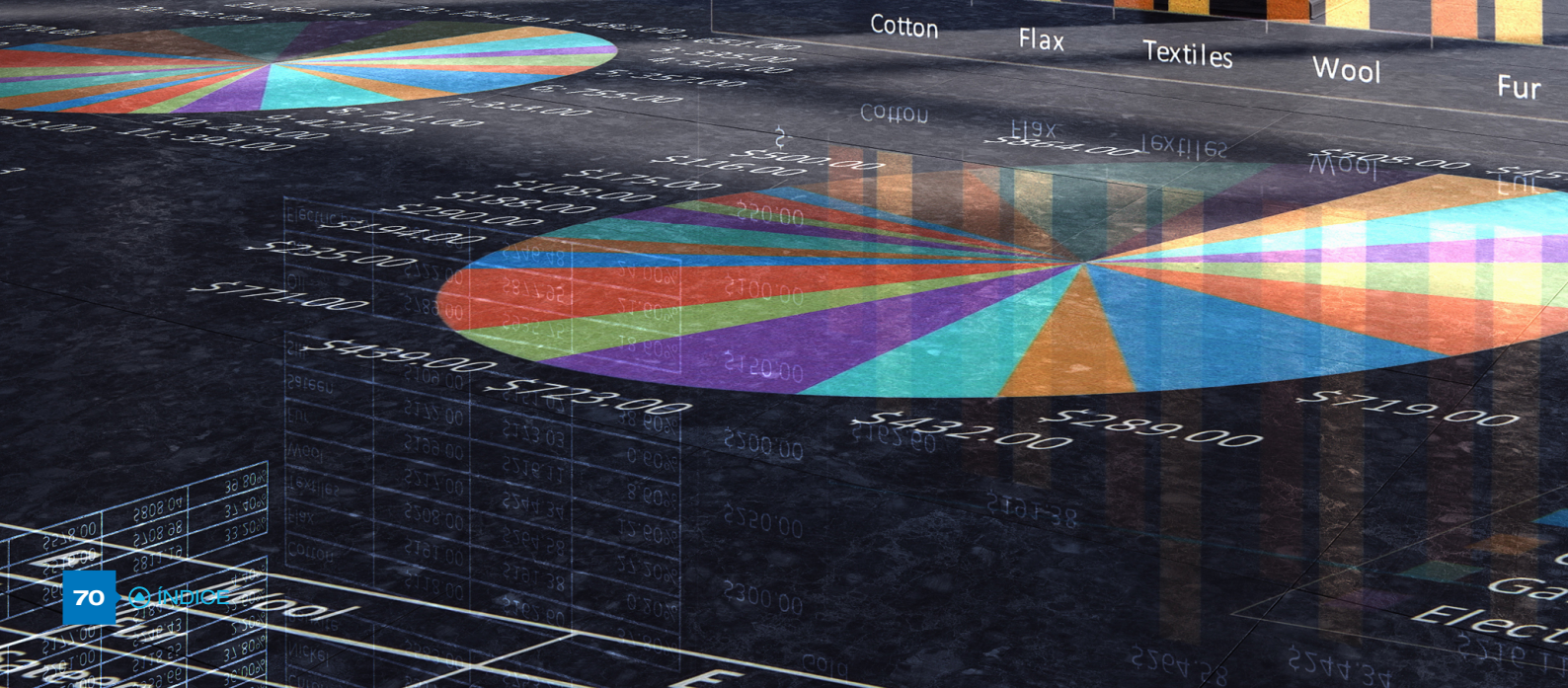


permite identificar similitudes entre las unidades de estudio y determinar los valores atípicos.



Representación gráfica multivariante del seguimiento de ejecución de proyectos de inversión en Perú, 2021

Multivariate Graphic Representation of the Monitoring of the Execution of Investment Projects in Peru, 2021

RESUMEN

El objetivo principal de este trabajo es mostrar las diferentes propuestas de gráficos multivariantes en seguimiento de proyectos de inversión tales como diagramas de caja y bigotes, gráficos de tallo y cajas, diagramas de dispersión, dendrograma, caras de Chernoff, curvas de Andrews, gráficos de estrella y gráficos de radar. Este estudio se justifica debido a que no existe mucha difusión de este tipo de gráficos y además no todos los softwares estadísticos permiten implementar estos, siendo una excepción a esto RStudio.

Se trata de un estudio de alcance exploratorio y diseño no experimental, transversal y descriptivo. La población está formada por 23 departamentos del Perú junto con la provincia constitucional del Callao. Debido a que la información para el departamento de Tumbes resultó estar incompleta, se optó por eliminar esta. Por tanto, la muestra final estuvo conformada por 23 observaciones y 7 variables con información sobre al seguimiento de ejecución de proyectos de inversión en Perú en el año 2021. Se hizo uso de técnicas gráficas de análisis multivariado utilizando el entorno de desarrollo integrado RStudio para su procesamiento.

Los resultados exponen diversas tendencias, conglomerados y análisis descriptivo de las 23 observaciones de la muestra respecto al seguimiento de ejecución en proyectos de inversión a través de los diversos gráficos multivariantes.

Podemos concluir que, mediante estos gráficos, se pudieron identificar similitudes entre las unidades de estudio, así como agrupamientos y valores atípicos (*outliers*). Además, se puede describir cómo fue la ejecución de proyectos de inversión en cada uno de los departamentos y la provincia del Callao que forman parte de la muestra.

ABSTRACT

The main objective of this work is to show the different proposals of multivariate graphs in the monitoring of investment projects such as box and whisker plots, stem and box plots, scatter plots, dendrogram, Chernoff faces, Andrews curves, graphs of star and radar charts. This study is justified because there is not much dissemination of this type of graphics, and neither all statistical software allows to implement them RStudio is an exception.

It is an exploratory study with a non-experimental, cross-sectional, and descriptive design. The population was made up of 23 departments of Peru together with the Constitutional Province of Callao. Since the information for the department of Tumbes turned out to be incomplete, it had decided to eliminate it. Therefore, the final sample consisted of 23 observations and 07 variables with information on monitoring the execution of investment projects in Peru in the year 2021. Multivariate analysis graphical techniques had used using the RStudio integrated development environment for their prosecution.

The results show various trends, clusters, and descriptive analysis of the 23 observations of the sample regarding the monitoring of execution in investment projects through the different multivariate graphs.

We can conclude that through these graphs it was possible to identify similarities between the study units as well as groupings and atypical values (outliers). In addition, it is possible to describe how was the execution of investment projects in each of the departments and the province of Callao that are part of the sample.



Palabras Claves

Gráficos multivariantes, Proyectos de inversión, análisis descriptivo, RStudio, conglomerados.

Key words

Multivariate charts, Investment projects, descriptive analysis, Rstudio, clustering.

INTRODUCCIÓN

La consecución de mayores niveles de productividad en el uso de recursos públicos tiene relevancia tanto en un sentido microeconómico y macroeconómico [13]. En las instituciones gubernamentales es de vital importancia conocer la forma cómo se están ejecutando los presupuestos de inversión a través de su ejecución en años anteriores, presupuesto de apertura, presupuesto modificado, devengado, avance y ejecución total [12]. Es por ello que casi siempre estamos enfrentando situaciones de estudio multidimensionales, referidas a información naturalmente compleja. Familiarizarnos con tal información requiere un análisis exploratorio de datos exhaustivo, no solamente de la forma unidimensional y bidimensional, sino también multidimensional [18]. Si bien desde hace un tiempo atrás se han desarrollado diversos procedimientos gráficos para representar datos multivariantes, actualmente, con la disponibilidad de *softwares* estadísticos, lo cual nos permite implementarlos de manera más sencilla, se sugiere aplicar estos para complementar las estrategias básicas de análisis exploratorio de la información. Las representaciones gráficas novedosas pueden ayudarnos a alcanzar una mejor comprensión del problema de interés [18]. El objetivo de este trabajo es, por tanto, presentar una aplicación de algunas de las propuestas de gráficos estadísticos multivariantes, tales como: diagrama de caja y bigotes, gráfico de tallo y hojas, diagramas de dispersión, dendrograma, caras de Chernoff, curvas de Andrews, gráficos de estrella y gráficos de radar, mediante el entorno de desarrollo integrado RStudio. A través de estos gráficos, se puede buscar similitudes entre las unidades de estudio, buscar agrupamientos, las variables que predominan en las observaciones, existencia de datos atípicos, entre otros.

FUNDAMENTOS

Si bien en diversos textos académicos se muestra la forma en cómo se elaboran dichos gráficos multivariantes, lo mejor sería emplear los *softwares* disponibles para realizar ello, aunque no en todos podemos es posible hacerlos, al menos de los que se proponen en este trabajo, siendo la excepción a todo el *software* R.

En el trabajo de Schiattino & Silva [18], referido a gráficos multivariantes aplicados al sistema de salud de Chile, se hace uso de varios tipos de estos gráficos que fueron implementados en los *softwares* R, Stata y SAS. El análisis gráfico puede sugerir algunas hipótesis, refutar otras y ayudar a la interpretación de resultados complejos [18].

Ocampo y Ospina [14] destacan también la importancia que tienen los gráficos pictóricos cuando se requiere hacer un análisis comparativo gráfico. Además, mencionan que «existen diferentes artículos que argumentan sobre la importancia de añadir profundidad a un gráfico en esencia 2D, ya que este es el estilo preferido por la mayoría de los periódicos y empresas» (p.14). También, de acuerdo con Sánchez [17], afirman que las herramientas gráficas se pueden utilizar de dos maneras: para hacer análisis de datos y para transmitir información a otras personas. A pesar de que, en dicho trabajo, «se presentan las principales herramientas gráficas para analizar datos de investigación clínica» (p.110), esto fácilmente se puede extender a otros campos de estudio, como por ejemplo economía, finanzas, contabilidad, etc.

En el trabajo de Correa & González [4], ellos destacan la importancia de hacer una buena comunicación de los datos a través de los gráficos estadísticos y afirman que «a pesar de la reconocida importancia este proceso no siempre se realiza de la menor manera» (p.5).

El análisis multivariado busca cómo se comportan un grupo de variables, pero de manera conjunta, pudiendo encontrarse así datos atípicos. Rubio [16] afirma que «una de las formas de realizar el análisis es mediante gráficos que representen dicha información» (p. 5).

Por todo lo descrito anteriormente, con este trabajo se quiere contribuir de alguna manera a una mayor difusión de estas herramientas gráficas estadísticas, así como su interpretabilidad y utilidad, no solo en el campo de la economía, sino también en otros en donde se requiera hacer un análisis exploratorio multivariado.

No se encontraron mayores limitaciones puesto que la información se encuentra disponible en el portal de Transparencia Económica del Ministerio de Economía y Finanzas, siendo esta de acceso público. Sin embargo, podríamos mencionar que dicha información se actualiza todos los días, razón por la cual los resultados que se mostrarán aquí son referentes al instante en el cual fueron tomados. Dentro de esta base de datos clasificada por departamentos, la información correspondiente al departamento de Tumbes resultó estar incompleta y es por ello que se descartó.

METODOLOGÍA

Tipo y diseño de investigación

Tam *et al.* (2008, p.145) afirman que «existen una multiplicidad de clasificaciones de tipos y métodos de investigación, e incluso ambigüedad y confusión de términos y conceptos» [19]. Para este trabajo, nos basaremos en la clasificación dada por Hernández *et al.* [8], puesto que se trata de una investigación cuantitativa.

Así, este estudio es de tipo o alcance descriptivo, ya que busca especificar propiedades y características importantes de cualquier fenómeno que se analice. Describe tendencias de un grupo o población [8]. Para nuestro caso, se describirá las características para cada departamento con relación a las variables del proyecto de inversión. Además, según Hernández, Fernández y Baptista [8] el diseño es no experimental (no hay manipulación de variables), transversal (se toman los datos en un solo instante de tiempo) y descriptivo (indaga la incidencia de los niveles de una o más variables en la población).

Población y muestra

La población está formada por los 24 departamentos del Perú junto la provincia constitucional del Callao, de la cual se tiene información sobre su ejecución en años anteriores (2018 y 2019), presupuesto institucional de apertura (PIA), presupuesto institucional modificado (PIM), devengado, avance y ejecución total. En toda investigación, no siempre se tiene una muestra ya que en algunos casos queremos efectuar un censo que incluya a todos los casos [8]. En este trabajo se tomará a toda la población siendo su unidad de estudio un departamento o la provincia del Callao. Con respecto a la información del departamento de Tumbes, la cual no está completa, Dagnino [6] dice que «no hay

ninguna forma totalmente satisfactoria para el manejo de los datos faltantes», ya que estos pueden generar sesgo en la información. Si bien es cierto que existen técnicas para lidiar con esto, como es el caso de la imputación basada en la media de los otros valores, se optó por eliminar dicha observación ya que los presupuestos asignados pueden diferir mucho entre sí.

Técnicas de recolección de datos

Para este trabajo se recurrió a una fuente de datos secundaria. Tal como lo define Malhotra [11] «las fuentes de datos secundarios externos publicados incluyen agencias gubernamentales federales, estatales y locales». La información utilizada se tomó de la página de Transparencia Económica Perú del Ministerio de Economía y Finanzas. Esta contiene información sobre el seguimiento de la ejecución de proyectos de inversión por departamentos del año 2021. Esta información se actualiza diariamente por lo que esta difiere de la actual. Los datos con los cuales se trabajaron fueron del 8 de abril de 2021.

Análisis de los datos

El análisis de información se hizo a través de gráficos univariados y multivariados para cada una de las variables. De acuerdo con Peña [15] «el primer paso de cualquier análisis multivariante es representar gráficamente las variables individuales» (p.15), además continúa afirmando que «estas representaciones son útiles para detectar asimetrías, heterogeneidad, datos atípicos, etc.»(p.107).

Así, podemos dividir el análisis gráfico en dos grupos: univariados (gráficos de tallos y hojas, diagramas de cajas y diagramas de dispersión) y multivariados (dendrograma, gráficos de estrella, curvas de Andrews, caras de Chernoff y gráficos de radar).

Para realizar todo lo anteriormente mencionado se hizo uso del entorno de desarrollo integrado RStudio versión 4.1.0 el cual resulta ser muy versátil y de un manejo relativamente sencillo.

RESULTADOS

Como se hizo mención en el apartado sobre la población y muestra, se optó por eliminar la observación correspondiente al departamento de Tumbes, puesto que esta se encontraba incompleta, ya que, de haber seguido con el análisis, nos produciría sesgo en la información [6]. Además, debido a que todas las variables están expresadas en el orden de los millones a excepción de la variable avance que se expresa en porcentajes, se llevó a cabo un escalamiento para cada variable en el rango de 1 a 10 ya que en algunos de los gráficos se requiere que las observaciones sean positivas. Si se hubiera aplicado una estandarización normal, esta podría generar observaciones negativas. «De hecho, la tipificación genera valores con signo negativo y positivo respecto a la diferencia con el promedio» (p. 173) [1].

Tabla 1

Valores transformados mediante escalamiento para el seguimiento de ejecución de proyectos de inversión en Perú, 2021

Departamento	Ejecución 2019	Ejecución 2020	PIA	PIM	Devengado	Avance	Ejecución total
Amazonas	1.98	4.51	3.54	3.11	3.29	5.45	2.39
Áncash	1.47	2.60	3.25	3.34	2.59	3.64	1.56
Apurímac	2.77	3.41	2.43	2.21	2.84	6.38	2.87
Arequipa	10.00	6.75	10.00	10.00	6.64	3.43	10.00
Ayacucho	5.16	4.83	4.94	6.18	4.87	4.21	5.31
Cajamarca	4.10	2.21	4.89	6.10	2.18	1.00	3.80
Callao	3.71	1.00	1.00	1.00	1.00	3.33	3.20
Cusco	7.87	5.70	8.97	7.77	10.00	8.09	8.04
Huancavelica	1.43	2.81	3.65	3.97	1.54	1.10	1.53
Huánuco	2.90	3.98	3.02	2.89	3.22	5.76	3.10
Ica	1.18	1.74	3.64	3.51	2.28	2.86	1.14
Junín	3.47	3.86	4.13	5.05	4.52	4.83	3.63
La Libertad	5.16	2.68	5.78	3.65	2.59	3.28	4.84
Lambayeque	4.65	4.20	3.46	3.57	2.96	4.10	4.68
Lima	2.23	4.04	2.04	2.26	3.69	8.55	2.53
Loreto	2.44	6.05	8.73	7.84	7.01	5.09	3.19
Madre de Dios	1.00	2.02	1.36	1.21	1.36	4.31	1.00
Moquegua	3.52	2.71	1.48	1.18	1.87	6.53	3.37
Pasco	3.56	2.51	2.66	2.51	1.11	1.26	3.35
Piura	6.39	10.00	7.01	9.37	8.15	4.98	7.44
Puno	2.40	4.44	5.10	4.34	3.31	3.79	2.74
San Martín	5.12	3.70	2.54	2.25	4.19	10.00	5.04
Tacna	2.84	4.25	3.04	2.57	3.88	8.09	3.12

Fuente: Transparencia Económica, [12], 2021

Diagrama de tallo y hojas

Lo primero que se hizo fue el análisis gráfico univariado, empezando por representar a cada una de las variables por medio de un diagrama de tallo y hojas. Tal como afirma Curts *et al.* (1988), «Tukey (1977) ideó el diagrama de tallo-y-hoja como un híbrido donde se combinan los aspectos visuales de un histograma con la información numérica que proporciona una tabla de distribución de frecuencias» (p.32) [5].

0	34556788999
1	111235666
2	04
3	1

Figura 1. Diagrama de tallo y hojas de ejecución en 2019
Fuente: elaboración propia.

0	3790233448
2	011233457249
4	9

Figura 2. Diagrama de tallo y hojas de ejecución en 2020
Fuente: elaboración propia.

0	1448
2	01244567780556
4	088
6	06

Figura 3. Diagrama de tallo y hojas de PIA
Fuente: elaboración propia.

0	345
2	2224578012257
4	301
6	2349

Figura 4. Diagrama de tallo y hojas de PIM
Fuente: elaboración propia.

0	11223333344444
0	555688
1	02

Figura 5. Diagrama de tallo y hojas de devengado
Fuente: elaboración propia.

0	5669
1	000011223334
1	56699
2	03

Figura 6. Diagrama de tallo y hojas de avance
Fuente: elaboración propia.

0	45669
1	001222233447889
2	79
3	6

Figura 7. Diagrama de tallo y hojas de ejecución total
Fuente: elaboración propia.

En la figura 1 sobre la ejecución presupuestal del año 2019, se observa la mayor tendencia en el tallo cero seguido de los tallos uno hasta tres que muestra solamente un valor; además, se puede señalar que la variabilidad o dispersión en relación con cada uno de los datos se orienta hacia la derecha de manera decreciente; en el caso de la figura 2 sobre la ejecución para el año 2020, la tendencia de los tallos cero como dos (que tienen una presencia de pares) cuentan con valores similares en cantidades desde 0 hasta 8 en el primer caso y nueve para el segundo tallo; en el caso del tercer tallo (4), solamente se cuenta con un valor de 9, siendo una situación atípica para esta variable. Con referencia a la figura 3 sobre el presupuesto institucional de apertura (PIA), los tallos se presentan desde cero hasta seis (presentados en pares) en la que se observa la forma de distribución apuntalada (platicúrtica) frente a la presencia de cuatro valores para cero, seguido de puntuaciones desde 0 hasta 8 y para el caso de los tallos cuatro y seis con tres y dos datos partiendo de cero hasta ocho con saltos directos en cada uno de los casos. En cuanto a la figura 4, la tendencia que se observa es de manera apuntalada desde el tallo cero con la presencia de tres datos y el caso de los tallos cuatro y seis que muestran de tres a cuatro apariciones y la diferencia es para el tallo dos con valores desde 2 hasta 8 como puntuación máxima. Para la figura 5 sobre devengado, la dispersión cuenta con un sesgo marcado hacia la derecha con puntuaciones presentadas para el tallo cero en dos partes que cuenta con 16 datos en la primera parte y seis para la segunda mitad, además se observa que el tallo uno presenta dos valores. La figura 6 sobre los avances muestra una distribución mesocúrtica con apariciones del tallo cero hasta dos contando el tallo uno con dos partes: doce en la primera mitad y cinco en la segunda parte, luego de la primera mitad va disminuyendo la frecuencia llegando a dos datos en el tallo dos. En el caso de la ejecución total que se presenta en la figura 7, se observa un sesgo hacia la derecha con valores de los tallos desde cero con la menor cantidad de apariciones y uno con la mayor frecuencia absoluta, luego en caída para dos y tres.

Gráfico de caja y bigotes

De acuerdo con Díaz & Morales [7], un gráfico de caja y bigotes consiste de «una caja, y guiones o segmentos. Se dibuja una línea a través de la caja que representa la mediana. El extremo inferior de la caja es el primer cuartil (Q_1) y el superior el tercer cuartil (Q_3)» (p.10).

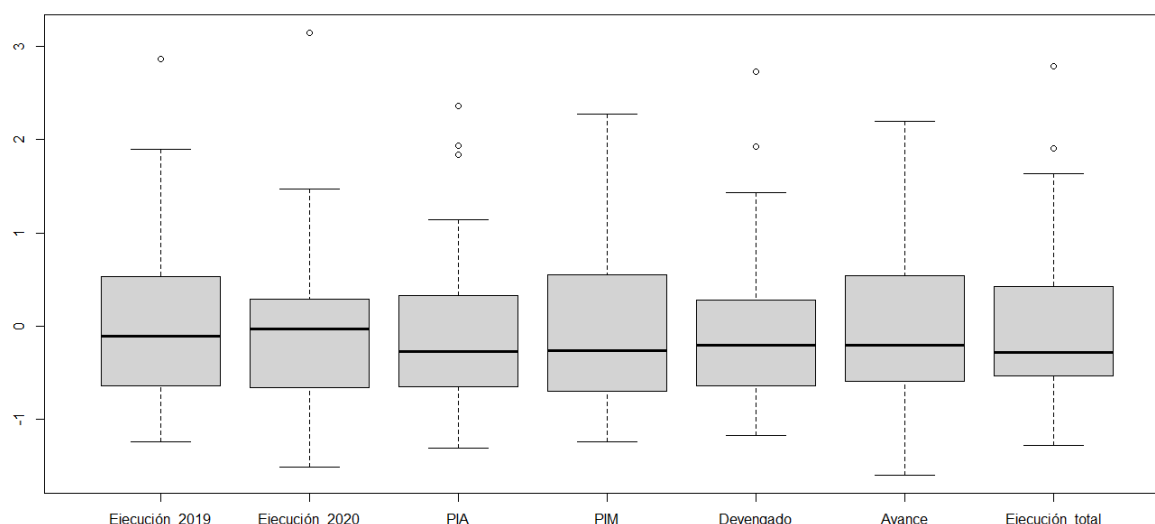


Figura 8. Gráfico de caja y bigotes para el seguimiento de ejecución de proyectos de inversión en Perú, 2021
Fuente: elaboración propia.

En la figura 8, se puede observar en primer lugar la supuesta distribución normal de los valores para cada una de las variables y su presencia en los diversos cuartiles, así como los valores atípicos para cada una de las variables producto del análisis. En relación con la distribución de la normalidad para los cuartiles, la variable devengado presentan mejor distribución de los cuartiles entre los valores del rango intercuartil con dos valores atípicos en el extremo superior, luego se encuentra el PIA con distribuciones más uniformes para los cuartiles y con tres valores atípicos que se presentan en los extremos superiores. Los valores que presentan mayor cantidad porcentual en el segundo cuartil en adelante están considerados en la ejecución del año 2019 y la ejecución total con un valor atípico en cada uno de los casos; en cuanto al mismo sentido del segundo cuartil, con mayor porcentaje se encuentra el PIM y el avance que no presentan valores atípicos.

En cuanto a los valores con mayor presencia por debajo del segundo cuartil, se ubica la ejecución del año 2020 y con un valor atípico en el extremo superior. Asimismo, la variable que cuenta con mayor cantidad de valores atípicos es PIA, con tres. Con respecto al sesgo, la ejecución del año 2019, el PIM y el avance presentan sesgos hacia la derecha.

Diagrama de dispersión

Los diagramas de dispersión «son gráficos en los cuales se representan los individuos u objetos por puntos asociados a cada par de coordenadas (valores de cada par de variables)» (p.9) [7]. Estos gráficos buscan una posible asociación lineal entre dos pares de variables. Se implementó este tipo de gráficos en uno solo conocido como matriz de dispersión de datos.

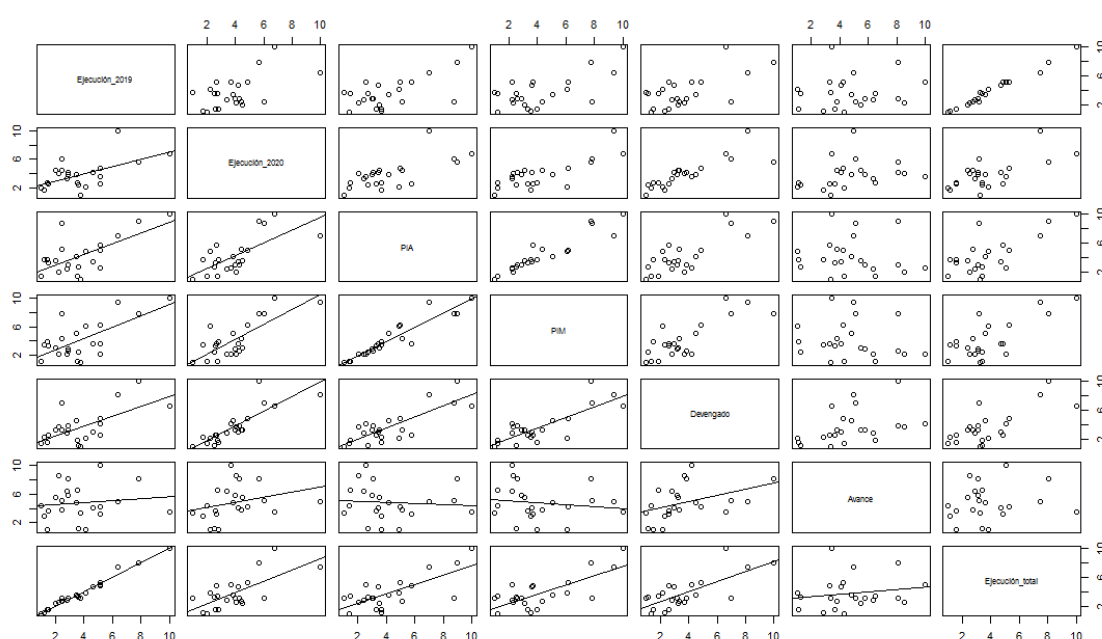


Figura 9. Matriz de dispersión para el seguimiento de ejecución de proyectos de inversión en Perú, 2021
Fuente: elaboración propia.

De acuerdo con la figura 9, se puede observar una cierta asociación lineal entre las variables PIM y PIA; con respecto a las variables PIM y devengado, existe un indicio de alguna asociación lineal entre ellas y en el resto de las variables comparadas de dos a dos se observa una dispersión marcada entre ellas.

Dendrograma

Seguidamente se hizo un análisis gráfico multivariado empezando por un dendrograma. De acuerdo con Peña [15] «el dendrograma

o árbol jerárquico, es una representación gráfica del resultado del proceso de agrupamiento en forma de árbol» (p.245). La principal utilidad de este tipo de gráfico es buscar similitudes entre las observaciones para buscar formar agrupamientos mediante el uso de una distancia ultramétrica [15]. Por medio de la función hclust del paquete stats de R, se puede implementar un dendrograma [3]. A continuación, se muestra el dendrograma para las variables en estudio:

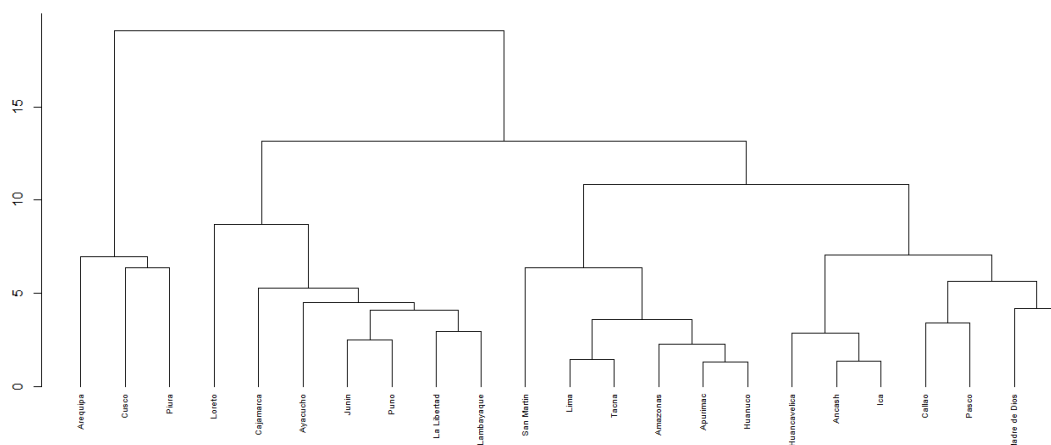


Figura 10. Dendrograma para el seguimiento de ejecución de proyectos de inversión en Perú, 2021
Fuente: elaboración propia.

El nivel de similitud se mide en el eje vertical (alternativamente se puede mostrar el nivel de distancia) y las diferentes variables se especifican en el eje horizontal. La figura 10 muestra la existencia de dos conglomerados iniciales que tienen un corte en la puntuación 13 en el nivel de similitud; se puede observar, luego, que hay cuatro subgrupos bien conformados al tener la línea de corte en el valor 9 para cada uno de los conglomerados.

Podemos utilizar la técnica de agrupamiento no supervisado llamada k-means para verificar lo mostrado por el dendrograma. El coeficiente de aglomeración nos permite medir esto; valores cercanos a 1 sugieren una buena agrupación [10].

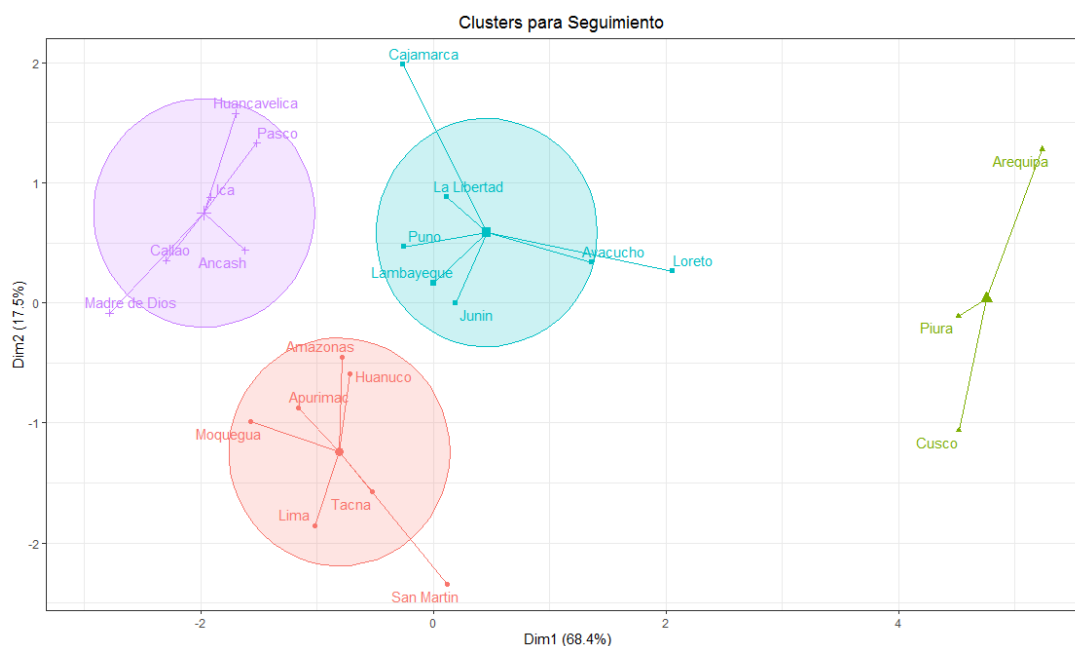


Figura 11. Clústeres para el seguimiento de ejecución de proyectos de inversión en Perú, 2021
Fuente: elaboración propia.

De acuerdo con la figura 11, se pueden apreciar los 4 conglomerados que nos sugería el dendrograma. El valor del coeficiente de aglomeración resultó ser 0.8086, lo cual nos indicaría que los agrupamientos son buenos. Cabe resaltar que se proyectaron sobre un plano bidimensional dichos conglomerados para poder observarlos. Por ejemplo, el clúster formado por Arequipa, Cusco y Piura es el que más se diferencia de los demás y se caracteriza por presentar los mayores valores para la ejecución 2020, PIA y PIM.

Caras de Chernoff

Otro tipo de gráficos estadísticos utilizados en análisis multivariados con las caras de Chernoff. Estos asocian a cada variable una característica del rostro como la longitud de la nariz, tamaño de los ojos, forma de los ojos, ancho de la boca entre otros [7]. Se buscarán similitudes entre cada una de las observaciones por medio de estos gráficos.

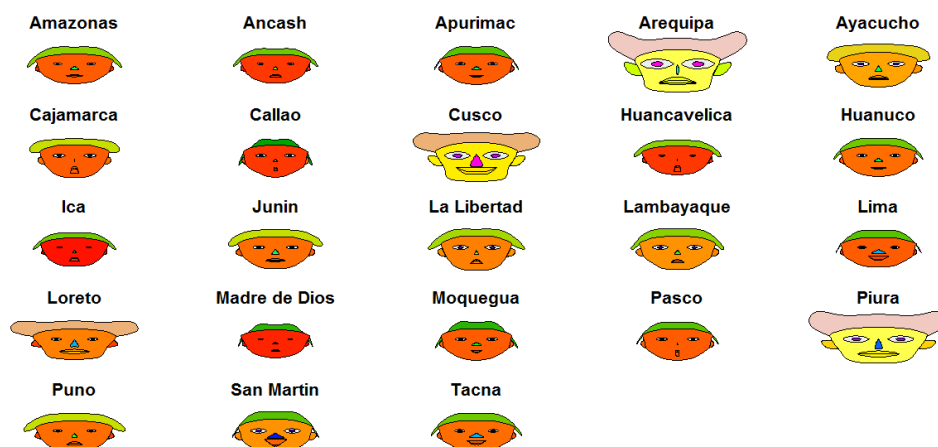


Figura 12. Caras de Chernoff para el seguimiento de ejecución de proyectos de inversión en Perú, 2021

Fuente: elaboración propia.

En la figura 12, se observa que el PIM se presenta con mayor similitud para los departamentos de Arequipa y Piura; en el caso del PIA y el PIM, se cuenta con similitudes para los departamentos de Cusco y Loreto. En el caso Apurímac con Tacna, los valores de la ejecución del año 2019 y la PIA cuentan con similitudes próximas en cada caso. En cuanto a los departamentos de Amazonas y Huánuco, las similitudes halladas se encuentran en la ejecución del año 2019, el PIA, el PIM y el devengado, como aspectos más relevantes. El departamento de Moquegua y la provincia constitucional del Callao son los que presentan mayores similitudes en la Ejecución presupuestal del año 2019.

Curvas de Andrews

Las curvas de Andrews, propuestas por Andrews (1972), se definen como las combinaciones lineales de observaciones de

un conjunto de datos multivariado. Cada una de ellas se asocia a un coeficiente de una serie de Fourier [7].

$$x_i(t) = \frac{1}{\sqrt{2}}x_{i1} + \text{sen}(t)x_{i2} + \cos(t)x_{i3} + \text{sen}(2t)x_{i4} + \dots$$

$$, t \in [-\pi, \pi]$$

Es importante el orden en el cual se colocan las variables puesto que, si hay un número grande de variables, las últimas no contribuirán mucho en la forma de la curva. Así, al hacer un análisis a través de curvas de Andrews, será necesario ordenar previamente las variables según la cantidad de información que aporta cada una de ellas.

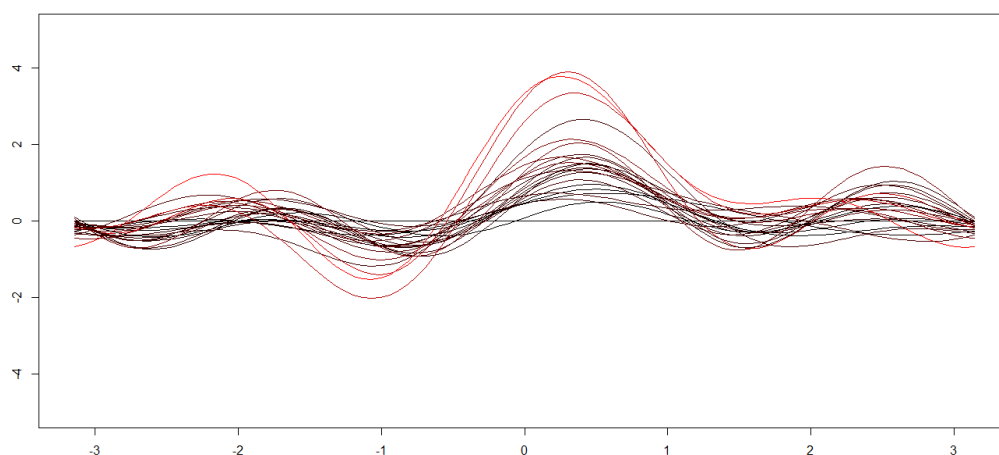


Figura 13. Curvas de Andrews para el seguimiento de ejecución de proyectos de inversión en Perú, 2021

Fuente: elaboración propia.

Los resultados de las variables escaladas para las estimaciones de los datos para cada una de las siete variables en estudio se han considerado como parte del análisis de seguimiento a través de la curva de Andrews. En la figura 13, se puede observar que la magnitud de cada variable estudiada por los diversos departamentos, incluyendo la provincia constitucional del Callao, de forma individual afecta las frecuencias, la amplitud y la periodicidad de la función de forma individual.

Gráfico de estrella

Los gráficos de estrella se utilizan para representar datos multivariados. Balzarini *et al.* [2] afirman que, para construir este tipo de gráficos, «una observación p-dimensional es representada en el plano por una estrella construida sobre un círculo con p radios o rayos igualmente espaciados que nacen desde el centro del círculo. La longitud de la estrella sobre cada radio representa el valor de la variable asociada al rayo» (p.16). Así, se realizó también la representación gráfica para cada variable.

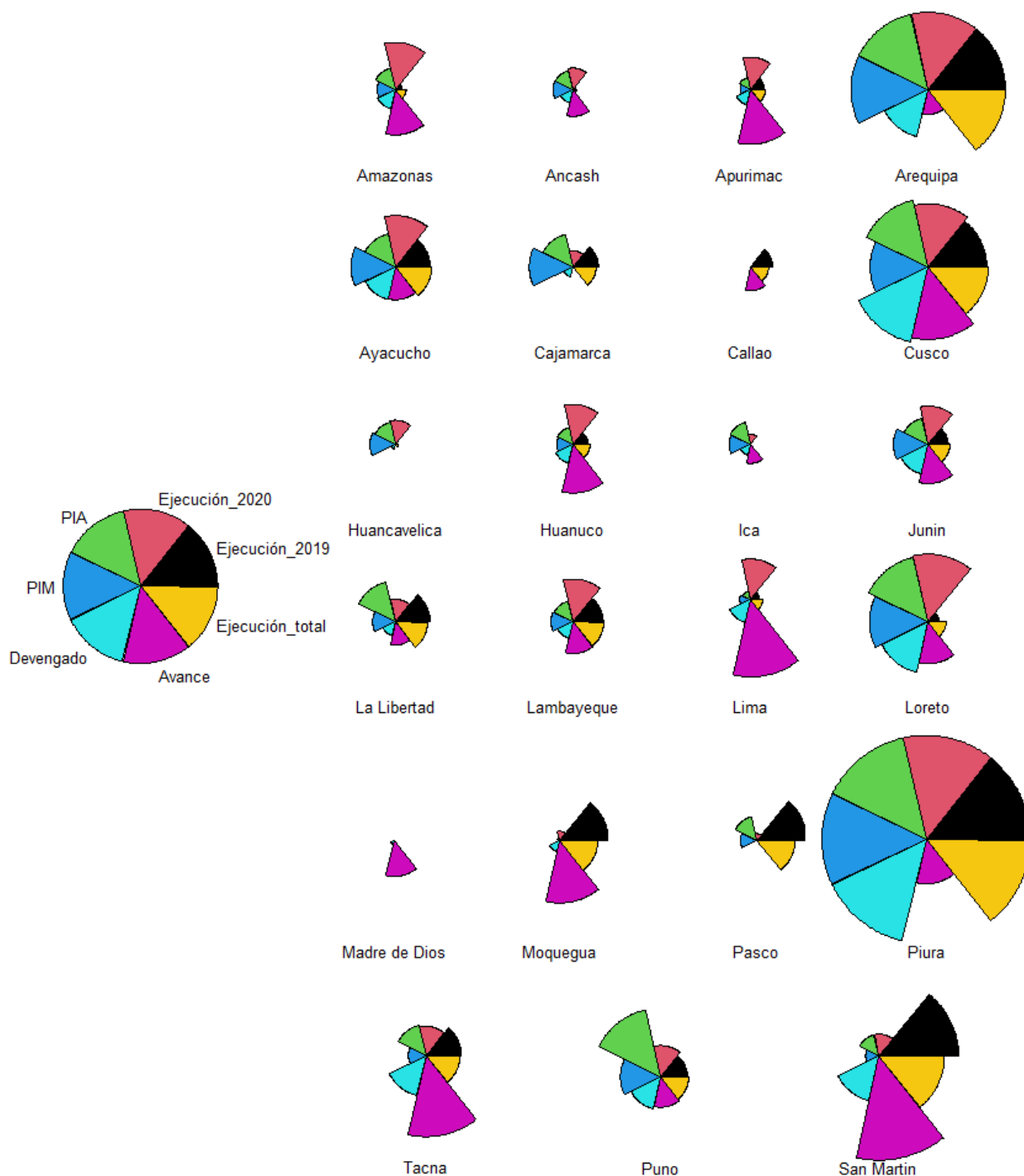


Figura 14. Gráficos de estrella para el seguimiento de ejecución de proyectos de inversión en Perú, 2021
Fuente: elaboración propia.

Los resultados del gráfico de estrella nos permiten reconocer los valores atípicos o similares que se encuentran dentro de cada unidad de evaluación y los lugares que son muestreados para los resultados multivariados, donde la figura 14 es una representación de existencia de valores atípicos en los departamentos Amazonas, Huánuco, Lima, Madre de Dios, San Martín, Áncash y, en menor escala, para Junín. En cuanto a la ejecución total, se observan valores atípicos en el departamento de Piura. En la ejecución del año 2019, el departamento de Ica presenta valores por debajo de los normales al igual que en Huancavelica. En la ejecución del año 2020, los departamentos de Amazonas, Áncash, Cajamarca, Provincia del Callao y La Libertad son los que presentan valores atípicos por debajo de las puntuaciones normales. En cuanto al PIA, La Libertad cuenta con valores atípicos y con el caso de los devengados y la PIM no se determinan muchas variaciones entre valores atípicos.

Gráfico de radar

Es similar al gráfico de estrella, pero colocando una variable en cada estrella y no para cada unidad de observación, donde son identificadas dichas unidades como rayos concéntricos [18]. Esto representa una desventaja, ya que, en caso de contar con muchas variables u observaciones, este se vería muy saturado. De acuerdo con la *Guía para representar datos estadísticos* del INEI [9], «Se basan en la aplicación de ejes radiales con origen común y escalas estandarizadas. Cada eje se utiliza para indicar el valor de un indicador específico. Regularmente se representan indicadores distintos con valores previamente estandarizados para fines de comparación» (p.45).

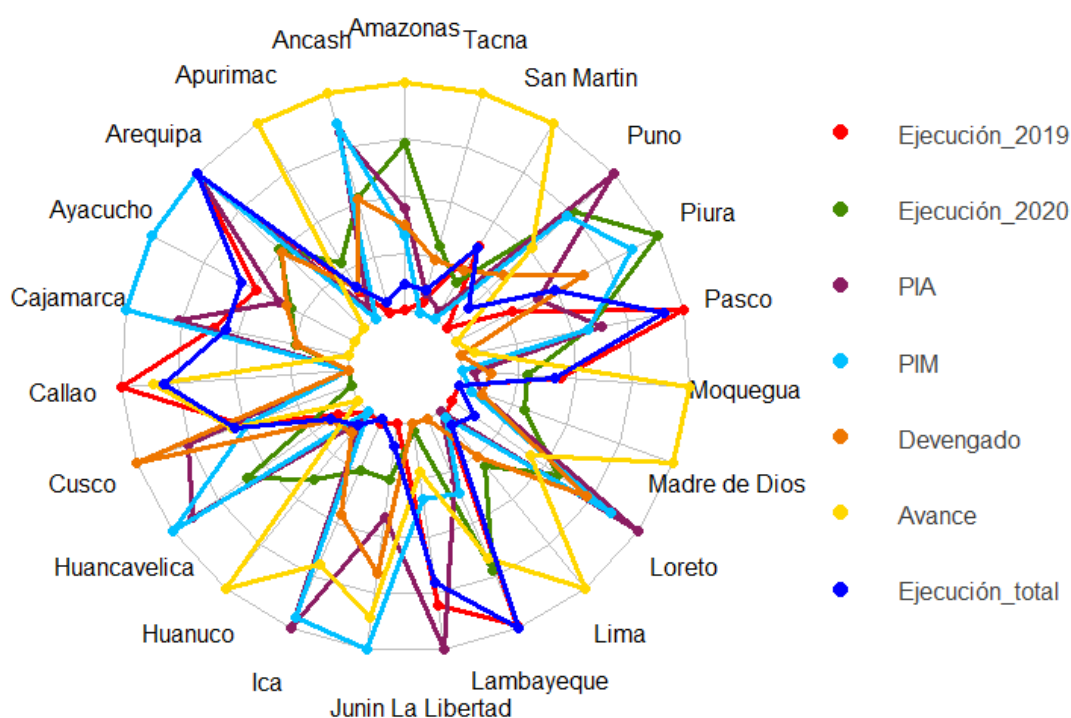


Figura 15. Gráficos de radar para el seguimiento de ejecución de proyectos de inversión en Perú, 2021
Fuente: elaboración propia.

En la figura 15, podemos observar el gráfico de radar para las 7 variables estudiadas. Por ejemplo, se aprecia que con respecto a la variable avance los departamentos con los valores más altos son Amazonas, Áncash, Apurímac, Tacna, San Martín, Moquegua, Madre de Dios, Lima y Huánuco. Para el caso de la variable PIM, los departamentos que presentan los mayores valores son Arequipa, Ayacucho, Cajamarca, Ica y Junín.

CONCLUSIONES

- El análisis observacional multivariante aplicado a las variables macroeconómicas mediante gráficos permite identificar rápidamente las similitudes entre las unidades de estudio, con ello se puede desarrollar su clasificación o agrupación en clústeres; además de determinar los valores atípicos.

- Si bien las representaciones gráficas son de bastante utilidad a un nivel exploratorio, es necesario llevar a cabo pruebas de hipótesis para confirmar lo que muestran los gráficos, esto no es parte de los objetivos del trabajo presentado aquí, pero se podría abordar en un futuro
- De acuerdo con los gráficos de caja y bigotes obtenidos para cada una de las variables estudiadas, lo más resaltante ha sido que los PIA y los devengados tienen aproximadamente una distribución simétrica, lo cual sería útil para realizar contrastes estadísticos; aunque en el resto de las variables las distribuciones no tienden a la simetría. El PIA y el devengado del departamento de Cusco resultaron atípico en ambos casos; esto nos indicaría que estos valores son muy diferentes para este departamento con respecto a los otros.

- Mediante los diagramas de dispersión, observamos que existe un cierto grado de relación directa entre el PIA y PIM, es decir, si aumenta la PIA, también aumenta el PIM. Algo similar sucede entre las variables PIM y devengado. Se podría aplicar un análisis de regresión lineal simple para encontrar una ecuación de regresión validando los supuestos de este modelo.
- El dendrograma nos mostró de que existen indicios de la conformación de 4 conglomerados entre los departamentos. Es por ello que se realizó un agrupamiento no supervisado (k-means) y mediante el coeficiente de aglomeración, cuyo valor es 0.8086. En ese sentido, podemos concluir que los agrupamientos son buenos. De todos ellos, el que se diferencia más fue el conformado por los departamentos de Arequipa, Cusco y Piura que presenta los mayores valores para la ejecución 2020, PIA y PIM.
- Tanto las caras de Chernoff como los gráficos de estrella nos permiten ver que las observaciones correspondientes a los departamentos de Arequipa, Piura y Cusco se diferencian bastante de los otros. Eso lo notamos también cuando se hizo el análisis de conglomerados, ya que estas comparten características similares.
- Por último, con respecto al gráfico de radar, notamos que los departamentos que más destacan en avance son Amazonas, Áncash, Apurímac, Tacna, San Martín, Moquegua, Madre de Dios, Lima y Huánuco, mientras que los que menos ejecutaron fueron Callao, Ayacucho, Arequipa, Huancavelica y Piura, lo cual no llevaría a tomar medidas al respecto, puesto que Arequipa y Piura fueron los que más ejecutaron en el año 2020, pero con respecto al avance hasta junio de 2021, son los que menor porcentaje presentan.

REFERENCIAS

- [1] Actis di Pasquale, E., & Balsa, J. (2017). *Revista de métodos cuantitativos para la economía y la empresa*, 23, 164-193.
- [2] Balzarini, M., Bruno, C., Córdoba, M., & Teich, I. (2015). *Herramientas en el análisis estadístico multivariado*. Córdoba: Universidad Nacional de Córdoba.
- [3] Carmona, F. (2018). *Un análisis con R. Datos multivariantes*. Departament d'Estadística.
- [4] Correa, J., & González, N. (2002). *Gráficos estadísticos con R*. Recuperado el 20 de junio de 2021, de <https://cran.r-project.org/doc/contrib/grafi3.pdf>
- [5] Curtis G., J., Alcántara L., L., & Chiappa C., X. (1987). Introducción al análisis exploratorio de datos multidimensionales. *Ciencias*, 30-35.
- [6] Dagnino, J. (2014). Datos faltantes (missing values). *Revista Chilena de Anestesiología*, 43, 332-334.
- [7] Díaz Monroy, L. G., & Morales Rivera, M. A. (2012). *Análisis estadístico de datos multivariados*. Bogotá: Universidad Nacional de Colombia.
- [8] Hernández, S. R., Fernández, C. C., & Baptista, L. P. (2014). *Metodología de la investigación*. México: McGraw-Hill Interamericana Editores.

- [9] INEI. (2009). *Guía para representar datos estadísticos*. Lima: Centro de Investigación y Desarrollo.
- [10] Jiménez, M. (2018). *Clustering jerárquico en R*. Obtenido de Rpubs: <https://rpubs.com/mjimcua/clustering-jerarquico-en-r>
- [11] Malhotra, N. (2008). *Investigación de mercados*. México: Pearson Educación.
- [12] MEF. (2021). *Transparencia económica*. Recuperado el 8 de mayo de 2021, de https://apps5.mineco.gob.pe/bingos/seguimiento_pi/Navegador/default.aspx
- [13] Mostajo Guardia, R. (2002). El sistema presupuestario peruano. *CEPAL*.
- [14] Ocampo García, J., & Ospina Betancur, J. (2013). Análisis comparativo de Colombia frente a los nueve países más desarrollados según el IEPG, mediante el análisis multivariado y los gráficos pictóricos. *Journal of Engineering and Technology*, 2(2), 8-20.
- [15] Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw-Hill.
- [16] Rubio Donet, J. L. (2018). *Detección de datos multivariados atípicos con series finitas de Fourier*. Lima: Universidad Nacional Agraria La Molina.
- [17] Sánchez Pedraza, R. (2000). Análisis de los datos mediante herramientas gráficas. *Revista de la Facultad de Medicina*, 48(2), 104-110.
- [18] Schiattino, I., & Silva, C. (2013). Representación gráfica de información multivariante. Aplicación al sistema de salud de Chile. *Revista Chilena de Salud Pública*, 17(2), 117-123.
- [19] Tam Málaga, J., Vera, G., & Oliveros Ramos, R. (2008). Tipos, métodos y estrategias de investigación científica. *Pensamiento y acción*, 5, 145-154.

ACERCA DEL AUTOR

José Luis Espinoza Melgarejo

Magíster en Docencia Universitaria e Investigación Pedagógica de la Universidad San Pedro de Chimbote (USP), Perú; licenciado en Matemática de la Universidad Nacional de Ingeniería (UNI), Perú. Docente de educación superior con más de 6 años de experiencia laborando en instituciones como IDAT, Tecsup y la Universidad Privada del Norte. Ha desarrollado diversos cursos en matemática, estadística, finanzas entre otros. Especialista en estadística y ciencia de datos que cuenta con estudios de posgrado en la Universidad Nacional Mayor de San Marcos (UNMSM), Perú y actualmente estudiante de doctorado en Estadística Matemática de la Universidad Nacional del Santa (UNS), Perú.

@ jespinozame@tecsup.edu.pe

Recibido: 27-06-21
Revisado: 26-08-21
Aceptado: 08-09-21



Esta obra está bajo una Licencia Creative Commons AtribuciónNoComercial 4.0 Internacional.