente pregunta: "¿Cuáles son los estudios oficiales de más alto nivel que ha finalizado Ud. (obteniendo la titulación oficial correspondiente)?" Las posibles respuestas van enumeradas del 1 al 9. La número 1 significa "No sabe leer (analfabeto)", la 2 "Sin estudios sabe leer", la 3 "Estudios Primarios incompletos (Preescolar)", la 4 "Enseñanza de Primer Grado (EGB 1ª etapa, Ingreso, etc.) (Estudió hasta los 10 años)", la 5 "Enseñanza de 2º Grado/1º Ciclo (EGB 2ª etapa, 4º Bachiller, Graduado Escolar, Auxiliar Administrativo, Cultura General, etc.) (Estudió hasta los 14 años)", la 6 "Enseñanza de 2º Grado/2º Ciclo (BUP, COU, FP1, FP2, PREU, Bachiller Superior, Acceso a la Universidad, Escuela de Idiomas, etc.)", la 7 "Enseñanza universitaria primer ciclo, carreras de 3 años (Grado universitario, Ingenierías Técnicas/ Peritaje, Diplomados, ATS, Magisterio, tres años de carrera, etc.)", la 8 "Enseñanza universitaria segundo ciclo, carreras de 4 a 6 años (Máster, Licenciaturas, Ingenierías superiores, Medicina, etc.)" y la 9 "Enseñanza universitaria tercer ciclo (Doctorado)" (FECYT, 2021). Es importante recordar que la media en esta variable de nivel de estudios se encuentra en la obtención de la Educación Secundaria Obligatoria (ESO). Esta variable es interesante de analizar debido a que permite que se lleve a cabo

un seguimiento de cuál es la predisposición hacia la vacuna de las personas con mayor o menor nivel de estudios, y ver si estos colectivos, a lo largo de la pandemia, han evolucionado su conducta.

La cuarta variable (con nombres "P7" en las dos primeras encuestas y "P10\_2" en la tercera) que se va a escoger es la que refleja el miedo a contagiarse del Coronavirus. Dicha variable se midió a través de la siguiente pregunta: "¿Hasta qué punto diría que sería grave para usted contraer el nuevo coronavirus?" Se responde en una escala en la que 1 significa "nada grave" y 7 "muy grave" (FECYT, 2021). Esta variable mide el miedo a contagiarse del COVID-19. A medida que transcurra la pandemia y se neutralice el número de muertos y de contagiados, la consideración como "muy grave" de contagiarse variará. Esta variable ha sido escogida ya que refleja la reacción de la juventud ante la incertidumbre provocada por la pandemia. De aquí se pueden obtener otras conclusiones derivadas de este miedo (o esta tranquilidad) ante la gravedad de la situación, como puede ser la confianza respecto al Gobierno. Es interesante ya que permite que se estudie, para las personas que están asustadas con contagiarse del COVID-19, si la vacuna es la solución idónea o no.

En las 4 variables escogidas se pueden obtener solo, como respuesta, un número aislado. Por lo tanto, dichas variables se caracterizan por ser cuantitativas. Además, son variables discretas, puesto que solo se pueden obtener determinados valores "aislados" y que no haya valores entre ellos que supongan una infinidad (Westreicher, 2021). Un ejemplo para comprender la diferencia entre una variable cuantitativa discreta de una continua es que el número de primos hermanos maternos que tienes es discreto, mientras que las diferentes marcas que han obtenido los alumnos de una clase en un examen de Educación Física de 100 metros lisos es continua.

El hecho de trabajar con 4 variables cuantitativas y discretas facilita el desarrollo de la investigación. No obstante, dado que el rango no es el mismo en las 4 variables, se estandarizarán las variables para poder trabajar con las 4 juntas a la hora de realizar los análisis *clustering*, transformando en las 4 variables su media a 0 y su desviación típica a 1 (Statologos, 2022).

Una vez se han escogido las variables a analizar y se han explicado las dos técnicas estadísticas a practicar y el procedimiento a seguir, se dará paso a la investigación. Se irá explicando, paso a paso, cada avance obtenido.

### 3. CLUSTERING Y REGRESIÓN LINEAL MÚLTIPLE EN RSTUDIO

#### 3.1. Preparar la base de datos.

El primer paso ha sido escoger la base de datos completa de la encuesta de julio 2020 de la FECYT y en RStudio, se ha extraído a un fichero “csv” (separado por comas). Después, en el programa Excel, se han “limpiado” los datos. Es decir, los valores “No sabe” y “No contesta”, asignados con “98” y “99”, se han asignados como valores perdidos “NA”, los cuales, para poder llevar a cabo la técnica *clustering* por medio de K-Means, se han suprimido después. Además, solo han permanecido aquellos datos correspondientes a los participantes cuya edad está comprendida entre 18 y los 26 años (inclusive).

El siguiente paso consiste en transformar a las variables a una misma escala. Normalizar, según Marín (2018) consiste en "ajustar los valores medidos en diferentes escalas respecto a una escala común" (Marín, 2018). Se debe recordar que la variable que hace referencia al nivel de estudios (D5) comprende valores del 1 al 9, mientras que el resto de las variables del 1 al 7. Por medio de la función “scale”, se escalan las variables, centrando la media en 0 y la varianza en 1 (Amesquita, 2022).

Una vez se han escalado las variables, por medio del algoritmo K-Means (K-Medias) se van a ir representando y analizando la relación entre la variable principal (P13\_1, la cual recoge la reticencia de la población hacia la vacuna) con el resto de las variables. A continuación de cada gráfico, se expondrán las conclusiones obtenidas acerca de los diferentes patrones de conducta observados en el análisis *clustering*.

Este proceso de “preparación” de la base de datos se va a llevar a cabo en las tres encuestas de una manera paralela.

#### 3.2 Encuesta Julio 2020.

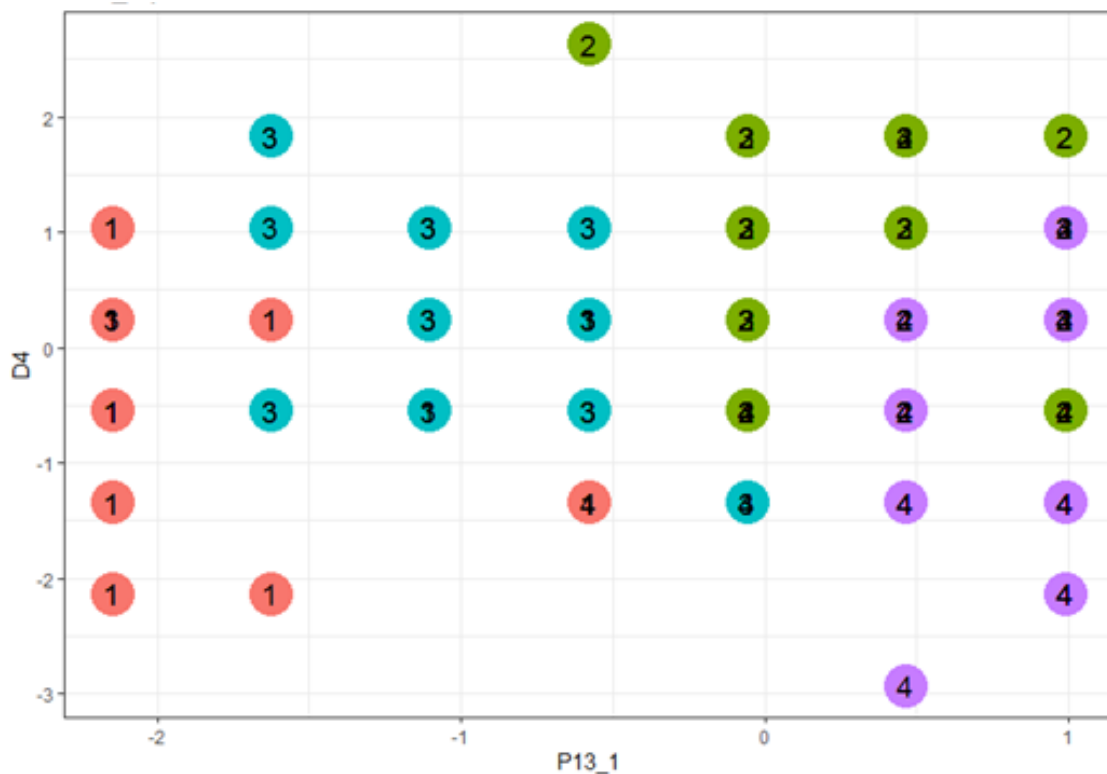
Se debe recordar que esta encuesta se realizó en la fase inicial de la pandemia, en la que la población acababa de salir de un confinamiento domiciliario de 3 meses, de forma que la incertidumbre y desconocimiento en las opiniones acerca de la vacuna era real.

Por medio de la función “ggplot” se van a representar las tres relaciones existentes entre la reticencia hacia la vacunación con el resto de las variables.

### 3.2.1. P13\_1 y D4. Reticencia hacia la vacuna e ideología política.

Se han utilizado 4 *clusters* en este análisis.

**Figura 5. Análisis *clustering* julio 2020 "Reticencia hacia la vacuna e ideología política" (FECYT, 2021).**



Se puede observar un grupo de jóvenes que no están a favor de ponerse la vacuna, representados en el *cluster* 1 (rojo) cuya ideología abarca desde el centro derecha hasta la extrema izquierda. En cambio, se puede observar un grupo de personas reflejado en el *cluster* 4 (morado) que siendo de ideología izquierda, están a favor de la vacuna. El *cluster* 2 (verde) refleja un grupo de jóvenes que, de ideología de "derechas", ni están a favor ni en contra de la vacuna.

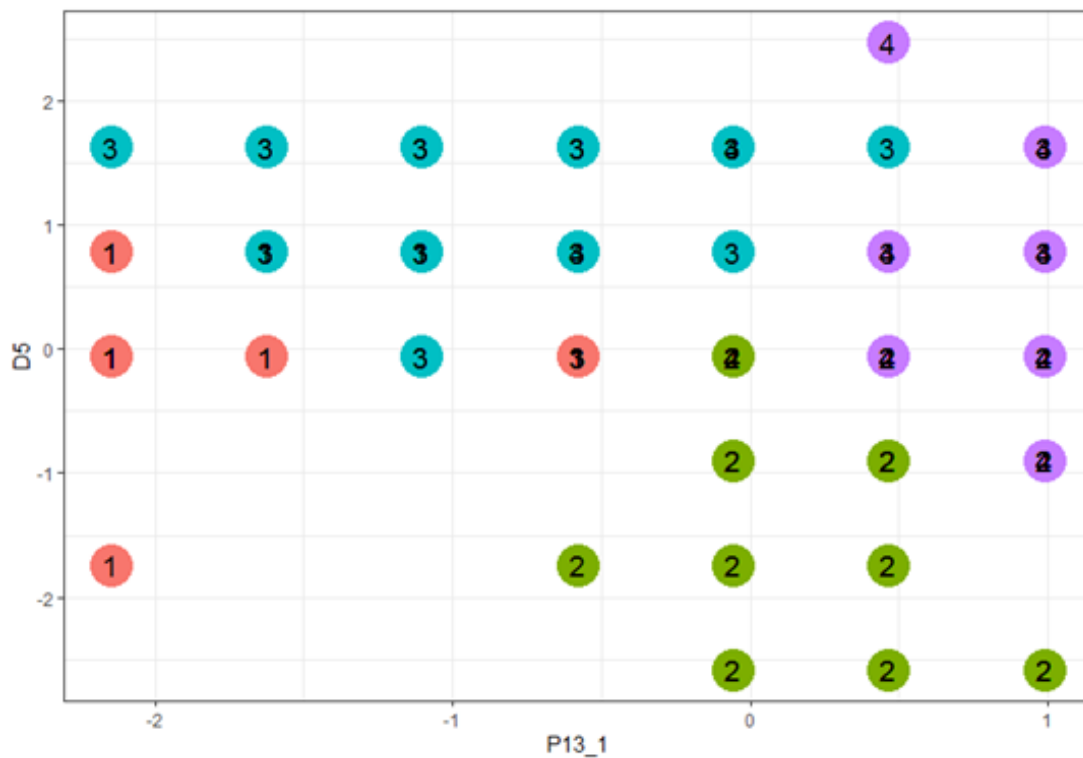
No obstante, es un gráfico muy disperso con datos heterogéneos. Una explicación para esto es que, como se ha dicho con anterioridad, al encontrarse la población en la fase inicial de la pandemia, recién salida de un confinamiento domiciliario y con un número de muertes acumulado de los últimos meses muy elevado, la incertidumbre entre los

jóvenes era real. De esta forma se puede explicar que jóvenes de todas las ideologías sean reticentes a la vacuna.

### 3.2.2. P13\_1 y D5. Reticencia hacia la vacuna y nivel de estudios.

Se han utilizado, de nuevo, 4 *clusters* para este análisis.

**Figura 6. Análisis *clustering* julio 2020 "Reticencia hacia la vacuna y nivel de estudios" (FECYT, 2021).**



En este análisis se pueden encontrar diversos patrones de comportamiento. Desde personas con un nivel de estudios inferior a la media que están a favor de la vacuna (*cluster 2* verde) hasta personas con un nivel de estudios superior a la media que también está a favor de la vacuna (*cluster 4* morado). Por otra parte, el *cluster 3* (azul) representa a aquel grupo de jóvenes que, teniendo un nivel de estudios superior a la media, algunos están en contra de recibir la vacuna y otros están a favor.

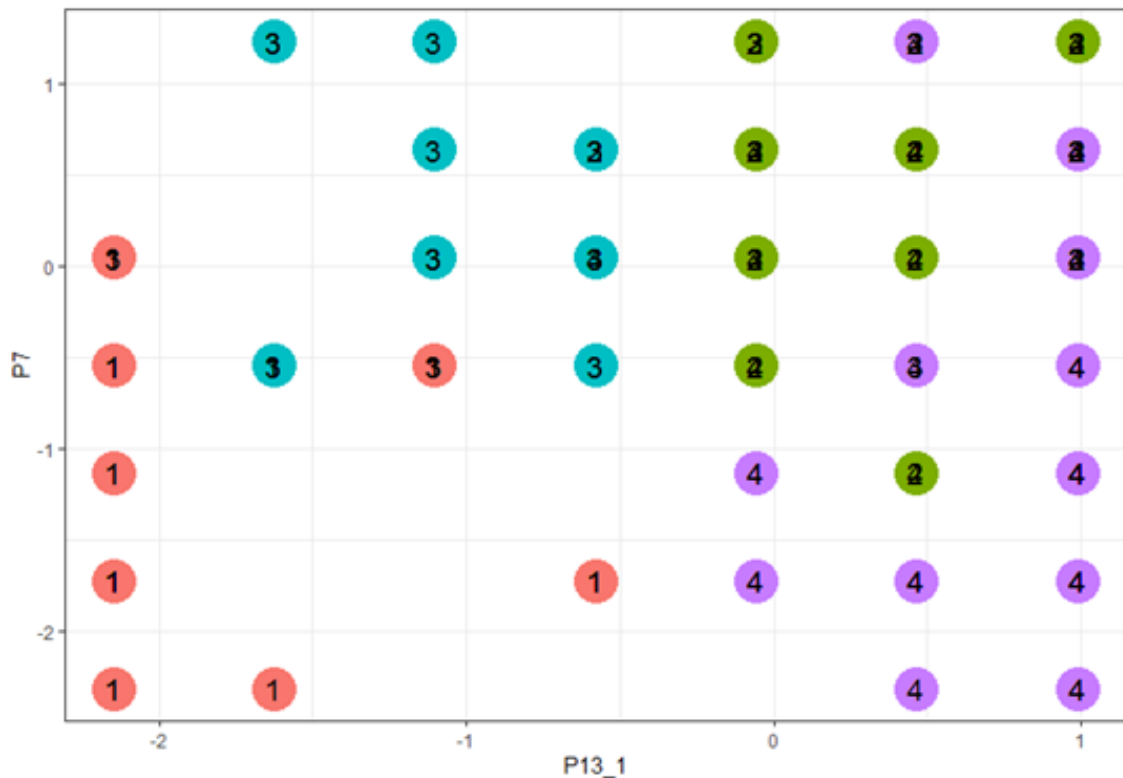
Es interesante recalcar, por último, que apenas existen personas con bajo nivel de estudios que, sin tener la ESO, estén en contra de administrarse la vacuna.



### 3.2.3. P13\_1 y P7. Reticencia a la vacuna y miedo a contagiarse del COVID-19.

Se han vuelto a utilizar 4 centroides en este análisis.

**Figura 7. Análisis *clustering* julio 2020 "Reticencia hacia la vacuna y miedo a contagiarse del COVID-19" (FECYT, 2021).**



El *cluster 1* (rojo), compuesto por personas que están en contra de recibir la vacuna, apenas tiene miedo de contagiarse del COVID-19, cosa que es razonable. En cambio, llama la atención ciertas observaciones del *cluster 3* azul que muestran la existencia de jóvenes que, considerando como “grave” contagiarse del Coronavirus, muestran reticencia hacia la vacuna. Esto refleja que esos jóvenes claramente no ven a la vacuna contra el Coronavirus como la solución idónea para combatir la pandemia.

En cambio, los *clusters 4* (morado) y *2* (verde), muestran a la gran parte de la población que está a favor de suministrarse la vacuna en cuanto antes, pero, respectivamente, los primeros consideran que no es grave del todo contagiarse mientras que los segundos sí que lo piensan.

### 3.2.4. Regresión lineal múltiple.

Por otra parte, para poder llegar a comprender de una manera más completa en qué medida puede afectar cada variable a la variable P13\_1 (reticencia hacia la vacunación), se ha llevado a cabo una regresión lineal múltiple por medio de la herramienta de análisis de datos RStudio (Anderson, 2019).

La misma base de datos con la que se ha trabajado el clustering se ha utilizado en este apartado.

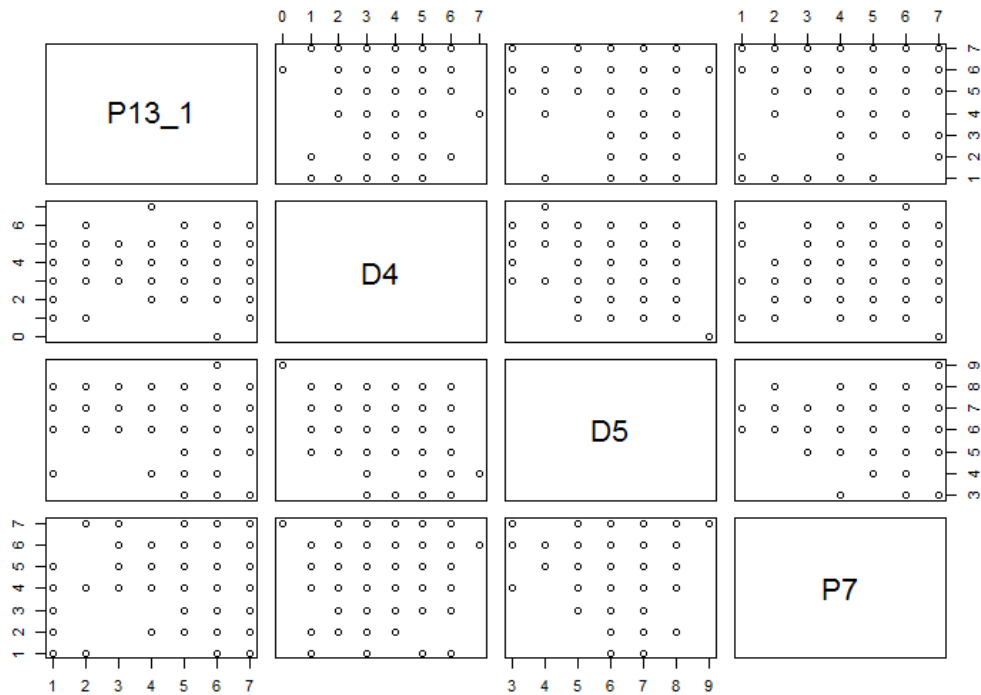
Primero se han obtenido la matriz de correlación de las variables y la matriz de dispersión, mostrándose a continuación:

**Figura 8. Matriz de correlación de las variables escogidas en la encuesta julio 2020.**

|       | P13_1   | D4      | D5      | P7      |
|-------|---------|---------|---------|---------|
| P13_1 | 1.0000  | -0.0031 | -0.2527 | 0.2928  |
| D4    | -0.0031 | 1.0000  | -0.0640 | 0.1660  |
| D5    | -0.2527 | -0.0640 | 1.0000  | -0.1485 |
| P7    | 0.2928  | 0.1660  | -0.1485 | 1.0000  |

Los coeficientes de correlación mostrados encuentran su valor entre -1 y 1. Cuanto más cercano a 1, significa que ambas variables crecen de manera directamente proporcional, en cambio, si es al -1, cuando una crece la otra decrece y viceversa. Cuanto más cercanas a 0, significa que apenas entre ambas variables hay correlación. En este caso, se aprecia que las correlaciones no son excesivamente “fuertes”. Hay dos relaciones lineales positivas: la primera entre el miedo a contraer el COVID-19 y la reticencia hacia la vacuna y la segunda entre el miedo a contraer el COVID-19 y la ideología política. El resto de las relaciones son negativas (Data Science Team, 2020).

**Figura 9. Matriz de dispersión de las variables escogidas en la encuesta julio 2020.**



Se puede observar una alta dispersión entre las variables. A continuación, se llevará a cabo la regresión lineal múltiple. Se expondrán los resultados obtenidos y se comentará su significado.

**Figura 10. Resultados de la regresión lineal múltiple de la encuesta julio 2020.**

```
Call:
lm(formula = P13_1 ~ D4 + D5 + P7, data = encuesta_julio2020_limpia)

Residuals:
    Min       1Q   Median       3Q      Max
-4.7334 -0.9660  0.2152  1.3781  3.5349

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.06954    0.91037   6.667 2.95e-10 ***
D4           -0.09417    0.10579  -0.890 0.374530
D5           -0.34847    0.11177  -3.118 0.002116 **
P7            0.30572    0.07927   3.857 0.000159 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.795 on 184 degrees of freedom
Multiple R-squared:  0.1342,    Adjusted R-squared:  0.1201
F-statistic: 9.509 on 3 and 184 DF,  p-value: 7.175e-06
```

El p-valor tiene un valor inferior a 0,05, lo que significa que los cambios en la variable P13\_1 encuentran relación (aunque sea un mínimo) con las variables predictoras de este modelo (Minitab, 2015). Por otra parte, el estadístico F, cuya cifra es 9,509, lo que indica que los coeficientes no son altamente significativos en este modelo de regresión. Además, el R<sup>2</sup>, que oscila entre 0 y 1 y cuando más cercano a 1 más explicativas son las variables independientes, no es muy alto al ser 0,1342 (Verdaguer, 2021).

La ecuación de regresión obtenida es la siguiente:

---


$$Y = 6,06 - 0,094 * D4 - 0,35 * D5 + 0,30 * P7 + e$$


---

6,06 es el valor que tendría la variable Y si el valor de las variables independientes fuese 0. -0,094 es el coeficiente de regresión de la variable D4, -0,35 el de la variable D5 (el más alto en valor absoluto, por lo tanto, el que más influye sobre la variable respuesta) y 0,30 el de la variable P7.

La *e* representa el residuo, es decir, la diferencia entre el valor observado y el valor que se ha obtenido del modelo. Al final del apartado, se van a exponer dos ejemplos en los que se aplica esta ecuación de regresión lineal.

Ahora, se estudiarán los intervalos de confianza de los coeficientes obtenidos, estableciendo el nivel de confianza en 95%:

**Figura 11. Intervalos de confianza en 95% para la encuesta julio 2020.**

|             | 2.5 %      | 97.5 %     |
|-------------|------------|------------|
| (Intercept) | 4.2734430  | 7.8656381  |
| D4          | -0.3028957 | 0.1145471  |
| D5          | -0.5689918 | -0.1279444 |
| P7          | 0.1493317  | 0.4621089  |

Esto quiere decir que, a un nivel de confianza del 95%, el valor del coeficiente de la variable D4 se encuentra entre -0,30 y 0,11, y el modelo practicado lo ha fijado en 0,094. En el caso de la variable D5, el valor del coeficiente se encuentra entre -0,56 y -0,12, y el modelo practicado lo ha fijado en -0,35. En el caso de la variable P7, el valor del coeficiente se encuentra entre 0,149 y 0,46 y el modelo practicado lo ha fijado en 0,30 (Anderson, 2019).

Por último, se van a mostrar dos ejemplos de aplicación de la ecuación de regresión elaborada para la encuesta de julio 2020:

- Observación 1: un joven respondió “4” en la pregunta D4 (ideología política centro-derecha), “6” en D5 (tiene el bachiller) y 2 en P7 (no considera casi nada grave el contraer el Coronavirus). Según el modelo elaborado:

$$6,06 - 0,094*4 - 0,35*6 + 0,30*2 = 4,184$$

El valor de la observación (P13\_1) fue 4 (indiferente ante el estar seguro o no de recibir la vacuna), y el valor obtenido gracias al modelo es 4,184.

El residuo en este caso es de -0,184.

- Observación 2: un joven respondió “5” en la pregunta D4 (de ideología de “derechas”), “6” en la pregunta D5 (con bachillerato) y “7” en la pregunta P7 (con mucho miedo a contagiarse del Coronavirus). Según el modelo elaborado:

$$6,06 - 0,094*5 - 0,35*6 + 0,30*7 = 5,59$$

El valor de la observación (P13\_1) fue 5 (parcialmente seguro de recibir la vacuna), y el valor obtenido gracias al modelo 5,59.

El residuo en este caso es de -0,59.

### **3.2.5. Conclusión**

A partir de los tres análisis realizados por medio de *clustering* se pueden encontrar ciertos patrones de conducta entre los jóvenes en relación con la disposición o no a vacunarse. Por lo general, los grupos encontrados son considerablemente homogéneos entre sí. Sin embargo, se empiezan a distinguir ciertos patrones de comportamiento entre los jóvenes. Esta dispersión se aprecia en el modelo de regresión lineal múltiple ejecutado.

### **3.3. Encuesta enero 2021.**

En este apartado se va a replicar el mismo proceso llevado a cabo en el apartado anterior, pero con los datos obtenidos de la base de datos de la segunda encuesta, realizada en enero de 2021. Para ello, lo primero que se va a hacer es "preparar" el dataset en Excel para poder trabajar con él en RStudio.

Se debe recordar que, en enero de 2021, ya se había comenzado a administrar la vacuna a la población. Además, la curva de fallecidos por COVID-19 drásticamente

decreció en ese mes de vacunación. Esta aparente "esperanza" se pudo ver reflejada en los patrones de conducta de los jóvenes, en concreto en su predisposición a vacunarse.

Sin embargo, se debe recordar que el número de años medio para desarrollar una vacuna que sea completamente novedosa es de, al menos, 4 años (Asociación Española de Vacunología, 2020). Partiendo de que la vacuna contra el COVID-19 se desarrolló en menos de un año desde el inicio de la pandemia, es normal que las personas puedan tener desconfianza en esta vacuna tan novedosa.

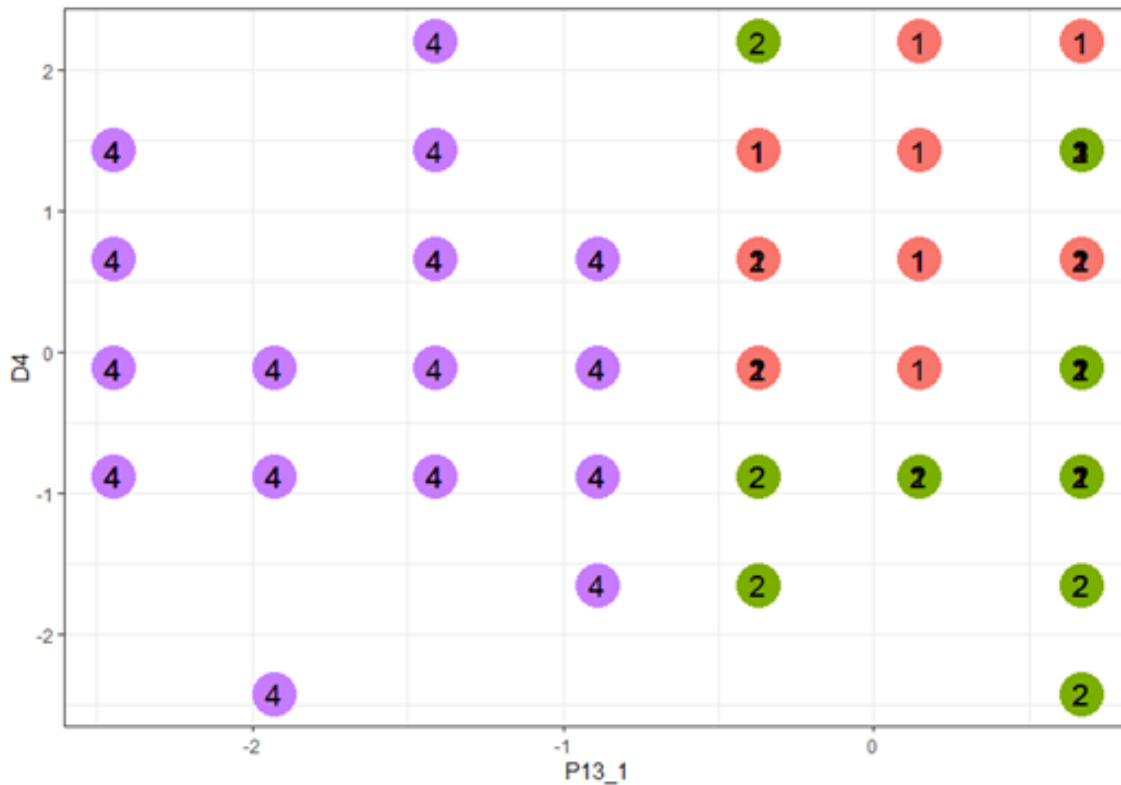
En esta encuesta, como el nivel de contagio había aumentado entre la población durante la tercera ola, la pregunta acerca del "miedo" o conciencia de gravedad frente a contagiarse del COVID-19 cambió. Es decir, primero se preguntó si el individuo entrevistado había superado el Coronavirus o no. Si la respuesta era un "Sí" o "Creo que sí, pero no ha sido confirmado con un test", se procedía a preguntar: "En su caso personal, ¿qué nivel de riesgo considera que tiene Ud. de sufrir complicaciones de salud en el caso de volver a contagiarse con el coronavirus? Por favor, use una escala de 1 a 7, en la que 1 es riesgo bajo de sufrir complicaciones de salud y 7 riesgo alto de sufrir complicaciones de salud". En el caso de que la respuesta a la pregunta primera fuera "No", se procedía a preguntar: "En su caso personal, ¿qué nivel de riesgo considera que tiene Ud. de sufrir complicaciones de salud en el caso de contagiarse con el coronavirus? Por favor, use una escala de 1 a 7, en la que 1 es riesgo bajo y 7 riesgo alto" (FECYT, 2021). A pesar del matiz de haber superado el COVID-19 o no, estas dos preguntas se van a interpretar como una sola en la que se pregunta al individuo acerca de su percepción de la gravedad a la hora de contagiarse o volver a contagiarse del Coronavirus. De una escala del 1 al 7, 1 es poco grave y 7 muy grave. Dicha variable se ha denominado "P7".

Además, otra diferencia de este apartado con el anterior es que aquí, en algunos análisis, se van a asignar diferente número de *clusters* en lugar de 4 con la intención de clasificar de una manera óptima a los jóvenes en función del patrón de comportamiento que siguen en cada análisis.

### 3.3.1. P13\_1 y D4. Retención hacia la vacuna e ideología política.

En este análisis de *clustering* se decidió asignar solo 3 centroides.

Figura 12. Análisis *clustering* enero 2021 "Retención hacia la vacuna e ideología política" (FECYT, 2021).

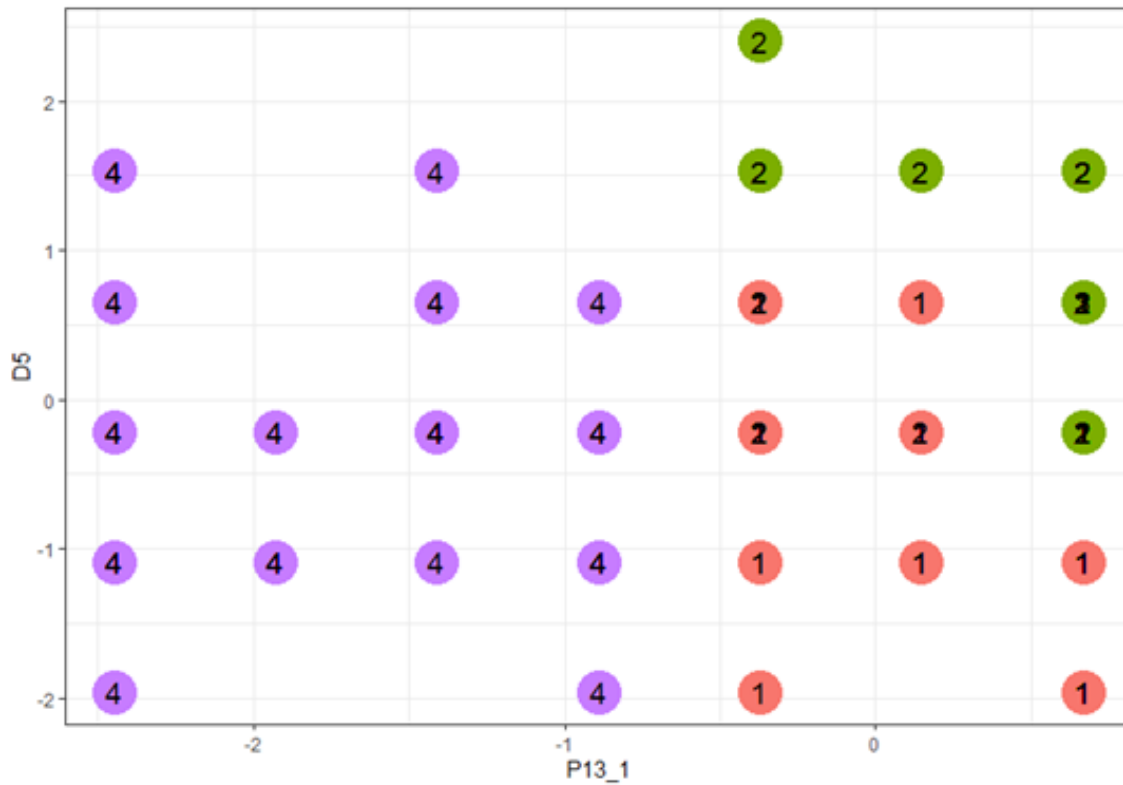


Se pueden diferenciar claramente tres conductas diferentes de los jóvenes. El *cluster* 1 (rojo) representa a un grupo de jóvenes que, siendo de ideología más conservadora (derecha), no están en contra de vacunarse contra el Coronavirus. En cambio, *cluster* 4 (morado) representa a aquel grupo de personas que, abarcando todas las ideologías, pero predominando la de centro, están completamente en contra de ponerse la vacuna. Por otra parte, el *cluster* 2 (verde) es el espejo del *cluster* 1, ya que refleja a aquel grupo de jóvenes que, de ideología progresista, están a favor de vacunarse en cuanto antes contra el COVID-19.

### 3.3.2. P13\_1 y D5. Reticencia hacia la vacuna y nivel de estudios.

Se volvió a utilizar la misma cantidad de *clusters* que en el anterior.

Figura 13. Análisis *clustering* enero 2021 "Reticencia hacia la vacuna y nivel de estudios" (FECYT, 2021).



Hay personas con todo tipo de nivel de estudios que están completamente en contra de la vacunación (*cluster* 4 morado). Es curioso que el espacio de jóvenes con un nivel de estudios inferior a la media y contrarios a la vacuna contra el COVID-19 se ha ocupado en comparación a julio 2020.

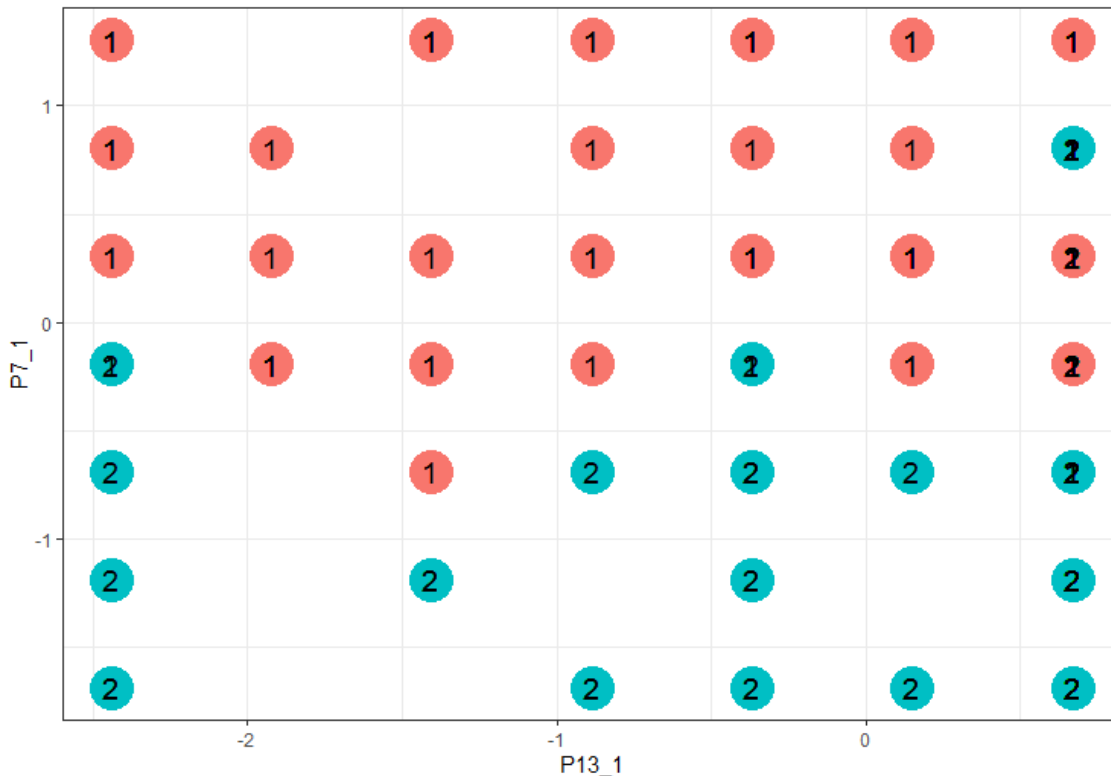
En cambio, el *cluster* 2 (verde) refleja la parte de un grupo de jóvenes que, teniendo un nivel de estudios superior a la media, están a favor de recibir la vacuna contra el Coronavirus.



### 3.3.3 P13\_1 y P7. Reticencia hacia la vacuna y miedo a contagiarse del COVID-19.

En este apartado se han utilizado tan solo dos *clusters*.

**Figura 14. Análisis *clustering* enero 2021 "Reticencia hacia la vacuna y el miedo a contagiarse de COVID-19" (FECYT, 2021).**



Es interesante en este gráfico que hay una clara preponderancia de los jóvenes que consideran grave padecer de COVID-19 (*cluster* 1 rojo) sobre los jóvenes que no (*cluster* 2 azul). Además, hay jóvenes de todo tipo (en cuanto al miedo a padecer de Coronavirus) que están a favor de la vacuna.

### 3.3.4. Regresión lineal múltiple.

Al igual que en la regresión practicada en el apartado anterior, se ha utilizado la misma base de datos con la que se ha trabajado el *clustering* en este apartado.

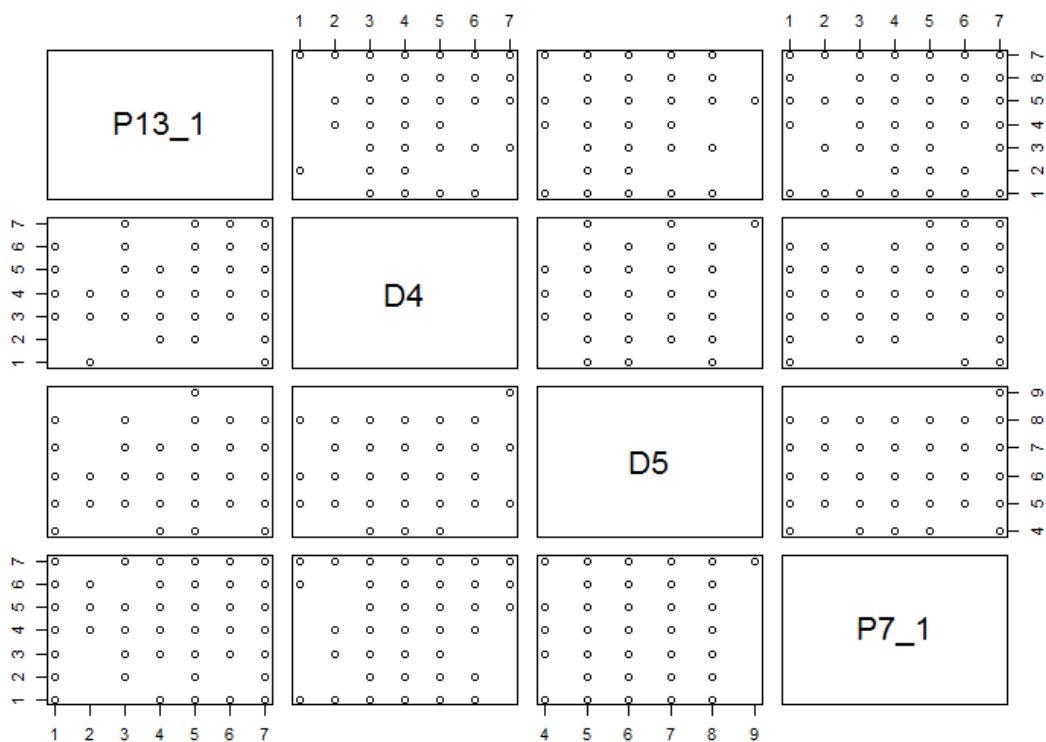
Primero se han obtenido la matriz de correlación de las variables y la matriz de dispersión, mostrándose a continuación:

**Figura 15. Matriz de correlación de las variables escogidas en la encuesta enero 2021.**

|       | P13_1   | D4      | D5      | P7_1    |
|-------|---------|---------|---------|---------|
| P13_1 | 1.0000  | -0.0478 | 0.0782  | -0.0602 |
| D4    | -0.0478 | 1.0000  | -0.0430 | 0.0766  |
| D5    | 0.0782  | -0.0430 | 1.0000  | -0.2309 |
| P7_1  | -0.0602 | 0.0766  | -0.2309 | 1.0000  |

Aunque es una relación débil, es curioso que la relación entre la variable D5 (nivel de estudios) y la reticencia hacia la vacunación se ha tornado de negativa a positiva en comparación a la encuesta anterior. Además, la relación entre el miedo a contagiarse del COVID-19 y la reticencia hacia la vacuna ahora son inversamente proporcionales. Es evidente que la dispersión sigue presente, ante la mayoría de coeficientes cercanos a 0.

**Figura 16. Matriz de dispersión de las variables escogidas en la encuesta enero 2021.**



Se puede volver a observar una alta dispersión entre las variables. A continuación, se realizará la regresión lineal múltiple. Se expondrán los resultados obtenidos y se comentará su significado.

**Figura 17. Resultados de la regresión lineal múltiple de la encuesta enero 2021.**

```
Call:
lm(formula = P13_1 ~ D4 + D5 + P7_1, data = encuesta_enero2021_limpia)

Residuals:
    Min       1Q   Median       3Q      Max
-5.0328 -0.7435  1.0595  1.3092  1.7130

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.44192    1.03837   5.241  4.4e-07 ***
D4            -0.06228    0.11047  -0.564   0.574
D5             0.11221    0.12735   0.881   0.379
P7_1          -0.04000    0.07309  -0.547   0.585
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.935 on 182 degrees of freedom
Multiple R-squared:  0.009725, Adjusted R-squared:  -0.006598
F-statistic: 0.5958 on 3 and 182 DF, p-value: 0.6185
```

El p-valor en este modelo es superior a 0,05, lo que concluye que este modelo no es explicativo, ya que la variable dependiente no encuentra de manera respaldada su variación en los predictores (Minitab, 2015). Por otra parte, el estadístico F es de 0,59, siendo una cifra muy baja, significando esto la poca precisión de este modelo. Esto se justifica además con un R<sup>2</sup> muy cercano a 0 (0,00972) (Verdaguer, 2021).

La ecuación de regresión obtenida es la siguiente:

---


$$Y = 5,44 - 0,06 \cdot D4 - 0,11 \cdot D5 - 0,04 \cdot P7\_1 + e$$


---

5,44 es el valor de la ordenada en el origen, es decir, el valor de Y si todas las X fuesen 0 (Amat, 2016). -0,06 es el coeficiente de regresión de la variable D4, -0,11 el de D5 y -0,04 el de P7. Es curioso que la variable con coeficiente superior es el nivel de estudios. Además, el valor del coeficiente del miedo a contraer el Coronavirus ha descendido notablemente.

Al final del apartado, se van a volver a exponer otros dos ejemplos en los que se aplica esta ecuación de regresión lineal.

Ahora, se estudiarán los intervalos de confianza de los coeficientes obtenidos, estableciendo el nivel de confianza en 95%:

**Figura 18. Intervalos de confianza en 95% para la encuesta enero 2021.**

|             | 2.5 %      | 97.5 %    |
|-------------|------------|-----------|
| (Intercept) | 3.3931225  | 7.4907123 |
| D4          | -0.2802518 | 0.1556887 |
| D5          | -0.1390558 | 0.3634783 |
| P7_1        | -0.1842075 | 0.1042035 |

Esto quiere decir que, a un nivel de confianza del 95%, el valor del coeficiente de la variable D4 se encuentra entre -0,28 y 0,15, y el modelo practicado lo ha fijado en -0,06. En el caso de la variable D5, el valor del coeficiente se encuentra entre -0,13 y 0,36, y el modelo practicado lo ha fijado en -0,11. En el caso de la variable P7, el valor del coeficiente se encuentra entre -0,18 y 0,10 y el modelo practicado lo ha fijado en -0,04.

(Anderson, 2019).

Por último, se van a mostrar dos ejemplos de aplicación de la ecuación de regresión elaborada para la encuesta de enero 2021:

- Observación 1: un joven respondió “5” en la pregunta D4 (), “5” en D5 () y “6” en P7\_1 (). Según el modelo elaborado:

$$5,44 - 0,06*5 - 0,11*5 - 0,04*6 = 4,35.$$

El valor de la observación (P13\_1) fue 4 (indiferente ante el estar seguro o no de recibir la vacuna), y el valor obtenido gracias al modelo es 4,35.

El residuo en este caso resultó ser -0,35.
- Observación 2: un joven respondió “2” en la pregunta D4 (de ideología centro-izquierda), “7” en D5 (con un nivel de estudios de grado universitario de 3 años) y “1” en P7\_1 (considera que es medianamente grave el contagiarse de COVID-19). Según el modelo elaborado:

$$5,44 - 0,06*2 - 0,11*7 - 0,04*1 = 4,51.$$

El valor de la observación (P13\_1) fue 5 (parcialmente a favor de recibir la vacuna), y el valor obtenido gracias al modelo es 4,51.

El residuo en este caso resultó ser 0,49.

### **3.3.5 Conclusión**

En este análisis se puede observar que los patrones de conducta de los jóvenes han evolucionado en los 5 meses de diferencia que separan la primera de la segunda encuesta. No solo han tenido tiempo para madurar su decisión de cara a apostar por la vacuna como solución o no, sino que encima la tercera ola ha caracterizado a los resultados de esta encuesta como tendentes a ser favorables, cada vez más, a la vacuna. La incertidumbre tras la tercera ola ante la decisión de si confiar en la vacuna o no se ha visto reflejada en el modelo de regresión realizado, en el cual las variables se han dispersado en comparación al modelo de la encuesta de julio 2021. Es cierto que el número de fallecidos en la tercera ola ha afectado a la variable que mide el grado de miedo a contraer el COVID-19.

### **3.4 Encuesta mayo 2021**

En este apartado se realizará el último análisis *clustering* de esta investigación. Para ello, se replicará el proceso llevado a cabo en los análisis de las dos encuestas anteriores.

Recordando el contexto de esta encuesta, nos encontramos en el mes de mayo de 2021, en el cual ya se había comenzado a administrar la vacuna contra el Coronavirus desde enero de ese año (Sanidad, 2021).

Antes de trabajar en RStudio con ello, se va a preparar la base de datos en Excel. Para ello, se van a escoger las mismas 4 variables que solo afecten a los jóvenes entrevistados cuya edad se encuentre entre los 18 y los 26 años (inclusive).

Cabe destacar que, en cuanto a la variable relacionada con el miedo a contagiarse o no del COVID-19, hay dos preguntas que se han mezclado en una sola variable. Esto se debe a que, primero, se ha preguntado si se ha padecido de Coronavirus. Si la respuesta es “Sí”, se ha preguntado si se considera grave padecer “otra vez” de los síntomas del COVID-19. En cambio, si la respuesta es “No”, la pregunta es prácticamente igual, pero sin el matiz “otra vez”. En ambos casos, las posibilidades de respuesta vuelven a ser desde el 1 (“Nada grave”) hasta el 7 (“Muy grave”) (FECYT, 2021).

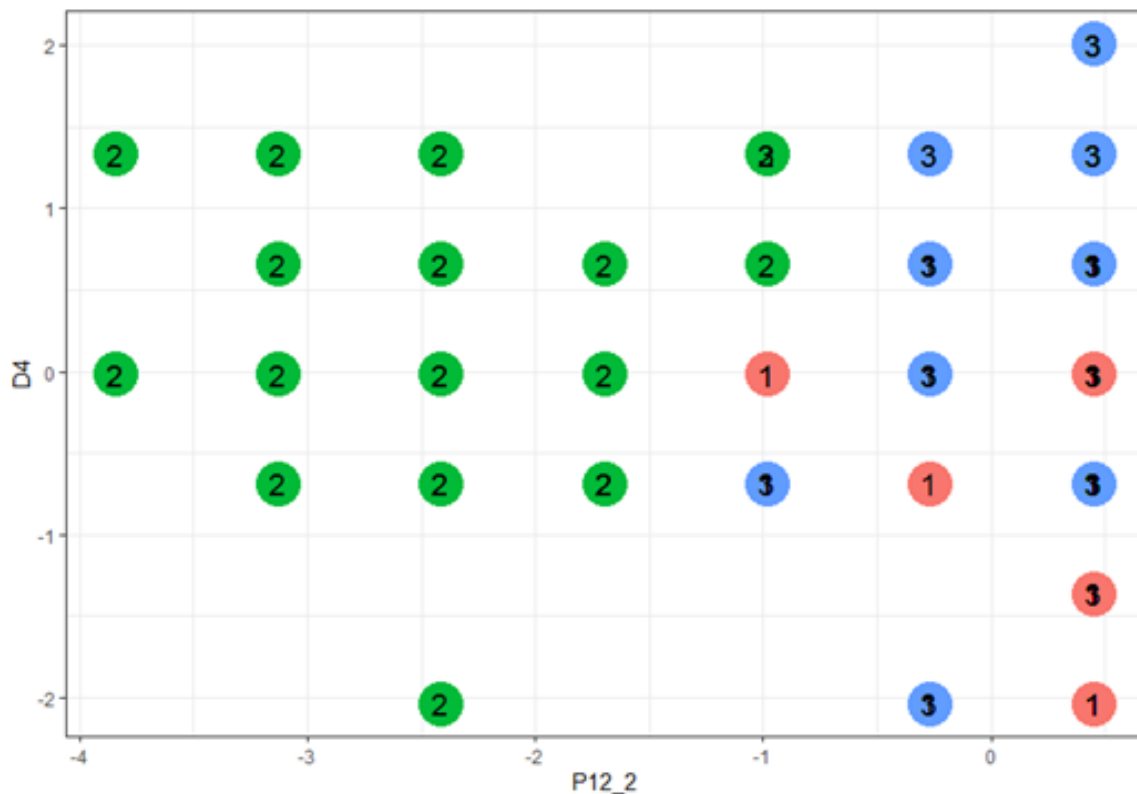
Por otra parte, debido a que el porcentaje de vacunados entre la población comenzó a crecer, antes de preguntar al entrevistado si se pondría la vacuna

inmediatamente o no, se preguntó si se había recibido la vacuna. En el caso en el que la respuesta fuera "Sí", se ha interpretado como un 7 ("Totalmente seguro de recibir la vacuna contra el Coronavirus") para la base de datos de este apartado. El razonamiento que se ha seguido para hacer esto ha sido basado en que "no hay persona más segura de ponerse la vacuna que la que ya se la ha puesto". Por último, los "98" y "99" que equivalen a "No sabe / No contesta" se han suprimido de la base de datos (FECYT, 2021).

### 3.4.1. P12\_2 y D4. Reticencia hacia la vacuna e ideología política.

El análisis *clustering* se ha decidido realizar con tres centroides.

**Figura 19. Análisis *clustering* mayo 2021 "Reticencia hacia la vacuna e ideología política" (FECYT, 2021).**



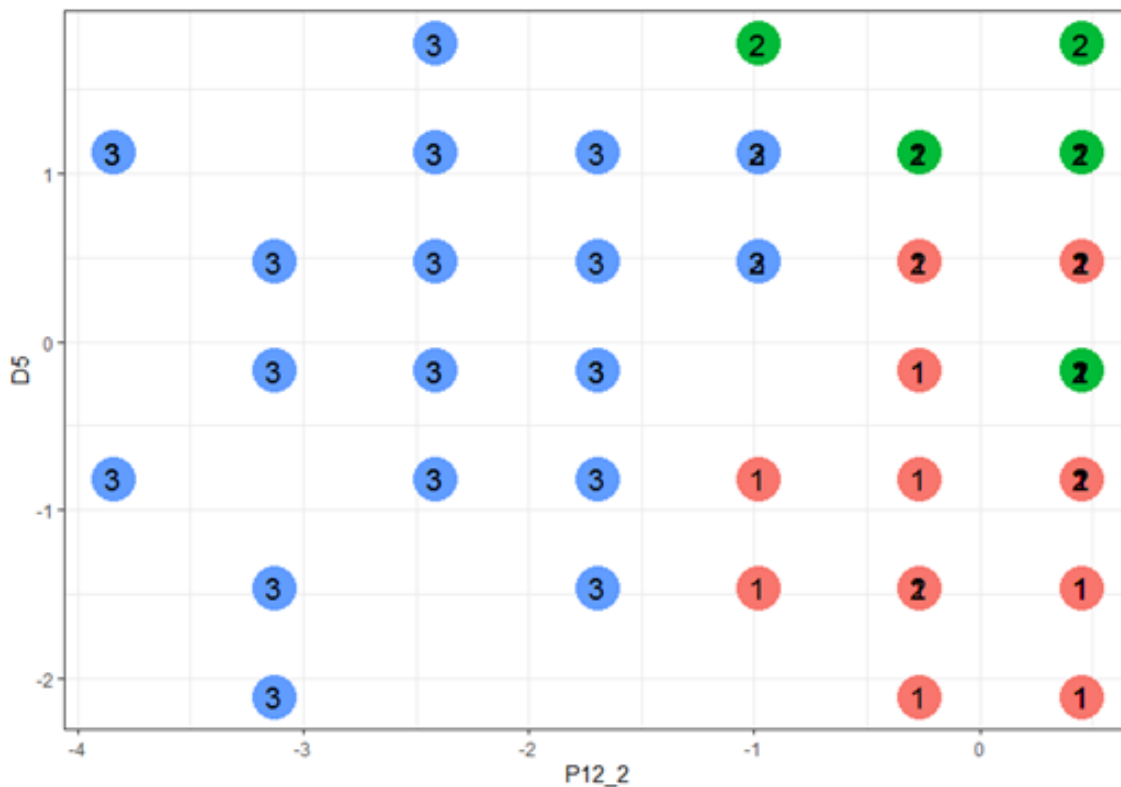
Los *clusters* 3 (azul) y el 1 (rojo) representan a personas que, estando a favor de la vacuna, los primeros representan a los jóvenes de ideología de "derechas" y los segundos de "izquierdas".

Las personas que están en contra de recibir la vacuna son, mayoritariamente, de ideología de centro (aunque también abarcan a jóvenes de izquierda y de "derechas") lo que significa que no hay un claro colectivo joven antivacunas de una ideología marcada.

### 3.4.2. P12\_1 y D5. Reticencia hacia la vacuna y nivel de estudios.

Como el anterior *clustering*, se han asignado tres centros.

**Figura 20. Análisis *clustering* mayo 2021 "Reticencia hacia la vacuna y nivel de estudios" (FECYT, 2021).**



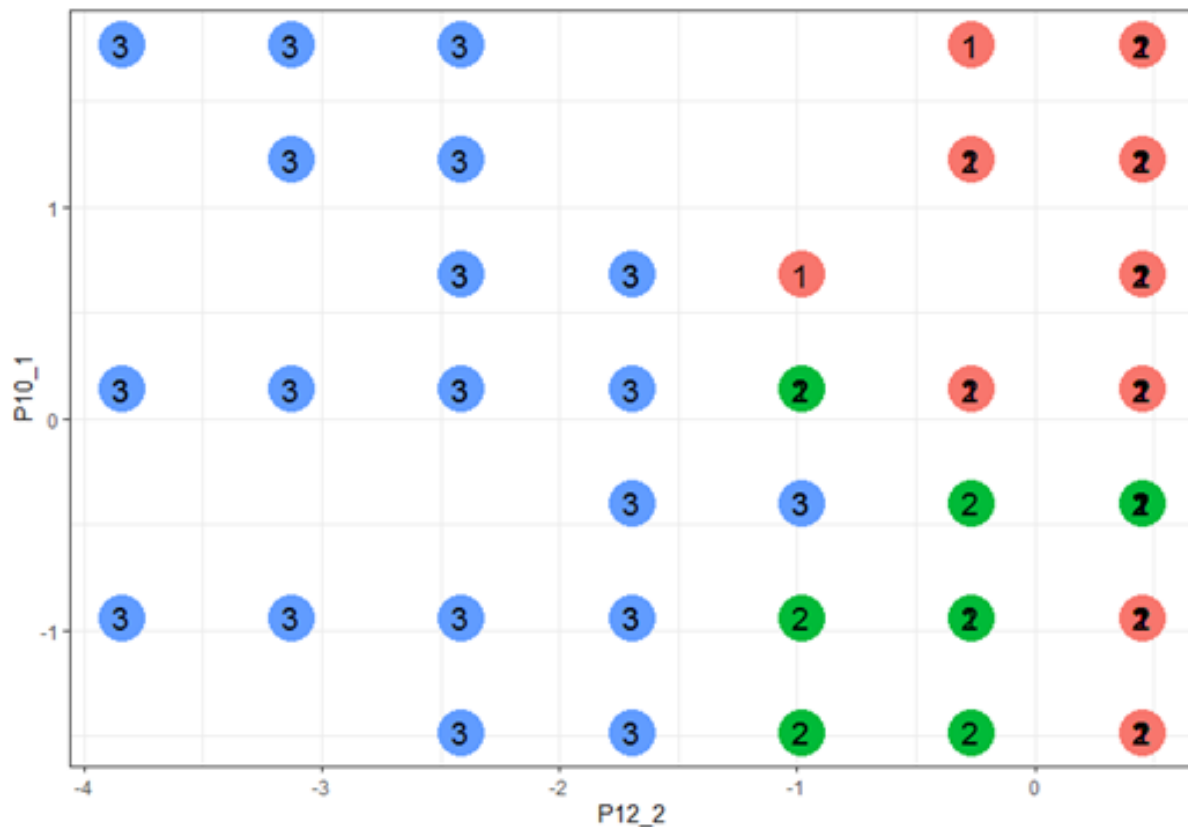
Primero, cabe destacar un mayor número de personas con intención de vacunarse, ya sean las personas jóvenes del *cluster 1* (rojo) con un nivel de estudios inferior a la media (la ESO), como los del *cluster 2* (verde) con estudios universitarios superiores. Esto se explica con el avance de la vacunación en masa a la población.

El *cluster 3* (azul) representa a una gran parte de la población joven muy dispersa, ya que abarca a personas de todo tipo de nivel de estudios. Sin embargo, son personas que no están seguras de recibir la vacuna.

### 3.4.3. P12\_1 y P10\_1. Reticencia hacia la vacuna y miedo a contagiarse de COVID-19.

Por último, se volvieron a utilizar 3 clusters para este análisis.

**Figura 21. Análisis *clustering* mayo 2021 "Reticencia hacia la vacuna y miedo a contagiarse de COVID-19" (FECYT, 2021).**



Es un análisis *clustering* cuyos datos son dispersos, pero se pueden obtener patrones de comportamiento de ellos. Primero, el *cluster 2* (verde) es el núcleo de aquellos entrevistados que no tienen miedo a contagiarse del Coronavirus y que están a favor de la vacuna. En cambio, hay otro grupo de individuos que, teniendo miedo a los efectos de padecer COVID-19, están a favor de la vacuna (*cluster 1* rojo).

No obstante, es curioso el colectivo del *cluster 3* (azul), el cual representa a una amplia parte de los jóvenes entrevistados que están en contra de recibir la vacuna. En dicho grupo, se encuentran desde jóvenes con mucho miedo a padecer de Coronavirus pero que se niegan a vacunarse, hasta jóvenes que no tienen ningún miedo o que mantienen cierta indiferencia ante la gravedad o no de contagiarse.



### 3.4.4. Regresión Lineal Múltiple.

Por último, se volverá a practicar una regresión lineal múltiple utilizando los datos con los que se ha trabajado elaborando el análisis *clustering* en este apartado.

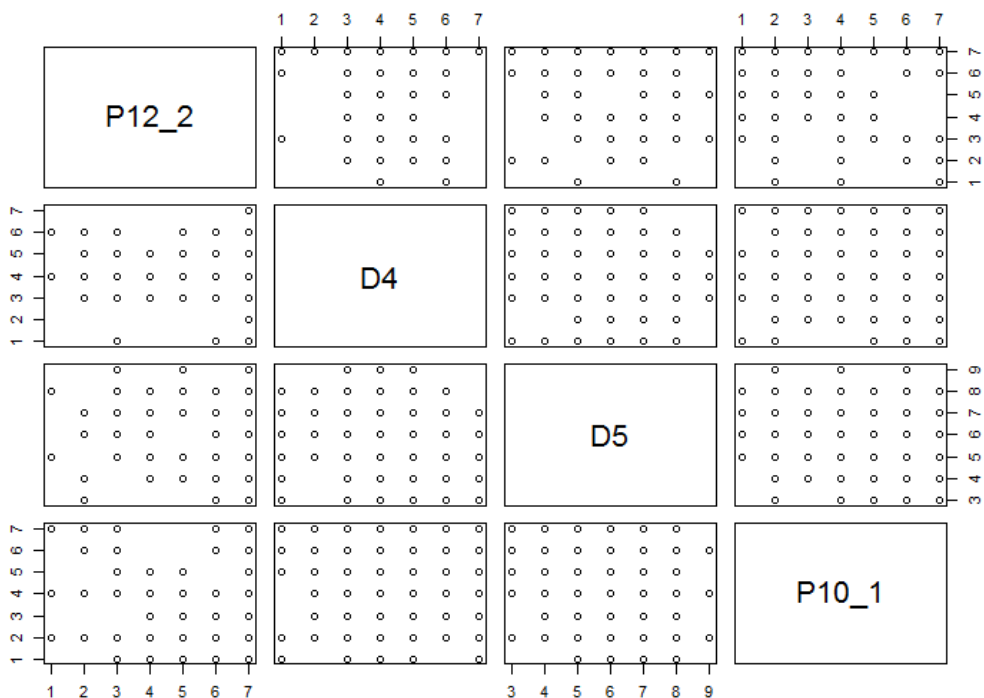
Primero se han obtenido la matriz de correlación de las variables y la matriz de dispersión, mostrándose a continuación:

**Figura 22. Matriz de correlación de las variables escogidas en la encuesta mayo 2021.**

|       | P12_2   | D4      | D5      | P10_1   |
|-------|---------|---------|---------|---------|
| P12_2 | 1.0000  | -0.0105 | -0.0519 | -0.0624 |
| D4    | -0.0105 | 1.0000  | -0.2314 | -0.0069 |
| D5    | -0.0519 | -0.2314 | 1.0000  | -0.2293 |
| P10_1 | -0.0624 | -0.0069 | -0.2293 | 1.0000  |

El hecho de que todos los coeficientes sean negativos significa que todas las relaciones entre todas las variables son inversamente proporcionales. Cabe destacar que la relación con más consistencia es aquella entre el nivel de estudios (D5) y la ideología política (D4).

**Figura 23. Matriz de dispersión de las variables escogidas en la encuesta mayo 2021.**



Se puede volver a observar una alta dispersión entre las variables. A continuación, se llevará a cabo el último modelo de regresión lineal múltiple. Se expondrán los resultados obtenidos y se comentará su significado.

**Figura 24. Resultados de la regresión lineal múltiple de la encuesta mayo 2021.**

```
Call:
lm(formula = P12_2 ~ D4 + D5 + P10_1, data = encuesta_mayo2021_limpia)

Residuals:
    Min       1Q   Median       3Q      Max
-5.5142  0.3939  0.5590  0.6583  0.9982

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.14641    0.63225   11.303  <2e-16 ***
D4           -0.02718    0.06763   -0.402    0.688
D5           -0.06953    0.06661   -1.044    0.298
P10_1        -0.06075    0.05430   -1.119    0.265
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.402 on 204 degrees of freedom
Multiple R-squared:  0.009306, Adjusted R-squared: -0.005263
F-statistic: 0.6387 on 3 and 204 DF, p-value: 0.5909
```

El p-valor es de 0,59, significando esto que, otra vez, el modelo no es explicativo, y que los cambios en la reticencia hacia la vacuna entre los jóvenes no encuentran explicación en el modelo compuesto por su nivel de estudios, ideología política y miedo a contagiarse (Minitab, 2015). El estadístico F es, otra vez, cercano a 0, lo que lleva a pensar que el modelo es poco útil. Por último, el R<sup>2</sup> es muy cercano a 0 otra vez (0,0093), confirmando lo expresado con anterioridad (Verdaguer, 2021).

La ecuación de regresión obtenida es la siguiente:

---


$$Y = 7,14 - 0,03 \cdot D4 - 0,07 \cdot D5 - 0,06 \cdot P10\_1 + e$$


---

7,14 es el valor de la ordenada en el origen. -0,03 el coeficiente de regresión de la variable D4, -0,07 el de la variable D5 y -0,06 el de la variable P10\_1. Otra vez, el coeficiente más alto (en valor absoluto) vuelve a ser el del nivel de estudios.

Al final del apartado, se van a volver a exponer otros dos ejemplos en los que se aplica esta ecuación de regresión lineal.

Ahora, se estudiarán los intervalos de confianza de los coeficientes obtenidos, estableciendo el nivel de confianza en 95%:

**Figura 25. Intervalos de confianza en 95% para la encuesta mayo 2021.**

|             | 2.5 %      | 97.5 %     |
|-------------|------------|------------|
| (Intercept) | 5.8998327  | 8.39298834 |
| D4          | -0.1605163 | 0.10616259 |
| D5          | -0.2008680 | 0.06180664 |
| P10_1       | -0.1678150 | 0.04631024 |

Esto quiere decir que, a un nivel de confianza del 95%, el valor del coeficiente de la variable D4 se encuentra entre -0,16 y 0,10, y el modelo practicado lo ha fijado en -0,03. En el caso de la variable D5, el valor del coeficiente se encuentra entre -0,20 y 0,06, y el modelo practicado lo ha fijado en -0,07. En el caso de la variable P10\_1, el valor del coeficiente se encuentra entre -0,16 y 0,04 y el modelo practicado lo ha fijado en -0,06.

(Anderson, 2019).

Por último, se van a mostrar dos ejemplos de aplicación de la ecuación de regresión elaborada para la encuesta de mayo 2021:

- Observación 1: un joven respondió “2” en la pregunta D4 (ideología prácticamente de extrema “izquierda”), “6” en D5 (con bachillerato) y “2” en P10\_1 (apenas considera grave contagiarse del Coronavirus). Según el modelo elaborado:

$$7,14 - 0,03*2 - 0,07*6 - 0,06*2 = 6,54.$$

El valor de la observación (P12\_2) fue 7 (totalmente a favor de la vacuna), y el valor obtenido gracias al modelo 6,54.

El residuo en este caso resultó ser 0,54.

- Observación 2: un joven respondió “3” en la pregunta D4 (de ideología centro-izquierda), “7” en D5 (con un nivel de estudios de grado universitario de 3 años) y “4” en P10\_1 (considera que es medianamente grave el contagiarse de COVID-19). Según el modelo elaborado

$$7,14 - 0,03*3 - 0,07*7 - 0,06*4 = 6,32.$$

El valor de la observación (P12\_2) fue 5 (parcialmente a favor de la vacuna), y el valor obtenido gracias al modelo 6,32.

El residuo en este caso resultó ser -1,34.

### **3.4.5 Conclusión.**

Se observa la influencia de los 4 meses transcurridos de vacunación en masa posicionando cada vez más a los jóvenes a estar a favor de ella. Sin embargo, también se aprecia cierta incertidumbre en aquellos grupos que siguen sin estar a favor. De nuevo, el hecho de que el modelo sea tan disperso se debe a que la variabilidad de la reticencia hacia la vacuna entre los jóvenes no encuentra explicación lógica entre un modelo compuesto por estas tres variables independientes.

## **4. RELACIÓN ENTRE LOS DATOS OBTENIDOS CON LA REALIDAD**

Una vez se han llevado a cabo los análisis de *clustering* sobre las tres diferentes encuestas de la FECYT, y se han comentado los diferentes patrones de conducta identificados, se buscará una explicación a la evolución de estos patrones. Dicha explicación se basará en eventos social ocurrido, la implementación de las medidas gubernamental para combatir el COVID-19 en ese momento o la evolución de la tasa de contagios o fallecidos a lo largo de la pandemia.

A continuación, siguiendo el mismo orden que en el análisis *clustering*, se analizará cada patrón de comportamiento de los jóvenes, encuesta a encuesta.

### **4.1 Clustering.**

#### **4.1.1 Ideología política y la reticencia hacia la vacunación.**

En julio de 2020 en España y en el mundo, imperaba una sensación de incertidumbre con el COVID-19. Recién desconfinado el país, se procedió a la primera encuesta de la FECYT. En ella ya se comenzó a preguntar por la vacuna, aspecto que intranquilizó a parte de la población. En la primera encuesta se ve reflejado que, en cuanto a los jóvenes de ideología de "derechas", había tanto pro-vacunas como anti-vacunas. Curiosamente, fueron jóvenes de ideología de "izquierdas" los que más en contra estuvieron de la vacunación. Esto es interesante ya que, la confianza o no en la vacunación deriva indirectamente en la confianza en el Gobierno, ya que son los principales administradores de la vacuna. Dado que el Gobierno es de ideología más cercana al

número 1 que al 7 en la escala de la variable D4, lo normal sería que fueran los jóvenes que no comulguen con sus ideas los que estuvieran en contra de ello. Sin embargo, también se debe recordar que había un grupo de jóvenes de ideología de extrema izquierda proclives a la vacuna. Es decir, la incertidumbre del inicio de la pandemia se reflejó en la heterogeneidad de los resultados obtenidos.

Más adelante, en octubre de 2020 se implantó el toque de queda, prohibiendo estar fuera del domicilio desde prácticamente media noche hasta la madrugada (Hernández y de la Rosa, 2021, p.64). Esta medida afectó mucho socialmente a los jóvenes, los cuales, cuando en enero de 2021 se les volvió a preguntar en la segunda encuesta por la vacuna, las respuestas sufrieron un cambio. Es evidente que el cambio del escenario social provoca cambios a su vez en los patrones de comportamiento. En esta segunda encuesta, los jóvenes conservadores fueron los que más a favor estaban de la vacunación, creciendo respecto a julio 2020. No obstante, también había jóvenes de "derechas" en contra de recibir la vacuna. Este descontento se materializó justo en octubre de 2020 cuando, el segundo partido de la oposición, el cual comulga más con aquellos jóvenes cuya respuesta tiende hacia el 7 y no hacia el 1 en la pregunta de la ideología política, presentó una moción de censura al Gobierno progresista demostrando descontento y "fracaso" del Gobierno a la hora de la gestión de la pandemia (García-Rendueles, 2021). A su vez, había un pequeño grupo dominante de jóvenes de ideología de "izquierdas" que se mantuvo a favor de la vacunación. En cambio, los jóvenes de ideología de centro-izquierda se mantuvieron en su postura antivacunas. Es cierto que en España el proceso de inmunización de la población con doble dosis se retrasó más de la cuenta en comparación con otros países. Entre esto último y que, primero, se administró la vacuna al personal sanitario y a las personas consideradas más vulnerables, hizo que el sector juvenil se sintiera relevado a una posición secundaria, pudiendo provocar así su descontento (Picazo, 2021).

Sin embargo, como se puede observar en el gráfico representado en la introducción, fue a partir del mes de enero después de la tercera ola cuando las muertes por Coronavirus empezaron a descender. Esto significa un aumento de la esperanza en la población, haciendo que se asociara la vacunación con la supervivencia al COVID-19. Esto se reflejó claramente en la encuesta de mayo de 2021. Hasta este mes, el decretado Estado de Alarma mantenía el toque de queda nocturno en algunas regiones del país (FECE, 2021). Sin embargo, entre el descenso de muertos a niveles impensables meses

antes y el crecimiento de la población inmunizada influyó que, cuando se preguntó a los jóvenes, por lo general creciera el colectivo pro-vacunas, independientemente de la ideología. No obstante, se debe recordar que fueron las personas de ideología de centro los más disidentes a vacunarse en esta segunda encuesta.

Por último, en mayo de 2021 cuando se realizó la tercera encuesta, el colectivo juvenil a favor de la vacuna se mantuvo estable. En cambio, fue la sección centro-derecha de los jóvenes los menos partidarios a recibir la vacuna contra el Coronavirus. El descontento con el Gobierno de "izquierdas" es la principal explicación a este motivo. No obstante, se debe destacar una excepción marginal de algún que otro joven que, siendo de ideología de extrema izquierda, estuvo completamente en contra de recibir la vacuna.

Es importante recordar que, como se ha especificado antes, el periodo medio para el desarrollo de una vacuna, antes de 2020, era de unos 4 años (AEV, 2020), de forma que puede ser comprensible que no se confíe a 100% en un Gobierno que administra una vacuna desarrollada en apenas 1 año, independientemente de la ideología de dicho Gobierno.

Por lo tanto, es entendible que, ante la incertidumbre y confusión del principio de la pandemia, se haya evolucionado a una postura común de ambas ideologías a favor de la vacuna, demostrando así que lo lógico no debería tener que estar politizado. No obstante, no se debe olvidar el alto nivel de dispersión de las observaciones, haciendo que la heterogeneidad entre datos pertenecientes al mismo *cluster* sea considerable.

#### **4.1.2 Nivel de estudios y reticencia hacia la vacunación**

Anteriormente se ha comentado repetidas veces que la incertidumbre no ha ayudado mucho a las personas a actuar de una manera homogénea. Esto se ha reflejado en los jóvenes en cuanto a su opinión por la vacuna y su nivel de estudios.

Se debe recordar que hay muchos factores que afectan a ambas variables, ya sea la confianza que se tiene en el Gobierno o incluso el compromiso social de la población a la hora de estar al día con las nuevas noticias del Coronavirus. Un dato interesante es saber cuántos jóvenes leen el periódico cada día, síntoma de tener un mínimo de curiosidad por la actualidad del momento, iniciativa intelectual y pretensiones de ampliar tu conocimiento en áreas, hasta 2020, desconocidas por la mayoría. A pesar de estar

fechado en 2012, es representativo que, según un estudio de la Universidad de Jaume I de Castellón, solo un 28,8% de los jóvenes entre 16 y 30 años consumen prensa, ya sea escrita u on-line (Casero-Ripollés, 2012, p. 153). Atendiendo a otro estudio de Statista de 2021, el porcentaje de jóvenes entre los 14 y los 34 años que leen diariamente el periódico son 17,1% (Orús, 2022).

En la primera encuesta de la FECYT, se puede observar que, aunque la sección a favor de la vacuna la componen personas cuyo nivel de estudios abarca todos los niveles, solo son personas con niveles superior a la media (haber alcanzado la ESO) los que están en contra de la vacuna. Es decir, es interesante que, ante una pandemia mundial de la que no se sabe apenas nada, las personas sin estudios no se oponen a la solución propuesta por las autoridades. Esto puede ser un síntoma de lo que se ha explicado con anterioridad. La conclusión a la que se puede llegar es que puede ser que una persona con bajo nivel de estudios sea menos crítica y se preocupe menos por la trascendencia de una vacuna cuyo desarrollo ha sido excesivamente rápido (ONUSIDA, 2021).

Más adelante, en la segunda encuesta celebrada en enero de 2021, es importante destacar que tuvo lugar un cambio en la actitud de ciertos grupos de jóvenes. En concreto, mientras que los jóvenes con un nivel superior de estudios se mantuvieron en una conducta a favor de la vacunación contra el Coronavirus, fueron las personas con un nivel de estudios inferior a la media los que adoptaron en esta encuesta posturas en contra de la vacunación. Una explicación que puede tener esto es que, mientras que las personas con nivel superior de estudios pudieron acabar considerando en enero 2021, después de todo, a la vacuna como la única solución para terminar con la pandemia, aquellos con un nivel inferior de estudios desarrollaron su desconfianza hacia el Gobierno y vacuna con posterioridad al resto de jóvenes. No obstante, no se debe olvidar que hubo casos extremos de pequeños grupos de jóvenes con un nivel de estudios superior en contra de la vacuna. Además, se debe recordar que el grupo de personas con un nivel de estudios inferior a la media que en julio de 2021 estuvo a favor de la vacuna, mantuvo su posición pro-vacuna.

Sin embargo, tras el descenso acentuado de fallecidos por Coronavirus a medida que crecía el porcentaje de personas con la pauta completa administrada durante los meses de enero a mayo de 2021, en la tercera encuesta se demostró un crecimiento de los jóvenes de todo tipo de nivel de estudios que estaban a favor de vacunarse. No obstante,

permaneció un grupo de jóvenes que no eran partidarios de la vacuna, aunque el nivel de estudios de ese grupo era muy heterogéneo, abarcando todos los niveles.

Vuelve a demostrarse que, también en el ámbito del nivel de estudios, personas de todos los niveles tienden a acabar, por lo general, adoptando una postura común en favor de la vacunación contra el COVID-19.

Por último, y relacionado con ambas variables (ideología política y nivel de estudios) se debe mencionar un estudio llevado a cabo por la Universidad Autónoma de Madrid, la Universidad Complutense de Madrid y la Universidad de Queensland (Australia) (2020), en el que se estudia la ideología política de las personas no partidarias de la vacunación, con fecha en 2018. Dicho estudio obtuvo ciertas conclusiones que se expondrán a continuación. Primero, tan solo el 3,3% de las personas entrevistadas confía apenas en la efectividad de la vacunación infantil. Además, se determinaron 6 patrones de conducta ideológica de personas antivacunas. Dichos patrones fueron, de mayor a menor medida:

1. Personas de corriente ideológica progresista y alto nivel educativo.
2. Jóvenes con bajo nivel educativo pero con una ideología extrema, ya sea de “derechas” o “izquierdas”.
3. Personas de bajo nivel educativo e ideológicamente centradas.
4. Personas de nivel medio educativo e ideológicamente centradas.
5. Personas de nivel educativo alto e ideológicamente centradas.

(Citado en UCM, 2020).

Si se comparan los resultados obtenidos por dicho estudio con los obtenidos por la investigación realizada a partir de las encuestas de la FECYT, encontramos similitudes. Principalmente, es cierto, como se ve reflejado en la encuesta de mayo 2021, que las personas con ideología moderada son los principales antivacunas de todos. Además, como se ve reflejado, es cierto que hay personas de todos los niveles educativos contrarios a las vacunas.



#### **4.1.3. Miedo hacia contagiarse de COVID-19 y reticencia hacia la vacunación.**

En la primera encuesta realizada se pudo encontrar colectivos de jóvenes de diversas clases en cuanto al miedo de padecer Coronavirus. Primero, es normal que hubiera jóvenes que, a favor de recibir la vacuna contra el COVID-19 en cuanto antes, consideraban que era tanto mucho como poco grave padecer sus síntomas. Sin embargo, cabe destacar aquel grupo que, siendo completamente contrario a recibir la vacuna, pero, a su vez, consideraban que era “muy grave” contagiarse del COVID-19. Esto demuestra que la confianza por la vacuna, en algunos colectivos, era nula completamente, ya que hasta personas que tenían considerable miedo por contagiarse del Coronavirus no se pondrían en ese momento la vacuna.

Sin embargo, ante la recaída durante la Navidad de 2020-2021, en la que el número de contagiados y fallecidos se incrementó en grandes cifras, la incertidumbre ante la vacuna creció considerablemente, provocando una mayor dispersión de los datos en la encuesta de enero 2021. El hecho de que la vacuna solo se había administrado a los profesionales sanitarios y a personas consideradas vulnerables significaba que los jóvenes iban a ser los últimos en recibirla, de forma que la espera a ello hizo crecer la heterogeneidad de las respuestas, habiéndolas de todo tipo. Había jóvenes tanto contrarios como favorables a la vacuna cuyo miedo a la vacuna oscilaba desde el miedo rotundo al nulo.

No obstante, es evidente que una encuesta realizada en un escenario como fue el de mayo 2021 iba a ser muy diferente a la encuesta de enero 2021. No se puede comparar al escenario español tras la tercera ola, en la que el número de hospitalizaciones y muertos aumentó (Expansión, 2022; Datadista, 2022). A pesar de que la dispersión de los datos es alta, se diferencian patrones de comportamiento en el *clustering*. Primero, hay un gran colectivo de personas que, totalmente a favor de vacunarse contra el Coronavirus, no tienen ningún miedo y piensan que no es grave contagiarse de él. Esto puede suponer una relajación por parte de la población a la hora de contagiarse o no debido al crecimiento exponencial de la administración de la vacuna en España. Por otra parte, sigue habiendo jóvenes que, estando a favor de la vacuna, siguen teniendo miedo de contagiarse. Por último, es sorprendente cómo, después de año y dos meses de pandemia, sigue habiendo personas que, considerando como “muy grave” el padecer COVID-19, son contrarios a la vacunación. Esto significa que la vacuna, incluso habiendo sido demostrado que la tercera

ola fue frenada por el incremento de la vacunación en masa, no había adquirido la confianza del total de la población en mayo de 2021. Esto se debe a que, a pesar de ser de utilidad en el corto plazo, se seguía teniendo miedo de sus “posibles efectos desconocidos en el largo plazo”. Sin embargo, se debe recordar que se ha demostrado que el 90% de los efectos adversos que puede provocar cualquier vacuna aparecen a lo largo del primer mes tras la vacunación (Villar García, 2022, p.81).

#### **4.2 Regresión lineal múltiple.**

Los tres modelos practicados en RStudio han resultado ser poco fructíferos debido a la alta dispersión de los datos estudiados. La principal conclusión que se puede obtener es que, en líneas generales, la variabilidad de la reticencia hacia la vacunación entre los jóvenes no encuentra explicación lógica en un modelo de regresión compuesto por las variables predictoras escogidas para esta investigación.

## 5. CONCLUSIONES

1. Tras el inicio de una pandemia de la magnitud del Coronavirus en marzo de 2020, el ser humano se ha encontrado en una situación sin precedente en la historia reciente, y en ella poder estudiar el comportamiento esperado de ciertos colectivos es relevante (Guterres, 2020).

2. Siendo la vacuna contra el COVID-19 la principal solución de cara a poner fin a la pandemia y combatir el número de contagios, es muy importante estudiar, respecto a los jóvenes entre 18 y 26 años, su predisposición o no a recibir la vacuna, y a qué se debe su actitud. Los aspectos que se han escogido que se consideran los más influyentes con la decisión de los jóvenes han sido la ideología política, el nivel de estudios y el tener miedo o no a contagiarse del Coronavirus.

3. El hecho de tener a disposición herramientas de Machine Learning que permiten procesar grandes volúmenes de datos es una ventaja a la hora de llevar a cabo técnicas estadísticas de modelaje y análisis de macrodatos, como puede ser *clustering* o regresión lineal múltiple.

4. Debido a la incertidumbre generada entre la población, la dispersión de los datos es muy considerable a la hora de analizarlos y estudiarlos por medio de la técnica de *clustering* y regresión lineal múltiple en RStudio.

5. A pesar de haber cierta predisposición a lo largo de la pandemia reflejada en las tres encuestas hacia la vacunación y pérdida de miedo al COVID-19, sigue habiendo colectivos de la población joven que no están del todo de acuerdo con recibir la vacuna contra el Coronavirus. Entre esos colectivos se han encontrado patrones en común relacionados con su ideología o nivel de estudios. Cabe destacar que el hecho de que el Gobierno vigente a lo largo de la pandemia fuera de corte progresista ha influido sobre la decisión de los jóvenes a vacunarse o no. Sin embargo, se ha demostrado que el comportamiento más común adoptado por el joven medio a lo largo de la pandemia ha sido el aumentar su confianza hacia la vacuna contra el COVID-19.

6. Los tres modelos de regresión lineal múltiple han ayudado a concluir que la incertidumbre entre los jóvenes es real debido a la alta dispersión de los datos obtenidos.

## 6. BIBLIOGRAFÍA

- Anderson, Y., (2019). *Regresión lineal múltiple con R*. Github. [https://yuasaavedraco.github.io/Docs/Regresi%C3%B3n lineal m%C3%BAltip e con R.html#introducci%C3%B3n](https://yuasaavedraco.github.io/Docs/Regresi%C3%B3n%20lineal%20m%C3%BAltip%20e%20con%20R.html#introducci%C3%B3n)
- Asociación Española de Vacunología (2020). *¿Cuánto tiempo lleva fabricar una vacuna o cómo conseguir lo imposible?* <https://www.vacunas.org/cuanto-tiempo-lleva-fabricar-una-vacuna-o-como-conseguir-lo-imposible/#:~:text=El%20record%20en%20desarrollar%20una,o%20la%20econ om%C3%ADa%20puede%20tolerar>
- Amat, J., (2016). *Introducción a la Regresión Lineal Múltiple*. Ciencia de datos. [https://www.cienciadedatos.net/documentos/25\\_regresion\\_lineal\\_multiple](https://www.cienciadedatos.net/documentos/25_regresion_lineal_multiple)
- Amesquita, D., (2022) *¿Cómo normalizar datos entre 0 y 1?* Statologos. <https://statologos.com/normalizar-datos-entre-0-y-1/>
- Aprende Machine Learning (2020). *Aprendizaje por refuerzo*. <https://www.aprendemachinelearning.com/aprendizaje-por-refuerzo/>
- Arroyo, J (2021). *La segunda ola del Covid-19 deja un exceso de mortalidad de 26.186 casos*. Redacción médica. <https://www.redaccionmedica.com/secciones/sanidad-hoy/segunda-ola-covid-19-deja-exceso-mortalidad-casos-3517>
- BBC News Mundo (2022). *Covid: por qué la OMS dice que ómicron puede significar el fin de la pandemia en Europa*. BBC. <https://www.bbc.com/mundo/noticias-internacional-60111876>
- Carrasquilla-Batista, A., Chacón-Rodríguez, A., Núñez-Motero, K., Gómez-Espinoza, O., Valverde, J. y Guerrero-Barrantes, M., (2016). *Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal*. Tecnología en Marcha. [https://www.scielo.sa.cr/scielo.php?script=sci\\_arttext&pid=S0379-39822016000900033](https://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S0379-39822016000900033)
- Casero-Ripollés, A., (2012). *Beyond Newspapers: News Consumption among Young People in the Digital Era*. Scipedia, p. 153. [https://www.scipedia.com/wd/images/f/f2/Draft\\_Content\\_470717419-26690.pdf](https://www.scipedia.com/wd/images/f/f2/Draft_Content_470717419-26690.pdf)
- COVID-19 en España. (2022, abril). En *Google Noticias*. <https://news.google.com/covid19/map?hl=es&mid=%2Fm%2F06mkj>

- COVID-19 en Todo el mundo. (2022, abril). En *Google Noticias*. <https://news.google.com/covid19/map?hl=es>
- DATADISTA (2022). *Evolución de la vacunación contra la COVID-19 en España*. <https://www.datadista.com/coronavirus/evolucion-de-la-vacunacion-en-espana/>
- Data Science Team (2020). *¿Qué es una Matriz de Correlación?* <https://datascience.eu/es/matematica-y-estadistica/que-es-una-matriz-de-correlacion/>
- Expansión (2022). *España - COVID-19 - Crisis del coronavirus*. Datosmacro. <https://datosmacro.expansion.com/otros/coronavirus/espana>
- Fagella, D. (2020). *What is Machine Learning?* EMERJ. <https://emerj.com/ai-glossary-terms/what-is-machine-learning/>
- FECE (2021). *FIN DE ESTADO DE ALARMA – NUEVAS RESTRICCIONES*. <https://fece.org/blog/2021/05/11/fin-de-estado-de-alarma-nuevas-restricciones/>
- FECYT (2021). *Un 83% de la población española confía en la vacunación contra la covid, 25 puntos más que en enero*. Ministerio de Ciencia e Innovación. <https://www.fecyt.es/es/noticia/un-83-de-la-poblacion-espanola-confia-en-la-vacunacion-contra-la-covid-25-puntos-mas-que-en>
- García-Rendueles, M., (2021). *Mociones de censura en España. Análisis de los discursos de los políticos en el hemiciclo y las portadas de la prensa escrita de la quinta moción (2020)*. Universidad de Valladolid. [https://uvadoc.uva.es/bitstream/handle/10324/51346/TFG\\_F\\_2021\\_106.pdf?sequence=1&isAllowed=y](https://uvadoc.uva.es/bitstream/handle/10324/51346/TFG_F_2021_106.pdf?sequence=1&isAllowed=y)
- Gallardo. *Capítulo 3 Métodos Jerárquicos de Análisis Cluster*. Universidad de Granada. <https://www.ugr.es/~gallardo/pdf/cluster-3.pdf>
- González, A. (2021). *Reportaje en profundidad: ¿Todo bajo control? Los problemas de la vacuna contra la COVID-19*. Universidad de Sevilla. [https://idus.us.es/bitstream/handle/11441/126285/PER\\_GONZALEZLOPEZ\\_TFG.pdf?sequence=1&isAllowed=y](https://idus.us.es/bitstream/handle/11441/126285/PER_GONZALEZLOPEZ_TFG.pdf?sequence=1&isAllowed=y)
- Guterres, A., (2020). *"Esta es, sobre todo, una crisis humana que llama a la solidaridad"*. Naciones Unidas. <https://www.un.org/es/coronavirus/articles/humanitarian-crisis-coronavirus-solidarity>

- Hernández, A., y de la Rosa, M.T. (2021). *EL ESTADO DE ALARMA EN ESPAÑA EN LA PANDEMIA DE LA COVID-19*, p.54. Universidad de La Laguna. <https://riull.ull.es/xmlui/bitstream/handle/915/24153/EL%20ESTADO%20DE%20ALARMA%20EN%20ESPANA%20EN%20LA%20PANDEMIA%20DE%20OLA%20COVID-19.pdf?sequence=1>
- Larrañaga, P., Inza, I. y Moujahid, A. (2012). *Tema 14. Clustering*. Universidad del País Vasco. <https://www.google.com/search?q=universidad+ehu&oq=universidad+ehu&aqs=edg..69i57j69i64.5237j0j1&sourceid=chrome&ie=UTF-8>
- Marín, Y.D., (2018) *Reescalar, normalizar y variables indicadoras*. RPubs. <https://rpubs.com/ydmarinb/429761>
- Martos, F.A., Martín, M., Zambrana, J.L. y Bianchi, J.L. (2021). *REFLEXIONES SOBRE EL FUTURO MODELO DE ATENCIÓN AL FINAL DE LA QUINTA OLA*. Actualidad médica. <https://semst.org/wp-content/uploads/2022/01/Actualidad-Medica-COVID-2021.pdf#page=39>
- Mateos, A. y González, E. (2021). *Tres CCAA, entre las 5 regiones europeas con más muertes en la primera ola*. Redacción médica. <https://www.redaccionmedica.com/secciones/sanidad-hoy/tres-ccaa-entre-las-5-regiones-europeas-con-mas-muertes-en-la-primera-ola-5748>
- Matús-Abásolo, C., Nemeth-Kohanszky, M.E. e Inostroza-Tapia, M.A. (2020). *Atención de Pacientes en Tratamiento de Ortodoncia Durante la Pandemia COVID-19 (SARS-CoV-2). Presentación de un Algoritmo*. Scielo. [https://scielo.conicyt.cl/scielo.php?pid=S0718-381X2020000400489&script=sci\\_arttext](https://scielo.conicyt.cl/scielo.php?pid=S0718-381X2020000400489&script=sci_arttext)
- Minitab (2015). *Cómo interpretar los resultados del análisis de regresión: p-valores y coeficientes*. Addlink. [https://www.addlink.es/noticias/minitab/2540-como-interpretar-los-resultados-del-analisis-de-regresion-p-valores-y-coeficientes#:~:text=Un%20p%2Dvalor%20bajo%20\(%3C,en%20la%20variabl e%20de%20respuesta](https://www.addlink.es/noticias/minitab/2540-como-interpretar-los-resultados-del-analisis-de-regresion-p-valores-y-coeficientes#:~:text=Un%20p%2Dvalor%20bajo%20(%3C,en%20la%20variabl e%20de%20respuesta).
- Nájera, R. (2021). *Así se eliminó la poliomielitis en España*. Gobierno de España. <https://www.vacunacovid.gob.es/voces-expertas/asi-se-elimino-la-poliomielitis-en-espana>

- Orellana-Centeno, J.E. y Guerrero Sotelo, R.N. (2022). *La variante SARS-CoV-2 ómicron*. Revista ADM, p.28. <https://www.medigraphic.com/pdfs/adm/od-2022/od221e.pdf>
- Orús, A., (2022). *Distribución porcentual de individuos que leyeron periódicos de tirada diaria en España en 2021, por grupo de edad*. Statista. <https://es.statista.com/estadisticas/476755/distribucion-de-lectores-de-periodicos-en-espana-por-edad/>
- ONUSIDA (2021). *¿Cómo se ha podido dar tan rápido con una vacuna contra la COVID-19?* <https://www.unaids.org/es/resources/presscentre/featurestories/2021/february/20210209-how-was-a-covid-19-vaccine-found-so-quickly>
- Pérez Abreu, M.R., Gómez Tejada, J.J., y Dieguez Guach, R.A. (2020). *Características clínico-epidemiológicas de la COVID-19*. Revista Habanera de Ciencias Médicas, p.2. <https://www.medigraphic.com/pdfs/revhabciemed/hcm-2020/hcm202e.pdf>
- Phillips, A., (2020). *¿Cuándo nació la idea de la juventud?* La Vanguardia. <https://www.lavanguardia.com/historiayvida/historia-contemporanea/20200823/27078/xxxx.html>
- Picazo, J.J., (2021). *Vacuna frente al COVID-19*. National Library of Medicine. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8638770/>
- Punz, H. (2022). *Austria confirma un nuevo récord de contagios de COVID-19, más de 17.000 en las últimas 24 horas*. Europapress. <https://www.europapress.es/internacional/noticia-austria-confirma-nuevo-record-contagios-covid-19-mas-17000-ultimas-24-horas-20220112122305.html>
- Recuero de los Santos, P. (2021). *Tipos de aprendizaje en Machine Learning: supervisado y no supervisado*. Telefónica Tech. <https://empresas.blogthinkbig.com/que-algoritmo-elegir-en-ml-aprendizaje/>
- RTVE (2021). *Sánchez asegura que la Constitución es "la hoja de ruta del Gobierno progresista"*. <https://www.rtve.es/noticias/20211205/sanchez-constitucion-hoja-ruta-del-gobierno-progresista/2236868.shtml>
- Rueda, J.J., (2020). *Más seguridad, menos libertad*. Heraldo. <https://www.heraldo.es/noticias/opinion/2020/04/04/mas-seguridad-menos-libertad-jose-javier-rueda-la-firma-1367923.html>

- Russo, C., Ramón, H., Alonso, N., Cicerchia, B., Esnaola, L. y Tessore, J.P. (2016). *Tratamiento Masivo de Datos Utilizando Técnicas de Machine Learning*. UNNOBA.  
[https://repositorio.unnoba.edu.ar/xmlui/bitstream/handle/23601/107/1\\_resource.pdf?sequence=1](https://repositorio.unnoba.edu.ar/xmlui/bitstream/handle/23601/107/1_resource.pdf?sequence=1)
- Sanidad (2021). *La campaña de vacunación contra el COVID-19 llega a su segundo mes en España con un total de 1.243.783 personas inmunizadas*. La Moncloa.  
<https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/sanidad14/Paginas/2021/270221-campana-vacunacion.aspx>
- Sanitaria (2022). *¿Cuáles han sido las características de la tercera ola en España?* Redacción médica. <https://www.redaccionmedica.com/recursos-salud/faqs-covid19/cuales-han-sido-las-caracteristicas-de-la-tercera-ola-covid-en-espana>
- Silva, C., Jadue, J., Crivelli, A., y Buccioni, E., (2017). Capítulo 5. Regresión lineal. *Revistas Universidad Santiago de Chile*.  
<https://www.revistas.usach.cl/ojs/index.php/contribuciones/article/view/3026/2750>
- Statologos (2022). *Cómo estandarizar datos en R (con ejemplos)*.  
<https://statologos.com/estandarizar-datos-en-r/>
- TeleSUR (2021). *Covid-19 causa la mayor crisis demográfica en España desde 1941*. <https://www.telesurtv.net/news/covid-muertes-espana-sobrerregistro-madrid-20210617-0010.html>
- Tovar, J. (2021). *Coronavirus en abril en España: control de la cuarta ola y aumento de la vacunación*. Effe: Salud. <https://efesalud.com/coronavirus-abril-espana-contencion-control-cuarta-ola/>
- Universidad Complutense de Madrid (2020). *La polarización ideológica influye más que la educación en la reticencia hacia las vacunas infantiles*. Oficina de Transferencia de Resultados de Investigación (OTRI).  
<https://www.ucm.es/otri/noticias-la-polarizacion-ideologica-influye-mas-que-la-educacion-en-la-reticencia-hacia-las-vacunas-infantiles>
- Verdaguer, F., (2019). *Regresión lineal en Excel: Aprende, calcula, interpreta*. Ninja Excel. <https://www.ninjaexcel.com/formulas-y-funciones-de-excel/regresion-lineal/>

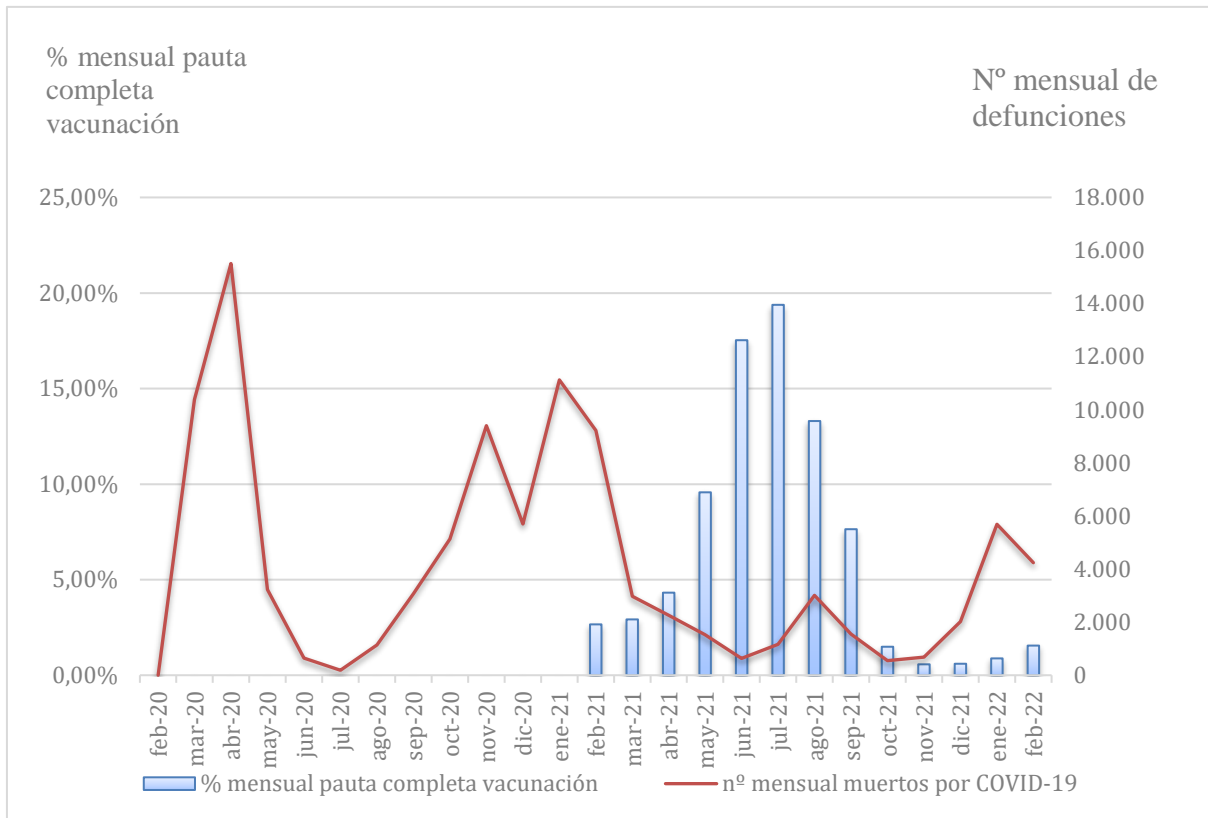


- Vilá-Baños, R., Rubio-Hurtado, M.J., Berlanga-Silvente, V., Torrado-Fonseca, M., (2014). *Cómo aplicar un cluster jerárquico en SPSS*. Revista de Innovación y Búsqueda de Educación. <http://diposit.ub.edu/dspace/bitstream/2445/65577/1/628893.pdf>
- Villardón, V., (2007). *INTRODUCCION AL ANÁLISIS DE CLUSTER*. Universidad de Salamanca. [https://d1wqtxts1xzle7.cloudfront.net/32334872/cluster-with-cover-page-v2.pdf?Expires=1654528966&Signature=hS-Xishyd3tQOoappL3QRzTPpTzSRfJpDBftdpOJH85V0hv67juDIUzCX6poniJ3MJ4ZwIYhts4LGIwx~nzn5lr1y2RT3e2-LMuHLpO8YQe9JOesWjQn0KSB~mEqWEDsnUO3O7yf2eVkeK6WJs33ckCsKVHAVsqny~jgPgnnzfNRxtI9FazKrrfdSc9~QoeQO5R9z~hUXt6mXvh1EOexAbiXxXQNLoXOYa-SPHG3evHmBjZdxAWHPhxm9kxRk0XUsMypXJ4LFkE8xtZJzOXzDfn5RG~5ObBKOIETjA0iXO-~tLhsvAWXkcG7c4F2cIxEGvGAg~6bLLDi2a8s1N8KA\\_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA](https://d1wqtxts1xzle7.cloudfront.net/32334872/cluster-with-cover-page-v2.pdf?Expires=1654528966&Signature=hS-Xishyd3tQOoappL3QRzTPpTzSRfJpDBftdpOJH85V0hv67juDIUzCX6poniJ3MJ4ZwIYhts4LGIwx~nzn5lr1y2RT3e2-LMuHLpO8YQe9JOesWjQn0KSB~mEqWEDsnUO3O7yf2eVkeK6WJs33ckCsKVHAVsqny~jgPgnnzfNRxtI9FazKrrfdSc9~QoeQO5R9z~hUXt6mXvh1EOexAbiXxXQNLoXOYa-SPHG3evHmBjZdxAWHPhxm9kxRk0XUsMypXJ4LFkE8xtZJzOXzDfn5RG~5ObBKOIETjA0iXO-~tLhsvAWXkcG7c4F2cIxEGvGAg~6bLLDi2a8s1N8KA_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA)
- Villar García, J., (2022). *Sobre los bomberos a los que les daba miedo el fuego*. Archivos de prevención, p. 81. <https://archivosdeprevencion.eu/index.php/apr1/article/view/218/126>
- Westreicher, G., (2021). *Variable Cuantitativa*. Economipedia. <https://economipedia.com/definiciones/variable-cuantitativa.html>

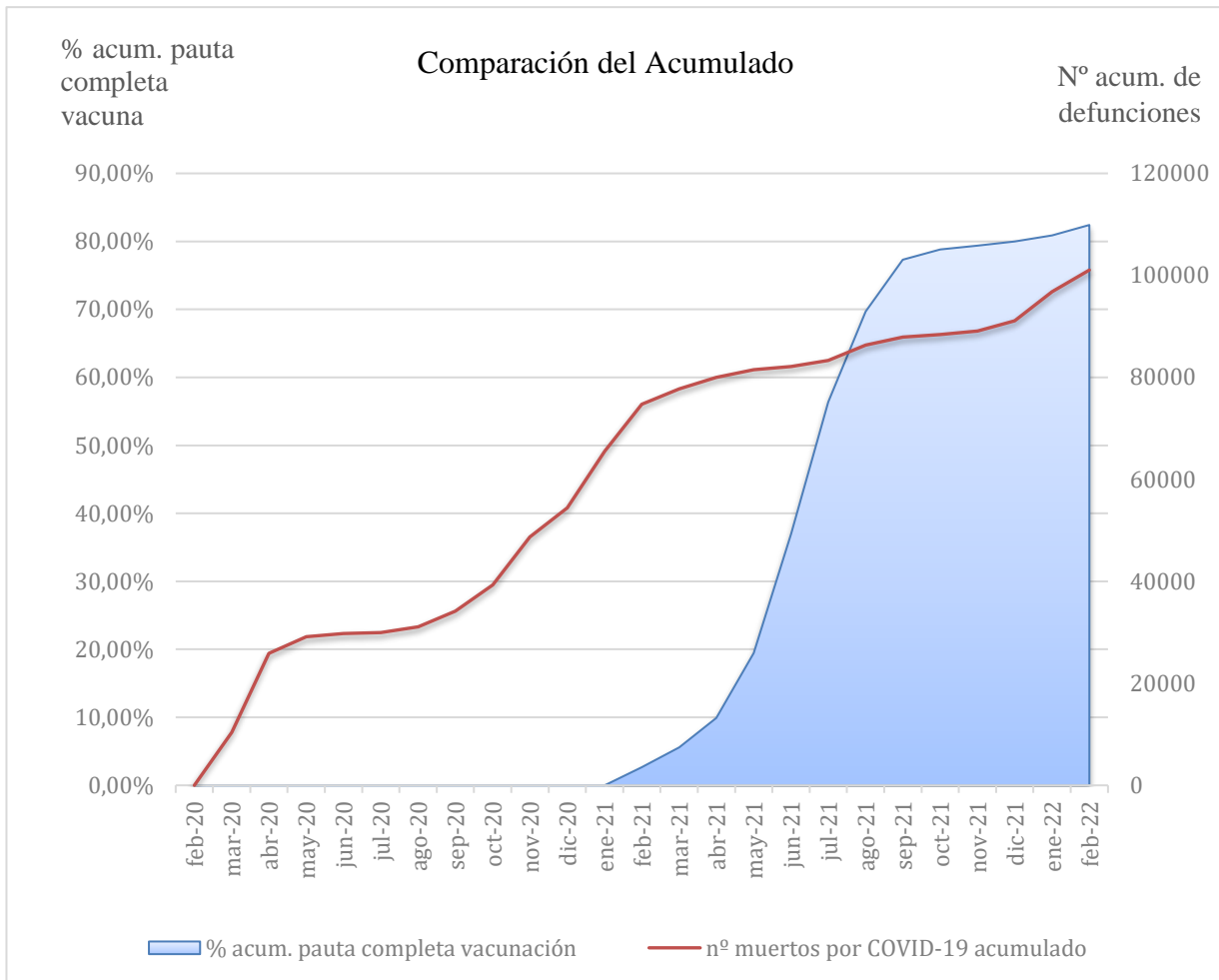
## ANEXO

### 1. Figuras (Gráficos y Tablas a lo largo de la investigación).

#### 1.1. Figura 1. Comparación del número de fallecidos con el porcentaje de vacunación de la población española en cifras mensuales a lo largo de la pandemia (Expansión, 2022; Datadista, 2022).



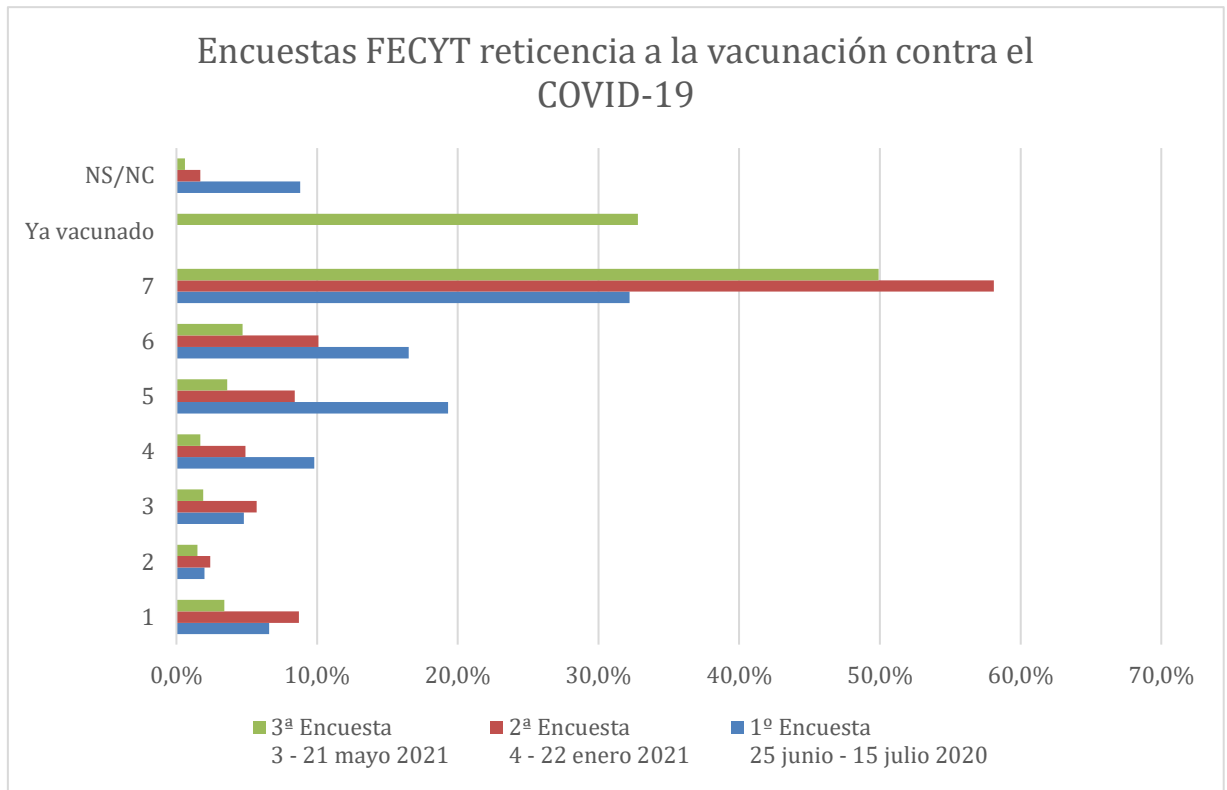
**1.2. Figura 2. Comparación del número de fallecidos con el porcentaje de vacunación de la población española en cifras acumuladas a lo largo de la pandemia (Expansión, 2022; Datadista, 2022).**



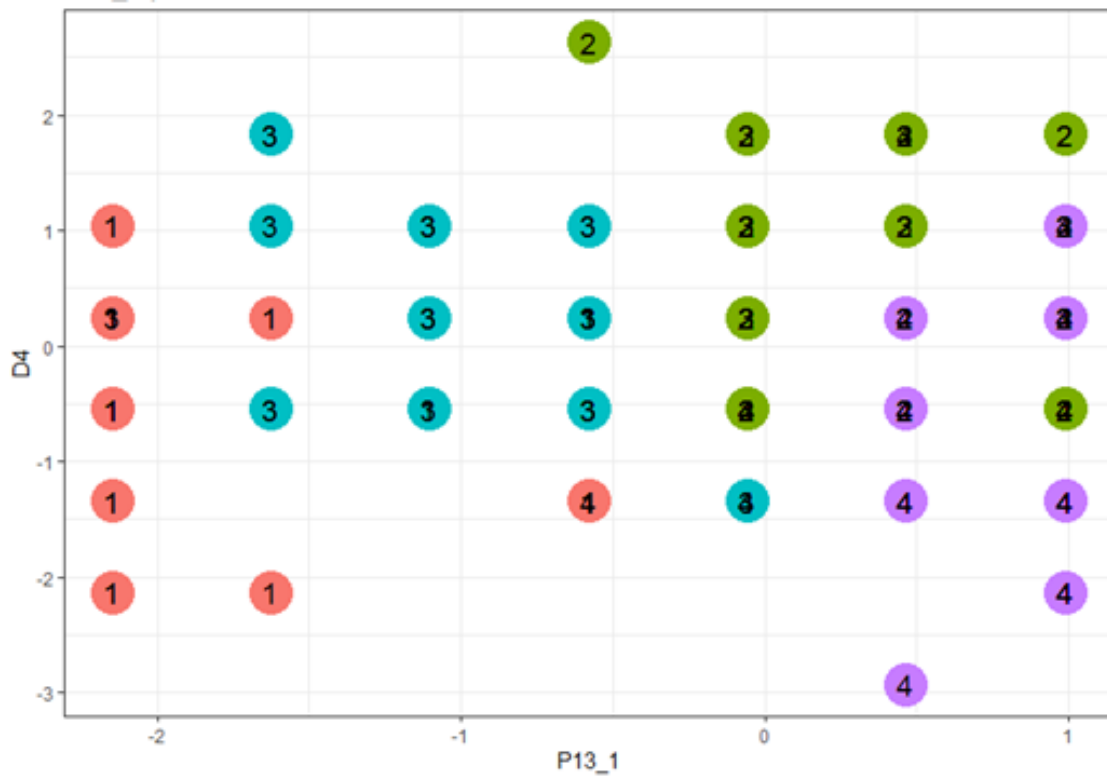
**1.3. Figura 3. Evolución de las respuestas globales en las tres encuestas a la pregunta acerca de la reticencia hacia la vacunación (FECYT, 2021).**

| ESCALA      | 1 <sup>º</sup> Encuesta<br>25 junio - 15 julio 2020 | 2 <sup>a</sup> Encuesta<br>4 - 22 enero 2021 | 3 <sup>a</sup> Encuesta<br>3 - 21 mayo 2021 |
|-------------|---|--|---|
| 1           | 6,6%  | 8,7%   | 3,4%  |
| 2           | 2,0%  | 2,4%   | 1,5%  |
| 3           | 4,8%  | 5,7%   | 1,9%  |
| 4           | 9,8%  | 4,9%   | 1,7%  |
| 5           | 19,3%   | 8,4%   | 3,6%  |
| 6           | 16,5%   | 10,1%  | 4,7%  |
| 7           | 32,2%   | 58,1%  | 49,9%                                       |
| Ya vacunado | 0,0%  | 0,0%   | 32,8%                                       |
| NS/NC       | 8,8%  | 1,7%   | 0,6%  |
| Muestra (N) | 2.100   | 2.108  | 2.100                                       |

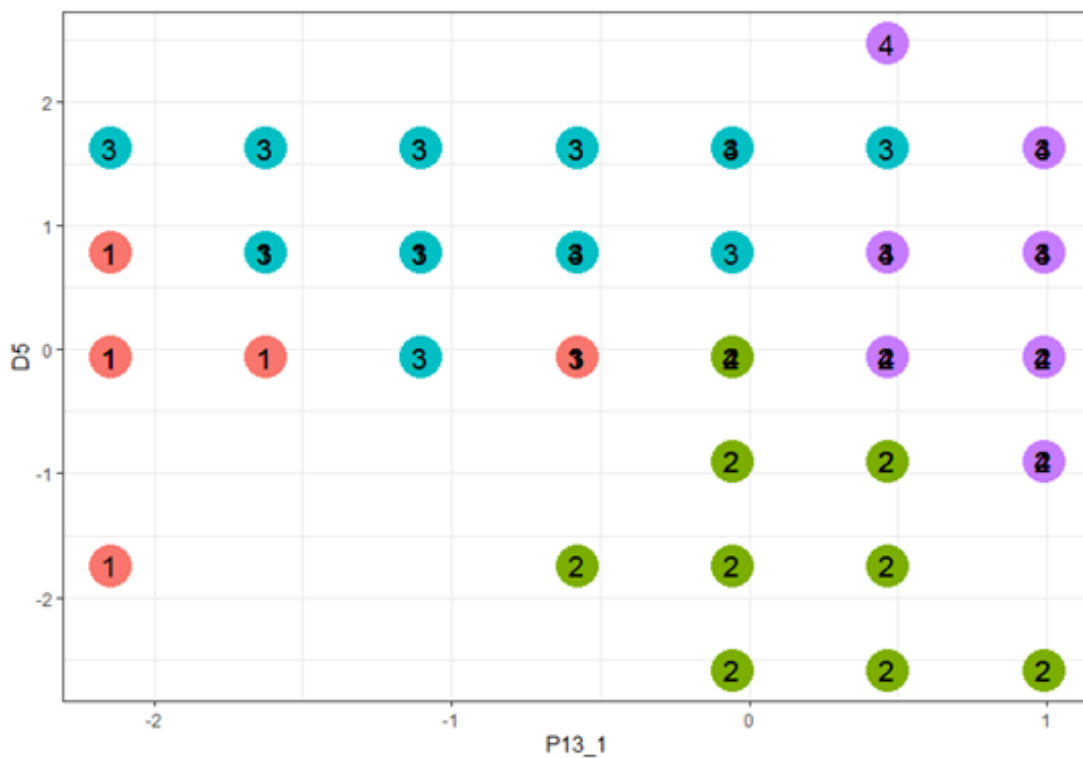
**1.4. Figura 4. Gráfico con las respuestas globales en las tres encuestas a la pregunta acerca de la reticencia hacia la vacunación (FECYT, 2021).**



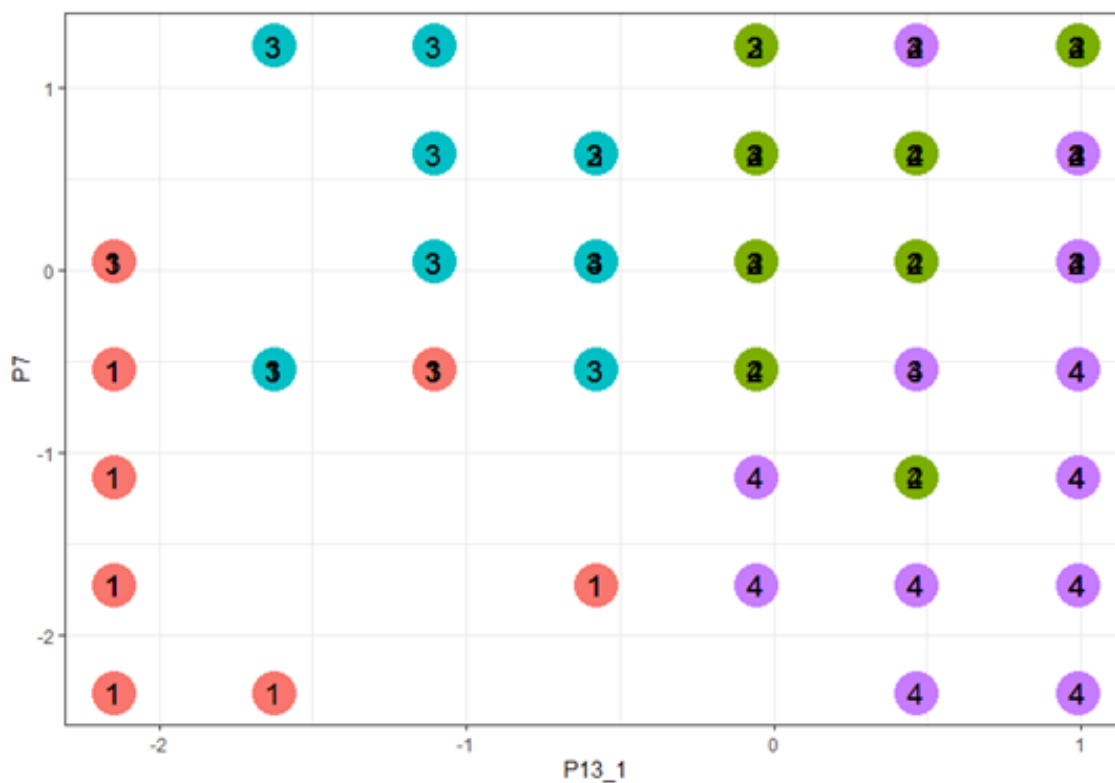
1.5. Figura 5. Análisis *clustering* julio 2020 "Reticencia hacia la vacuna e ideología política" (FECYT, 2021).



1.6 Figura 6. Análisis *clustering* julio 2020 "Reticencia hacia la vacuna y nivel de estudios" (FECYT, 2021).



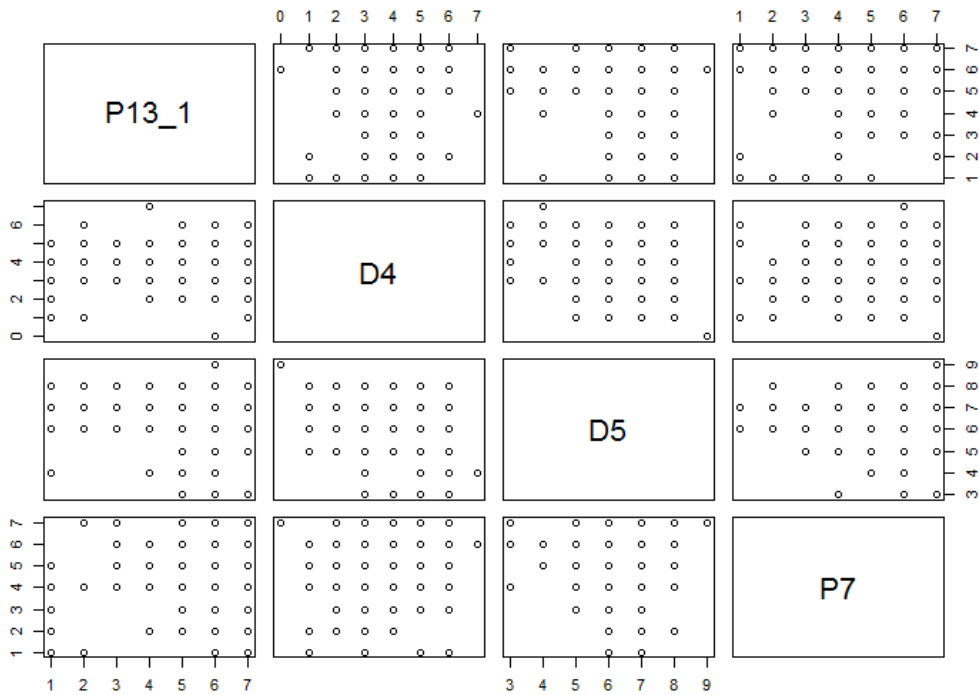
1.7. Figura 7. Análisis *clustering* julio 2020 "Reticencia hacia la vacuna y miedo a contagiarse del COVID-19" (FECYT, 2021).



1.8. Figura 8. Matriz de correlación de las variables escogidas en la encuesta julio 2020.

|       | P13_1   | D4      | D5      | P7      |
|-------|---------|---------|---------|---------|
| P13_1 | 1.0000  | -0.0031 | -0.2527 | 0.2928  |
| D4    | -0.0031 | 1.0000  | -0.0640 | 0.1660  |
| D5    | -0.2527 | -0.0640 | 1.0000  | -0.1485 |
| P7    | 0.2928  | 0.1660  | -0.1485 | 1.0000  |

**1.9. Figura 9. Matriz de dispersión de las variables escogidas en la encuesta julio 2020.**



**1.10. Figura 10. Resultados de la regresión lineal múltiple de la encuesta julio 2020.**

```
Call:
lm(formula = P13_1 ~ D4 + D5 + P7, data = encuesta_julio2020_limpia)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-4.7334 -0.9660  0.2152  1.3781  3.5349
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.06954    0.91037   6.667 2.95e-10 ***
D4           -0.09417    0.10579  -0.890 0.374530
D5           -0.34847    0.11177  -3.118 0.002116 **
P7            0.30572    0.07927   3.857 0.000159 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

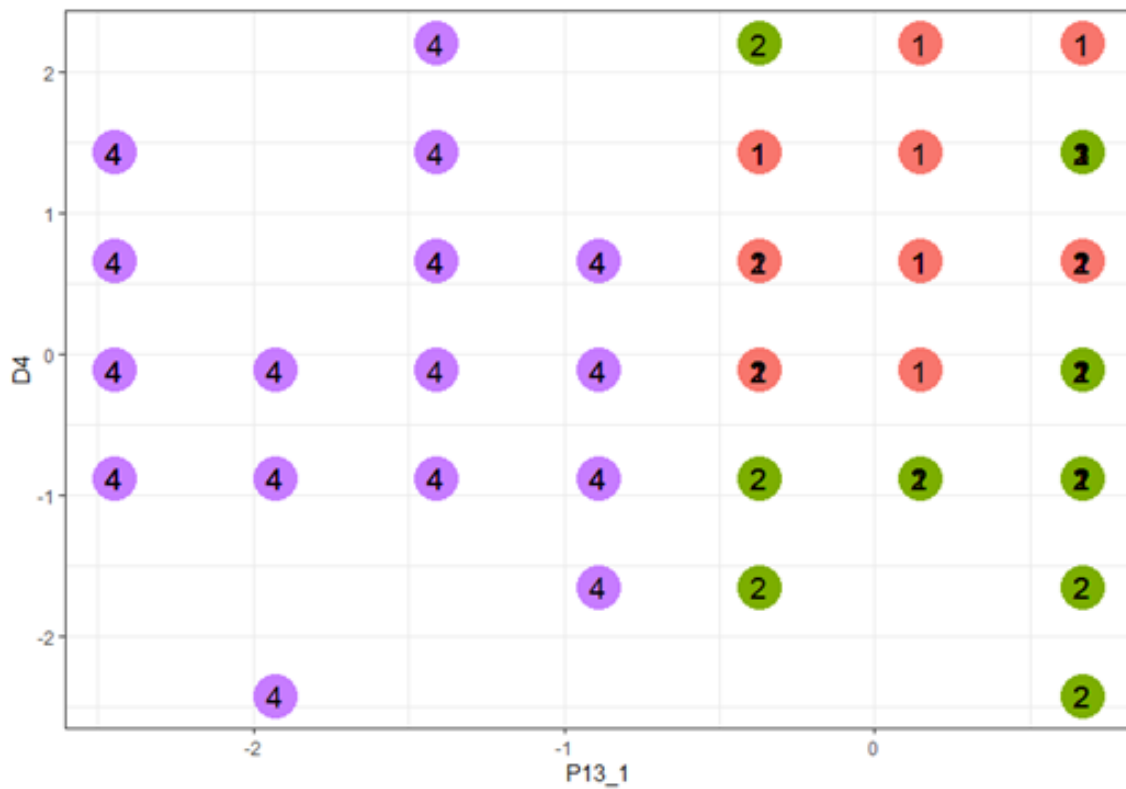
```
Residual standard error: 1.795 on 184 degrees of freedom
Multiple R-squared:  0.1342,    Adjusted R-squared:  0.1201
F-statistic: 9.509 on 3 and 184 DF,  p-value: 7.175e-06
```



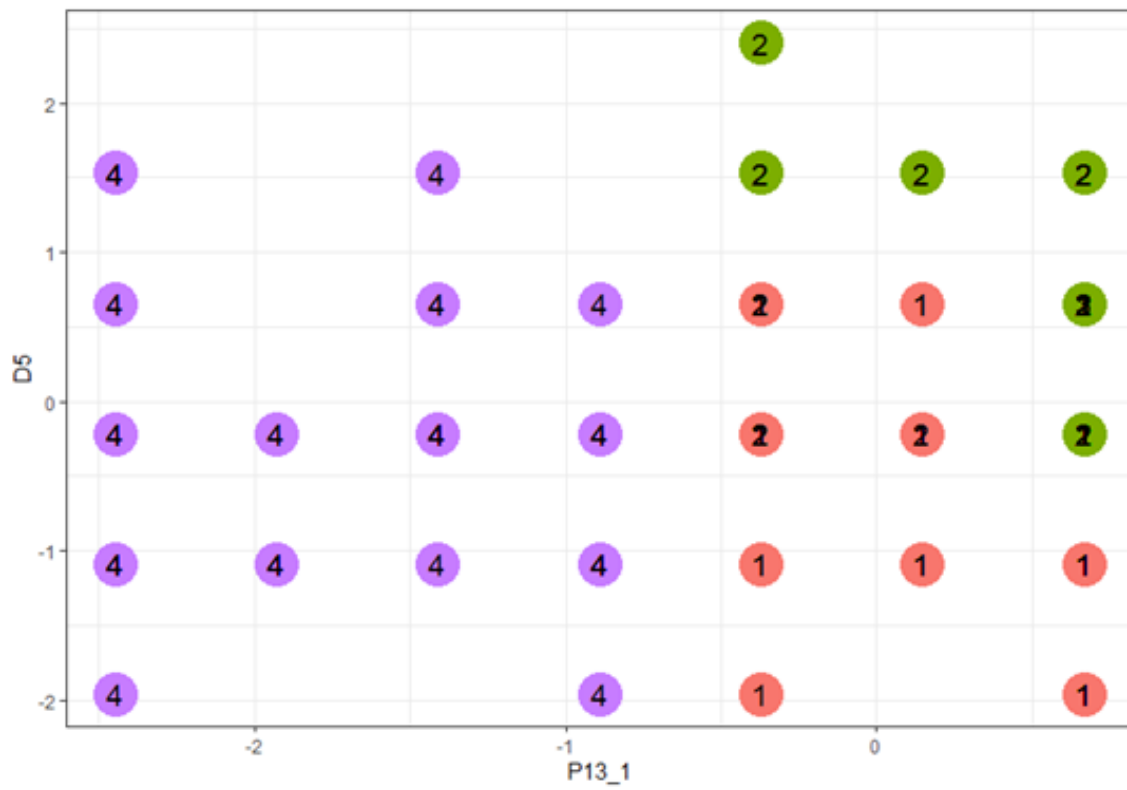
1.11. Figura 11. Intervalos de confianza en 95% para la encuesta julio 2020.

|             | 2.5 %      | 97.5 %     |
|-------------|------------|------------|
| (Intercept) | 4.2734430  | 7.8656381  |
| D4          | -0.3028957 | 0.1145471  |
| D5          | -0.5689918 | -0.1279444 |
| P7          | 0.1493317  | 0.4621089  |

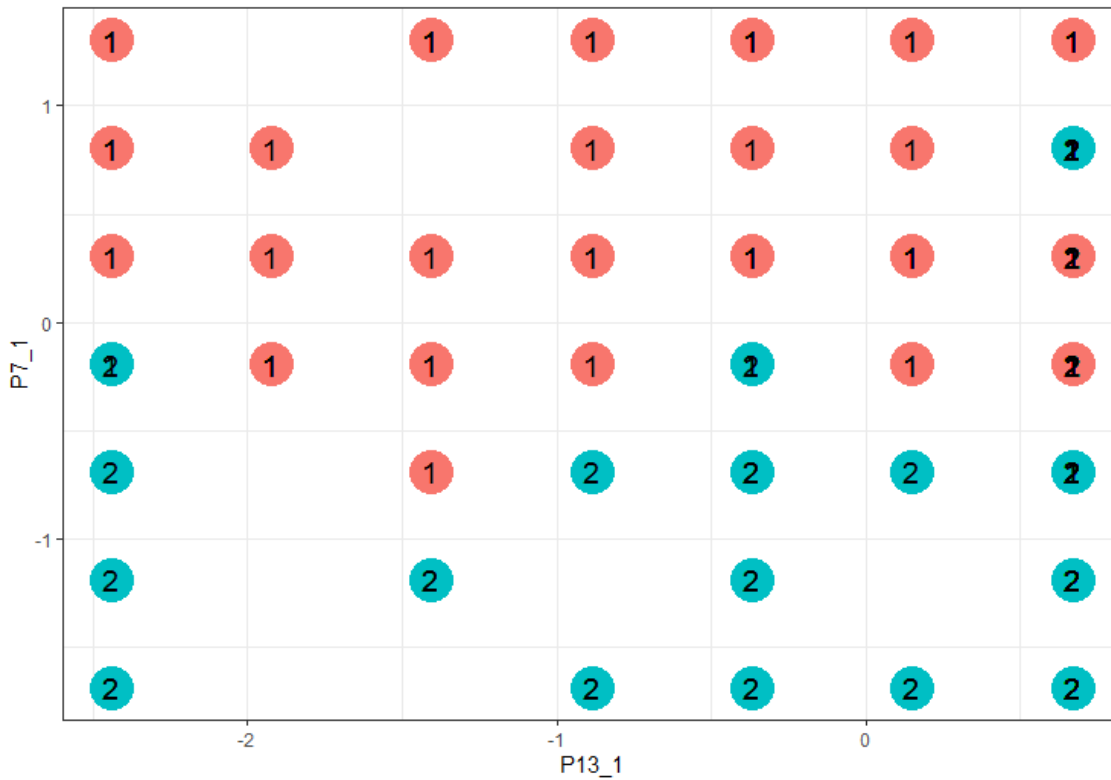
1.12. Figura 12. Análisis *clustering* enero 2021 "Reticencia hacia la vacuna e ideología política" (FECYT, 2021).



1.13. Figura 13. Análisis *clustering* enero 2021 "Reticencia hacia la vacuna y nivel de estudios" (FECYT, 2021).



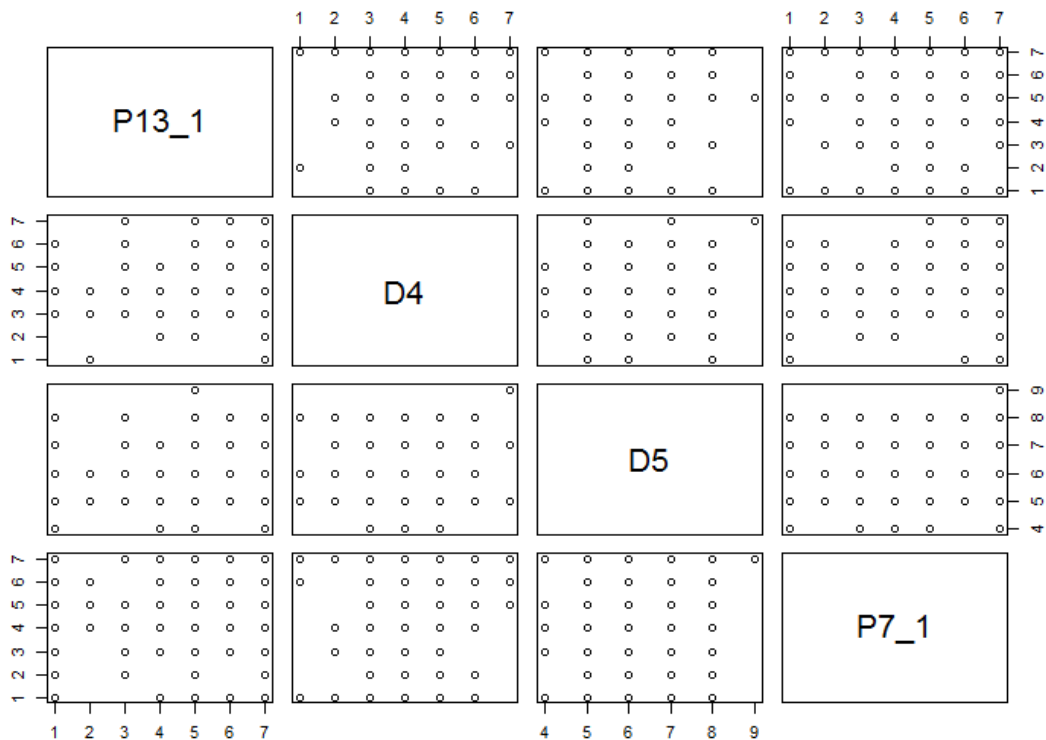
1.14. Figura 14. Análisis *clustering* enero 2021 "Reticencia hacia la vacuna y el miedo a contagiarse de COVID-19" (FECYT, 2021).



1.15. Figura 15. Matriz de correlación de las variables escogidas en la encuesta enero 2021.

|       | P13_1   | D4      | D5      | P7_1    |
|-------|---------|---------|---------|---------|
| P13_1 | 1.0000  | -0.0478 | 0.0782  | -0.0602 |
| D4    | -0.0478 | 1.0000  | -0.0430 | 0.0766  |
| D5    | 0.0782  | -0.0430 | 1.0000  | -0.2309 |
| P7_1  | -0.0602 | 0.0766  | -0.2309 | 1.0000  |

1.16. Figura 16. Matriz de dispersión de las variables escogidas en la encuesta enero 2021.



1.17. Figura 17. Resultados de la regresión lineal múltiple de la encuesta enero

```
Call:
lm(formula = P13_1 ~ D4 + D5 + P7_1, data = encuesta_enero2021_limpia)

Residuals:
    Min       1Q   Median       3Q      Max
-5.0328 -0.7435  1.0595  1.3092  1.7130

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.44192    1.03837    5.241  4.4e-07 ***
D4           -0.06228    0.11047   -0.564   0.574
D5            0.11221    0.12735    0.881   0.379
P7_1         -0.04000    0.07309   -0.547   0.585
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

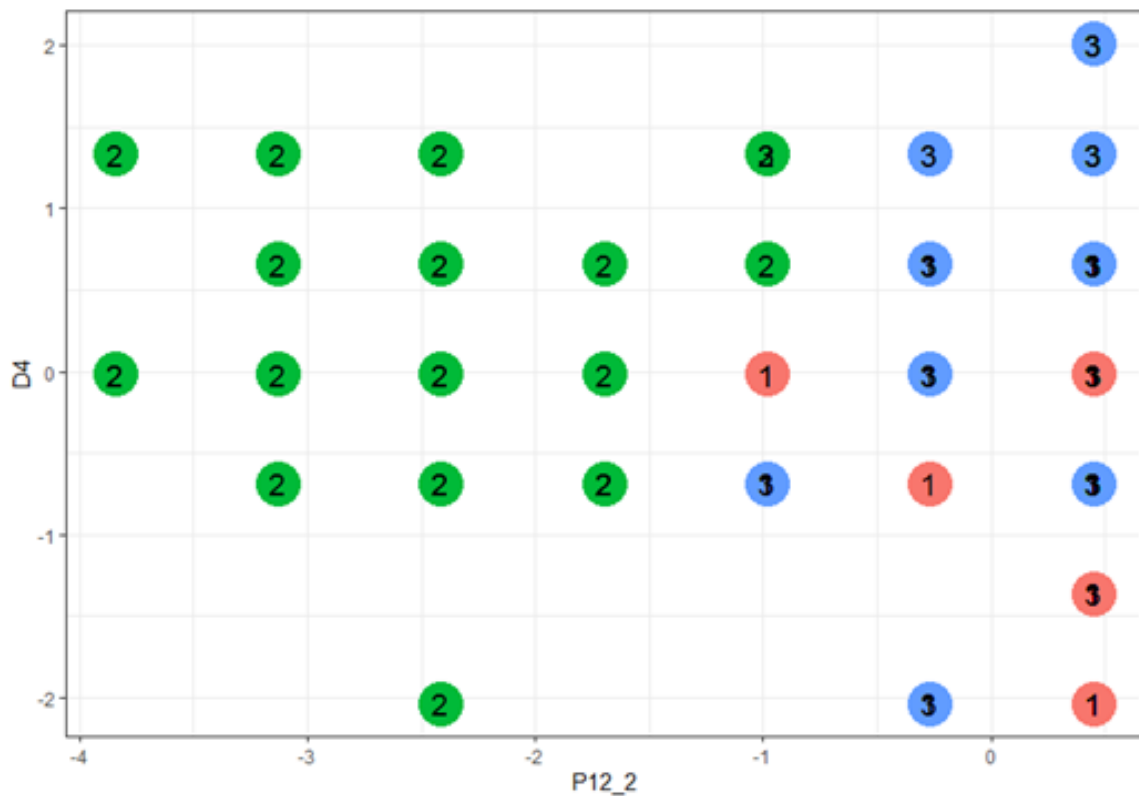
Residual standard error: 1.935 on 182 degrees of freedom
Multiple R-squared:  0.009725, Adjusted R-squared:  -0.006598
F-statistic: 0.5958 on 3 and 182 DF,  p-value: 0.6185
```

2021.

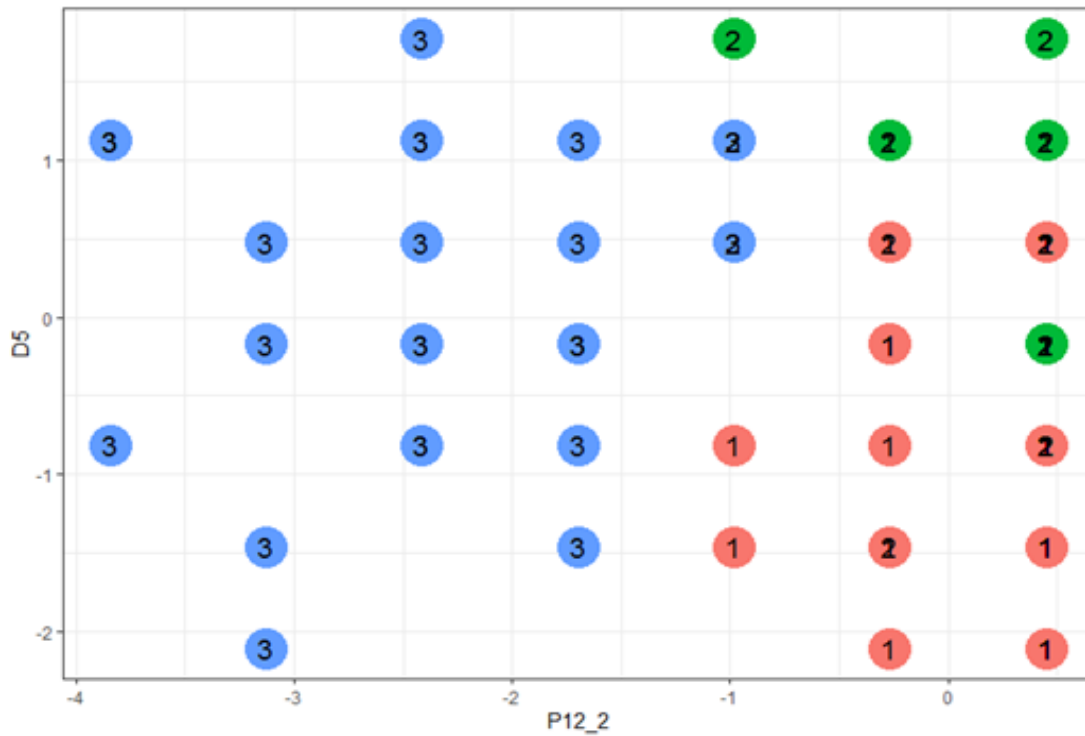
1.18. Figura 18. Intervalos de confianza en 95% para la encuesta enero 2021.

|             | 2.5 %      | 97.5 %    |
|-------------|------------|-----------|
| (Intercept) | 3.3931225  | 7.4907123 |
| D4          | -0.2802518 | 0.1556887 |
| D5          | -0.1390558 | 0.3634783 |
| P7_1        | -0.1842075 | 0.1042035 |

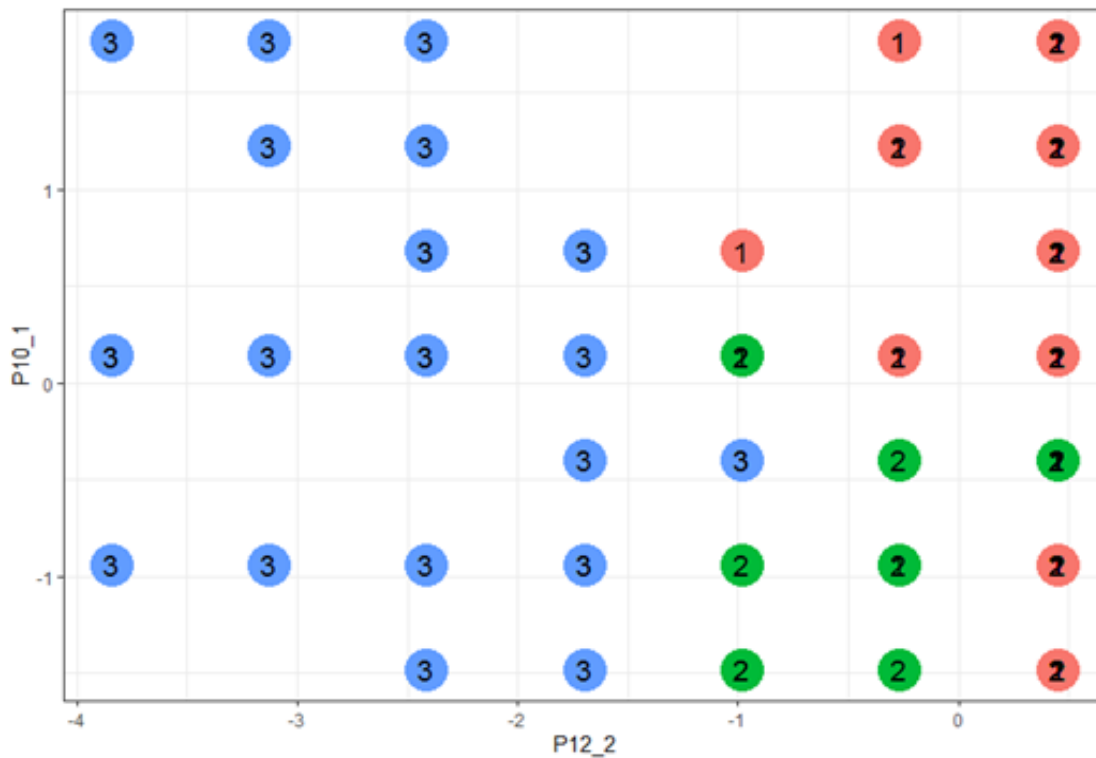
1.19. Figura 19. Análisis *clustering* mayo 2021 "Reticencia hacia la vacuna e ideología política" (FECYT, 2021).



1.20. Figura 20. Análisis *clustering* mayo 2021 "Reticencia hacia la vacuna y nivel de estudios" (FECYT, 2021).



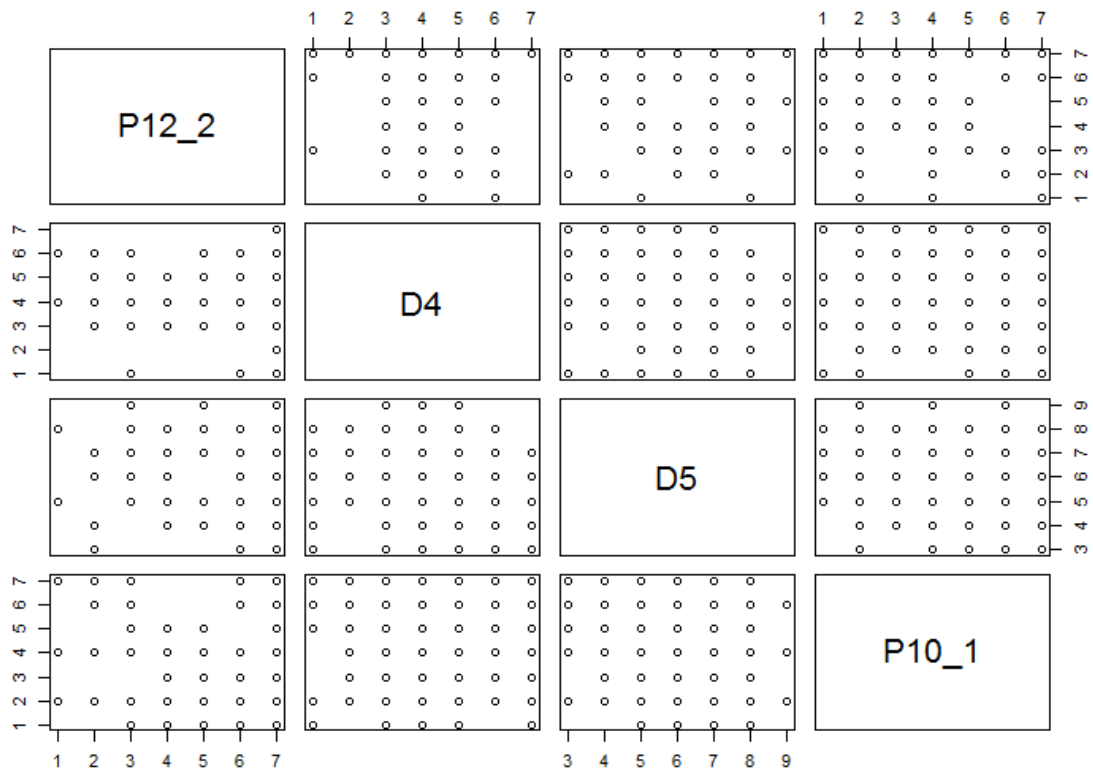
1.21. Figura 21. Análisis *clustering* mayo 2021 "Reticencia hacia la vacuna y miedo a contagiarse de COVID-19" (FECYT, 2021).



**1.22. Figura 22. Matriz de correlación de las variables escogidas en la encuesta mayo 2021.**

|       | P12_2   | D4      | D5      | P10_1   |
|-------|---------|---------|---------|---------|
| P12_2 | 1.0000  | -0.0105 | -0.0519 | -0.0624 |
| D4    | -0.0105 | 1.0000  | -0.2314 | -0.0069 |
| D5    | -0.0519 | -0.2314 | 1.0000  | -0.2293 |
| P10_1 | -0.0624 | -0.0069 | -0.2293 | 1.0000  |

**1.23. Figura 23. Matriz de dispersión de las variables escogidas en la encuesta mayo 2021.**



**1.24. Figura 24. Resultados de la regresión lineal múltiple de la encuesta mayo 2021.**

```
Call:
lm(formula = P12_2 ~ D4 + D5 + P10_1, data = encuesta_mayo2021_limpia)

Residuals:
    Min       1Q   Median       3Q      Max
-5.5142  0.3939  0.5590  0.6583  0.9982

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.14641    0.63225   11.303  <2e-16 ***
D4            -0.02718    0.06763   -0.402    0.688
D5            -0.06953    0.06661   -1.044    0.298
P10_1        -0.06075    0.05430   -1.119    0.265
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.402 on 204 degrees of freedom
Multiple R-squared:  0.009306, Adjusted R-squared: -0.005263
F-statistic: 0.6387 on 3 and 204 DF, p-value: 0.5909
```

**1.25. Figura 25. Intervalos de confianza en 95% para la encuesta mayo 2021.**

```
              2.5 %    97.5 %
(Intercept)  5.8998327  8.39298834
D4           -0.1605163  0.10616259
D5           -0.2008680  0.06180664
P10_1       -0.1678150  0.04631024
```

**2. Scripts de RStudio. Clustering.**

**2.1 Encuesta julio 2020.**

```
library(tidyverse)
library(factoextra)
library(cowplot)
library(ggpubr)
library(cluster)
```

#se lee primero el documento .sav de la encuesta global y se transforma a excel



```

fecyt_jul20 <- read_sav("fecyt_jul20.sav")
View(fecyt_jul20)
df <- data.frame(fecyt_jul20)
write.csv(df, "julio excel entero 2020")

# se limpia en excel de valores perdidos y se filtra la edad

View(encuesta_julio2020_limpia)

summary(encuesta_julio2020_limpia)

juventud_julio2020 = scale(encuesta_julio2020_limpia, center = TRUE, scale
= TRUE)
summary(juventud)

#llamamos a la nueva base de datos "juventud_julio2020"

juventud_julio2020 = as.data.frame(juventud_julio2020)

clusters1encuesta1 = kmeans(juventud_julio2020,centers=4,nstart = 50)
clusters1encuesta1

#ahora se van representando en clusters de 2 variables viendo el grado de
afectación de las 3 variables restantes sobre la variable P13_1
juventud_julio2020 = juventud_julio2020 %>% mutate(cluster =
clusters1encuesta1$cluster)

#el primero es la relación entre la reticencia vacunal (X=P13_1) y la ideología
política (Y=D4)

(ggplot(juventud_julio2020, aes(x = P13_1, y = D4)) +

```

```

geom_point(aes(color=as.factor(cluster)), size=10)+
geom_text(aes(label = cluster), size = 5) +
theme_bw() +
theme(legend.position = "none")+
labs(title = "P13_1 y D4")
)

```

#el segundo es la relación entre la reticencia hacia la vacuna (X=P13\_1) y el nivel de estudios de los jóvenes (Y=D5)

#cada vez que se calcula un nuevo cluster, se va a la linea de arriba de clusters1 encuesta1 = kmeans(juventud\_julio2020,centers=4,nstart = 50) y en número de "centers" se introduce el número de clusters deseados

```

(ggplot(juventud_julio2020, aes(x = P13_1, y = D5)) +
  geom_point(aes(color=as.factor(cluster)), size=10)+
  geom_text(aes(label = cluster), size = 5) +
  theme_bw() +
  theme(legend.position = "none")+
  labs(title = "P13_1 y D5")
)

```

#el tercero es la relación entre P13\_1 y el "miedo" (la conciencia de gravedad) de contagiarse del Covid-19

#se usan 4 centroides

```

(ggplot(juventud_julio2020, aes(x = P13_1, y = P7)) +
  geom_point(aes(color=as.factor(cluster)), size=10)+
  geom_text(aes(label = cluster), size = 5) +
  theme_bw() +
  theme(legend.position = "none")+
  labs(title = "P13_1 y P7")
)

```

)

## 2.2 Encuesta enero 2021

```
library(tidyverse)
```

```
library(factoextra)
```

```
library(cowplot)
```

```
library(ggpubr)
```

```
library(cluster)
```

```
#se muestra la base de datos global y se transforma a .csv para manipularla  
en excel
```

```
View(fecyt_ene21)
```

```
df <- data.frame(fecyt_ene21)
```

```
write.csv(df, "enero excel entero 2021")
```

```
# se limpia en excel de valores perdidos y se filtra la edad
```

```
View(encuesta_enero2021_limpiar)
```

```
summary(encuesta_enero2021_limpiar)
```

```
#se normaliza la base de datos
```

```
juventud_enero2021 = scale(encuesta_enero2021_limpiar, center = TRUE,  
scale = TRUE)
```

```
summary(juventud_enero2021)
```

```
#llamamos a la nueva base de datos normalizada "juventud_julio2020"
```

```
juventud_enero2021 = as.data.frame(juventud_enero2021)
```

```
clusters2encuesta1 = kmeans(juventud_enero2021,centers=3,nstart = 50)
```

```
clusters2encuesta1
```

```
#ahora se van representando en clusters de 2 variables viendo el grado de  
afectación de las 3 variables restantes sobre la variable P13_1
```

```
juventud_enero2021 = juventud_enero2021 %>% mutate(cluster =  
clusters2encuesta1$cluster)
```

```
#el primero es la relación entre la reticencia vacunal (X=P13_1) y la ideología  
política (Y=D4)
```

```
(ggplot(juventud_enero2021, aes(x = P13_1, y = D4)) +  
  geom_point(aes(color=as.factor(cluster)), size=10)+  
  geom_text(aes(label = cluster), size = 5) +  
  theme_bw() +  
  theme(legend.position = "none")+  
  labs(title = "P13_1 y D4")  
)
```

```
#el segundo es la relación entre la reticencia hacia la vacuna (X=P13_1) y el  
nivel de estudios de los jóvenes (Y=D5)
```

```
(ggplot(juventud_enero2021, aes(x = P13_1, y = D5)) +  
  geom_point(aes(color=as.factor(cluster)), size=10)+  
  geom_text(aes(label = cluster), size = 5) +  
  theme_bw() +  
  theme(legend.position = "none")+  
  labs(title = "P13_1 y D5")  
)
```

#el tercero es la relación entre P13\_1 y el "miedo" (la conciencia de gravedad) de contagiarse del Covid-19

#en este clustering se han utilizado dos centroides, en la línea de arriba  
clusters2encuesta1 = kmeans(juventud\_enero2021,centers=3,nstart = 50),  
en "centers" se fijan 2 centroides

```
(ggplot(juventud_enero2021, aes(x = P13_1, y = P7_1)) +  
  geom_point(aes(color=as.factor(cluster)), size=10)+  
  geom_text(aes(label = cluster), size = 5) +  
  theme_bw() +  
  theme(legend.position = "none")+  
  labs(title = "P13_1 y P7")  
)
```

### 2.3 Encuesta mayo 2021

```
library(tidyverse)
```

```
library(factoextra)
```

```
library(cowplot)
```

```
library(ggpubr)
```

```
library(cluster)
```

#se abre la base de datos global de la encuesta de mayo, que es .sav y se convierte a .csv para filtrarla en excel

```
fecyt_may21 <- read_sav("fecyt_may21.sav")
```

```
View(fecyt_may21)
```

```
df <- data.frame(fecyt_may21)
```

```
write.csv(df, "mayo excel entero 2021")
```

# se limpia en excel de valores perdidos y se filtra la edad

```

View(encuesta_mayo2021_limpia)

summary(encuesta_mayo2021_limpia)

#se normaliza la base de datos

juventud_mayo2021 = scale(encuesta_mayo2021_limpia, center = TRUE,
scale = TRUE)
summary(juventud_mayo2021)

#llamamos a la nueva base de datos normalizada "juventud_julio2020"

juventud_mayo2021 = as.data.frame(juventud_mayo2021)

clusters3encuesta1 = kmeans(juventud_mayo2021,centers=3,nstart = 50)
clusters3encuesta1

#ahora se van representando en clusters de 2 variables viendo el grado de
afectación de las 3 variables restantes sobre la variable P13_1
juventud_mayo2021 = juventud_mayo2021 %>% mutate(cluster =
clusters3encuesta1$cluster)

#el primero es la relación entre la reticencia vacunal (X=P12_2) y la ideología
política (Y=D4)

(ggplot(juventud_mayo2021, aes(x = P12_2, y = D4)) +
  geom_point(aes(color=as.factor(cluster)), size=10)+
  geom_text(aes(label = cluster), size = 5) +
  theme_bw() +
  theme(legend.position = "none")+

```

```
labs(title = "P12_2 y D4")
)
```

#el segundo es la relación entre la reticencia hacia la vacuna (X=P13\_1) y el nivel de estudios de los jóvenes (Y=D5)

```
(ggplot(juventud_mayo2021, aes(x = P12_2, y = D5)) +
  geom_point(aes(color=as.factor(cluster)), size=10)+
  geom_text(aes(label = cluster), size = 5) +
  theme_bw() +
  theme(legend.position = "none")+
  labs(title = "P12_2 y D5")
)
```

#el tercero es la relación entre P13\_1 y el "miedo" (la conciencia de gravedad) de contagiarse del Covid-19

```
(ggplot(juventud_mayo2021, aes(x = P12_2, y = P10_1)) +
  geom_point(aes(color=as.factor(cluster)), size=10)+
  geom_text(aes(label = cluster), size = 5) +
  theme_bw() +
  theme(legend.position = "none")+
  labs(title = "P12_2 y P10_1")
)
```

### **3. Scripts de RStudio. Regresión lineal múltiple.**

#### **3.1 Encuesta junio 2020**

```
#RLM en julio 2020
```

```
library(readxl)
```

```

encuesta_julio2020_limpiar <- read_excel("encuesta_julio2020_limpiar.xlsx")

View(encuesta_julio2020_limpiar)

#se obtiene la matriz de correlación entre las variables

cor(encuesta_julio2020_limpiar)

#se reducen a 4 decimales

round(cor(encuesta_julio2020_limpiar), 4)

#se obtiene la matriz de dispersión

plot(encuesta_julio2020_limpiar)

#ecuación de regresión lineal múltiple

RLMjulio2020 <- lm(P13_1 ~ D4 + D5 + P7, data = encuesta_julio2020_limpiar)

summary(RLMjulio2020)

#por último, se estudiará el intervalo de confianza para los coeficientes que
se han obtenido

confint(RLMjulio2020, conf.level=0.95)

```

### 3.2 Encuesta enero 2021.

```
#RLM en enero 2021
```



```

library(readxl)

encuesta_enero2021_limpiar <-
read_excel("encuesta_enero2021_limpiar.xlsx")

View(encuesta_enero2021_limpiar)

#se obtiene la matriz de correlación entre las variables

cor(encuesta_enero2021_limpiar)

#se reducen a 4 decimales

round(cor(encuesta_enero2021_limpiar), 4)

#se obtiene la matriz de dispersión

plot(encuesta_enero2021_limpiar)

#ecuación de regresión lineal múltiple

RLMenero2021 <- lm(P13_1 ~ D4 + D5 + P7_1, data =
encuesta_enero2021_limpiar)

summary(RLMenero2021)

#por último, se estudiará el intervalo de confianza para los coeficientes que
se han obtenido

confint(RLMenero2021, conf.level=0.95)

```

### 3.3 Encuesta mayo 2021

```
#RLM en mayo 2021

library(readxl)

encuesta_enero2021_limpiar <-
read_excel("encuesta_mayo2021_limpiar.xlsx")

View(encuesta_mayo2021_limpiar)

#se obtiene la matriz de correlación entre las variables

cor(encuesta_mayo2021_limpiar)

#se reducen a 4 decimales

round(cor(encuesta_mayo2021_limpiar), 4)

#se obtiene la matriz de dispersión

plot(encuesta_mayo2021_limpiar)

#ecuación de regresión lineal múltiple

RLMmayo2021 <- lm(P12_2 ~ D4 + D5 + P10_1, data =
encuesta_mayo2021_limpiar)

summary(RLMmayo2021)
```

#por último, se estudiará el intervalo de confianza para los coeficientes que se han obtenido

```
confint(RLMmayo2021, conf.level=0.95)
```