



Centro de
Investigación en
Matemáticas, A.C.

Análisis de la calidad de vida a través de diferentes herramientas estadísticas

Estadística multivariada

Alondra E. Matos Mendoza

María G. López Salomón

Monterrey, 5 de junio de 2023

Agenda

- 1 Introducción
- 2 Metodología
 - Descripción de la muestra
 - Descripción de la metodología
- 3 Resultados
 - Regresión lineal múltiple
 - MDS
- 4 Conclusiones



Introducción

En varios países, existe una creciente demanda de políticas públicas centradas en aspectos relevantes para las personas, buscando el bienestar de los ciudadanos y promoviendo el desarrollo personal y comunitario, por lo que garantizar la felicidad de los ciudadanos se considera una responsabilidad gubernamental y un indicador de progreso social.

En este contexto, el uso de técnicas estadísticas avanzadas ha demostrado ser prometedor para analizar la calidad de vida y la felicidad de manera más precisa y completa, mediante variables que permitan interpretar el estado de ánimo así como el bienestar de los ciudadanos.



Objetivos

- **Objetivo General**

- Analizar la calidad de vida utilizando un enfoque multidimensional que incorpore variables como el PIB per cápita (GDP), el apoyo social percibido, la esperanza de vida saludable, la libertad para tomar decisiones vitales, la generosidad y las proporciones de corrupción.

- **Objetivos Específicos**

- Identificar y cuantificar la influencia relativa de cada una de las variables consideradas en la calidad de vida de las personas a través del análisis de regresión multivariada.
- Explorar el potencial del escalamiento multidimensional clásico (MDS) en el análisis de la calidad de vida.



Descripción de la base de datos

Los datos fueron provistos por Kaggle, los cuales se obtuvieron del Informe Mundial de la Felicidad del 2019. La base contiene 156 registros de países, y cada registro proporciona:

- **Score:** Índice de felicidad.
- **GDP per cápita (Gross domestic product).**
- **Apoyo social.**
- **Esperanza de vida saludable.**
- **Libertad de tomar decisiones.**
- **Generosidad.**
- **Percepciones de la corrupción.**

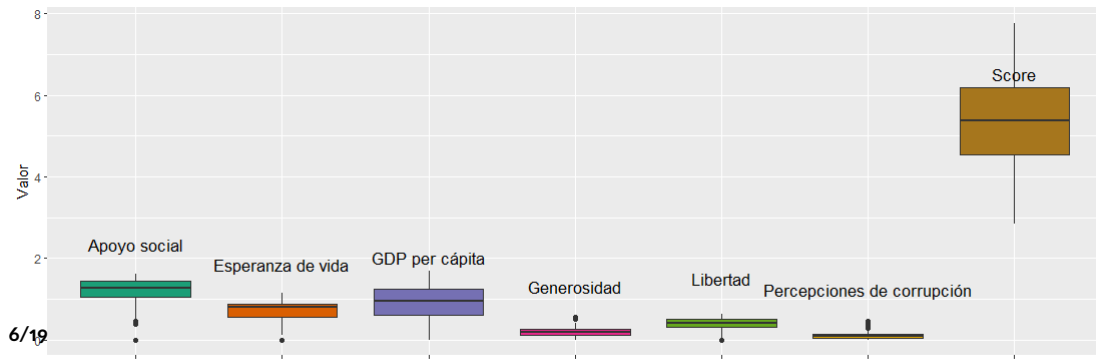
El informe detallado de cómo se estimaron los valores de estas variables se encuentra disponible en: <<https://worldhappiness.report/ed/2019/changing-world-happiness/>>



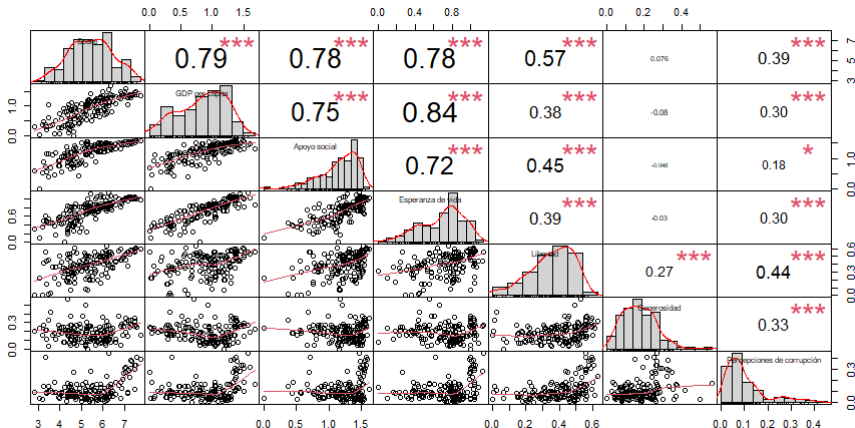
Descripción de la muestra

	Score	GDP per cápita	Apoyo social	Esperanza	Libertad	Generosidad	Percepciones de la corrupción
Mínimo	2.85	0.00	0.00	0.00	0.00	0.00	0.00
1er Cuantil	4.54	0.60	1.05	0.54	0.30	0.10	0.04
Mediana	5.38	0.96	1.27	0.78	0.41	0.17	0.08
Promedio	5.40	0.90	1.20	0.72	0.39	0.18	0.11
3er Cuantil	6.18	1.23	1.45	0.88	0.50	0.24	0.14
Máximo	7.76	1.68	1.62	1.14	0.63	0.56	0.45

Figura: Resumen estadístico



Metodología



Metodología

Ajuste de un modelo de regresión lineal.

- **Verificación de supuestos:**
 - **Homocedasticidad:** Gráfico de residuos estudentizados vs valores predichos;
 - **Normalidad:** Prueba de Anderson-Darling;
 - **Independencia:** Prueba de Durbin Watson;
- **Prueba de la razón de verosimilitud.**
- **Reajuste del modelo.**



Metodología

Escalamiento Multidimensional (MDS)

- **Cálculo de las distancias euclidianas.**
- **Construcción de la configuración MDS mediante la solución clásica.**
- **Elección de la dimensión de la representación euclidiana de acuerdo al porcentaje de la variabilidad explicada.**
- **Aplicación de K-means.**



Resultados

Cofiente de determinación: $R^2 = 0.779$

Raíz del error cuadrático medio: $RMSE = 0.50$

Variable	coef	std err	t	Intervalos de confianza de 95 %				
				Individuales			Simultáneos	
				P> t	[0.025	0.975]	[0.025	0.975]
Constante	1.7952	0.211	8.505	0.000	1.378	2.212	0.9915	2.599
GDP per cápita	0.7754	0.218	3.553	0.001	0.344	1.207	-0.0556	1.6064
Apoyo social	1.1242	0.237	4.745	0.000	0.656	1.592	0.2221	2.0263
Esperanza de vida saludable	1.0781	0.335	3.223	0.002	0.417	1.739	-0.1958	2.3521
Libertad de decidir	1.4548	0.375	3.876	0.000	0.713	2.197	0.0255	2.8841
Generosidad	0.4898	0.498	0.984	0.327	-0.494	1.473	-1.4056	2.3852
Percepciones de corrupción	0.9723	0.542	1.793	0.075	-0.099	2.044	-1.093	3.0376

Figura: Estimaciones de los coeficientes del modelo



Resultados

Cofiente de determinación: $R^2 = 0.771$

Raíz del error cuadrático medio: $RMSE = 0.53$

Variable	coef	std err	t	Intervalos de confianza del 95					
				Individuales			Simultáneos		
				P> t	[0.025	0.975]	[0.025	0.975]	
Constante	1.8921	0.199	9.491	0.000	1.4982	2.286	1.2198	2.5644	
GDP per cápita	0.8105	0.216	3.745	0.000	0.3829	1.2382	0.0807	1.5404	
Apoyo social	1.0166	0.235	4.331	0.000	0.5528	1.4804	0.225	1.8082	
Esperanza de vida saludable	1.1414	0.337	3.384	0.001	0.475	1.8079	0.004	2.2789	
Libertad de decidir	1.8458	0.340	5.423	0.000	1.1733	2.5184	0.698	2.9936	

Figura: Estimaciones de los coeficientes del modelo sin las variables Generosidad y Percepciones de corrupción



Regresión lineal múltiple

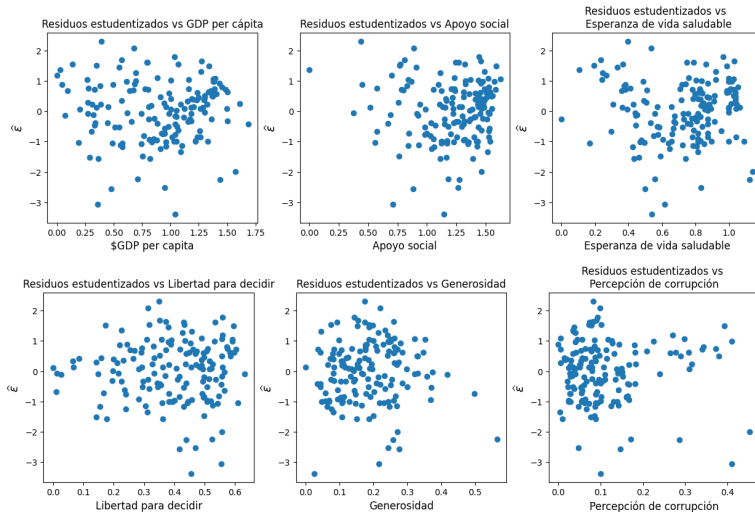


Figura: Gráfica de los residuales contra una variable predictora considerando el modelo completo



Resultados

Prueba de normalidad de Anderson-Darling: $\text{Valor } p = 0.11 > 0.05$

Prueba de autocorrelación: $1.5 < \text{Durbin} - \text{Watson} : 1.64 < 2.5$

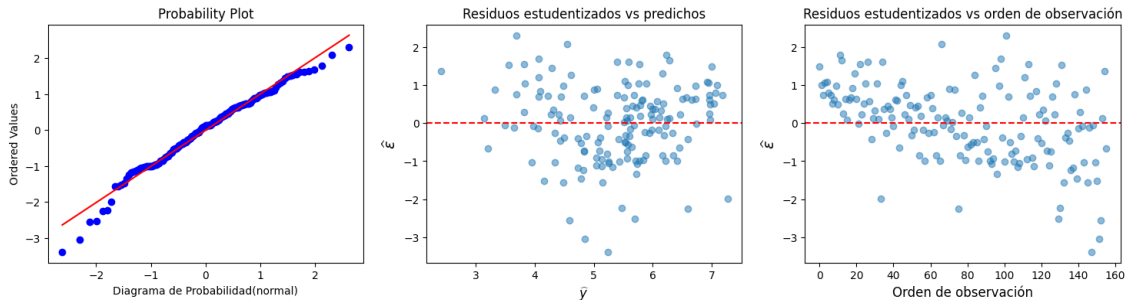


Figura: Verificación de supuestos del modelo completo

Resultados

Prueba de normalidad de Anderson-Darling: $\text{Valor } p = 0.27 > 0.05$

Prueba de autocorrelación: $1.5 < \text{Durbin} - \text{Watson} : 1.64 < 2.5$

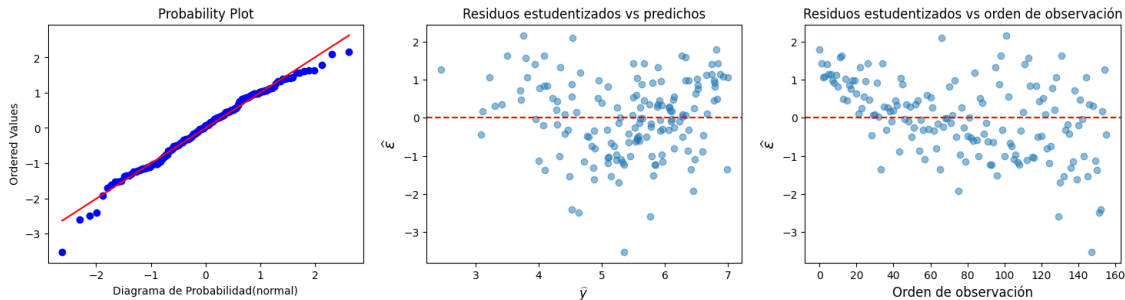


Figura: Verificación de supuestos del modelo sin las variables Generosidad y Percepciones de corrupción



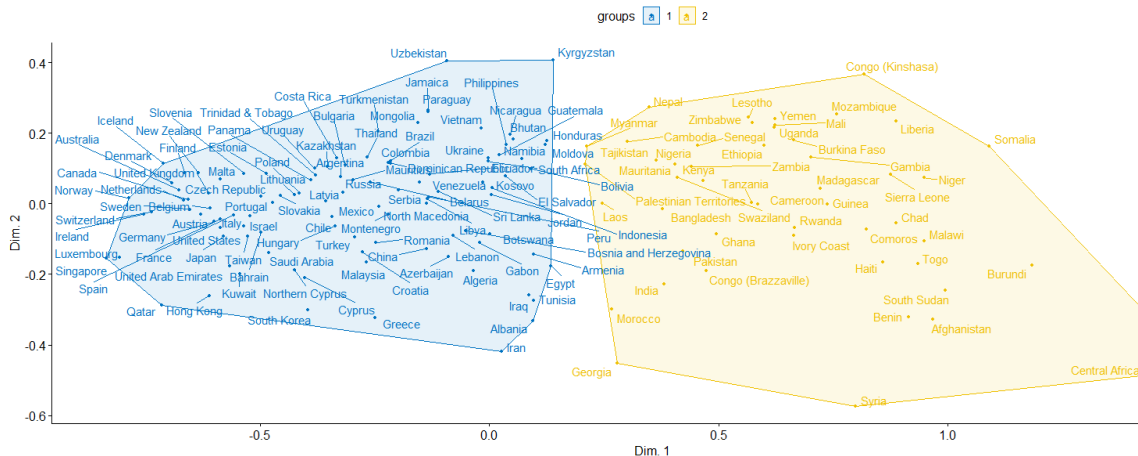


Figura: Representación en dos dimensiones obtenida mediante MDS

Grupo	GDP per cápita	Apoyo social	Esperanza de vida saludable	Libertad	Generosidad	Percepciones de corrupción
1	1.13	1.36	0.856	0.422	0.174	0.115
2	0.432	0.879	0.447	0.329	0.208	0.101

Tabela: Promedio de valores por grupo



Conclusiones

Se puede predecir el índice de felicidad (con una precisión aceptable) mediante el siguiente modelo lineal múltiple:

$$\begin{aligned}\text{Índice de felicidad} = & 1.89 + 0.81 * \text{GDP per cápita} + 1.01 * \text{Apoyo social} \\ & + 1.14 * \text{Esperanza de vida saludable} + 1.84 * \text{Libertad de decidir}\end{aligned}$$

Sin embargo, algunas variables predictoras están altamente correlacionadas, lo que aumenta la varianza del estimador y afecta el tamaño del intervalo de confianza.

La multicolinealidad se puede abordar con *Regresión Ridge* (*Regularización de Tikhonov*), que impone una penalización al tamaño de los coeficientes.




Conclusiones

Mediante MDS se puede representar las relaciones entre los los países en un espacio de dos dimensiones, en donde se observa la similitud de los países de acuerdo a su calidad de vida, determinado por las siguientes variables: GDP per cápita, Apoyo social, Esperanza de vida saludable, Libertad, Generosidad y Percepciones de corrupción.


No obstante, los clústeres resultantes de aplicar K-means en la configuración MDS de 2 dimensiones pueden superponerse en regiones, lo que sugiere la presencia de países con características compartidas. Esto puede hacer difícil asignar de manera precisa los países a un clúster específico.





Referencias

 Kemmler, G., Holzner, B., Kopp, M., Dünser, M., Greil, R., Hann, E., & Sperner-Unterweger, B. *Multidimensional Scaling as a Tool for Analysing Quality of Life Data*. [Rev. Quality of Life Research, 11, 3.] (2002).

 Potter, J., Cantarero, R., & Wood, H. *The Multi-Dimensional Nature of Predicting Quality of Life*. [Procedia- Social and Behavioral Sciences, 1877-0428.] (2012).

 Johnson, R. A., & Wichern, D. W. *Applied Multivariate Statistical Analysis*. Prentice Hall. (2007).

 Montgomery, D. C., Peck, E. A., & Vining, G. G. *Introduction to Linear Regression Analysis*. Wiley-Interscience. (2001).

 P. Hryoruk, S. Grygoruk, N. Khrushch, T. Hovorushchenko. *Using non-metric multidimensional scaling for assessment of regions' economy in the context of their sustainable development* [Khmelnyskyi National University, 11 Instytutaska Str. Khmenlnyskyi] (2020).

