

Tarea 01

Análisis de datos multivariados

- Adriana Haydeé Contreras Peruyero
- Alejandro Jiménez Palestino
- Andrés Peña Montalvo
- Arturo Sánchez González
- Jesus Alberto Urrutia Camacho

Parte 1. Exploración gráfica y datos faltantes

La base de datos `Mexico_bienestar.csv` contiene 33 indicadores de bienestar de todas las entidades federativas y los valores registrados a nivel nacional en el censo del 2015. En el archivo `Variables_bienestar_México.docx` se encuentra explicado cómo se obtuvo cada uno de estos indicadores. (Fuente: [<https://www.inegi.org.mx/app/bienestar/>]) Muestra la respuesta a continuación de cada pregunta y explica los valores o gráficos que muestres. No muestres toda la salida de R, sólo lo necesario para responder a cada pregunta. Puedes utilizar tablas, notas al pie de la figura, etc. Además de tu archivo PDF, sube el código completo (en `.txt` o `.R`) y el archivo con datos que utilizaste (en `.csv` o `.txt`).

1. Indica las escalas (nominal, ordinal, absoluta, intervalo, radio) en las que se han medido las siguientes variables:
 - Satisfacción con la vida
 - Contaminación del aire
 - Habitaciones por persona
 - Esperanza de vida al nacer
 - Habilidades de los estudiantes
 - Salud autorreportada
 - Tasa de obesidad
 - Satisfacción con tiempo para ocio
2. Explora la base de datos e indica el número de valores faltantes, así como las variables y las entidades federativas a las que corresponden dichos valores, y responde: ¿Qué tipo de mecanismo de pérdida crees que le corresponde a los datos faltantes de esta base de datos?
3. Deberás de darle algún tratamiento a esos datos faltantes para poder continuar (imputar datos, eliminar casos, ...). Realiza ese tratamiento y justifícalo. *Si imputaste datos, cómo lo hiciste y qué valores fueron? Si eliminaste casos, cómo lo hiciste y qué casos fueron?
4. Para las siguientes variables:
 - `Percepción.de.ausencia.de.corrupción.en.el.sistema.judicial`,
 - `Tasa.de.mortalidad.materna`, y
 - `Viviendas.con.acceso.a.servicios.básicos`, muestra graficada la función empírica de distribución (ecdf), boxplot, Normal qq-plot e histograma. *Describe con detalle y explica* tus gráficas. ¿Qué relación encuentras entre los ecdf's y los boxplots que graficaste? ¿Cuál de estas variables tiene sesgo negativo? ¿Cuál de estas variables tiene sesgo positivo? ¿Alguna de estas variables tiene distribución aproximadamente Normal?
5. Con la base sin datos faltantes, calcula la matriz de correlaciones (R). Muéstrala.
6. Con la base sin datos faltantes, calcula el vector de medias. Muéstralo.
7. Con la base sin datos faltantes, calcula la matriz de varianzas y covarianzas (S). Muéstrala.
8. Indica entre que pares de variables hay mayor correlación y entre que pares de variables hay menos correlación.

9. Indica las 3 variables con mayor variabilidad y las 3 variables con menor variabilidad.
10. Genera al menos 2 gráficos multivariados para comparar las entidades federativas. ¿Cuáles se parecen menos? Si tuvieras que formar grupos de universidades ¿Cuáles formarías y por qué?

Parte 2. Datos faltantes

Consulta libros, artículos y bibliografía seria, y completa la siguiente tabla. Anota las referencias al final.

Tratamiento de datos faltantes	¿En qué consiste?	¿En qué casos se recomienda usarlo? (datos perdidos tipo MCAR, MAR, MNAR)
Omisión Total		
Omisión Parcial		
Imputación con la Media		
Imputación Hot Deck		
Imputación usando regresión		
Imputación usando Algoritmo EM		
Imputación múltiple		

Parte 3. Algebra de matrices (1.5 puntos)

Considera la siguiente matriz. Para cada uno de los incisos, muestra las operaciones que realizaste para justificar tu respuesta.

$$A = \begin{pmatrix} 6 & 4 & 0 \\ 4 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

- (a) Obtener los valores propios (calcule el polinomio característico) y los vectores asociados a cada uno de dichos valores propios.
- (b) ¿Es A una matriz idempotente?
- (c) ¿Es A una matriz no singular?
- (d) ¿Cuánto vale la traza de la matriz A ?
- (e) ¿Cuanto vale el rango de la matriz A ?
- (f) Determina si esta matriz es definida positiva, semidefinida positiva o si no lo es.

Referencias

- [1] Friedberg, Stephen H., Insel, Arnold J., and Spence, Lawrence E., *Linear algebra*, 4th. ed., Prentice Hall, 2002.