**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**

**FACULTAD DE CIENCIAS SOCIALES**

América Latina y Europa: análisis de la calidad de vida acorde a aspectos socio - políticos y económicos

Nombres y códigos:

* ÁVALOS SÁNCHEZ, Ariana Nicole (20221729)
* VALVERDE ALATRISTA, Andrea Sofía (20222754)

Tipo de evaluación: Entrega 4 - Trabajo Grupal

Curso: Estadística para el Análisis Político 2 - POL231

Horario: 689B

Profesor: José Manuel Magallanes

Jefe de Práctica: Chiara Zamora Mendoza

SEMESTRE 2024-2

**ÍNDICE**

[I. TEMA DE INVESTIGACIÓN 3](#_heading=h.dtcoiowz3i3y)

[II. JUSTIFICACIÓN DEL TEMA 3](#_heading=h.qsg2wgfdzf8z)

[a. Variable dependiente 3](#_heading=h.150t2nfdukoy)

[b. Variable indicada como predictora 5](#_heading=h.9hjc2lda6ia3)

[c. Variable de control 6](#_heading=h.gjdgxs)

[III. HIPÓTESIS 6](#_heading=h.3zf2308ln9j4)

[IV. METODOLOGÍA 7](#_heading=h.3xi9j6myk7xo)

[V. LIMPIEZA DE LA BASE DE DATOS 9](#_heading=h.fo0chipo56f8)

[VI. HALLAZGOS 10](#_heading=h.c6y6ko2lzw7y)

[a. Análisis Factorial 10](#_heading=h.9qb8o22779m1)

[b. Análisis Clúster 12](#_heading=h.b2952x5e6049)

[i. Estrategia de partición 12](#_heading=h.azb28ou017q)

[1. PAM 13](#_heading=h.nwy24kcsdbgg)

[ii. Estrategia jerárquica 13](#_heading=h.4ph3tjifmwrd)

[1. AGNES 13](#_heading=h.cwbsr73hx0cj)

[2. DIANA 14](#_heading=h.3b50od69o1tl)

[c. Análisis bivariado 14](#_heading=h.ags5b1ibhr4l)

[i. Numérica con numérica 14](#_heading=h.lst8zriwzeqz)

[1. Calidad de vida con población 14](#_heading=h.8nfugr3p95hw)

[2. Calidad de vida con PBI per cápita 15](#_heading=h.2p78awph4i3p)

[3. Calidad de vida con tasa de inflación 16](#_heading=h.qblyw6rhk46)

[4. Calidad de vida con división administrativa 17](#_heading=h.uonhcqrm7eh3)

[ii. Numérica con categórica 17](#_heading=h.oy9tg965tect)

[1. Calidad de vida y régimen político 17](#_heading=h.lb2hdorgy5qa)

[2. Calidad de vida con estructura legislativa 18](#_heading=h.omqbypxmhwky)

[3. Calidad de vida con continente 19](#_heading=h.ne1hu3bluhw2)

[d. Análisis multivariado 20](#_heading=h.99qd82r5327)

[VII. CONCLUSIONES 21](#_heading=h.29ra6hrzwbf7)

[VIII. BIBLIOGRAFÍA 22](#_heading=h.y0v67an43z6t)

# TEMA DE INVESTIGACIÓN

El presente trabajo de investigación tiene como objetivo el análisis de un grupo de factores que pueden o no tener una influencia respecto a la calidad de vida. Para ello se ha considerado dos espacios geográficos distintos: Europa y Latinoamérica ya que el primero en cuestión es considerado como un territorio con países desarrollados; mientras que el segundo, con países en vías de desarrollo. Asimismo, se ha considerado importante evaluar si la calidad de vida es afectada por aspectos gubernamentales, sociales y económicos debido a que no se ha evaluado mucho el caso. Por lo tanto, la pregunta a la que se busca dar respuesta en el presente trabajo es ¿qué tanto impacto pueden tener diversos criterios sociales, económicos y políticos (gubernamentales) en la calidad de vida de los países latinoamericanos y europeos? ¿Estos criterios son los que definen que sean países desarrollados o en vías de desarrollo?

# JUSTIFICACIÓN DEL TEMA

## Variable dependiente

La calidad de vida es una variable multidimensional, pero que, tal como se señala en un artículo de Quality of Life, se puede estudiar bajo la agrupación de dos enfoques: subjetivo y objetivo. En este caso, se aplicará el segundo, el cual está asociado con factores externos que influyen en la calidad de vida. En este sentido, se operacionaliza la variable según el acceso a los servicios básicos que brinda el Estado, así como educación, salud, vivienda, tecnología y empleabilidad. Tal información será analizada en base a los datos que brinda el Banco Mundial, correspondientes al año 2022.

En primer lugar, la variable calidad de vida se operacionaliza en función a la tasa bruta de matrícula de los estudiantes en educación primaria. Ello implica el porcentaje de estudiantes matriculados en educación primaria en relación con la población total en edad de asistir a dicho nivel académico. Este indicador puede superar el 100%, puesto que incluye a todos los estudiantes, ya sean superiores o inferiores a la edad oficial por repeticiones de grados o ingresos tardíos al nivel primario. En esta línea, este porcentaje está relacionado con el acceso a la educación, y que contribuye en la calidad de vida, debido a que, al ser el primer nivel educativo de los niños, les permitirá poder desarrollar habilidades básicas (lectura, comprensión, matemáticas), que serán clave para su desarrollo laboral posterior. También, contribuye a una reducción en la desigualdades, ya sean económicas o sociales, por las herramientas académicas que brinda la educación a los estudiantes, permitiéndoles mejorar su calidad de vida en función a sus capacidades desarrolladas.

Asimismo, conforme al acceso a los servicios de salud, se emplea el indicador del porcentaje total de causas de muerte por enfermedades transmisibles y condiciones maternales, prenatales y en materia de nutrición. Dicho indicador es relevante para medir la calidad de vida en base a cómo es que las personas no acceden a un servicio de salud. Dichos datos evidencian una proporción de las causas de mortalidad, las cuales pueden ser prevenidas, tratadas y superables con una disponibilidad de servicios de salud, y de esta forma, mejorar la calidad de vida de las personas.

En tercer lugar, otro indicador para medir la calidad de vida es el porcentaje de la población que utiliza al menos servicios básicos de saneamiento. Esta operacionalización está relacionada con el acceso a servicios de vivienda, donde las instalaciones sanitarias, así como los inodoros, lavamanos, duchas, entre otros, son importantes para mejorar la calidad de vida de las personas. Ello debido a que, a través de la provisión de servicios de saneamiento, se puede prevenir diferentes tipos de enfermedades en relación a la higiene, que está vinculado con la protección de la salud pública, y conservar el medio ambiente.

Además, en base al acceso de servicios de tecnología, se emplea el porcentaje de suscripciones a telefonía fija para medir la calidad de vida. Dicho indicador hace referencia a la suma de números activos de líneas telefónicas fijas análogas. Este servicio de comunicación es relevante hoy en día por la interconectividad global, ya que ofrece una conexión estable y seguridad de datos, es de mayor accesibilidad por su costo, y permite mantener contacto con las demás personas. En este sentido, satisface algunas necesidades de las personas que pueden abarcar desde la vida social hasta la laboral.

Por último, se operacionaliza la calidad de vida de acuerdo al porcentaje de la fuerza laboral desempleada, que abarca a la proporción de personas que están sin algún trabajo, pero están disponibles y buscándolo. Este indicador está relacionado con el acceso de servicios de empleabilidad, el cual implica programas o recursos de apoyo por parte del Estado para apoyar a las personas en el ámbito laboral. En este aspecto, estos servicios no solo ayudan a que haya una reducción del desempleo, sino también a que se potencien las habilidades laborales, sociales de las personas, y contribuyen a generar una mayor estabilidad económica.

## Variable indicada como predictora

Por otra parte, se presentan cuatros variables que pueden impactar en la calidad de vida de las personas: régimen político, estructura legislativas, población y división administrativa, que fueron obtenidas por la base de datos de Central Intelligence Agency (Cia. gov).

Primero, el **régimen político** (reg\_pol) ayudaría a examinar la calidad de vida mediante su influencia en la manera en que se implementan las medidas estatales y se establece la estructura social. Por lo tanto se puede abordar la firmeza, transparencia y gestión general frente a la cobertura de necesidades básicas y la proyección y enfoque respecto a las demandas sociales.

En segundo lugar, la **estructura legislativa** (cam) abordaría la eficiencia legislativa para la resolución de problemáticas sociales que impactarían en la calidad de vida. Por lo tanto, se la podría apreciar como un filtro propicio o no en la aprobación de leyes para mejorar la calidad de vida de los ciudadanos. Asimismo, acorde a Cuadros (2017), se explica que el rol del poder legislativo es fundamental para la aplicación de las leyes abordando la seriedad de la situación y el compromiso con la aplicación, formulación y el respeto de leyes puesto a que las medidas que se tomen impactan de forma directa a los ciudadanos.

Así mismo, la **densidad poblacional** (pobla)**,** que implica la cantidad de habitantes en un determinado lugar, puede influir de forma positiva como negativa en la calidad de vida. Ello dado que si bien en áreas donde hay una mayor densidad poblacional debería haber un mayor acceso a servicios básicos, en muchas ocasiones, se puede observar un exceso en el uso de estos servicios o una falta acceso. No obstante, en otros casos, se puede evidenciar una mejor disponibilidad de los servicios básicos en relación con sus densidades poblacionales respectivas.

La últimavariable consideradaes la **división administrativa** (div\_admi) que corresponde a cómo está dividido el territorio de cada país; por ejemplo, en el caso de Perú cuenta con 24 departamentos. En ese sentido, se podría decir que a una mayor división administrativa puede haber una menor calidad de vida debido a que se puede presentar la centralización. Tal es el caso de Perú que es evidente la centralización en Lima y las grandes diferencias con los otros departamentos. Sin embargo, no se afirma que en todos los casos se vea con la misma magnitud ni que se evidencie una relación directa entre la división administrativa y la calidad de vida de las personas.

## Variable de control

Para el respectivo análisis, también se ha empleado dos variables de control: PBI per cápita y tasa de inflación, los cuales fueron extraídos de Central Intelligence Agency (Cia. gov), con respecto al año 2022.

Acorde a un estudio realizado por el Banco de Desarrollo de América Latina y el Caribe se ha estimado que gestionar bien el agua puede generar beneficios económicos, mientras que una mala gestión puede causar pérdidas en el **PIB per cápita** (PBI\_perk). Acorde a ello, se ha de interpretar que para que haya una buena calidad de vida se puede asociar un buen manejo para el financiamiento de los servicios básicos. Ello generaría un impacto en la calidad de vida de las personas tanto en el proceso de generar valor como en el resultado respecto a la calidad de vida. (se va a saber que es un aspecto que influye en la sociedad)

Por último, la **tasa de inflación** (tasa\_infla) es variable, pero esta impacta en la calidad de vida ya que afecta a la economía y, por ende, al poder adquisitivo. En ese sentido, la inflación siempre va a ver que tiene un impacto en la sociedad y consecuentemente en la calidad de vida. En ese sentido, se ha de entender que a mayor porcentaje de inflación habrá un impacto en los precios, desvalorización de la moneda y un efecto negativo en la sociedad. En el caso contrario, cuando se tiene una baja tasa de inflación, todo es más accesible a la ciudadanía y, por lo tanto, podría mantener o mejorar la calidad de vida.

# HIPÓTESIS

En principio, se espera obtener resultados positivos que demuestren que la calidad de vida de los países varía según el tipo de régimen político, la estructura legislativa (unicameral o bicameral), el PBI per cápita con el que cuentan, la tasa de inflación, la masa poblacional y la división administrativa territorial. Pues se considera que hay posibilidades de que un régimen político sí puede influir en la calidad de vida de los ciudadanos así como la estructura legislativa ya que es el proceso por el cual pasa la normativa que impacta en la vida de las personas. Asimismo, se ha considerado que el PBI per cápita y la tasa de inflación, al reconocerlas como variables de control, van a impactar en la calidad de vida de los ciudadanos. También se propone que la masa poblacional puede afectar a la calidad de vida pues se debe proveer más servicios a más personas. Finalmente, se ha considerado que a mayor o menor proporción de divisiones territoriales se vea afectada la calidad de vida ya que, en su mayoría, se habla de países con zonas centralizadas que perjudican otras áreas geográficas por lo que se podría considerar que a mayor distribución, mejor calidad de vida.

# METODOLOGÍA

La metodología aplicable en el presente trabajo de investigación consiste en un análisis comparativo multidimensional puesto a que se abordan temas político - gubernamentales tales como el régimen político, la estructura legislativa y la división administrativa territorial. Sin embargo, se aborda a su vez temas sociales como la cantidad de la población del territorio, además de incluir la distribución territorial previamente mencionada así como aspectos económicos como el PBI per cápita y la tasa de inflación.

Para el análisis estadístico de cada uno de estos componentes se ha optado, como se ha visto en las clases teóricas y prácticas, el uso del programa Rstudio que permitirá un análisis profundo, tanto del análisis descriptivo como de las correlaciones, que permitirá llegar a conclusiones frente a la hipótesis planteada. Por lo tanto, acorde a ello se ha optado, en principio, seleccionar las fuentes confiables que construirán las bases del trabajo, estas son Cia.gov y el Banco Mundial como las fuentes principales para hacer el índice de calidad de vida y para recopilar información respecto a las variables predictoras. Posteriormente, se delimitó el año que sería analizado, este es el año 2022 debido a que, en muchos casos, no había información actualizada al 2023 y, también, porque permite una evaluación post pandemia. En ese sentido, todas las variables, incluidas las que se usaron para armar el indicador de calidad de vida, son hasta el año 2022.

Además, para llegar a las conclusiones se ha hecho uso de distintas pruebas y códigos en el procesamiento de datos en el programa en mención. Para el logro del análisis deseado, se realizó un índice de calidad de vida a través del Análisis Factorial Exploratorio (EFA). Para realizarlo, primero se necesitó recopilar los datos para, posteriormente, estandarizarlos. Ello permitió que se realice el análisis factorial, el cual se demuestra a través de una gráfica de correlaciones y la realización de las pruebas tales como Barletto e identificar si la matriz era singular. Posteriormente, se deseaba hacer el análisis de clústers a través de la estrategia de partición como jerárquica.

Adicionalmente, según las variables seleccionadas entre numéricas y categóricas, se procedió a los análisis correspondientes a través de modelos de asociación / correlación y los modelos de regresión. En relación a las variables numéricas se realizaron fórmulas para representar las correlaciones entre la calidad de vida y la población, el PBI per cápita, la tasa de inflación y la división administrativa. Adicionalmente, a cada una de estos análisis se le realizó dos pruebas: la de pearson como la de spearman para analizar su significancia y la fuerza frente a la calidad de vida. Asimismo, cada uno de estos fue representado en gráficos de puntos. En el caso de las variables categóricas, a las que se les factorizó, también se realizó regresiones para la correlación lineal a las cuales se les aplicó la prueba T y Mann Whitney (la única variable a la que no se le pudo realizar es al régimen político (reg\_pol)). Al igual que las variables numéricas, estas también se les representó en gráficos con la diferencia que fue en boxplot o gráfico de cajas. Como última etapa del análisis estadístico deseado se realizó un modelo lineal: primero uno general que incluye todas las variables con las que se desea encontrar una influencia en calidad de vida y, posteriormente, se desagregó acorde afinidad. En el primer caso se realizó una evaluación sobre cómo influye cada variable en la calidad de vida en el que se ve cómo cada régimen político puede tener un impacto. Asimismo, se realizaron dos modelos más en el que se distingue el impacto de una estructura legislativa unicameral y la otra con una bicameral.

Finalmente, en el desglosamiento de cada uno se agrupó por criterios semejantes: primero, según el tipo de variable (numérica o categórica) en el cual se evalúa régimen político con estructura legislativa que, a su vez, están relacionadas a cómo se toman las decisiones; y, segundo, campo de acción que involucra, en las variables numéricas, las relacionadas al aspecto económico (tasa de inflación y PBI per cápita) y el campo social (población y distribución administrativa territorial). Cada uno de estos da como resultado distintas puntuaciones al ver qué tanta influencia tienen con respecto a la calidad de vida. En ese sentido, permite la delimitación más meticulosa según el campo de acción.

# LIMPIEZA DE LA BASE DE DATOS

La realización de la base de datos tuvo dos partes: la primera consistió en la construcción de una base de datos que contenga los indicadores que construirían la variable en cuestión: la calidad de vida. Esta se construyó en el mismo programa Rstudio y tuvo diversos procesos: 1) juntar varias bases de datos que construya una sola, 2) la recategorización de algunas variables, 3) la limpieza de los n.a. y 4) la estandarización de los datos.

La primera etapa consistió en importar cinco bases de datos del Banco Mundial y juntar cada una con las previas a través del comando “merge”. Respecto a la recategorización de algunas variables se dio en dos casos en específico: empleabilidad y salud. En el primero, debido a que los datos hacían referencia al porcentaje de la fuerza laboral desempleada cuando lo que se quería era la cantidad de personas empleadas; es decir, que a mayor empleabilidad, mejor calidad de vida. En el segundo caso, se presentaba la misma situación solo que respecto a la tasa de mortalidad, por lo cual se buscaba identificar que a menor probabilidad de mortandad había mayor calidad de vida. Por lo tanto, se tuvo que sacar el máximo y mínimo con la función summary y poder recategorizar los datos en cada variable.

En referencia a la limpieza de los n.a se realizó a través de la función “na.omit” ya que, para crear el indicador de calidad de vida a través del análisis factorial, se debía descartar todas las variables con datos perdidos. El proceso final de la primera base de datos es la estandarización de los datos que consistía en que todas las variables que construirían el índice de calidad de vida se encuentren en un mismo rango. Es a través de este que se crea un nuevo data frame llamado “datos\_std” que, únicamente, contiene todos los datos estandarizados.

Respecto a la segunda base de datos, esta fue construida manualmente por las autoras del presente trabajo. Esta base consistió en la recopilación de los datos de cada país para saber las características acorde a lo que se desea encontrar puesto a que no hay una base de datos que conglomere cada una de las variables. En ese sentido, al igual que la previa, se tuvo que realizar la limpieza de la misma con el comando “na.omit” y que se pueda realizar las pruebas para realizar los clústers y los demás análisis. Se ha de considerar que en cada proceso de limpieza de n.a. se reducía la cantidad de observaciones a analizar. Entre los países europeos y latinoamericanos se llegó a un total de 67 países, pero con las limpiezas realizadas para cada etapa del análisis estadístico para la comprobación de la hipótesis, la data final contiene 45 países entre países de América Latina y países de Europa.

# HALLAZGOS

## Análisis Factorial

Para crear la variable dependiente calidad de vida, se empleó el análisis factorial exploratorio (EFA), mediante el cual se obtuvó un índice en función a cuatro variables medibles: matricula\_edu22, mortalidad\_rec, saneamiento, telef\_fija22, y empleo\_rec.

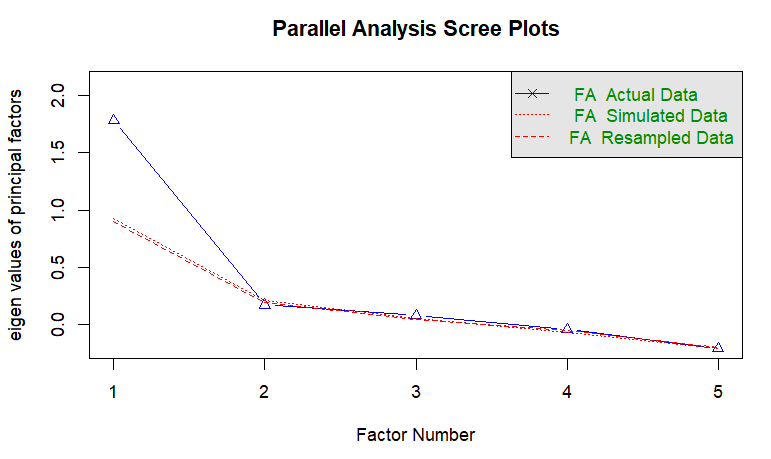
Para ello, primero, se calculó y graficó las correlaciones entre esas variables, en donde se halló que, siguiendo el criterio de Cohen, la variable matricula\_edu presenta una correlación débil y directa con las variables tele\_fija22 y saneamiento. No obstante, con la variable empleo\_rec manifiesta una relación inversa, es decir, mientras más porcentajes de alumnos matriculados en el nivel primario, el porcentaje de población empleada disminuye ligeramente, o viceversa. Asimismo, su correlación con mortalidad\_rec es no relevante, puesto que se encuentra en el intervalo de -0.1 y 0.

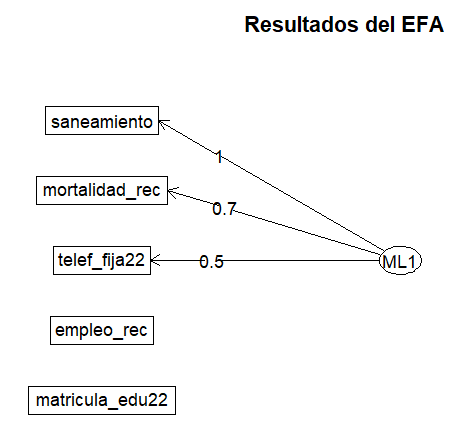
Cabe resaltar que las variables empleo\_rec y mortalidad\_ rec ya se encuentran recategorizadas para sus usos en la operacionalización de la calidad de vida. En este sentido, la primera equivale a un porcentaje de desempleo, donde su proporción más alta implica una mayor empleabilidad, y su más bajo porcentaje a una menor. De igual forma, la proporción mayor de mortalidad, señala una menor probabilidad de morir.

Además, la variable empleo\_rec presenta una correlación débil e inversa con saneamiento y mortalidad\_rec. En otras palabras, cuanto se incremente el porcentaje de población empleada, se reduce levemente el porcentaje de la población que utiliza al menos servicios básicos de saneamiento, y el porcentaje total de muertes no relacionadas por enfermedades transmisibles y condiciones maternales, prenatales y en materia de nutrición. También, muestra una correlación no relevante e inversa con tele\_fija 22.

En adición, la variable tele\_fija 22 evidencia una correlación moderada y positiva con saneamiento y mortalidad\_rec. En esta aspecto, mientras mayor es el porcentaje de suscripciones a telefonía fija, mayor es el porcentaje de la población que utiliza servicios básicos de saneamiento, y el porcentaje total de causas de muerte no vinculadas por enfermedades transmisibles y condiciones maternales, prenatales y en materia de nutrición, o viceversa.

Incluso, la correlación entre saneamiento y mortalidad\_rec es una fuerte y positiva, es decir, cuanto más incremente el porcentaje de la población que utiliza servicios básicos de saneamiento, mayor es el porcentaje total de muertes no relacionadas por enfermedades transmisibles y condiciones maternales, prenatales y en materia de nutrición.

Segundo, se aplicó el índice KMO (Kaiser - Meyer - OLkin) para evaluar si los datos permiten factorizar, es decir, son idóneos para realizar un análisis factorial. Sobre este, lo aceptable es que los valores de cada variables medibles sean mayores a 0.6. No obstante, solo la variable de tele\_fija22 presenta un valor de 0.75, y las demás se encuentran en el rago de 0.47 a 0.58, siendo así datos menos adecuados para el presente análisis. 

Tercero, para verificar si la matriz de correlación es adecuada para realizar un análisis factorial, se aplican dos pruebas: Bartlett y singular. Sobre la primera prueba, sirve para comprobar la hipótesis nula, que señala que la matriz de correlación es una matriz de identidad, es decir, que las variables no están correlacionadas, por lo que se desea obtener como resultado FALSE. De igual manera, con la segunda prueba, la hipótesis nula indica que la matriz de correlación es una matriz singular, lo cual implica una redundancia de algunas variables, a lo que se desea como resultado FALSE. Finalmente, una vez pasada ambas pruebas, se determina en cuántos factores se redimensionan las variables medibles, donde el resultado fue de una variable latente: un factor. Esta se puede visualizar el tercer gráfico con el título de “Resultados del EFA”. 

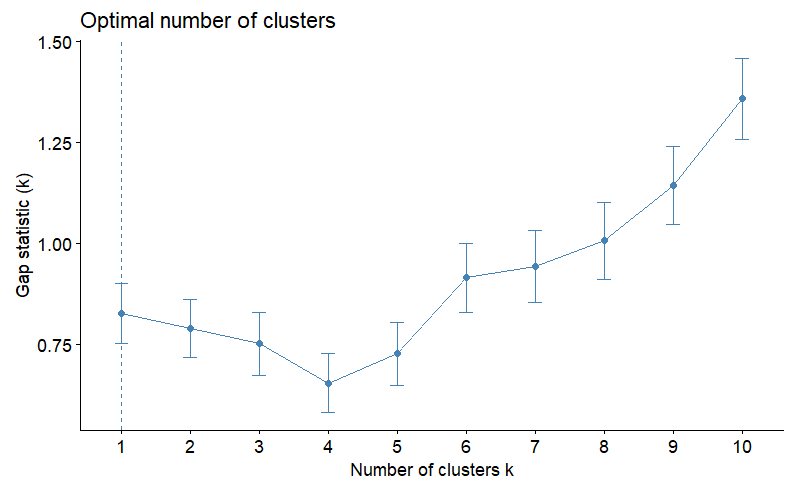
## Análisis Clúster

Posterior al Análisis Factorial Exploratorio (EFA) se intentó con el Análisis Clúster para clasificar a los casos u observaciones (los países) en subconjuntos según las similitudes que presenten y justifiquen su pertenencia a los grupos asignados. En ese sentido, se probó las dos estrategias de análisis: la de partición y la jerárquica. Es importante resaltar que es R quien propone la cantidad de clústers para los datos y, aunque uno mismo pueda asignar un número de clústers, el ideal es seguir lo recomendado por el programa. Asimismo, es importante recordar que el Análisis Clúster tiene como finalidad el análisis de las sombras de las tres estrategias: PAM, Agnes y Diana. Aquella que tenga el resultado de sombra más alta y pocos datos mal agrupados es el ideal. Sin embargo, para el presente trabajo académico, a pesar de haberse intentado, no se obtuvo los resultados esperados por lo cual no se pudo agrupar en clústers. En ese sentido, se presentarán las gráficas de los clústers recomendados por R según cada estrategia.

### Estrategia de partición

En la estrategia de partición se encuentra la prueba PAM que es uno de los medios que nos puede decir bajo qué estrategia es posible agrupar los datos. En este se agrupan datos no solapados en el cual el objetivo es dividir los datos dentro de un clúster con mucha semejanza y que los demás clústers sean muy diferentes de uno.

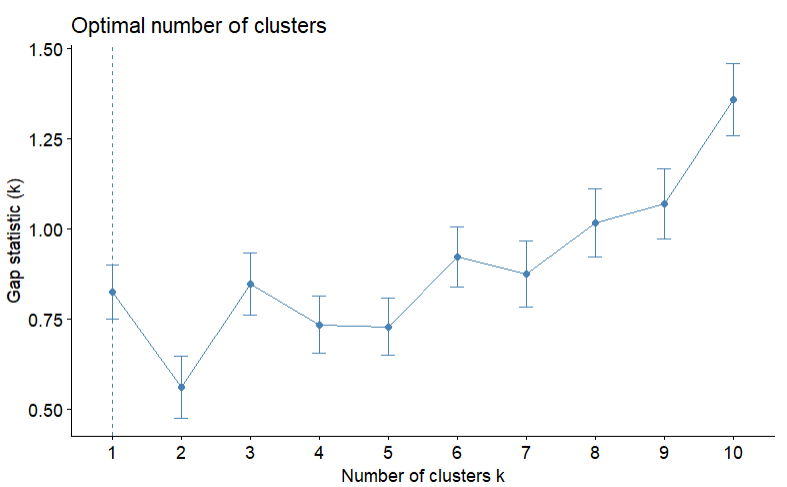
#### PAM

Al codificar acorde a PAM, R mostró en la gráfica cuántos cluster son recomendables hacer, el resultado fue salió 1. Acorde a lo estudiado, no es posible hacer un único clúster y que el presente resultado representaría que no hay varianzas significativas entre las observaciones estudiadas (países). Por lo tanto, ello se interpreta como una homogeneidad entre los casos de estudio y, por ende, es recomendable un único clúster.

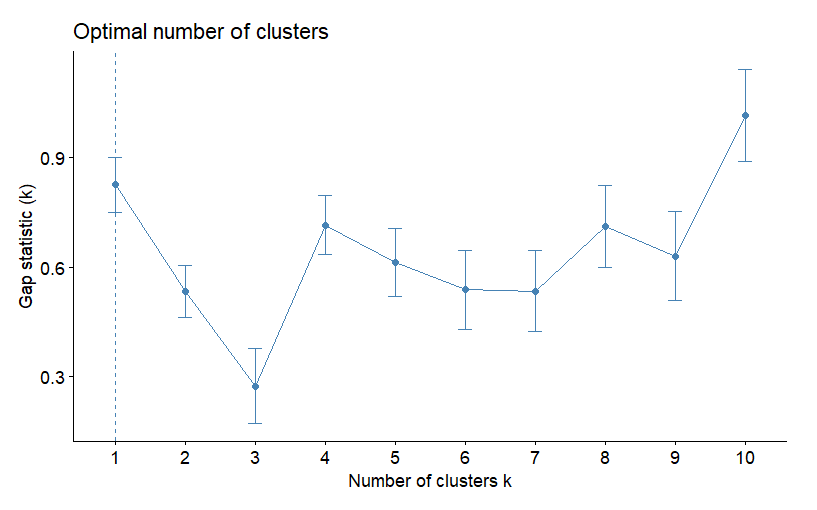
### Estrategia jerárquica

En el análisis de la estrategia jerárquica se suele agrupar los datos según semejanzas, pero acorde a una consecutividad; es decir, se juntan por similitudes y se van fusionando progresivamente. En este caso se identifica que los clústers son asignados según sus similitudes y los otros grupos no son tan distantes de uno sino que cumple con una secuencia progresiva.

#### AGNES

Al realizar el Análisis Clúster a través de la estrategia jerárquica, primero con Agnes, el resultado fue igual que con PAM en la estrategia de partición. La recomendación de R también fue de realizar un único clúster que demuestra, nuevamente, que los datos (países) en el análisis son muy semejantes a los datos analizados por lo que no es viable agruparlos en varios clústers. 

#### DIANA

La última estrategia jerárquica realizada fue la de Diana. Sin embargo, al igual que los dos análisis previos, recomendó un solo clúster por lo que llevó a que en el análisis no se realice la agrupación de países por clústers. Ello debido a que, como se ha mencionado previamente, no es viable realizar clústers con uno solo. A pesar de que en el análisis uno puede asignar el número que le parezca, siempre el ideal es seguir la recomendación de R ya que brindaría con mayor exactitud los resultados. 

## Análisis bivariado

El análisis bivariado consiste en la evaluación de dos variables y cómo una puede influir o asociarse con otra variable. Por lo tanto, en este tipo de análisis se busca interpretar la correlación entre variables (una dependiente y otra independiente) para predecir o llegar a conclusiones. En el presente trabajo se ha realizado el análisis bivariado, primero, entre variables numéricas y, posteriormente, entre variables numéricas con categóricas. Al tener el índice de calidad de vida (cal\_vida) que es numérica, se ha realizado con cada una de las variables el análisis correspondiente.

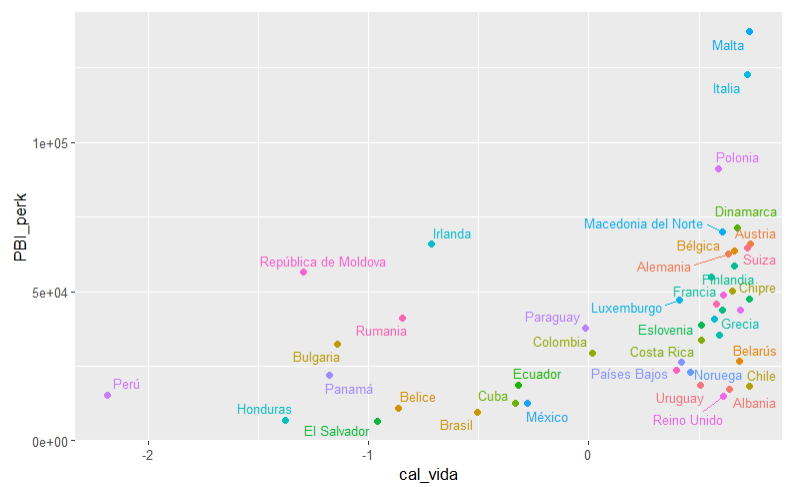
### Numérica con numérica

#### Calidad de vida con población

En el análisis de calidad de vida con la población de los países, en ambos casos variables numéricas, se obtuvo como resultado que la mayoría de los países se concentran en el mayor puntaje de calidad cuando hay una menor densidad poblacional. Entre los datos que sobresalen en la gráfica se encuentra Perú que, si bien no cuenta con una densidad poblacional grande, su índice de calidad de vida es muy bajo. Asimismo, se encuentra Brasil que, a pesar de ser el país con mayor densidad poblacional, se podría decir (superficialmente) que se encuentra dentro del promedio del índice de calidad de vida. Finalmente, se debe reconocer que la mayoría de países que resaltan entre una población promedio a baja y con un índice de calidad de vida favorable se encuentra la mayoría de países europeos y Chile. Mientras que en el promedio se logra identificar un mayor porcentaje de países latinoamericanos.

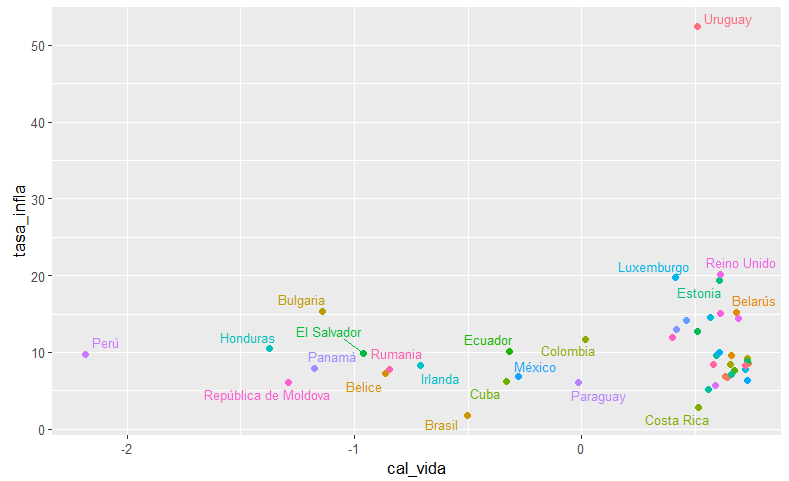
Además de la gráfica se realizó dos pruebas, la de Pearson y la de Spearman. Los resultados de Pearson demostró que su p - value es de 0.63 que es mucho mayor al valor promedio (0.05) de significancia de cuánto explica una variable. Por lo tanto este resultado demostraría que no hay una evidencia suficiente para explicar la correlación y, por ende, no se puede concluir que entre la variable de calidad de vida con la densidad poblacional haya una correlación. Asimismo, se demostró que el estimado de correlación es de -0.07 que representa la fuerza de significancia, en este caso al no haber significancia, también se concluye que sería muy débil. En el caso de Spearman se obtuvo de resultado que el p-value era de 0.79 por lo que tampoco demuestra que haya una correlación entre las variables analizadas. Por lo tanto, se concluye que las dos pruebas estadísticas demuestran que no hay una correlación significativa entre las variables.

#### Calidad de vida con PBI per cápita

La segunda variable analizada junto a calidad de vida es el PBI per cápita, una de las variables de control. En este caso, lo primero que se logra apreciar es que, nuevamente, Perú se encuentra muy lejos de los demás datos lo que demuestra que teniendo un bajo PBI\_per cápita el índice de calidad de vida de Perú es, a su vez, bajo. Asimismo, se encuentra entre los países con mayor PBI per cápita a Malta, Italia y Polonia que también cuentan con un alto índice de calidad de vida. Asimismo, se ha identificado que la mayoría de países latinoamericanos se encuentran en el promedio del índice de calidad de vida, a pesar de tener un PBI per cápita relativamente bajo. Entre las excepciones de países latinoamericanos se encuentra Uruguay, Chile y Costa Rica. 

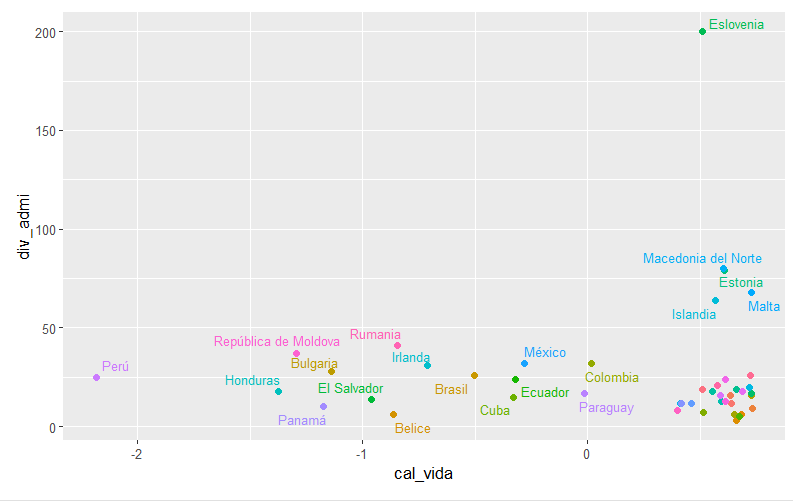
En el presente análisis también se realizaron las pruebas de Pearson y Spearman de los cuales se tuvo como resultado en el p - value de 0.003 y 7.246962e- respectivamente lo que demuestra que sí hay una correlación significativa. Asimismo, acorde a la fuerza de la correlación se obtuvo 0.43 y 0.56 respectivamente que demuestra que la fuerza de la correlación entre las variables es moderada en el caso de Pearson y fuerte en el caso de Spearman. Por lo tanto, se concluye que el PBI per cápita sí influye en la calidad de vida de las personas de manera significativa.

#### Calidad de vida con tasa de inflación

La siguiente variable a analizar es la de calidad de vida con la tasa de inflación, ambas también numéricas a las cuales también se les aplicó la prueba de Pearson y Spearman. En el gráfico realizado también se encuentra Perú como el que tiene el menor índice de calidad de vida a pesar de tener una tasa de inflación promedio. Por el contrario, se encuentra a Uruguay que tiene una tasa de inflación alta y cuenta con una calidad de vida muy favorable. Asimismo, se percibe que la mayoría de países se concentra en el área de mayor calidad de vida con una tasa de inflación entre el 5 y el 10%. 

Respecto a las pruebas realizadas, el resultado dado en R posterior a la fórmula realizada, dio como resultado que el p - value de las variables fue de 0.34 en el caso de Pearson y de 0.86 en el caso de Spearman. Estos resultados demuestran que no hay una correlación significativa entre las variables analizadas. Por lo tanto, se concluye que la tasa de inflación no influye en la calidad de vida de los países a pesar de haberla considerado una variable de control.

#### Calidad de vida con división administrativa

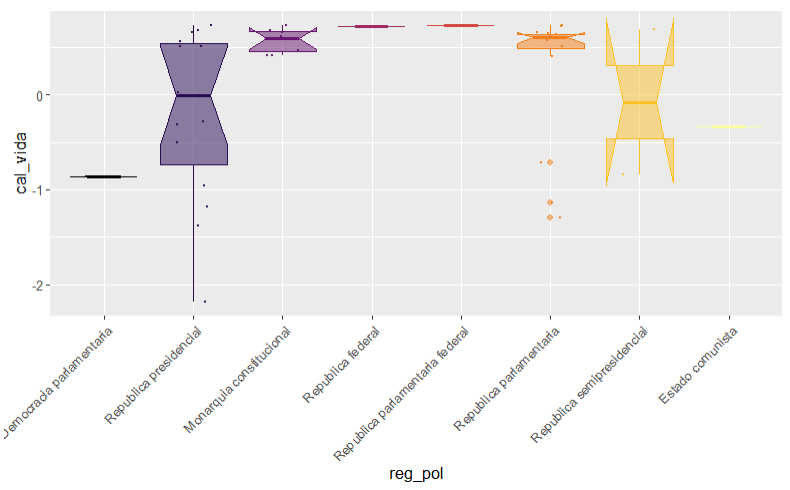
La última variable numérica con la que se analiza el índice de calidad de vida realizado es la división administrativa de un territorio. En el gráfico se puede apreciar una mayor dispersión de los puntos, siendo los más llamativos Perú y Eslovenia. En el caso de Perú por encontrarse nuevamente con menor calidad de vida y con una división territorial aparentemente con el promedio. Por el contrario, Eslovenia cuenta con un índice alto de calidad de vida, pero con una distribución territorial fuera de lo común ya que excede los 180 acorde el gráfico. 

Al igual que las demás variables, se realizaron las pruebas de Pearson y Spearman las cuales demuestran que entre la división administrativa territorial y la calidad de vida no hay una correlación significativa y débil. En el caso de Pearson se obtuvo como resultado 0.75, muy por encima del p - value deseado de 0.05; por el contrario, con la prueba de Spearman se obtuvo un p - value de 0.23, menos que en la prueba previa, pero aún así sobre el ideal.

### Numérica con categórica

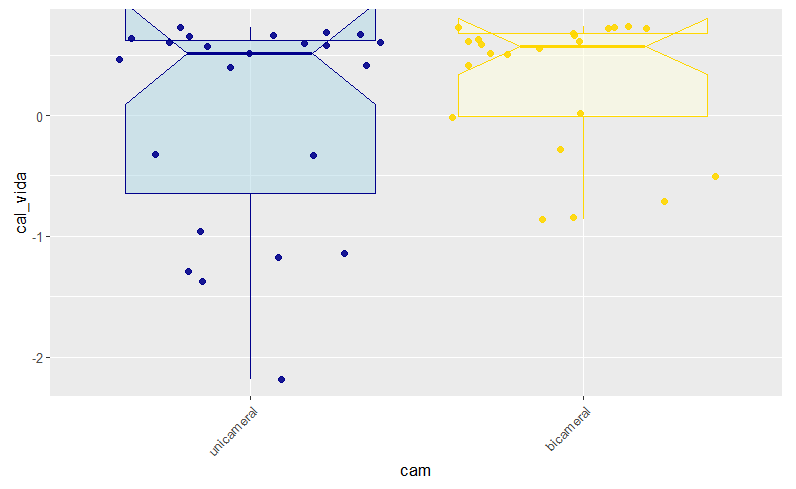
En esta sección se optó por evaluar también según el continente puesto a que la variable de régimen político, al tener varios niveles, no fue factible realizar las pruebas deseadas como la prueba T y la prueba de Mann - Whitnery. Por lo tanto, los gráficos obtenidos son 3: régimen político (reg\_pol), estructura legislativa (cam) y continente (CONTINENTE).

#### Calidad de vida y régimen político

Si bien no se pudo realizar las pruebas deseadas, se pudo graficar cómo se ve la distribución de los datos. Es importante resaltar que la presente variable de análisis, régimen político, se factorizó y se realizó la tabla de resultados. En esta se obtuvo como resultado que hay un solo país que cuenta con una Democracia Parlamentaria, uno con una República parlamentaria federal, también con República federal y un país con un Estado comunista. Asimismo, hay solo dos países que cuentan con una República semipresidencial, dos con una monarquía constitucional, quince países en una República presidencial y dieciséis en una República Parlamentaria.

En el gráfico se consigue identificar que hay una mayor variabilidad en los regímenes de República Presidencial pues en dicho caso hay mayor dispersión de los datos que equivale a una mayor heterogeneidad. Asimismo, se puede ver que en el caso de la Monarquía Constitucional y el de República Parlamentaria hay una mayor heterogeneidad aunque en el último tipo de régimen cuentan con más datos atípicos. Finalmente, se percibe que hay una mayor heterogeneidad en una República semipresidencial por el largo de la caja que representa a los datos.

#### Calidad de vida con estructura legislativa

La variable de estructura legislativa también se factorizó y se encontró que veintitrés países cuentan con una estructura legislativa unicameral y veintidós países cuentan con uno bicameral. Acorde a ello, se puede apreciar en el gráfico que hay una mayor heterogeneidad de datos en el caso de los países con una estructura legislativa unicameral a diferencia de los que tienen la bicameral. Asimismo, se percibe que la mayoría de datos se concentra en la parte superior de calidad de vida, que es la mayor puntuación obtenida de los índices. Pero también se demuestra que la mayoría de países que tienen una puntuación más débil o baja respecto a calidad de vida se encuentra en el sistema unicameral. 

Acorde a las pruebas a realizar que son la prueba T y el Mann-Whitnery se obtuvo como resultado que en el caso de la primera prueba el p - value fue de 0.209, mientras que en la segunda prueba es de 0.207. En ambos casos se concluye que no hay una correlación significativa entre la calidad de vida con la estructura legislativa que pueda tener un país. Por lo tanto, no se puede asumir que el proceso en el que se formulan, crean y aprueban leyes y que puede generar un impacto en la calidad de vida de las personas sea cierta.

#### Calidad de vida con continente

Finalmente, por curiosidad y para poner en práctica las pruebas vistas en el curso se optó por el análisis de cómo influencia ser de un continente en la calidad de vida. Acorde al gráfico se demuestra que en el caso de Europa sí hay una mayor calidad de vida, hay una homogeneidad de datos y cuentan con valores atípicos. Por el contrario, en América Latina se encuentra que hay datos demasiado dispersos, cuenta con datos extremos, pero no cuenta con datos atípicos.

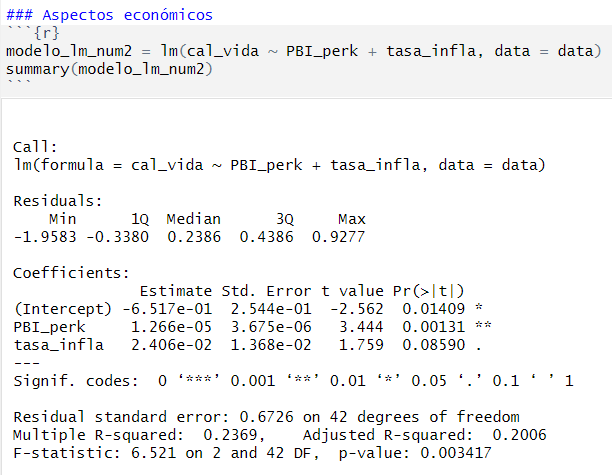
Según las pruebas realizadas se obtuvo como resultado que, acorde a la prueba T o student - T, el p - value es de 0.002 lo que demuestra que sí hay una correlación significativa. Ello demuestra que la calidad de la vida va acorde a si una persona es o no de un continente ya que este es el que define la calidad de vida al tener una correlación. Del mismo modo, la prueba de Mann - Whitnery dio un resultado aún más significativo al ser este de 0.0005. Por ende, en ambos casos se concluye que la calidad de vida se rige según al continente perteneciente. Por lo tanto, el ser parte de un continente u otro tiene un impacto en la calidad de vida en los distintos ámbitos en los que se evalúe.

## Análisis multivariado

Por último, se realiza un análisis múltiple para analizar la relación entre la variable calidad de vida (v. dependiente) y las seis variables independientes presentadas. Para ello, se realizan tres tipos de regresiones lineales múltiples: una general; que incluye todas las variables independientes; con las variables categóricas; con las variables numéricas, en donde se juntó a las variables según su vinculación con aspectos territoriales y económicos.

Con respecto a la primera regresión lineal múltiple, se puede evidenciar un p-value igual a 0.1814, por lo que no se puede rechazar la hipótesis nula, debido a que su p-value es mayor a 0.05. En este sentido, el modelo no es válido, lo que indica que las variables independientes de forma conjunta no tienen un efecto estadísticamente significativo sobre la variable dependiente (calidad de vida). No obstante, de forma individual, se muestra una significancia de 0.1 tanto al régimen político de democracia parlamentaria como a la tasa de inflación, y una de 0.05 al régimen político de república presidencial.

Así mismo, se presentó dificultades para poder evaluar de forma conjunta las categorías de unicameralidad o bicameralidad de la variable cam\_factor en la anterior regresión lineal múltiple. Ante ello, se aplicó dicha regresión dos veces, donde en cada una se analiza a las dos categorías de dicha variable independiente. De igual forma, en dichas regresiones, el p-value sigue siendo mayor a 0.05, lo que señala la invalidez del modelo.

Además, conforme a las regresiones lineales múltiples con las variables categóricas, se presenta una regresión con las variables reg\_pol y cam\_factor, en donde se analizan las ocho categorías de régimen político y de forma conjunta la de estructura legislativa. En esta regresión, el p-value sigue siendo mayor a 0.05 con un valor de 0.1257, por lo que el modelo no influye de forma estadísticamente significativa a la calidad de vida. Sin embargo, el régimen político de república presidencial muestra una significancia de 0.1. De igual modo que en la primera regresión múltiple, se realizó dos regresiones que incluyen cada una a unicameralidad o bicameralidad, donde el p-value sigue siendo 0.1257. 

Adicionalmente, se aplicó una regresión múltiple a las variables numéricas relacionados con aspectos territoriales: pobla y div\_admi. En esta regresión, el p-value es igual a 0.856, por lo que el modelo no es válido. Incluso, ninguna variable presenta algún nivel de significancia de forma independiente. No obstante, si se realiza una regresión lineal múltiple con las variables numéricas vinculadas a aspectos económicos: PBI\_ perk y tasa\_infla, el p-value es 0.0034. Ello señala que sí rechaza la hipótesis nula, por lo que el modelo sí es válido. También, el modelo, dado por el PBI per cápita y la tasa de inflación, explica en 23% la variabilidad de la calidad de vida. Asimismo, ambas variables independientes presentan un nivel de significancia estadística con respecto a la variable dependiente, siendo el PBI\_perk a un 0.01 y la tasa\_infla a un 0.1.

# CONCLUSIONES

Entre las principales conclusiones se ha encontrado que la variable que sí tiene un grado apropiado de significancia es el PBI per cápita, una variable de control. Esta se debe comprender como un aspecto fundamental respecto a la calidad de vida de los países, es decir que hay una correlación directa entre el PBI per cápita y la calidad de vida. Asimismo, se ha encontrado que, si bien hay variables que tienen un grado de significancia al 0.05 o al 0.1, el modelo en su totalidad no puede predecir ni cuenta con una correlación con la variable analizada: la calidad de vida. Finalmente, se ha identificado que una variable que no estaba considerada dentro de las principales, CONTINENTE, también es la otra variable que tiene una correlación muy fuerte con respecto a la calidad de vida de cada país.

Por lo tanto, en base a las pruebas y el profundo análisis estadístico realizado se puede responder a la pregunta principal que se planteó al principio del trabajo ¿qué tanto impacto pueden tener diversos criterios sociales, económicos y políticos (gubernamentales) en la calidad de vida de los países latinoamericanos y europeos? ¿Estos criterios son los que definen que sean países desarrollados o en vías de desarrollo?

La respuesta a la que se ha llegado es que la única variable que tiene un impacto en la calidad de vida es la variable del PBI per cápita (PBI\_perk) que en todas las pruebas ha demostrado un nivel alto nivel de significancia. Por el contrario, los otros aspectos tanto gubernamentales como sociales, no influyen en la calidad de vida ni tienen correlación entre sí. Por lo tanto, no se puede decir que, acorde a las variables seleccionadas, estas definan quiénes son países desarrollados ni cuáles no, ni que sean estas variables las que las posición entre los países desarrollados o los de vías de desarrollo.

# BIBLIOGRAFÍA

Banco Mundial (2022) *Inscripción escolar, nivel primario (% bruto)* <https://datos.bancomundial.org/indicator/SE.PRM.ENRR>

Banco Mundial (2022) *Causa de muerte, por enfermedades transmisibles y condiciones maternales, prenatales y en materia de nutrición (% del total)* <https://datos.bancomundial.org/indicador/SH.DTH.COMM.ZS>

Banco Mundial (2022) *People using at least basic sanitation services (% of population)* <https://datos.bancomundial.org/indicador/SH.STA.BASS.ZS>

Banco Mundial (2022) *Suscripciones a telefonía fija (por cada 100 personas)* <https://datos.bancomundial.org/indicator/IT.MLT.MAIN.P2>

Banco Mundial (2022) *Desempleo, total (% de la fuerza laboral total) (estimación modelada de la OIT)* <https://datos.bancomundial.org/indicator/SL.UEM.TOTL.ZS>

Banco de Desarrollo de América Latina y el Caribe (2018). *Los grandes beneficios económicos de gestionar bien el agua.*

<https://www.caf.com/es/conocimiento/visiones/2018/03/los-grandes-beneficios-economicos-de-gestionar-bien-el-agua/>

Central Intelligence Agency.

<https://www.cia.gov/the-world-factbook/countries/>[[1]](#footnote-0)

Cuadros, H. (2017). *La importancia del Poder Legislativo en el Perú desde el marco constitucional.* LP Derecho <https://img.lpderecho.pe/wp-content/uploads/2017/09/Importancia-del-Poder-Legislativo-2017.pdf>

Quality of Life Initiative (2023). *A Brief History of Quality of Life: From Aristotle to the Present, Evolution of a Concept at the Core of Human Development.* <https://unhabitat.org/sites/default/files/2024/02/quality_of_life_initiative_synthesis_paper_1_a_brief_history_of_quality_of_life.pdf>

1. La página Central Intelligence Agency (CIA.gov) es la fuente que se usó para armar la base de datos manualmente para cada país; es decir, en esta página se exploró los 67 países que se tuvieron desde el inicio del trabajo de investigación para el análisis estadístico correspondiente. [↑](#footnote-ref-0)