



**GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS**

**CURSO ACADÉMICO 2021-2022**

**TRABAJO DE FIN DE GRADO**

**Análisis estadístico multivariante del impacto del  
COVID-19 por Comunidades Autónomas**

**Multivariate statistical analysis of the impact of COVID-  
19 by Spanish regions**

AUTOR/A

**JOSÉ JULIÁN GÓMEZ GARCÍA**

TUTOR/A

**CARMEN MARÍA SORDO GARCÍA**

FECHA:

**24/02/2022**

## ÍNDICE

RESUMEN.....	3
1. INTRODUCCIÓN.....	4
2. MARCO TEÓRICO .....	4
2.1 CRISIS SANITARIA POR COVID-19: NATURALEZA E IMPACTO .....	4
2.1.1 Enfermedad COVID-19.....	4
2.1.2 Salud .....	5
2.1.3 Ámbito Económico .....	6
2.1.4 Empleo y desigualdad.....	7
2.1.5 Gestión Pública y financiamiento .....	8
2.2 ANÁLISIS MULTIVARIANTE.....	8
2.2.1 Análisis Factorial.....	8
2.2.3 Análisis clúster.....	9
2.2.4 Análisis anova.....	9
3. ANÁLISIS DE LOS DATOS Y RESULTADOS. ....	9
3.1 EXPOSICIÓN DE LAS VARIABLES.....	9
3.2 ANÁLISIS DESCRIPTIVO.....	10
3.3 ANÁLISIS FACTORIAL .....	17
3.4 ANÁLISIS CLUSTER.....	20
3.5 ANÁLISIS ANOVA.....	25
4. CONCLUSIONES.....	27
BIBLIOGRAFÍA.....	28

## **Análisis estadístico multivariante del impacto de la crisis COVID-19 en las comunidades autónomas españolas**

### **RESUMEN**

Este documento tiene como finalidad, comparar la situación de las 19 comunidades autónomas de España, durante el año 2020, año en que se produjo la pandemia del coronavirus. Para ello realizaremos un análisis detallado de datos multivariante: análisis factorial, análisis clúster, análisis de la varianza..., de manera que dichos datos sean estadísticamente significativos.

Realizaremos una evaluación de los datos obtenidos de cada comunidad autónoma, que incluirá referencias de todas las variables analizadas con anterioridad.

A partir de las conclusiones obtenidas, detectaremos que comunidades han conseguido sobrellevar mejor la pandemia, y cuales peor, además de detectar las amenazas y fortalezas de cada comunidad con la posibilidad de potenciar y corregir la situación de las mismas.

### **SUMMARY**

This document aims to compare the situation of the 19 autonomous communities of Spain, in 2020, the year in which the coronavirus pandemic occurred... To do this, we will perform a detailed analysis of multivariate data: factor analysis, cluster analysis, analysis of variance ..., so that these data are statistically significant.

We will carry out an evaluation of the data obtained from each autonomous community, which will include references of all the variables analyzed previously.

Based on the conclusions obtained, we will detect which communities have managed to cope better with the pandemic, and which worse, in addition to detecting the threats and strengths of each community with the possibility of enhancing and correcting their situation.

## 1. INTRODUCCIÓN

Vamos a realizar una investigación, con la finalidad de estudiar la situación social y económica, de cada una de las comunidades autónomas de España, derivadas de la pandemia producida por el virus COVID-19 (Coronavirus).

La investigación comenzará con la organización y agrupación de las CCAA en función de sus similitudes y diferencias en gestión administrativa de manera que se pueda dosificar la información. Seguidamente son identificadas las variables secundarias para la extracción de datos, las cuales fueron:

- Número de defunciones por el Covid-19, por CCAA.
- Número de defunciones por CCAA.
- Número de casos de coronavirus.
- Número de nacimientos por CCAA.
- PIB por comunidad autónoma.
- Población por comunidad autónoma.
- IPC por comunidad autónoma.
- Tasa de paro.
- Número de camas de hospital por CCAA.
- Presupuesto para sanidad por CCAA.

En el proceso de búsqueda y recolección, la principal fuente de información para la obtención de datos fue el INE (Instituto Nacional de Estadística) y la página web “datosmacro”. Para llevar a cabo el análisis de los datos fue implementado cuatro modalidades de análisis:

- Análisis descriptivo: se presentaron gráficamente la interacción de las CCAA en función a las variables secundarias, con el objetivo de ordenar toda la información y realizar una comparación preliminar que permita establecer posibles comportamiento tendencia.
- Análisis factorial: se buscó reducción de los datos observados con la finalidad de explicar las correlaciones entre las variables en términos de un número menor de variables no observadas llamadas factores, permitiendo explicar la máxima información contenida en los datos
- Análisis clúster: se agruparon las observaciones con el objetivo establecer el comportamiento de las CCAA en función de sus congruencias y discrepancias, de modo que se puedan puntualizar conclusiones de una manera más fácil y efectiva.
- Análisis anova: se realizaron comparativas entre las medias de los diferentes grupos del análisis previo.

Por ultimo fueron identificadas las problemáticas y fortalezas que existen en las zonas a estudiar para posteriormente establecer medidas que aplaquen las debilidades en la gestión de las Comunidades Autónomas antes una emergencia sanitaria.

## 2. MARCO TEÓRICO

### 2.1 CRISIS SANITARIA POR COVID-19: NATURALEZA E IMPACTO

#### 2.1.1 Enfermedad COVID-19

Los coronavirus (CoV) son una familia de virus extensa causantes de enfermedades respiratorias tanto humanas como animales que se manifiestan en varios niveles de gravedad, como el refriado común y otros padecimientos más severos como el síndrome respiratorio de Oriente Medio (MERS) y el síndrome respiratorio agudo severo (SRAS). (Organización Panamericana de la Salud, 2020). De acuerdo con la Organización Mundial de la Salud (OMS) (2020) el nuevo coronavirus conocido como COVID-19 es una enfermedad respiratoria causada por el virus SARS-CoV-2, la cual genera un padecimiento respiratorio que puede manifestarse de leve a moderado, por lo que diversas personas pueden recuperarse sin requerir tratamiento y atención médica especial, sin embargo, las personas mayores y pacientes de riesgo (personas con diabetes, padecimientos cardiovasculares, enfermedades respiratorias, cáncer o cualquier otra enfermedad subyacente) tienen una alta probabilidad de desarrollar un cuadro clínico más grave.

Este nuevo coronavirus fue descubierto en la ciudad Wuhan de China, cuando inicio un brote epidémico a finales del año 2019, siendo el hallazgo del virus SARS-CoV-2 anunciado oficialmente por el gobierno chino el 31 de diciembre del mismo año. No obstante, el 30 de enero de 2020 la OMS anunció la enfermedad como una emergencia pública a nivel internacional debido a la gran cantidad de casos en China y a su rápida propagación, llegando a 18 países de todo mundo para ese momento. La enfermedad por COVID-19 fue anunciada oficialmente por la OMS como pandemia el 11 de marzo de 2020, dada su alarmante propagación y la gravedad del cuadro clínico presentado; dicha enfermedad ha llegado a afectar en la actualidad a más de 190 países (Enríquez y Sáenz, 2021).

### **2.1.2 Salud**

LA pandemia por COVID-19 ha impactado significativamente a la sociedad en diversos niveles, ya que este paso de ser una emergencia sanitaria a representar una crisis que afectó al mundo entero. Es evidente, que, al tratarse de una crisis sanitaria, el primer aspecto que fue afectado fue el sistema sanitario de cada uno de los países, manifestando las debilidades que dicho sistema presenta ante situaciones emergentes de este calibre. Si bien, muchos de los países manifestaron dichos efectos en diferente tiempo entre 2020 y 2021, en menor o mayor medida, es innegable que en muchos de estos casos el sistema presentó un desbordamiento por la cantidad de casos a atender, escasez de suministros y medicamentos para atender a los pacientes con COVID-19 y la interrupción de los servicios para otros padecimientos que llevaron al colapso de los sistemas sanitarios de muchos de los países más afectados al ser ineficientes por la sobrecarga existente (Enríquez y Sáenz, 2021).

Según la OMS más del 90% de los servicios de salud de 216 países estudiados manifestaron las interrupciones del servicio de salud, el 50% manifestó las mayores problemáticas en la atención primaria de padecimientos comunes y retraso prolongado para la atención en padecimientos crónicos y el 20% problemas en la atención quirúrgica, crítica y de emergencia (FACME; Instituto de Salud y Estrategia, 2021).

Haciendo énfasis en el caso español, la crisis por COVID-19 ha reflejado lo debilitado que está el Sistema Nacional de Salud (SNS), la Salud Pública y una Atención Primaria, así como la alta dependencia de la población española a estos. La llegada de la pandemia ha significado un reto para la atención primaria en el país limitándose el acceso al mismo a casos de urgencia tanto en medicina como en enfermería, requiriendo improvisar, decidir y reorganizar todo el servicio, reprogramando citas previas de visitas, seguimiento de casos crónicos, estudios complementarios,

retrasando el control de vacunación de pacientes de atención primaria y transfiriendo la atención presencial a la atención virtual. Durante la primera ola, dependiendo de la región, la asistencia ambulatoria disminuyó entre 36 y 47%, las admisiones electivas entre 43% y 65% y la atención de urgencias entre 17% y 44%. Con respecto a los tiempos de espera, se ha registrado una disminución del 19,4% de pacientes en espera en consulta primaria, de 14% en la atención pacientes con cáncer y de 11,4% en pruebas diagnósticas, reflejando un aumento de tiempos de espera, tanto en la atención y diagnóstico como en las cirugías programadas. Según un estudio comparativo, durante el confinamiento se registró un descenso alarmante de cirugías pediátricas: se redujo en un 98% los registros de cirugías programadas, un 84% la presión asistencial desde urgencias y 82% las consultas externas ambulatorias (FACME; Instituto de Salud y Estrategia, 2021). Con respecto a las Comunidades Autónomas (CCAA), Felgueroso, et al. (2020) señalan que se destaca el aumento de la heterogeneidad entre estas tanto en situaciones COVID-19 como en no COVID-19, con algunas de las comunidades superando la media nacional en relación a nuevos contagios y fallecidos diarios, como es el caso de Castilla La Mancha, Castilla y León y Madrid; mientras que la capacidad de camas de hospitales y en UCI's presentan la mayor diferencia entre regiones, debido a los alto número de contagios pero con menor número de fallecidos en las regiones de mayor capacidad, como es el caso de Aragón. En relación a la prevalencia en hospitales y UCI's, usando como criterio que esta no exceda el 0% de la capacidad de camas, siete de las comunidades satisfacen este criterio en los lapsos más leves de la pandemia, siendo el caso contrario los meses críticos.

De acuerdo con Pinilla Domínguez y González López (2021), el COVID-19 ha revelado las grandes deficiencias del servicio sanitario y atención médica, evidenciando las debilidades y precariedad del sistema de salud pública, así como la vulnerabilidad ante emergencias sanitarias de esta magnitud. Sin embargo, esta autora también señala los aspectos positivos que ha dejado la pandemia, como la adaptación rápida de la red pública, el liderazgo en el sistema, la colaboración entre profesionales sanitarios y la creación de nuevos avances como la "telemedicina".

### **2.1.3 Ámbito Económico**

La pandemia ha dado origen a una recesión económica de gran magnitud, donde la principal estrategia por parte de los entes gubernamentales y autoridades sanitarias ha sido salvaguardar los sistemas de salud pública y la seguridad de la ciudadanía por medio del confinamiento de la población, impactando significativamente en el sistema económico perjudicando a todos los niveles económicos de la sociedad, desde las familias hasta las grandes empresas (Felgueroso, et al., 2020).

En relación al PIB nacional, este ha sido afectado debido a la disminución del consumo y aumento del ahorro de las familias, la pérdida de riqueza debido a la inestabilidad de los mercados financieros. En este sentido, las pérdidas asociadas al PIB están asociadas a causas exógenas al país, como la evolución epidemiológica global que determina la estabilidad de las actividades turísticas, principal sector económico español; causas endógenas, relacionadas al control de la enfermedad dentro de las fronteras españolas y, en tercer lugar, la caída producida por las restricciones económicas impuestas como las intervenciones y restricciones subóptimas (Pinilla Domínguez y González López, 2021). Bajo este panorama, España se posicionó como uno de los países europeos más afectados por la pandemia, considerando que en 2019 la contribución del sector turístico en el PIB fue mayor al 12%, reduciéndose a menos del 5% en el año 2020, reflejándose en una disminución del 11% en el PIB nacional. Otros sectores económicos, como el textil, confección y calzado, automoción y

componentes, petróleo y actividades, también fueron severamente afectados por la pandemia de la COVID-19; mientras que la distribución minorista de alimentos, la farmacéutica y la industria de las TIC (Tecnologías de Información y Comunicaciones) y el comercio electrónico incrementaron su demanda (Martínez, 2021; Observatorio Sectorial DBK, 2021). En relación a la variación del PIB de las diferentes provincias, aquellas zonas altamente dependientes del turismo, como las Islas Baleares, Málaga, Alicante, Cádiz, Tarragona y Valencia, manifestaron una caída abrupta del PIB, mientras que otras zonas han evidenciado un impacto moderado debido a la contribución de otros sectores en la economía regional (Martínez, 2021).

### **2.1.4 Empleo y desigualdad**

En el ámbito social, al paralelo que la situación económica, la pandemia ha generado una crisis social que ha aumentado el desempleo y la desigualdad económica en un 4% (Felgueroso y De la Fuente, 2021). Durante los primeros meses de confinamiento se presentaron profundas dudas con respecto a la reincorporación laboral por parte de aquellas personas que fueron despedidas, les fue recortada su jornada laboral o se tomaron vacaciones sin sueldo, generando una alta incertidumbre con respecto a las finanzas familiares; en términos generales, el nivel de formación escolar fue determinante en el empleo durante el periodo de confinamiento, debido a que trabajadores menos cualificados y asociados con puestos de trabajo que requiere de proximidad física fueron los principales afectados (Felgueroso, et al., 2020). Con respecto a esto, la pandemia dio paso al aumento del empleo bajo la modalidad teletrabajo, sin embargo, esta modalidad impulsa una brecha de desigualdad, ya que ocupaciones específicas por su naturaleza son más aptas para desarrollarse en teletrabajo (trabajadores con alto nivel de escolaridad), por lo que directivos, profesional técnico y asistencia o apoyo administrativo son los grupos que pueden trabajar por teletrabajo, mientras que mantenimiento, servicio, reparación, entretenimiento son los grupos que menos pueden trabajar bajo esta modalidad, tal distribución puede ser una fuerza generadora de desigualdad. En tal sentido, la pandemia ha impactado significativamente en el empleo de baja distribución salarial, siendo este colectivo una población vulnerable debido a los despidos, reducción de la jornada y la incapacidad de teletrabajar. Las regiones más afectadas por esta premisa, fueron las dependientes de las actividades turísticas y la poca concentración de personal cualificado como Madrid, las Islas Baleares y Canarias, mientras que Navarra, Asturias, Extremadura y Aragón registraron un crecimiento del empleo en la distribución salarial baja debido la expansión de ocupaciones elementales y actividades agrícolas. En el otro extremo, el empleo de la distribución salarial alta ha aumentado a nivel nacional (Felgueroso y De la Fuente, 2021).

A pesar de estas circunstancias, la tasa de paro apenas se vio afectada, subiendo 1,5 puntos a mediados del año 2020, mientras que el desempleo disminuyó 3,5 puntos; esta poca variación puede deberse a la implementación de políticas por parte el ERTE (Expediente de Regulación Temporal de Empleo) para mantener el empleo y por lo tanto, a pesar de las condiciones laborables, estas personas no pueden ser calificadas como desempleadas sino como ocupadas, sin embargo, tampoco podían buscar empleo debido a las restricciones del confinamiento, manteniendo la tasa de paro sin alterar pero disminuyendo el nivel de empleo nacional. Bajo este panorama. Si bien, estos indicadores de desempleo no describen la realidad durante la pandemia, el aumento en la alta de prestaciones de desempleo por parte del ERTE durante este periodo es un buen indicativo de la magnitud del impacto de la pandemia, llegando a un aumento de 90% en marzo y abril, resaltando que aproximadamente un 25% de los ocupados estuvo bajo algún tipo de ERTE, siendo los principales sectores afectados la hostelerías y las

actividades recreativas con 70% y el 60% de los empleados con reducción de su jornada o suspensión del contrato respectivamente (Ocaña, et al., 2021).

### **2.1.5 Gestión Pública y financiamiento**

La crisis sanitaria ha revelado las debilidades existentes en el sistema de salud público español, que, a pesar de tratarse de una crisis a nivel mundial de gran magnitud, es evidente y ha sido tema de discusión en la literatura actual, que los recortes y el poco del financiamiento e inversión en la salud pública, como consecuencia de la crisis económica de 2008, ha representado un factor decisivo en el desempeño y rendimiento eficaz del sistema sanitario. Bajo esta premisa, se destaca que el gasto corriente e inversión en la red pública del Sistema Nacional de Salud (SNS) fue gravemente afectada durante la crisis 2008-2013, sin poder recuperarse completamente en 2018, registrándose un aumento de gasto sanitario público total de 85% pero con una caída en la inversión de 16%, en comparación al año 2002, último año en que se realizaron transferencias de ámbito sanitario a las CC.AA; sin resultados de mejora para el 2019 y 2020 dadas las prórrogas de los presupuestos; por lo que el gasto público e inversión en España, a nivel nacional, es muy moderado, mientras que el nivel de gastos entre las CCAA presenta variabilidad manifestándose en una heterogeneidad con respecto a políticas de prevención y salud pública. (Pinilla Domínguez y González López, 2021; Trapero-Bertran y Lobo, 2021). Con las nuevas reformas presupuestarias de 2013 y la libertad de gestión de las CCAA, provocó hasta la fecha de 2019 un incremento en las desigualdades en la prestación de la atención médica, a pesar que todas las comunidades autónomas mantuvieron reducciones en el gasto público, creando en copago farmacéutico, reduciendo gastos en atención y diagnóstico especializados, bienes de capital y detener las retribuciones al personal sanitario, con el objetivo de mejorar la eficiencia del servicio, sin embargo, la literatura señala que dicha reducción incide en el deterioro de la calidad de los servicios ofrecidos y la reducción del consumo de servicios de salud (Ramón-Dangla, Rico-Gómez y Issa-Khozouz, 2021).

## **2.2 ANÁLISIS MULTIVARIANTE**

Los métodos de análisis de datos habituales (univariantes) han sido herramienta importante para el procesamiento de información siempre que estos datos solo abarquen una o pocas dimensiones, sin embargo, se vuelven poco eficientes cuando los datos a analizar situaciones de la realidad que contemplan una dimensionalidad compleja, el proceso de análisis de vuelve problemático. Por lo tanto, la aplicación de técnicas multivariantes se convierte en un candidato óptimo que permita reducir esta dimensionalidad y eliminar la multicolinealidad entre variables es un paso fundamental para poder realizar un análisis adecuado y eficaz de los datos. En tal sentido, el análisis multivariante engloba un conjunto de técnicas y métodos que buscan extraer de los datos las características más representativas del objeto bajo estudio a través de la proyección de variables en los datos de entrada y, en ocasiones, en los datos de salida (Muñoz, 2015).

### **2.2.1 Análisis Factorial**

El análisis factorial es una técnica eficaz para el análisis de relaciones complejas multidimensionales que se presentan en las investigaciones las ciencias de la administración; este tipo de análisis multivariante engloba un conjunto de métodos estadísticos que buscan definir la estructura subyacente en una matriz de datos. En términos generales, esta técnica de análisis busca enfocar la incógnita de cómo

procesar la estructura de las interrelaciones o correlaciones entre un conjunto amplio de variables, a través de la definición de una serie de dimensiones subyacentes comunes conocidas como factores. El objetivo del análisis factorial está dirigido a resumir y reducir el conjunto de datos, es decir, al momento de reducir los datos se obtienen las dimensiones subyacentes que, comparadas e interpretadas correctamente, permiten describir la situación bajo estudio en un número de elementos más pequeño que las variables originales (Mejía, 2017).

### **2.2.3 Análisis clúster.**

También conocida como análisis de conglomerado, esta técnica realiza un estudio de relaciones de interdependencia que busca describir la estructura subyacente de datos con el objetivo de identificar grupos de individuos, a diferencia del análisis factorial cuyo objetivo es identificar grupos de características similares. El análisis de clúster abarca un conjunto de técnicas que permiten reducir la complejidad de los datos a través de la creación de grupos mutuamente excluyentes, con el objetivo de representar la variabilidad del fenómeno bajo estudio o individuos a través de las similitudes de sus características (Meneses, 2019). En tal sentido, el análisis clúster agrupa los individuos en conglomerados basándose en las características o cualidades que los diferencian de los otros conglomerados, con el objetivo de maximizar la homogeneidad entre grupos a la vez que se maximiza la heterogeneidad de los agregados (Mejía, 2017).

### **2.2.4 Análisis anova.**

A diferencia del análisis factorial y el análisis de conglomerados, la técnica ANOVA busca analizar las relaciones de dependencia con la finalidad de plantear explicaciones y predicciones de las mismas, es decir, esta técnica es aplicada en escenarios donde se desea establecer relaciones simultáneas entre diversas variables que permitan describir y explicar la variabilidad de los grupos de individuos (Meneses, 2019).

## **3. ANÁLISIS DE LOS DATOS Y RESULTADOS.**

### **3.1 EXPOSICIÓN DE LAS VARIABLES**

La variable objetivo en la que centraremos nuestro análisis, será la Comunidad Autónoma. La elección de la misma radica en la importancia de la gestión administrativa y respuesta ante la situación pandemia presentada en el 2020, concretamente en el estudio se visualizarán la congruencia y discrepancias en las 19 CCAA que conforman el país. Para la realización de las comparaciones se efectuarán los cuatro análisis (descriptivo, factorial, clúster y anova) con la finalidad de identificar las deficiencias que existen en cada comunidad y establecer medidas que podrían llevarse a cabo para remediarlas.

Por la diversidad administrativas en las CCAA, fueron seleccionadas diez variables secundarias que proporcionaron información sobre el estado de la situación y nivel de respuesta ante la pandemia por COVID-19

A continuación, se describe cada una de las variables secundaria elegidas:

1. Número de defunciones provocadas por el Covid-19 (DefC-19): número de muertos por comunidad autónoma provocados por el Covid-19 durante el año 2020.

2. Número de muertos por comunidad autónoma (DefCCAA): número de defunciones totales en España, y por comunidad autónoma, durante el año 2020.
3. Número de casos de coronavirus por CCAA (CasosC-19): número de contagiados por el coronavirus en España, por comunidad autónoma, durante el año 2020.
4. Nacimientos en España (Nac20): número de nacimientos acaecidos en España, por comunidad autónoma, durante el año 2020.
5. PIB por comunidad autónoma en 2020 (PIB20): producto interior bruto por comunidad autónoma durante el año 2020.
6. Población Española (PobCCAA): población de cada comunidad autónoma a 31 de diciembre de 2020.
7. IPC 2020 (IPC20): Medida estadística que recoge cómo han evolucionado los precios de los bienes y servicios que consume una población residente en viviendas familiares en España. Indica la tasa de inflación de una zona, con él analizaremos la competitividad y el poder adquisitivo de cada comunidad autónoma.
8. Número de parados en España (Paro20): número total de desocupados, pertenecientes a la población activa de España, por comunidad autónoma durante el año 2020.
9. Número de camas de hospital por Comunidad Autónoma (CamasHospCCAA): referente al número de camas de hospital disponibles en España por Comunidad Autónoma.
10. Presupuesto para sanidad por Comunidad Autónoma (PreSanCCAA): presupuesto en miles de euros, que cada comunidad autónoma destina a sanidad.

A partir de la información obtenida por las variables secundarias, se obtiene el estado actual de la gestión administrativa en cada Comunidad Autónoma ante la pandemia del 2020.

### 3.2 ANÁLISIS DESCRIPTIVO.

En este apartado, vamos a requerir de las 8 variables elegidas anteriormente, las cuales han pasado por un estudio previo que nos permita determinar que son las mejores opciones para llevar a cabo todos y cada uno de los análisis, con la finalidad de obtener los mejores resultados y que exista el menor número de inconclusiones posibles.

En este apartado vamos a explicar mediante estadística descriptiva y gráficos el estado de las variables para cada comunidad autónoma.

**Tabla 1. Estadísticos descriptivos**

	Nº	Media	Mínimo	Máximo	Desviación. Típica
Número de muertos (COVID-19)	19	2884,89	407,00	11997,00	3397,16
Casos COVID-19	19	104582,21	3086,00	406900,00	119162,94
PIB 2020	19	58383,09	1445,93	214195,96	65145,21
Número de nacimientos	19	17812,37	771,00	64906,00	19729,84
Número de muertos	19	25873,79	601,00	79685,00	25193,81
Población CCAA	19	2497410,26	84202,00	8464411,00	2588362,16
IPC CCAA	19	121,52	106,51	142,54	10,47

Número de Parados CCAA	19	195768,42	9800,00	907200,00	232905,16
N.º Camas Hospital CCAA	19	5905,53	160,00	26059,00	6288,16
Presupuesto Sanidad CCAA	19	3421904,11	1171,00	10824520,00	3206298,43

**Fuente: elaboración propia a partir de datos del INE, y DATOS-MACRO.**

En la tabla 1, podemos observar diferentes estadísticos descriptivos para cada una de las variables seleccionadas en la investigación. En la primera columna nos encontramos con las variables, en la segunda el número de comunidades autónomas, en la tercera, la media estadística de las variables los datos de cada variable, en la cuarta y quinta, tenemos el máximo y mínimo valor de queda muestra estudiada para cada variable, y en la quinta columna tenemos la desviación típica de cada muestra, que nos marca la variación o la dispersión de cada una de ellas.

A priori, observamos gran variabilidad entre las variables, aunque en muchos de os casos se debe a la diferente densidad de población existente entre las diferentes comunidades autónomas. Esto es, en cuanto a las variables relacionadas con la demografía, sí que existe proporcionalidad entre los datos analizados y la población de las CCAA, sin embargo en cuanto al IPC y el nº de parados, sí que observamos la gran desigualdad que existe entre las comunidades autónomas.

A continuación, vamos a diferenciar la información por comunidades, para darle un sentido más específico a las variables estudiadas. Para ello utilizaremos diferentes gráficos de elaboración propia, a partir de la información obtenida de cada variable.

En el gráfico 1, tendremos desagregado en número de muertos por COVID-19, durante 2020, por CCAA, donde podremos observar como las comunidades con mayor población tendrán la mayor mortalidad y viceversa.

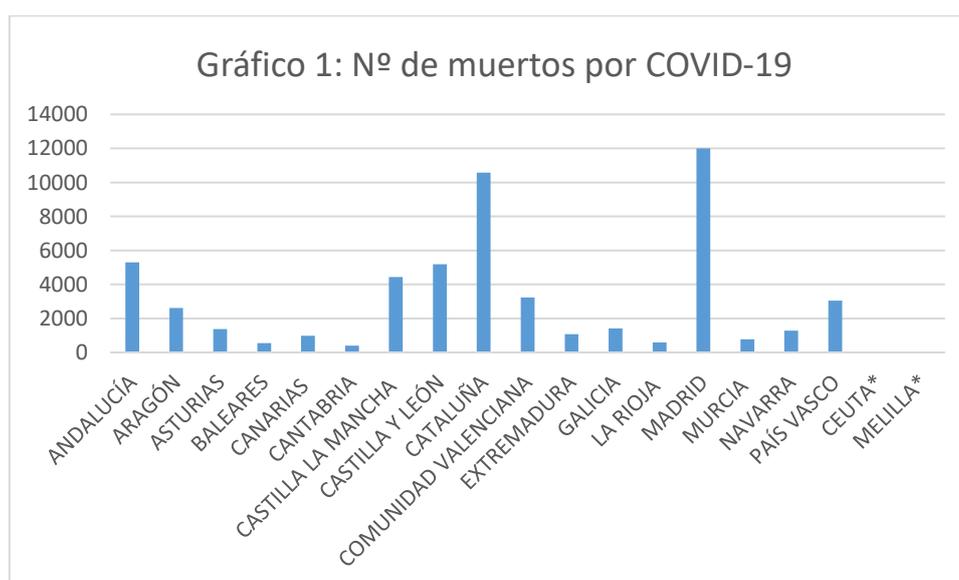


Figura 2. Cantidad de Muertos por COVID-19 (Elaboración propia a partir del INE y DATOS-MACRO)

En el gráfico 2, mostraremos en número de infectados por el Coronavirus en cada CCAA durante 2020. Podemos observar como el número de casos es proporcional al número de muertos en cada CCAA, salvo alguna excepción como por ejemplo Murcia o Islas Baleares.

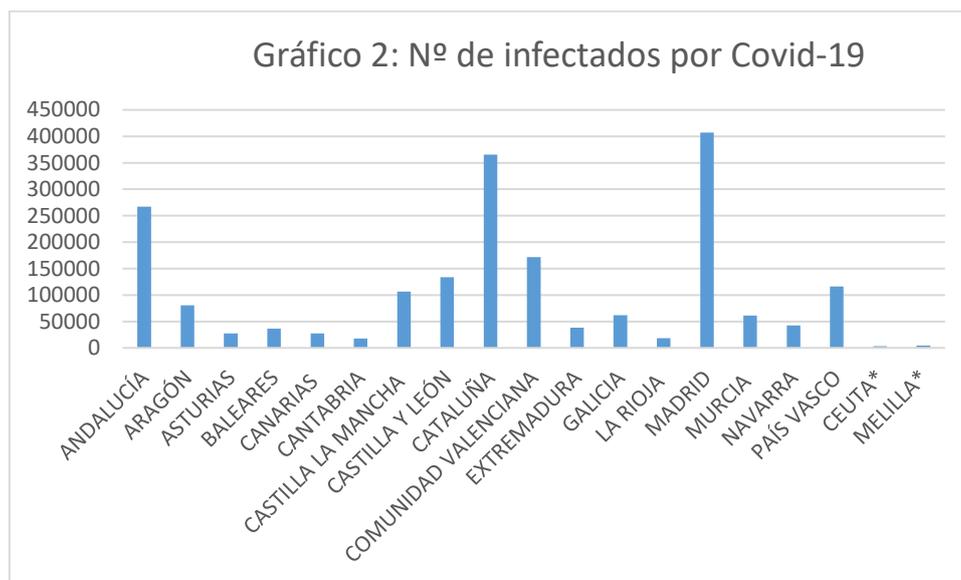


Figura 3. Cantidad de Infectados por Covid-19 (Elaboración propia a partir del INE y DATOS-MACRO)

El gráfico 3, muestra el PIB de cada comunidad autónoma en millones de euros. Decir que es un 10,8% inferior al año 2019, año pre-pandemia. La tendencia sigue siendo, a grandes rasgos la misma, es decir, las comunidades con mayor población son las que tienen un mayor PIB, destacando por encima de todas ellas Madrid y Cataluña.

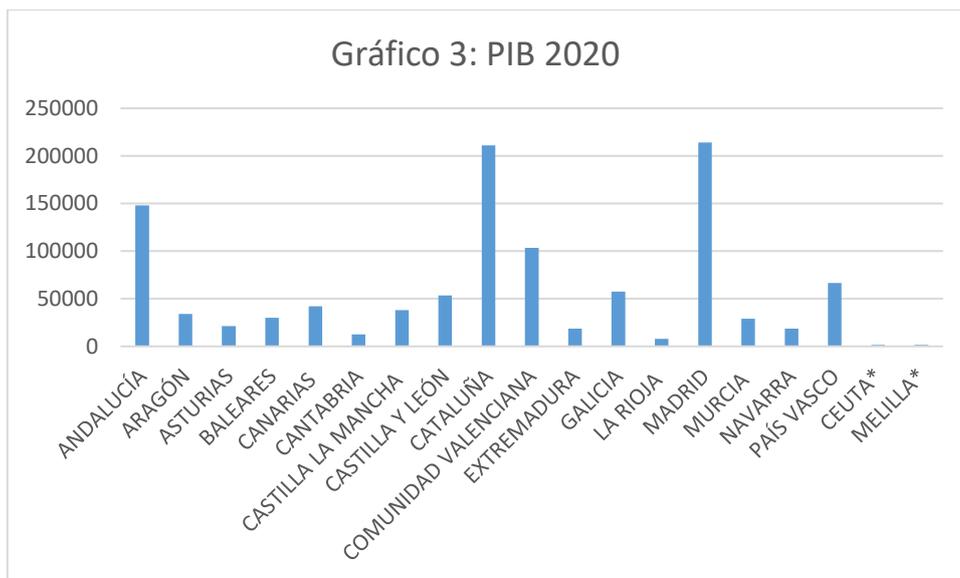


Figura 4. PIB 2020 (Elaboración propia a partir del INE y DATOS-MACRO)

En el gráfico número cuatro observamos el nº de nacimientos acaecidos durante el año 2020. Aquí sí que observamos proporcionalidad con la población de cada CCAA, sin embargo, también comprobamos como siendo Andalucía la más poblada, en cuanto a indicadores económica sí que se encuentra más atrás que las demás, a pesar de ser las más poblada.

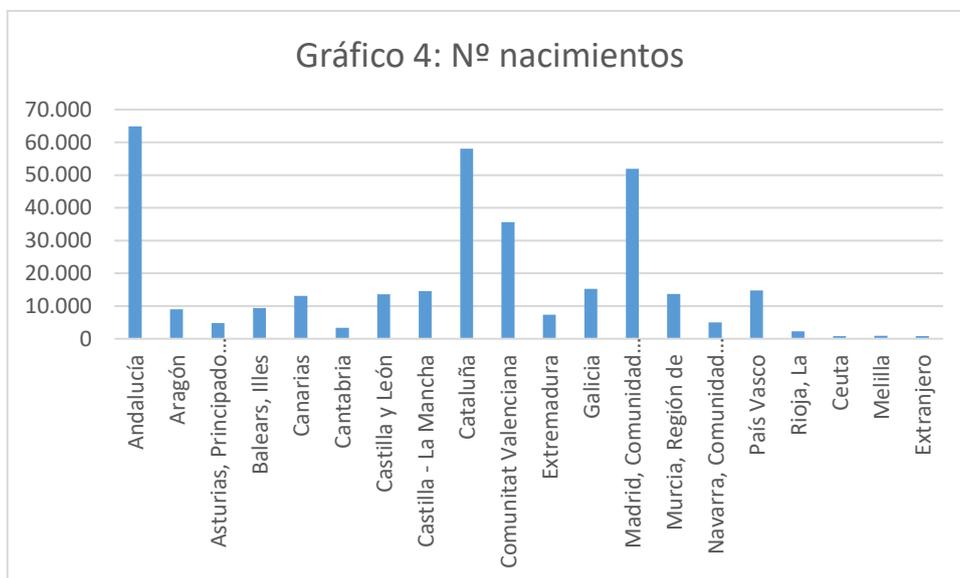


Figura 5. –Cantidad de Nacimientos (Elaboración propia a partir del INE y DATOS-MACRO)

Ahora expondremos el gráfico cinco, que mostrará la población de cada comunidad autónoma. Es la variable que hace de guía, en términos de proporcionalidad generalmente, para las demás.

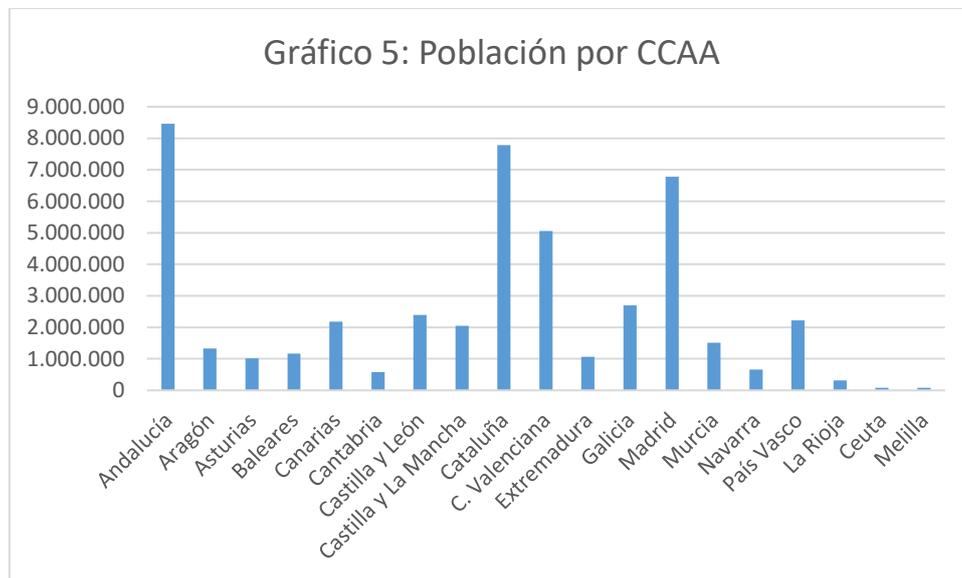


Figura 5. Cantidad poblacional por Comunidad Autónoma (Elaboración propia a partir del INE y DATOS-MACRO)

A continuación, vamos a mostrar en el gráfico número seis, el número de defunciones generales que hubo en España durante el año 2020. Otra vez más, volvemos a observar una clara relación con gráficas anteriores (Población; Casos Covid-19; Fallecidos por Covid-19).

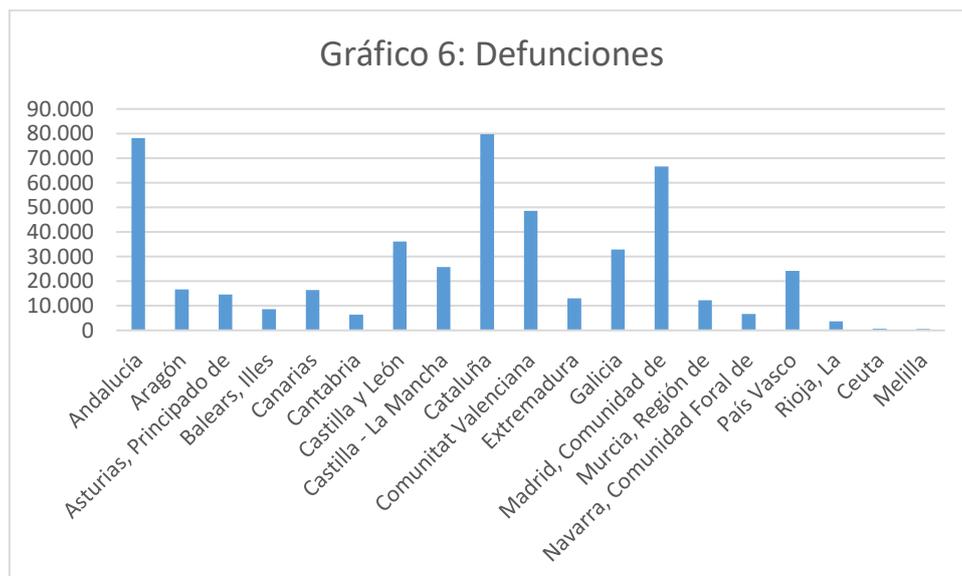


Figura 6. Defunciones Acaecidas en el 2020 (Elaboración propia a partir del INE y DATOS-MACRO)

En la Figura 7 se expone el índice de precios al consumo por CCAA, durante el año 2020. Donde, nos encontramos con una menor variación entre los valores del IPC para las CCAA. Destacándose los valores más elevados en Madrid, Cataluña, Ceuta, Melilla e Islas Baleares. Por el contrario, los índices más bajos los encontramos en Extremadura y Castilla la Mancha.

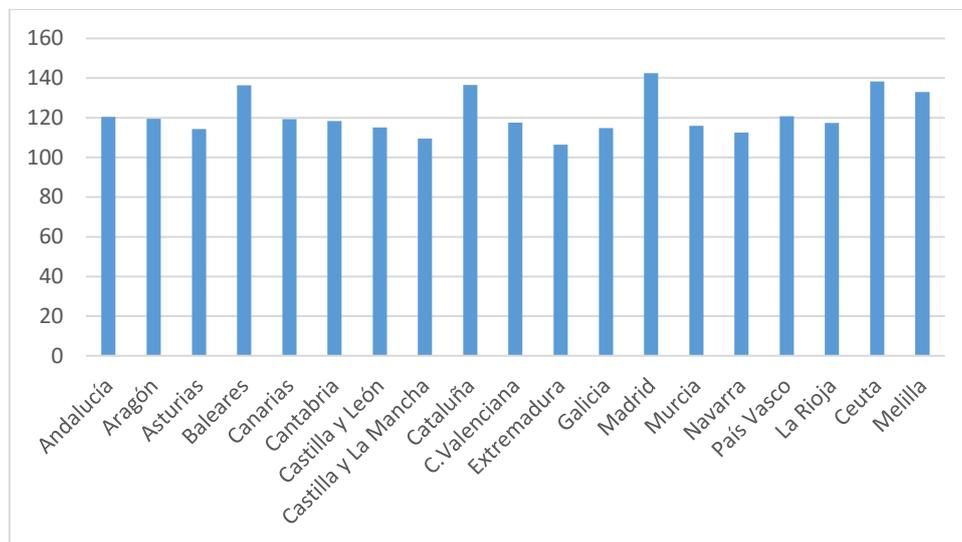


Figura 7. IPC por Comunidad Autónoma (Elaboración propia a partir del INE y DATOS-MACRO)

Ahora expondremos el gráfico número 8, que nos expondrá el número de personas en paro por comunidad autónoma durante el año 2020. Volvemos a encontrar proporcionalidad con la población, sin embargo encontramos un caso extremo, Andalucía, con un nivel de parados demasiado alto pese a ser la comunidad autónoma más poblada. Destacar el aumento de parados general, debido al efecto del Coronavirus.

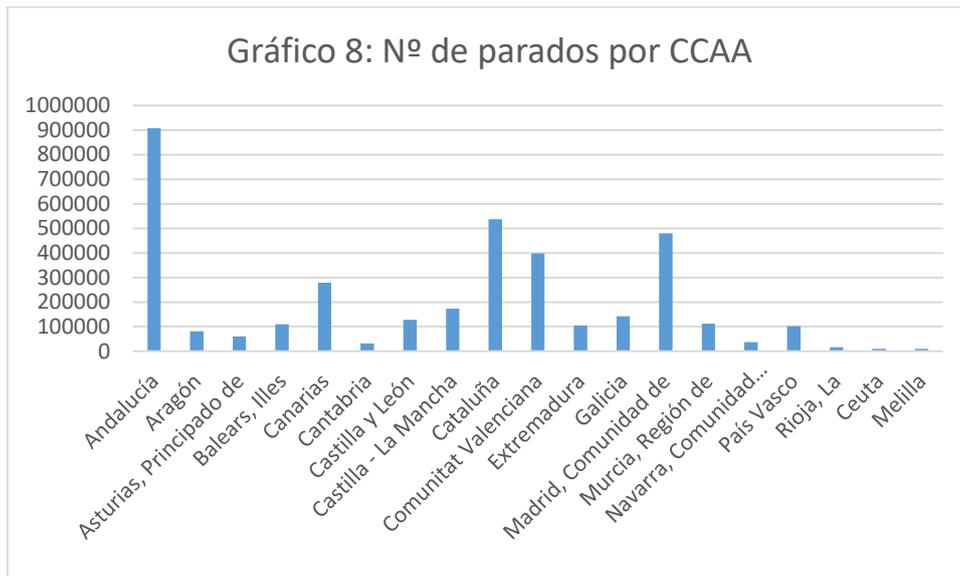


Figura 8. Cantidad de Parados durante la pandemia por COVID-19 (Elaboración propia a partir del INE y DATOS-MACRO)

Pasamos con los dos últimos gráficos. En el 9º gráfico, tenemos representadas el número de camas de hospital existentes por comunidad autónoma, donde podemos observar, que, pese a no ser la comunidad autónoma más poblada, Cataluña sí que es la que mayor número de camas de hospital ostenta, y además con mucha diferencia con respecto de las demás.

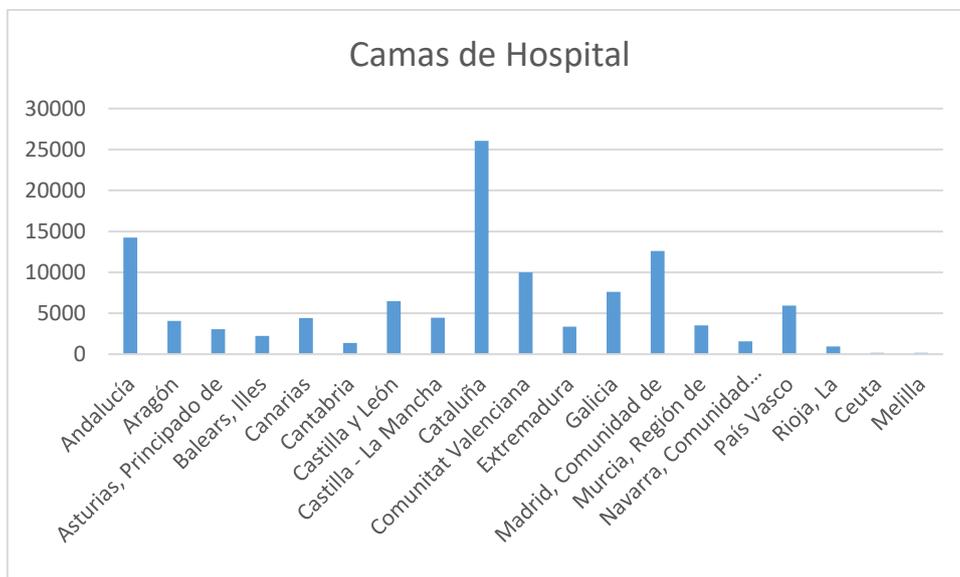


Figura 9. Cantidad de Camas de Hospital (Elaboración propia a partir del INE y DATOS-MACRO)

Finalmente, en el gráfico número 10, tenemos el presupuesto destinado a sanidad, por cada comunidad autónoma. Sigue un poco la corriente de las demás variables en cuanto a proporcionalidad, destacando por encima de las demás Andalucía, Madrid, o Cataluña.



Figura 10. Presupuesta de Sanidad asignado a las CCAA (Elaboración propia a partir del INE y DATOS-MACRO)

### 3.3 ANÁLISIS FACTORIAL

Previo al procedimiento del análisis factorial, se realizó el acondicionamiento de los datos estudiados por medio de la estandarización de las variables, de modo que los datos presentan las mismas dimensiones para su comparación (ver Anexo A).

Para la prueba de normalidad multivariante por el método Henze-Zirkler se obtuvo un p value de  $1,06e-11$ , el cual es menor a alfa de 0,05. De acuerdo con las consideraciones se rechaza la  $H_0$ , no existe normalidad multivariante.

Los resultados obtenidos de la matriz de correlaciones para cada muestra de las variables y la determinación de los indicadores de aplicabilidad del AFE por medio del contraste de esfericidad de Bartlett arrojaron un p value menor al alfa (ANEXO A), lo que confirma la aplicabilidad del análisis factorial

En la evaluación del descarte de las variables a través de la medida de adecuación muestral de Kaiser, Meyer y Ocklin (KMO) determinó que los valores obtenidos son superiores al valor mínimo aceptable de 0,5. Estableciendo la aceptación de todas las variables ya que los resultados se encuentran entre 0,51 y 0,92.

#### 3.3.1 Extracción de Número de Factores

Para selección del método de extracción fueron comparados el método de componentes principales (ACP) con el método de componentes principales iteradas o ejes principales (CPI). De ambos métodos se obtuvieron 3 factores, la Tabla 2 muestra los resultados de cada método

Tabla 2. Comparación del método de componentes principales y el método de componentes principales iteradas o ejes principales

ANÁLISIS ESTADÍSTICO MULTIVARIANTE DEL IMPACTO DEL COVID-19 POR COMUNIDADES AUTÓNOMAS

Método	VAR (%)			RMSR
	Factor 1	Factor 2	Factor 3	
ACP	69	19	8	0,02
CPI	69	19	5	0,01

En base a la teoría se debe seleccionar el método con resultado de RMSR más bajo y proporción de varianza más elevada. Como ambos métodos arrojaron un RMSR pequeño fue seleccionado el método ACP por poseer la proporción de varianza más alta.

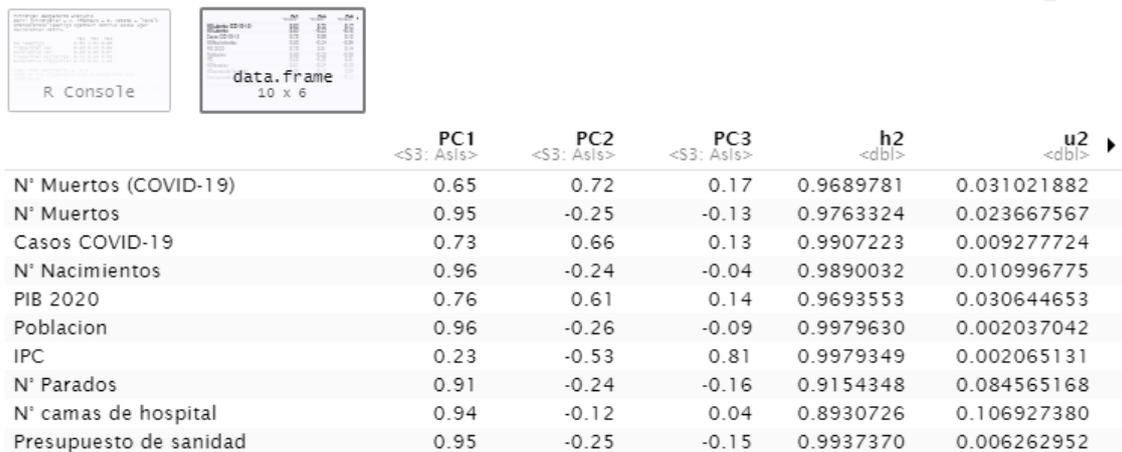


Figura 11. Resultados del método de componentes principales (ACP) por RStudio.

La figura 11, muestra los resultados obtenidos por el método ACP. Dichos resultados demuestran que el modelo es eficaz y los factores son representativos de las variables de estudio. Además de establecer cuáles variables están asignadas a cada uno de los factores:

**Factor 1: PC1**

Este factor está constituido por N° de muertos, N° de nacimientos, población, N° de parados, N° de camas de hospital y el presupuesto de sanidad, siendo el factor que explica la correlación existente entre ellas. Basado en esto podemos asumir que el PC1 describe el ámbito poblacional, así como los aspectos económicos dependientes de él, como las variables del sistema de salud, aspecto que es similar al análisis descriptivo. Las variables que tienen mayor impacto sobre este factor son el número de nacimientos y la población.

**Factor 2: PC2**

Este factor está constituido por N° de muertos (COVID19), Casos COVID-19 y PIB2020, explicando directamente la relación existente entre el impacto de la pandemia y la economía de cada región, por lo que a mayor PIB mayor es el número de casos COVID y muertos asociados a la enfermedad en la mayoría de las CCAA. La variable que mayor

impacto tiene en este factor es el número de muertos por COVID-19 y los casos COVID-19, de igual manera se observa que las variables asociadas al PC1 y el IPC tiene una relación inversa con respecto al PC2.

**Factor 3: PC3**

Este factor está constituido únicamente por el IPC, sin embargo, lo destacable de este factor es que tiene una relación inversa con el N° de muertos, N° de nacimientos, población, N° de parados y el presupuesto de sanidad.

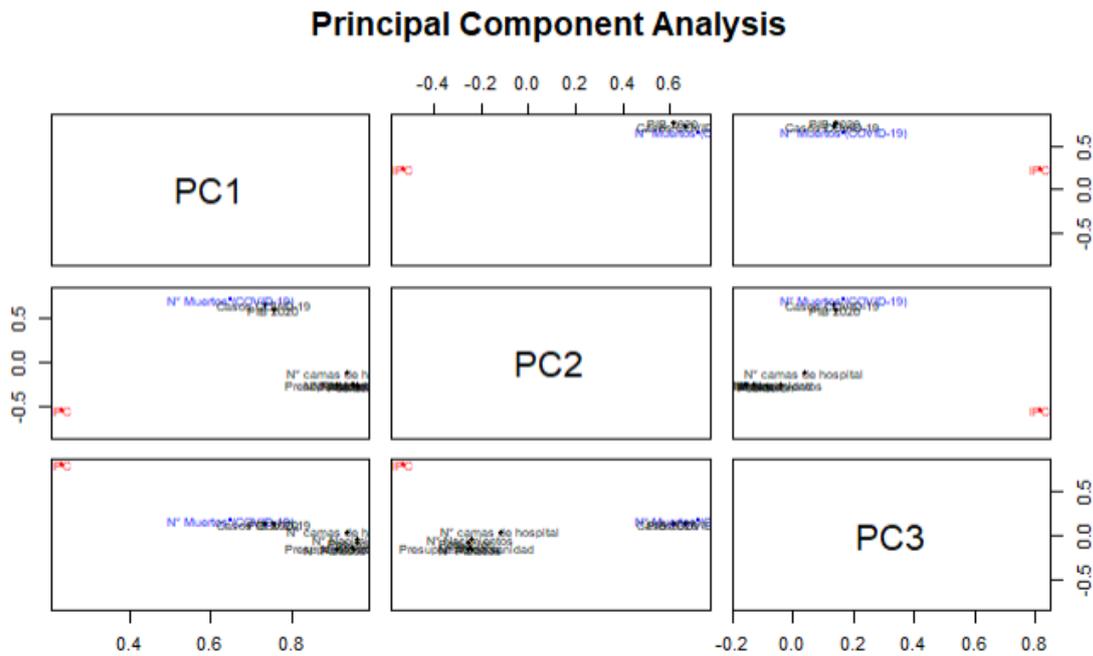


Figura 12. Representación gráfica de componentes extraídos

**3.3.2 Determinación de las Puntuaciones Factoriales**

A partir del modelo creado y la ejecución del software estadístico, se determinaron las puntuaciones factoriales y la matriz de carga visualizados en la Figura 13 y 14 respectivamente.

## ANÁLISIS ESTADÍSTICO MULTIVARIANTE DEL IMPACTO DEL COVID-19 POR COMUNIDADES AUTÓNOMAS

```
> acp1<- principal(DatosT[,2:11], nfactors = 3, rotate = "none", scores = T)
> acp1$scores
      PC1      PC2      PC3
[1,]  2.184175283 -0.451817913 -1.17703212
[2,] -0.412176093  0.136670617 -0.02553126
[3,] -0.633666903  0.005607594 -0.63866979
[4,] -0.547627532 -0.615100180  1.45889269
[5,] -0.253398081 -0.373906609 -0.45781544
[6,] -0.835763502 -0.135747290 -0.19693921
[7,] -0.009840841  0.249075624 -0.56781904
[8,] -0.067914778  0.712697870 -0.93007523
[9,]  2.496041988  0.496422161  1.50416430
[10,]  0.859098461 -0.089472805 -0.79110923
[11,] -0.620342917  0.159541419 -1.43255038
[12,] -0.034964978 -0.157630186 -0.82512456
[13,]  1.037711332 -2.349583402  0.63271108
[14,]  0.443133601  3.040596453  1.13532889
[15,] -0.745950002  0.230549637 -0.66523103
[16,] -0.241388558 -0.457879775 -0.27932167
[17,] -0.663954676  0.788293711  0.20968194
[18,] -0.968920722 -0.664226452  1.77444707
[19,] -0.984251083 -0.524090474  1.27199297
```

Figura 13. Puntuaciones Factoriales

```
> acp1<- principal(DatosT[,2:11], nfactors = 3, rotate = "none", scores = T)
> acp1$scores
      PC1      PC2      PC3
[1,]  2.184175283 -0.451817913 -1.17703212
[2,] -0.412176093  0.136670617 -0.02553126
[3,] -0.633666903  0.005607594 -0.63866979
[4,] -0.547627532 -0.615100180  1.45889269
[5,] -0.253398081 -0.373906609 -0.45781544
[6,] -0.835763502 -0.135747290 -0.19693921
[7,] -0.009840841  0.249075624 -0.56781904
[8,] -0.067914778  0.712697870 -0.93007523
[9,]  2.496041988  0.496422161  1.50416430
[10,]  0.859098461 -0.089472805 -0.79110923
[11,] -0.620342917  0.159541419 -1.43255038
[12,] -0.034964978 -0.157630186 -0.82512456
[13,]  1.037711332 -2.349583402  0.63271108
[14,]  0.443133601  3.040596453  1.13532889
[15,] -0.745950002  0.230549637 -0.66523103
[16,] -0.241388558 -0.457879775 -0.27932167
[17,] -0.663954676  0.788293711  0.20968194
[18,] -0.968920722 -0.664226452  1.77444707
[19,] -0.984251083 -0.524090474  1.27199297
```

Figura 14. Matriz de Carga

### 3.4 ANÁLISIS CLÚSTER

Una vez determinados los factores que representan las correlaciones existentes entre las variables, se procede al análisis clúster con el objetivo de segmentar la población formando grupos de individuos en conjuntos o segmentos llamados conglomerados. De esta manera se establecerán las diferencias y similitudes existentes entre las comunidades, considerando para su formación como variables los factores extraídos en el análisis factorial y la premisa de homogeneidad interna y heterogeneidad externa entre los grupos formados.

#### 3.4.1 Validación de la Segmentación

De acuerdo con el método de Hopkins el valor obtenido debe ser menor o próximo a 0,5 para ser aceptable. En la Figura 15 se apreció el resultado proporcionado por RStudio el cual indicó que la data es significativamente agrupable, válida para realizar segmentación.

```
> hopkins(DatosCluster[,c(2:4)], n = nrow(DatosCluster[,c(2:4)])) - 1)
$H
[1] 0.4333065
```

Figura 15. Método de Hopkins por RStudio

Para confirmar el resultado obtenido, se corroboró con el método Visual Assessment of Clúster Tendency (VAT) de manera que podamos tener un primer vistazo de la formación de conglomerados apreciado en la siguiente figura:

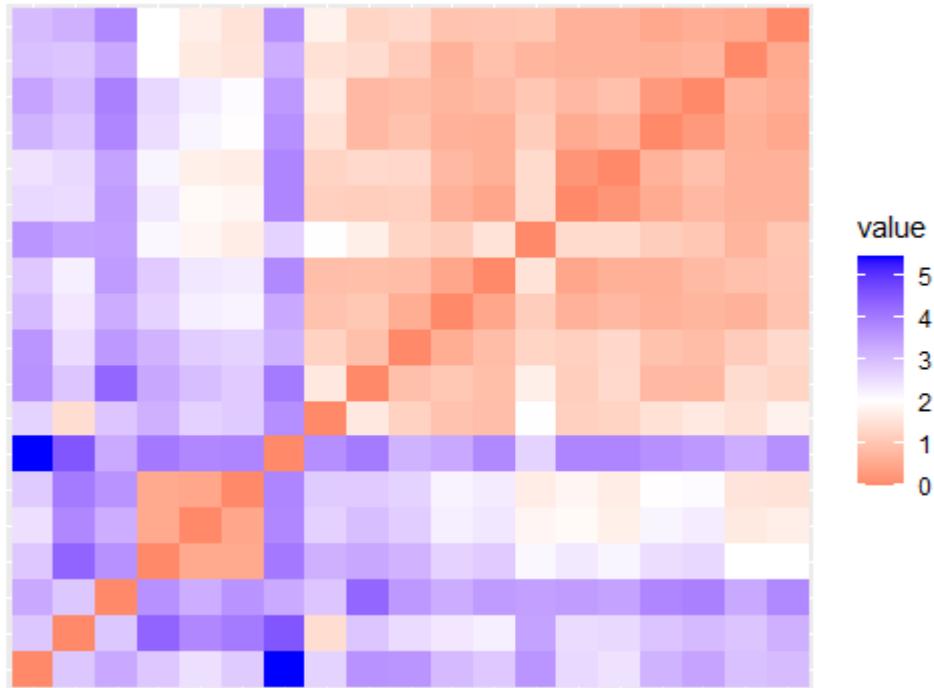


Figura 16 Método Visual Assessment of Cluster Tendency por RStudio

#### 4.4.2 Determinación del Número de Clúster

El software estadístico RStudio proporciona un algoritmo que ayuda a determinar la cantidad de conglomerados presentes en el análisis de los factores por medio del estadístico de GAP. En la Figura 17, se muestra la relación del estadístico de GAP y el número de conglomerados.

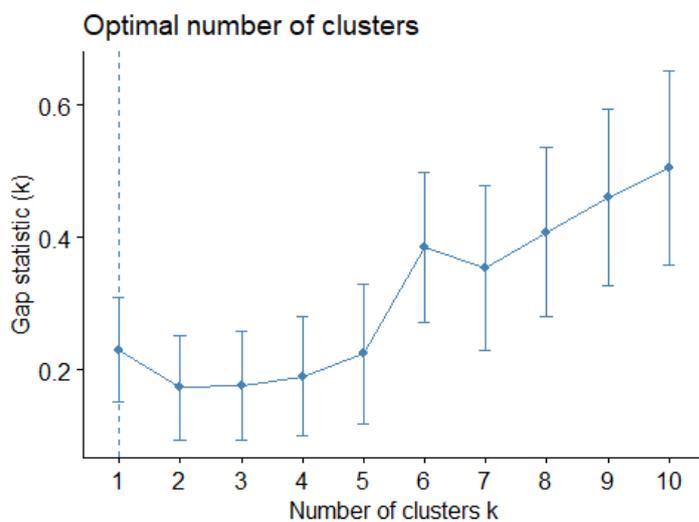


Figura 17. Número de Clúster por Estadístico de GAP

De acuerdo con el algoritmo, el software nos sugiere que se debe extraer 1 segmento de las CCAA, sin embargo, considerando la literatura y la población bajo estudio, se generarán modelos de 2 o 3 segmentos en los métodos de unión aplicables al análisis clúster. Por lo que se prosiguió con la determinación de los conglomerados a través de los métodos descritos en la sección metodológica.

- Método K-medoids (pam, Partición alrededor del medioide)

Por este método se obtuvo que el modelo más satisfactorio para los datos bajo estudio, es en formato por dos clústeres o grupos, el grupo 1 formado por 14 CCAA y el grupo 2 por 5 CCAA, visualizados en la siguiente figura:

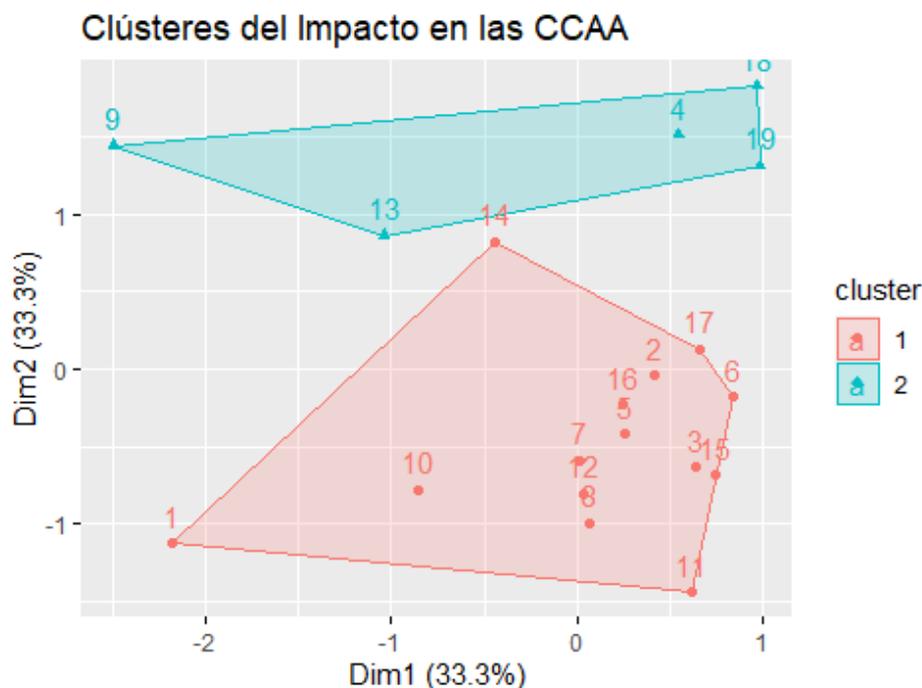


Figura 18. Formación de Conglomerados por Método K-medoids

De acuerdo con la representación numérica de los resultados del método empleado en la Figura 19; el grupo 1 está compuesto por: Andalucía, Aragón, Asturias, Canarias, Cantabria, Castilla y León, Castilla y La Mancha, Comunidad Valenciana, Extremadura, Galicia, Murcia, Navarra, País Vasco, La Rioja y el grupo 2 está conformado por: Baleares, Cataluña, Madrid, Ceuta y Melilla.

```
Medoids:
  ID      PC1      PC2      PC3
[1,]  7 -0.009840841  0.2490756 -0.567819
[2,]  4 -0.547627532 -0.6151002  1.458893
Clustering vector:
[1] 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 2
objective function:
  build  swap
1.187928 1.117489
```

Figura 19. Representación Numérica de la Formación de Conglomerados por Método K-medoids

- Método de Ward

Por este método se determinó que el modelo más satisfactorio para los datos bajo estudio es en formato por tres clústeres, el grupo 1 formado por 3 CCAA, el grupo 2 por 13 CCAA y el grupo 3 por 3 CCAA, observados en la siguiente figura:

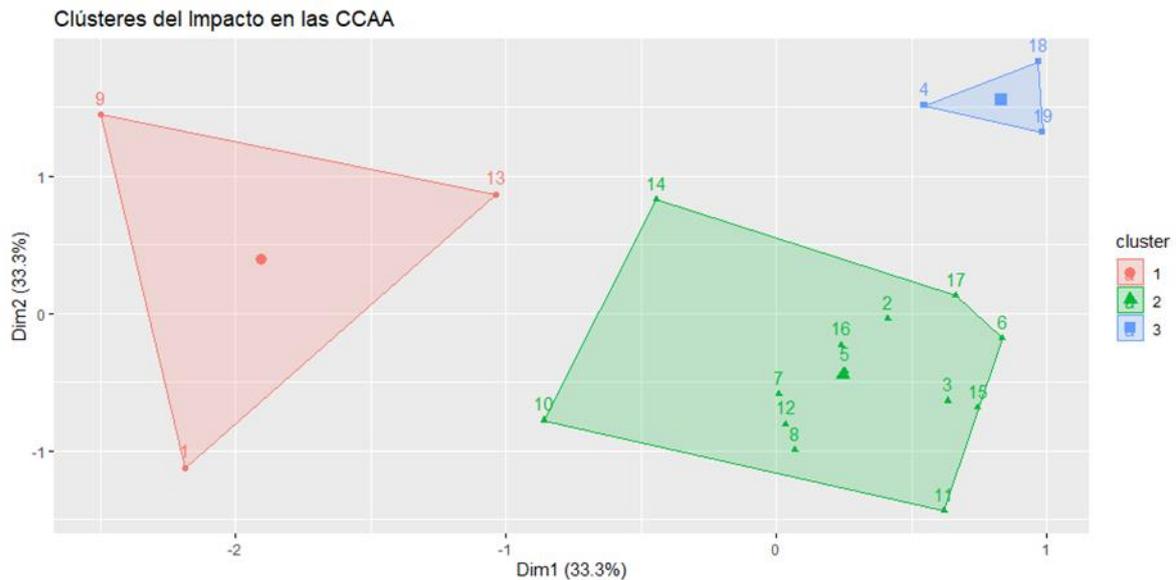


Figura 20 Formación de Clúster por Método de Ward

En base a la información suministrada por en el dendograma de la formación de clúster visualizado en la Figura 21. El grupo 1 está compuesto por: Andalucía, Cataluña y Madrid. El grupo 2 está compuesto por: Aragón, Asturias, Canarias, Cantabria, Castilla y León, Castilla y La Mancha, Comunidad Valenciana, Extremadura, Galicia, Murcia, Navarra, País Vasco, La Rioja. El grupo 3 está conformado por: Las Islas Baleares, Ceuta y Melilla.

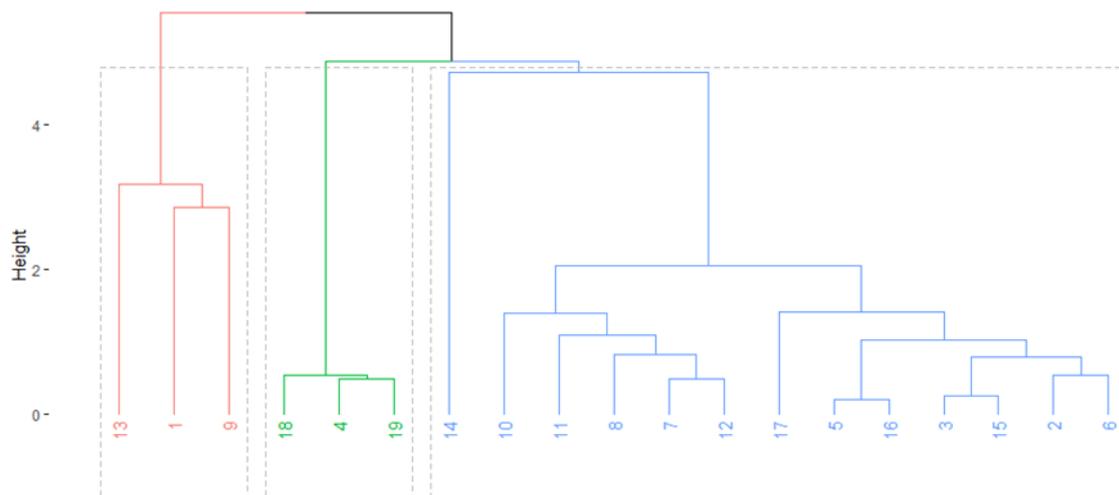


Figura 21. Dendograma de Clúster por Método de Ward

Considerando las similitudes y diferencias existentes en las comunidades, así como el efecto de los factores extraídos, con respecto a aspectos demográficos, económicos y de bienestar social (PC1, PC2 y PC3) sobre cada variable, se establece que el mejor modelo es el generado por el Método Ward. Se determina entonces los clústeres de nuestro análisis:

**Clúster 1:** este grupo presenta similitudes en relación con la población, siendo las comunidades con mayor población de España y por ende manifestando similitudes en relación a las variables demográficas y económicas relacionadas a la proporción poblacional. En este grupo de evidencia los hallazgos encontrados en el análisis factorial, resaltando el impacto del COVID-19 dado la cifra de habitantes en la zona a pesar de que se cuenta con un alto presupuesto de sanidad, sin embargo, la tasa de paro y el IPC también es significativa en estas regiones, catalogando a dichas zonas como vulnerables ante la crisis social y económica derivada de la pandemia. Tal relación se valida con la segmentación interna que se observa en el dendograma, el cual realiza una segmentación entre la comunidad de Madrid, Andalucía y Cataluña, quienes presentan mayor similitud con respecto al desarrollo económico-social e impacto del COVID-19.

**Clúster 2:** este grupo a pesar de sus similitudes está compuesto por pequeños segmentos que son más parecidos entre sí que otros. En general, presentan un desarrollo socioeconómico moderado y un bajo impacto del COVID-19 en comparación al clúster 1. Castilla y León, Castilla y La Mancha, la comunidad Valenciana, Extremadura y Galicia presentan similitudes demográficas y económicas, con un bajo desarrollo económico y social dado que presentan PIB y tasa de paro alto, así como una población similar entre ellas, con alto número de nacimientos y muertes; también se destaca que estas comunidades han presentado un alto número de contagios, pero con menor proporción de muertes en comparación a las comunidades del clúster 1. En cuanto a Aragón, Asturias, Cantabria, Canarias, Navarra, País Vasco y La Rioja, estas presentan un buen desarrollo económico y social con bajos niveles de IPC, un PIB moderado y una tasa de paro baja, el impacto de COVID es proporcional a la población con un buen presupuesto de sanidad y número de camas a excepción de La Rioja donde se registró un alto impacto de la enfermedad. Por último, en este clúster se encuentra la Murcia quien es una comunidad con un buen desarrollo económico y social, quien presenta al igual que las demás comunidades IPC bajo, pero una tasa de paro moderada y el PIB de la zona es el más alto de España y presentó un alto impacto del COVID debido al bajo nivel de inversión en el sistema de salud.

**Clúster 3:** este grupo presenta similitudes con respecto a la baja población, alto IPC, PIB y tasa de paro lo que manifiesta un desarrollo socioeconómico bajo, no obstante, presentan pocos casos y muertes asociados a la pandemia, siendo el impacto de la misma principalmente a nivel económico. Se puede decir que han sido las comunidades que mejor han manejado la crisis sanitaria o que han sido menos afectadas por ella, dado que presentan una cantidad baja de muertos y casos de COVID en comparación a otras comunidades, así como un presupuesto acorde a la población existente.

### 3.5 ANÁLISIS ANOVA

Con base a los resultados del análisis clúster, se segmentó la población bajo estudio estableciendo diferencias y semejanzas determinando la heterogeneidad intergrupala, sin embargo, para validar dicho modelo es necesario constatar que no existe homogeneidad intergrupala a partir del análisis de varianza ANOVA, para verificar que los grupos creados en el análisis clúster están correctamente segmentados.

Inicialmente se realizó la comparación de las medias para visualizar la existencia de diferencias entre ellas. Tal como se aprecia en la Figura 22, la comparación indica la existencia de diferencias significativas entre los conglomerados. La representación gráfica de estos resultados se encuentra en el Anexo B

```
> colMeans(c1[,1:3])
      PC1      PC2      PC3
1.9059762 -0.7683264  0.3199478
> colMeans(c2[,1:3])
      PC1      PC2      PC3
-0.2474715  0.3160305 -0.4203982
> colMeans(c3[,1:3])
      PC1      PC2      PC3
-0.8335998 -0.6011390  1.5017776
```

Figura 22. Comparación de Medias Significativas

#### 4.5.1 Test de Homogeneidad de Varianzas de Levene

Realizada la prueba, los resultados visualizados en la Figura 23 indican la presencia de heterocedasticidad entre los grupos. Ya que los valores de p value obtenidos fueron mayores al nivel de significancia por lo que fue rechazada la hipótesis nula demostrando que existe heterogeneidad intergrupala.

```
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 2  0.7828 0.4739
  16
```

(a)

```
> leveneTest(Datosward$PC2, Datosward$dendograma, center = median)
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 2  1.1198 0.3506
  16
```

(b)

```
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 2  1.5551 0.2415
  16
```

(c)

Figura 23. Teste de Homogeneidad de Varianza de Levene. a) PC1, b)PC2, c)PC3

#### 4.5.2 Determinación del Anova

De acuerdo con los resultados obtenidos por el software, mostrados en el Figura 24. En el modelo MPC1 la p value es significativamente menor a nivel de significancia 0,05; por lo que se rechazó la hipótesis nula y se concluyó que existen distintas medias entre los grupos. Sin embargo, en MPC2 y MPC3 el p value es superior a 0,05, por lo cual no se rechaza la hipótesis nula indicando que la media entre grupos es igual. No obstante, dado que el ANOVA es sensible a distribuciones no normales procedemos a ejecutar un test de Kruskal-Wallis para confirmar este resultado.

```
> MPC1<- aov(PC1 ~ dendograma, data= Datosward)
> summary(MPC1)
      Df Sum Sq Mean Sq F value Pr(>F)
dendograma  1 11.258  11.258   28.39 5.55e-05 ***
Residuals  17  6.742   0.397
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> MPC2<- aov(PC2 ~ dendograma, data= Datosward)
> summary(MPC2)
      Df Sum Sq Mean Sq F value Pr(>F)
dendograma  1  0.042  0.0419   0.04  0.844
Residuals  17 17.958  1.0564
> MPC3<- aov(PC3 ~ dendograma, data= Datosward)
> summary(MPC3)
      Df Sum Sq Mean Sq F value Pr(>F)
dendograma  1  2.095  2.0951   2.239  0.153
Residuals  17 15.905  0.9356
~|
```

Figura 24. Determinación del ANOVA

#### 4.5.3 Test de Kruskal Wallis

Realizado el test estadístico con resultados mostrados en el Figura 25, se observó que los P value de los 3 factores evaluados indican que al menos 1 grupo presenta media diferente por lo cual se rechaza la hipótesis nula y se concluye que la segmentación realizada en el estudio es válida.

```
> kruskal.test(Datosward$PC1, Datosward$dendograma)

Kruskal-wallis rank sum test

data:  Datosward$PC1 and Datosward$dendograma
Kruskal-wallis chi-squared = 9.9239, df = 2, p-value = 0.006999
> kruskal.test(Datosward$PC2, Datosward$dendograma)

Kruskal-wallis rank sum test

data:  Datosward$PC2 and Datosward$dendograma
Kruskal-wallis chi-squared = 7.0623, df = 2, p-value = 0.02927
> kruskal.test(Datosward$PC3, Datosward$dendograma)

Kruskal-wallis rank sum test

data:  Datosward$PC3 and Datosward$dendograma
Kruskal-wallis chi-squared = 6.9053, df = 2, p-value = 0.03166
```

Figura 25. Test de Kruskal Wallis

## 5 CONCLUSIONES.

Comparando los estadísticos de las diferentes variables estudiadas en el análisis descriptivo, se observó un mayor impacto de la pandemia en el ámbito de bienestar, salud y sistema. Como Andalucía y Cataluña poseen mayor población presentaron mayor número de casos y muertes registradas por la enfermedad. Sin embargo, Cataluña arrojó un mayor número de muertos en Cataluña; duplicando a Andalucía a pesar que mantuvieron una cantidad proporcional de casos. Por otro lado, Murcia y La Rioja con baja población, mostraron un alto número de contagios y muertes. Siendo Murcia la comunidad con mayor número de muertes asociadas al covid sobrepasando las muertes generales de la región, a pesar de contar con un presupuesto y número de camas similar a otras regiones con igual número de habitantes. A su vez, en Murcia se destaca con un alto número de parados e IPC bajo.

Del análisis factorial fueron extraídos tres factores. El Factor 1 formado por N° de muertos, N° de nacimientos, población, N° de parados, N° de camas de hospital y el presupuesto de sanidad; establece que a mayor población mayor número de muertos en general, número de nacimientos, camas de hospital, presupuesto. El Factor 2 constituido por N° de muertos (COVID19), Casos COVID-19 y PIB 2020, explica que a mayor valor de Factor 1 e IPC mayor es la afectación del Factor 2, impidiendo el desarrollo económico y la capacidad de afrontar una emergencia sanitaria. Por último, el Factor 3 conformado por el IPC, determina que a mayor N° de muertos, N° de nacimientos, población, N° de parados y el presupuesto de sanidad menor es este Factor.

Del Análisis clúster se conformaron 3 conglomerados. El Conglomerado 1 junta las comunidades con mayor cantidad poblacional y su relación proporcionalidad a variables demográficas y económicas (Andalucía, Cataluña y Madrid), es decir a nivel económico y social este grupo de comunidades tuvo mayor impacto ante la crisis sanitaria. El conglomerado 2 (Aragón, Asturias, Canarias, Cantabria, Castilla y León, Castilla y La Mancha, Comunidad Valenciana, Extremadura, Galicia, Murcia, Navarra, País Vasco, La Rioja) presentó similitudes en cuanto a valores moderados en variables económico-sociales y respuesta ante crisis sanitaria. El conglomerado 3 asocia las comunidades autónomas (Las Islas Baleares, Ceuta y Melilla.) que presentaron el impacto de la pandemia a nivel de desarrollo económico (baja población y alto IPC, PIB y cantidad de paro) y una mejor respuesta a nivel sanitaria (baja cantidad de muertos y casos por COVID-19).

## BIBLIOGRAFÍA

FACME; Instituto de Salud y Estrategia . (2021). *Impacto de la COVID-19 en lo NO-COVID-19. Estrategias para “re-priorizar” la atención NO-COVID en el Sistema Nacional de Salud*. Madrid: Si-Health.

INE, Instituto Nacional de Estadística. <https://ine.es/>

RStudio. <https://www.rstudio.com/products/rstudio/>

Enríquez, A., & Sáenz, C. (2021). *Primeras lecciones y desafíos de la pandemia de COVID-19 para los países del SICA. Serie Estudios y Perspectivas*. Ciudad de México: Comisión Económica para América Latina y el Caribe (CEPAL).

Felgueroso, F., & De la Fuente, A. (2021). *Estudios sobre la Economía Española - 2021/24. Aspectos económicos de la crisis del Covid-19: Boletín de seguimiento no. 11*. Madrid: Fundación de Estudios de Economía Aplicada (FEDEA).

Felgueroso, F., De la Fuente, A., Boscá, J. E., Costa i Font, J., Doménech, R., Ferri, J., y otros. (2020). *Estudios sobre la Economía Española - 2020/14. Aspectos económicos de la crisis del Covid-19. Boletín de seguimiento no. 2*. Madrid: Fundación de Estudios de Economía Aplicada (FEDEA).

Página Web Datos Macro. <https://datosmacro.expansion.com/>

Martínez, A. (2021). *La irrupción del COVID-19 en la economía española. Trabajo fin de Grado*. Universidad de Alicante.

Mejía, J. (2017). *Las ciencias de la administración y el análisis multivariante. Tomo II: las técnicas interdependientes*. Zapopan, Jalisco: Universidad de Guadalajara.

Meneses, J. (2019). *Introducción al análisis multivariante. Universitat Oberta de Catalunya*. FUOC PID\_00263801 .

Sordo García, Carmen María (2021). *Análisis Factorial*. Universidad de Cantabria.

Sordo García, Carmen María (2021). *Análisis Clúster*. Universidad de Cantabria.

Sordo García, Carmen María (2021). *Análisis Anova*. Universidad de Cantabria.

Muñoz, S. (2015). *Análisis Multivariante: soluciones eficientes e interpretables. Tesis Doctoral*. Universidad Carlos III de Madrid.

Observatorio Sectorial DBK. (2021). *COVID-19: Impacto en los Principales Sectores de la Economía Española*. Informa D&B.

Ocaña, C., Bandres, E., Culiá, E., Fernández, M. J., Malo, M. Á., Rodríguez, J. C., y otros. (2021). *Impacto social de la pandemia en España. Una evaluación preliminar*. Madrid: Funcas.

Organización Mundial de la Salud. (2020). *Coronavirus*. Recuperado el 25 de enero de 2022, de Sitio web de la OMS: <https://www.who.int/es/health-topics/coronavirus>

Organización Panamericana de la Salud. (2020). *Enfermedad por el Coronavirus (COVID-19)*. Recuperado el 25 de enero de 2022, de Sitio web de la OPS: <https://www.paho.org/es/enfermedad-por-coronavirus-covid-19>

Pinilla Domínguez, J., & González López, B. (2021). Impacto de la COVID-19 sobre la sanidad y la economía en España. // (100 ).

Ramón-Dangla, R., Rico-Gómez, M. L., & Issa-Khozouz, R. (2021). Mortalidad por COVID-19 en España durante la primera oleada de la Pandemia. Aproximación a través del Gasto Público Sanitario por Comunidades Autónomas. *Revista Esp Salud Pública* , 95 (e202103036).

Trapero-Bertran, M., & Lobo, F. (2021). El gasto público en servicios de prevención y salud pública en España antes de la COVID-19. Los datos nacionales. *Funcas. Cuadernos de Información Económica*. (280), 59-71.

Tusell, F. (2016). *Análisis Multivariante*. País Vasco.

Poza Lara, Carlos (2005): "*Análisis estadístico multivariante por comunidades autónomas*". Universidad Antonio de Nebrija.

## ANEXO A ANÁLISIS FACTORIAL

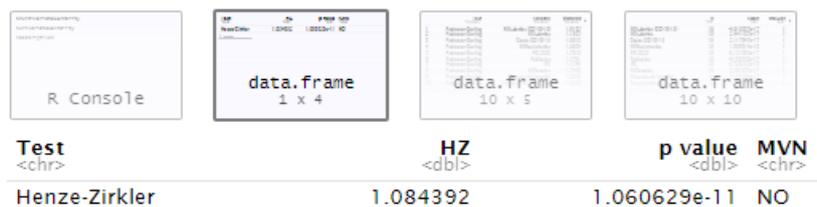
### Anexo A.1 Estandarización de datos

CCAA	N° Muertos COVID-19	N° Muertos	Casos COVID-19	N° Nacimiento	PIB 2020	Población	IPC	N° Parados	N° camas de hospital	Presupuesto de sanidad
1	0,71	2,08	0,14	0,02	0,14	0,23	-0,10	3,05	1,32	2,31
2	-0,08	-0,36	-0,20	-0,44	-0,38	-0,45	-0,20	-0,49	-0,29	-0,42
3	-0,44	-0,45	-0,65	-0,66	-0,57	-0,57	-0,69	-0,58	-0,46	-0,50
4	-0,69	-0,69	-0,57	-0,43	-0,43	-0,51	1,42	-0,37	-0,58	-0,53
5	-0,56	-0,38	-0,65	-0,24	-0,25	-0,12	-0,21	0,36	-0,24	-0,09
6	-0,73	-0,77	-0,73	-0,73	-0,70	-0,74	-0,30	-0,70	-0,72	-0,78
7	0,46	0,41	0,02	-0,21	-0,31	-0,04	-0,61	-0,29	0,09	0,03
8	0,68	0,00	0,24	-0,16	-0,08	-0,17	-1,14	-0,10	-0,23	-0,13
9	2,26	2,14	0,22	0,02	0,23	0,20	0,14	0,15	3,20	0,20
10	0,10	0,90	0,56	0,90	0,69	0,99	-0,39	0,87	0,65	1,03
11	-0,53	-0,51	-0,56	-0,53	-0,61	-0,55	-1,43	-0,39	-0,40	-0,52
12	-0,43	0,28	-0,36	-0,13	-0,01	0,08	-0,64	-0,23	0,27	0,21
13	-0,68	1,62	-0,72	1,73	-0,77	1,65	2,01	1,22	1,06	1,48
14	2,68	-0,54	2,54	-0,21	2,39	-0,38	-0,54	-0,36	-0,38	-0,47
15	-0,62	-0,76	-0,36	-0,65	-0,45	-0,71	-0,86	-0,68	-0,69	-0,71
16	-0,47	-0,06	-0,52	-0,16	-0,61	-0,11	-0,07	-0,40	0,00	0,17
17	0,05	-0,88	0,10	-0,79	0,12	-0,84	-0,40	-0,77	-0,79	-0,92
18	-0,85	-1,00	-0,85	-0,86	-0,87	-0,93	1,60	-0,80	-0,91	-1,06
19	-0,85	-1,00	-0,84	-0,86	-0,87	-0,93	1,10	-0,80	-0,91	-0,11

### Anexo A.2 Prueba de Normalidad Multivariantes por Henze-Kirkler. (a) Codificación en RStudio. b) Resultado de Henze-Kirkler

```
> DatosT<- DatosCCAA # Se crea un Data Frame
> DatosT<- scale(DatosT, center = T, scale = T)
> DatosT<- as.data.frame(DatosT)
> library(MVN)
> mvn(DatosT[2:11], subset = NULL, mvnTest = c("mardia","hz","royston", "dh", "energy"), covariance = TRUE, tol = 1e-25, alpha = 0.5, scale = FALSE, desc = TRUE, transform = "none", R = 1000, univariateTest = c("sw", "CVM", "Lillie", "SF", "AD"), univariatePlot = "none", multivariatePlot = "none", multivariateOutlierMethod = "none", bc = FALSE, bcType = "rounded", showOutliers = FALSE, showNewData = FALSE)
```

(a)



Test	HZ	p value	MVN
<chr>	<dbl>	<dbl>	<chr>
Henze-Zirkler	1.084392	1.060629e-11	NO

(b)

Anexo A.3 Matriz de Correlaciones

	N° Muertos (COVID-19)	N° Muertos	Casos COVID-19	N° Nacimientos	PIB 2020	Poblacion	IPC	N° Parados	N° camas de hospital	Presupuesto de sanidad
N° Muertos (COVID-19)	1.00	0.43	0.97	0.44	0.93	0.42	-0.10	0.37	0.54	0.41
N° Muertos	0.43	1.00	0.51	0.44	0.93	0.99	0.24	0.91	0.94	0.99
Casos COVID-19	0.97	0.51	1.00	0.44	0.93	0.99	0.24	0.91	0.94	0.99
N° Nacimientos	0.44	0.97	0.55	1.00	0.93	0.99	0.24	0.91	0.94	0.99
PIB 2020	0.93	0.53	0.98	0.93	1.00	0.99	0.24	0.91	0.94	0.99
Poblacion	0.42	0.99	0.52	0.93	0.99	1.00	0.24	0.91	0.94	0.99
IPC	-0.10	0.24	-0.07	0.93	0.99	1.00	0.24	0.91	0.94	0.99
N° Parados	0.37	0.91	0.50	0.93	0.99	1.00	0.24	0.91	0.94	0.99
N° camas de hospital	0.54	0.94	0.59	0.93	0.99	1.00	0.24	0.91	0.94	0.99
Presupuesto de sanidad	0.41	0.99	0.51	0.93	0.99	1.00	0.24	0.91	0.94	0.99
	N° Nacimientos	PIB 2020	Poblacion	IPC	N° Parados					
N° Muertos (COVID-19)	0.44	0.93	0.42	-0.10	0.37					
N° Muertos	0.97	0.53	0.99	0.24	0.91					
Casos COVID-19	0.55	0.98	0.52	-0.07	0.50					
N° Nacimientos	1.00	0.57	1.00	0.32	0.96					
PIB 2020	0.57	1.00	0.56	-0.03	0.54					
Poblacion	1.00	0.56	1.00	0.29	0.95					
IPC	0.32	-0.03	0.29	1.00	0.22					
N° Parados	0.96	0.54	0.95	0.22	1.00					
N° camas de hospital	0.91	0.63	0.92	0.30	0.80					
Presupuesto de sanidad	0.98	0.54	0.99	0.23	0.95					
	N° camas de hospital	Presupuesto de sanidad								
N° Muertos (COVID-19)	0.54	0.41								
N° Muertos	0.94	0.99								
Casos COVID-19	0.59	0.51								
N° Nacimientos	0.91	0.98								
PIB 2020	0.63	0.54								
Poblacion	0.92	0.99								
IPC	0.30	0.23								
N° Parados	0.80	0.95								
N° camas de hospital	1.00	0.92								
Presupuesto de sanidad	0.92	1.00								
Sample size										
[1] 19										

Probability values (Entries above the diagonal are adjusted for multiple tests.)

	N° Muertos (COVID-19)	N° Muertos	Casos COVID-19	N° Nacimientos	PIB 2020	Poblacion	IPC	N° Parados	N° camas de hospital	Presupuesto de sanidad
N° Muertos (COVID-19)	0.00	0.82	0.00	0.00	0.00	0.07	0.69	0.12	0.02	0.08
N° Muertos	0.06	0.00	0.44	0.06	0.06	0.07	0.32	0.00	0.02	0.00
Casos COVID-19	0.00	0.03	0.00	0.06	0.06	0.07	0.32	0.00	0.02	0.00
N° Nacimientos	0.06	0.00	0.01	1.00	0.01	0.07	0.32	0.00	0.02	0.00
PIB 2020	0.00	0.02	0.00	0.01	1.00	0.07	0.32	0.00	0.02	0.00
Poblacion	0.07	0.00	0.02	0.01	0.00	1.00	0.32	0.00	0.02	0.00
IPC	0.69	0.32	0.78	0.01	0.00	0.32	1.00	0.00	0.02	0.00
N° Parados	0.12	0.00	0.03	0.01	0.00	0.00	0.32	1.00	0.02	0.00
N° camas de hospital	0.02	0.00	0.01	0.01	0.00	0.00	0.32	0.00	1.00	0.00
Presupuesto de sanidad	0.08	0.00	0.03	0.01	0.00	0.00	0.32	0.00	0.00	1.00
	N° Nacimientos	PIB 2020	Poblacion	IPC	N° Parados					
N° Muertos (COVID-19)	0.80	0.00	0.85	1.00	1.00					
N° Muertos	0.00	0.39	0.00	1.00	0.00					
Casos COVID-19	0.34	0.00	0.39	1.00	0.44					
N° Nacimientos	0.00	0.25	0.00	1.00	0.00					
PIB 2020	0.01	0.00	0.33	1.00	0.36					
Poblacion	0.00	0.01	0.00	1.00	0.00					
IPC	0.18	0.89	0.23	0.00	1.00					
N° Parados	0.00	0.02	0.00	0.36	0.00					
N° camas de hospital	0.00	0.00	0.00	0.21	0.00					
Presupuesto de sanidad	0.00	0.02	0.00	0.34	0.00					
	N° camas de hospital	Presupuesto de sanidad								
N° Muertos (COVID-19)	0.36	0.88								
N° Muertos	0.00	0.00								
Casos COVID-19	0.20	0.44								
N° Nacimientos	0.00	0.00								
PIB 2020	0.11	0.36								
Poblacion	0.00	0.00								
IPC	1.00	1.00								
N° Parados	0.00	0.00								
N° camas de hospital	0.00	0.00								
Presupuesto de sanidad	0.00	0.00								

ANÁLISIS ESTADÍSTICO MULTIVARIANTE DEL IMPACTO DEL COVID-19 POR COMUNIDADES AUTÓNOMAS

Anexo A.4 Indicadores de Aplicabilidad del AFE (Bondad del ajuste). a) Contraste de Esfericidad de Bartlett, b) • Medida de Adecuación Muestral de Kaiser, Meyer y Ocklin (KMO)

```
> dim(DatosT) # Tamaño de la muestra = 19 CCAA
[1] 19 11
> cor.test.bartlett(r, n= 19)
$chisq
[1] 412.4709

$ p.value
[1] 7.283645e-61

$df
[1] 45
```

(a)

```
> KMO(r)
Kaiser-Meyer-olkin factor adequacy
Call: KMO(r = r)
Overall MSA = 0.66
MSA for each item =
  N° Muertos (COVID-19)      0.80
  N° Muertos                 0.64
  N° Nacimientos             0.62
  PIB 2020                   0.51
  IPC                        0.52
  N° Parados                 0.92
  Presupuesto de sanidad    0.78
  Casos COVID-19            0.52
  Poblacion                  0.61
  N° camas de hospital       0.78
```

(b)

Anexo A.5 Método para la Extracción de Factores. a) Método de Componentes Principales (ACP), b) Método de Componentes Principales iteradas o Ejes principales (CPI)



	PC1 <S3: AsIs>	PC2 <S3: AsIs>	PC3 <S3: AsIs>	h2 <dbl>	u2 <dbl>
N° Muertos (COVID-19)	0.65	0.72	0.17	0.9689781	0.031021882
N° Muertos	0.95	-0.25	-0.13	0.9763324	0.023667567
Casos COVID-19	0.73	0.66	0.13	0.9907223	0.009277724
N° Nacimientos	0.96	-0.24	-0.04	0.9890032	0.010996775
PIB 2020	0.76	0.61	0.14	0.9693553	0.030644653
Poblacion	0.96	-0.26	-0.09	0.9979630	0.002037042
IPC	0.23	-0.53	0.81	0.9979349	0.002065131
N° Parados	0.91	-0.24	-0.16	0.9154348	0.084565168
N° camas de hospital	0.94	-0.12	0.04	0.8930726	0.106927380
Presupuesto de sanidad	0.95	-0.25	-0.15	0.9937370	0.006262952

```
> acp<- principal(r, nfactors = 3, rotate = "none")
> acp
Principal Components Analysis
Call: principal(r = r, nfactors = 3, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix

          PC1  PC2  PC3
SS loadings  6.95 1.94 0.80
Proportion Var 0.69 0.19 0.08
Cumulative Var 0.69 0.89 0.97
Proportion Explained 0.72 0.20 0.08
Cumulative Proportion 0.72 0.92 1.00

Mean item complexity = 1.5
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is 0.02
```

(a)

	PA1 <S3: Axis>	PA2 <S3: Axis>	PA3 <S3: Axis>	h2 <dbl>	u2 <dbl>
N° Muertos (COVID-19)	0.65	0.71	0.10	0.9389116	0.061088401
N° Muertos	0.95	-0.26	-0.11	0.9739534	0.026046563
Casos COVID-19	0.74	0.67	0.08	1.0042652	-0.004265178
N° Nacimientos	0.97	-0.25	0.00	0.9953196	0.004680366
PIB 2020	0.75	0.61	0.09	0.9529671	0.047032886
Poblacion	0.96	-0.27	-0.06	1.0074240	-0.007423985
IPC	0.22	-0.43	0.67	0.6813568	0.318643188
N° Parados	0.90	-0.24	-0.13	0.8817250	0.118275047
N° camas de hospital	0.92	-0.12	0.07	0.8668869	0.133113111
Presupuesto de sanidad	0.96	-0.27	-0.14	1.0037671	-0.003767071

```
> cpi
Factor Analysis using method = pa
Call: fa(r = r, nfactors = 3, n.obs = 19, rotate = "none", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix

          PA1  PA2  PA3
SS loadings  6.91 1.87 0.53
Proportion Var 0.69 0.19 0.05
Cumulative var 0.69 0.88 0.93
Proportion Explained 0.74 0.20 0.06
Cumulative Proportion 0.74 0.94 1.00

Mean item complexity = 1.5
Test of the hypothesis that 3 factors are sufficient.

The degrees of freedom for the null model are 45 and the objective function was 29.82 with Chi Square of 412.47
The degrees of freedom for the model are 18 and the objective function was 15205693

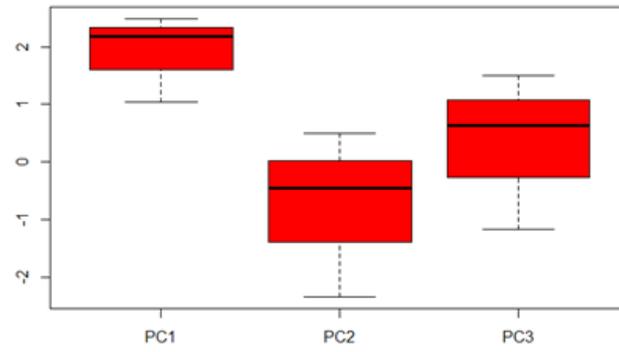
The root mean square of the residuals (RMSR) is 0.01
The df corrected root mean square of the residuals is 0.02
```

(b)

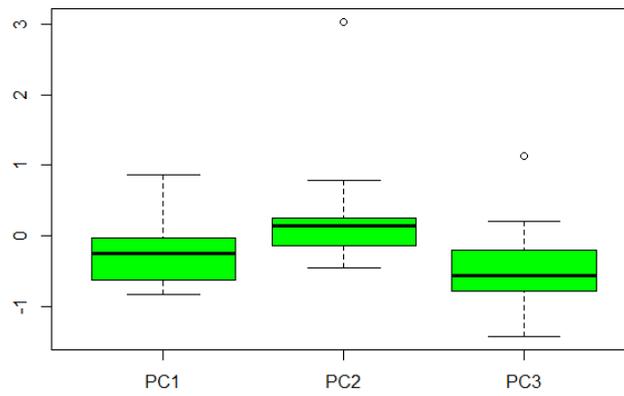
## ANEXO B ANÁLISIS ANOVA

Anexo B.1 Comparación de Medias Significativas. A) Clúster 1; b) Clúster 2; c) Clúster 3

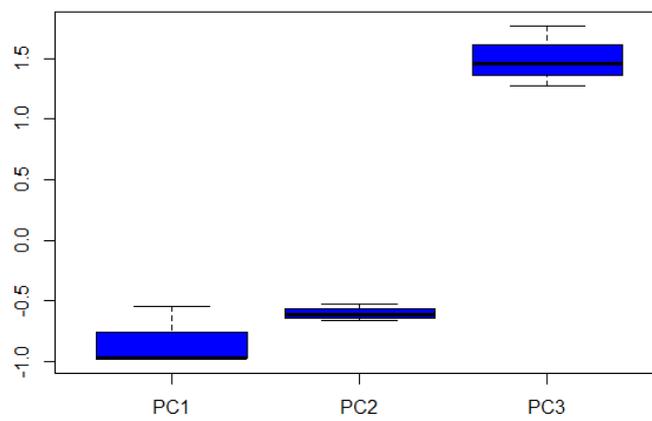
ANÁLISIS ESTADÍSTICO MULTIVARIANTE DEL IMPACTO DEL COVID-19 POR COMUNIDADES AUTÓNOMAS



(a)



(b)



(c)