

ANÁLISIS ESTADÍSTICO MULTIVARIADO

MAESTRÍA EN ECONOMETRÍA – MAESTRÍA EN ECONOMÍA

EXAMEN DOMICILIARIO

Fecha de entrega: lunes 23 de octubre de 2023.

Este trabajo está basado en los Indicadores de Desarrollo Mundial que publica el Banco Mundial. Estos indicadores comprenden una selección de variables económicas, sociales y ambientales conformada a partir de información del Banco Mundial y otras agencias. La base de datos completa cubre más de 1400 indicadores referidos a más de 200 economías a partir del año 1960 y se actualiza de manera periódica. Se puede acceder a estos datos través de este link:

<https://databank.worldbank.org/source/world-development-indicators#>

El objetivo de este trabajo es aplicar herramientas del análisis multivariado para realizar una descripción completa que permita comprender las similitudes y diferencias entre los países latinoamericanos y las interrelaciones entre las características analizadas en términos de desarrollo. Podrán consultar y utilizar los datos disponibles directamente desde la página web indicada más arriba o bien utilizar los datos de una consulta previa que abarca un período de 15 años recientes (años 2008 a 2022) y que se adjunta en formato CSV y Excel (Archivos en: [Data Extract from WBWDI.zip](#)).

Consignas del trabajo:

1. A partir de la base de datos disponible elige un subconjunto de 5 a 20 indicadores y un año de referencia para efectuar tu análisis. Esta será la matriz de datos que utilizarás durante todo el examen. Cualquier subconjunto de información es válido, no hay elecciones incorrectas; sin embargo, **verifica que la cantidad de observaciones disponibles supere a la cantidad de variables elegidas.**
2. Realiza un breve análisis descriptivo de la información disponible para el conjunto de variables y año elegidos. Calcula y describe distintas medidas de variabilidad global de los datos.
3. Calcula la matriz de variancias y covariancias y la matriz de correlaciones muestrales e interpreta brevemente sus resultados.
4. Realiza un análisis de componentes principales utilizando tanto la matriz de variancias y covariancias como la matriz de correlaciones. Describe de manera general los resultados obtenidos en cada caso y compara los resultados destacando diferencias y/o similitudes entre ambos casos. En particular, describe la proporción de variancia explicada por cada componente principal. Elige los resultados de uno de estos análisis para continuar con la próxima consigna.
5. Tomando en cuenta las primeras dos componentes principales del análisis elegido en el punto anterior calcula los coeficientes de correlación de cada componente principal con respecto a cada variable original. Interpreta la representación de las primeras 2 componentes principales en términos de su correlación con las variables que las definen.
6. Calcula el valor de las primeras 2 componentes principales para cada una de las observaciones (países). Realiza un breve análisis de estadística descriptiva de cada una de ellas.
7. Realiza un gráfico *biplot* para la representación conjunta de filas y columnas (equivalente a observaciones y variables) de la matriz de datos y describe los resultados obtenidos en términos de la distribución de las observaciones (países) en el subespacio de las primeras dos componentes.

8. Utilizando la representación *biplot* adecuada calcula una aproximación de dimensión 2 para la matriz de variancias y covariancias (o matriz de correlaciones, según corresponda) original e interpreta los resultados obtenidos.

Tomando en cuenta las primeras 2 (dos) componentes principales que representaban el mayor porcentaje de variancia explicada y que calculaste en los ítems anteriores, resuelve los siguientes puntos:

9. Realiza un análisis de clusters para hallar la jerarquía de agrupación de los países considerados en tu matriz de datos, en base a la información de las dos primeras componentes principales. Utiliza dos métodos distintos (puedes elegir entre los encadenamientos simple, completo, promedio, etc.).
10. En cada caso realiza una representación gráfica de los resultados obtenidos utilizando un dendrograma.
11. Interpreta los resultados obtenidos tomando en cuenta el gráfico *biplot* que realizaste para resolver las últimas consignas del primer trabajo práctico.
12. Resume brevemente las principales conclusiones del análisis.

Por último:

13. Efectúa un análisis factorial para describir la variabilidad común entre las variables de tu matriz de datos. El punto de partida será un modelo con un solo factor.
14. Estima la matriz de variancias y covariancias que surge del modelo factorial, la cual se descompone en variancia común (debido a los factores) y variancia específica. Compara estos resultados con respecto a la matriz *S* de variancias y covariancias muestral.
15. Repite los dos puntos anteriores agregando un nuevo factor al modelo. Se producen cambios en las comunales de las variables con respecto al modelo estimado en el punto 13? Determina cuál de las dos especificaciones del modelo factorial ($m = 1$ ó $m = 2$) resulta más adecuada para representar la estructura de asociación entre las variables. Es posible que en función de las variables y/o países elegidos en tu matriz de datos la especificación del modelo factorial con dos factores no sea la óptima. En ese caso agrega los comentarios que creas conveniente.
16. Interpreta brevemente los resultados obtenidos, tomando en cuenta las eventuales diferencias y semejanzas con los resultados del cálculo de componentes principales.

Comentarios generales:

En función de los indicadores y año elegidos es posible que en la matriz de datos encuentres valores faltantes (*missing values*) para uno o más países. Si esto ocurre puedes descartar estos países para trabajar con una matriz de datos completa. Nuevamente, verifica contar con una cantidad suficiente de observaciones para el análisis ($n > p$).

Tanto en el archivo Excel como en la página web mencionada encontrarán una breve definición de los indicadores disponibles.

El trabajo es estrictamente individual. La entrega se realizará por mail a las direcciones alejandra.clemente@utdt.edu y alejandra.clemente@hotmail.com hasta las 23:59hs de la fecha límite establecida. Para las entregas fuera de la fecha límite establecida se aplicará un descuento a la calificación total del trabajo del 10% por cada día de retraso.

Al redactar el mail de entrega se recomienda utilizar el asunto "FINAL AEM 2023". En el mismo se deberá adjuntar un único archivo comprimido (formatos .zip o .rar) que contenga un documento de reporte que describa las soluciones a las consignas del trabajo y los archivos auxiliares que respalden los cálculos efectuados (archivos de programas .do como así también los archivos .log de resultados que correspondan). Está permitido utilizar otras herramientas estadísticas alternativas a Stata; en todos los casos deberán adjuntar los códigos y/o archivos de trabajo.