CORPORACION UNIVERSITARIA MINUTO DE DIOS - UNIMINUTO

ESTADISTICA BASICA APLICADA CON PYTHON

Alexander Oviedo Fadul

Fundamentos para Inteligencia Artificial (NRC-197)

Semana 6

Docente: Breyner Alexander Parra Rojas

TABLA DE CONTENIDOS

1. Introduccion y Contexto
2. Descripcion del Dataset
3. Medidas de Posicion Central
4. Relacion entre Medias
5. Medidas de Posicion No Central (Cuantiles)
6. Medidas de Dispersion
7. Asimetria y Curtosis
8. Distribucion de Frecuencias
9. Resumen Estadistico Completo
10. Visualizaciones
11. Conclusiones
12. Referencias

Descripción del Trabajo

El presente script cumple con los requerimientos de la actividad de la Semana 6 del curso Fundamentos para Inteligencia Artificial (NRC-197). El script, desarrollado en Python mediante un cuaderno de Jupyter (Google Colab), implementa un análisis estadístico completo sobre el Dataset de Diabetes de la Universidad Estatal de Carolina del Norte (NCSU), disponible en https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html, con 442 pacientes y 10 variables predictoras estandarizadas mas una variable objetivo cuantitativa.

El análisis cubre todos los conceptos estadísticos solicitados: medidas de posición central, relación entre medias, medidas de posición no central, medidas de dispersión, asimetría y curtosis, distribución de frecuencias, y seis gráficos embebidos. Las referencias bibliográficas corresponden a los recursos asignados para la Semana 6.

El script fue ejecutado y verificado sin errores de forma automatizada mediante nbconvert. El cuaderno de Jupyter esta disponible en mi repositorio académico:

URL Google Colab:
https://colab.research.google.com/drive/1t0cVBHLYSnnHGhOO_wBuX_TQNI3mkDA9

Informacion del Dataset

Nombre: Diabetes Dataset (NCSU / scikit-learn)
Observaciones: 442 pacientes
Variables predictoras: 10 (edad, sexo, IMC, presion arterial, y 6 indicadores bioquimicos estandarizados)
Variable objetivo: Progresion de la enfermedad al anyo siguiente (rango 25-346)
Acceso: https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html
Referencia original: Efron, B., Hastie, T., Johnstone, I., y Tibshirani, R. (2004). Least angle regression. Annals of Statistics, 32(2), 407-499.

Referencias Bibliograficas (Normas APA 7.a Edicion)

Boschetti, A., y Massaron, L. (2018). *Python data science essentials: A practitioner's guide covering essential data science principles, tools, and techniques* (pp. 282-316). Packt Publishing.

Cady, F. (2020). *Data science: The executive summary — a technical book for non-technical professionals* (pp. 55-95). John Wiley & Sons.

Fuentes, A. (2018). *Become a Python data analyst: Perform exploratory data analysis and gain insight into scientific computing using Python* (pp. 121-137). Packt Publishing.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., y Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830. https://jmlr.org/papers/v12/pedregosa11a.html

Efron, B., Hastie, T., Johnstone, I., y Tibshirani, R. (2004). Least angle regression. *Annals of Statistics, 32*(2), 407-499. https://doi.org/10.1214/009053604000000067