Predicting the relationship between Hemodialysis Frequency and patient's demographics details, comorbidities, CKD severity, lab parameters, and Other Parameters.

Group-2 team members: Alekhya Jilla, Likhitha Kantipudi, Sai Shakti Rao Lendale, FNU Sahrash Fatima

Introduction:
End-stage renal disease (ESRD) is a terminal medical condition where kidney function ceases and hemodialysis or a kidney transplant is required to extend a patient's life. Chronic kidney disease (CKD) is a chronic condition that progresses to its terminal stage as ESRD when GFR is less than 15mL/min. The progression of ESRD from CKD is affected by various factors, like diabetes, glomerular diseases, vascular diseases, urinary tract obstruction, and tubulointerstitial disease. (Hashmi, 2023).
Hemodialysis is a medical procedure performed to clear waste products and excessive fluid for patients when kidney function ceases. It contains a dialyzer that circulates the patient's blood outside the body and balances electrolytes in the body. (National Institute of Diabetes and Digestive and Kidney Diseases, 2019). Patients with ESRD, typically require 3 hemodialysis sessions per week and 3-5 hours per session to completely clear toxins and filter the blood. (Cleveland Clinic, n.d.)

Aim:
To develop a model that predicts the relationship between patient's hemodialysis frequency based on their demographics, comorbidities, CKD severity, lab parameters, and other parameters.

Null Hypothesis:
H0: There is no significant association between hemodialysis frequency with one or more of the following patient factors:

- Demographics (age, gender, weight)
- Comorbidities (diabetes, hypertension, other)
- CKD severity (disease severity, cause of kidney failure)
- Lab parameters (Creatinine, Urea, Potassium, Hemoglobin, Hematocrit, and Albumin)
- Other parameters.

Alternate Hypothesis:
H1: There is a significant association between hemodialysis frequency and one or more of the following patient factors:

- Demographics (age, gender, weight)
- Comorbidities (diabetes, hypertension, other)
- CKD severity (disease severity, cause of kidney failure)
- Lab parameters (Creatinine, Urea, Potassium, Hemoglobin, Hematocrit, and Albumin)
- Other parameters.

Purpose:
This project seeks to predict the relationship between hemodialysis frequency with patient data like demographics, comorbidities, severity, lab parameters, mode, and process of hemodialysis. The results of this project can be quite helpful for clinical practitioners for better clinical decision-making and optimizing treatment plans that ultimately help in enhancing patient care.

Methodology:

- Data Collection and Storage: Includes obtaining relevant datasets from Kaggle that encompass the factors needed to perform data visualization. The obtained dataset is stored as an encrypted SQL file to ensure data integrity and data consistency.
- Data extraction and Preprocessing: Python will be used as a tool to extract required data from SQL and to handle missing values, possibly encode categorical data, and standardize data formats.
- Exploratory Data Analysis (EDA): This is to identify relationships between columns, detect outliers, and perform summary statistics that will be helpful for subsequent steps.

- Statistical Testing: Conducting normality testing like the Shapiro-Wilk test to assess the normal distribution of data. Statistical tests like hypothesis testing and correlation tests to establish relationships between variables.
- Data Visualization: the findings will be visualized using Histograms, scatter plots, heatmaps, etc. to gain insight into the analyzed data.
- Machine learning: prediction of the relationship between dependent and independent variables will be performed with machine learning regression models like linear regression. Data classification is done to classify variables that might impact the dependent variable.
- Final report: the findings are analyzed, and the relevant information is presented in the final report.

Data Description:

Dependent Variable	Variable type
Hemodialysis Frequency (per week)	Numerical

Independent Variable	Variable type	Independent Variable	Variable type
Age	Numerical	Dialysate Composition	Categorical
Gender	Categorical	Vascular Access Type	Categorical
Weight	Numerical	Dialyzer Type	Categorical
Co-morbidities (Diabetes, Hypertension)	Categorical	Kt/V	Numerical
Kidney Failure Cause	Categorical	URR	Numerical
Blood Pressure (Pre-Dialysis, During-Dialysis, Post-Dialysis)	Numerical	Urine Output(ml/day)	Numerical
Heart Rate	Numerical	Dry Weight(kg)	Numerical
Lab Parameters (Creatinine,Urea, Potassium, Hemoglobin,Hematocrit, and Albumin)	Numerical	Fluid Removal rate(ml/hour)	Numerical
Dialysis Duration (hours)	Numerical	Disease Severity	Categorical

Deliverables:
- Descriptive statistics
- Data visualizations charts
- Regression models and Classification models
- Final report

Expected Results:
Assuming that the alternate hypothesis is validated and there is a correlation between the frequency of hemodialysis sessions and various patient factors, including demographics, comorbidities, CKD severity, and laboratory parameters. Identifying these correlations may inform clinical decision-making and optimize patient management in CKD and hemodialysis therapy.

Team Members' Responsibilities:

- **Alekhya Jilla:** Data Extraction and Preprocessing, Statistical Testing.
- **Likhitha Kantipudi:** Data Collection and Storage**,** Exploratory Data Analysis.
- **Sai Shakti Rao Lendale:** Machine Learning Model.
- **FNU Sahrash Fatima:** Data Visualization, Final Report.

Timeline:

- 02/10/2024- 02/19/2024- Data Collection and Storage
- 02/20/2024-02/25/2024- Data Extraction and Preprocessing
- 02/26/2024-03/15/2024- Exploratory Data Analysis
- 03/16/2024-03/30/2024- Statistical Testing
- 03/31/2024-04/15/2024- Data Visualization
- 04/16/2024-04/30/2024-Machine Learning Model
- 05/01/2024-05/05/2024-Final report

References:

Hemodialysis Realtime Hospital Dataset. (n.d.). Www.kaggle.com. Retrieved February 19, 2024, https://www.kaggle.com/datasets/gskrdsolutions/hemodialysis-realtime-hospital-dataset

Hashmi, M. F. (2023, August 28*). End-Stage renal disease*. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK499861/

National Institute of Diabetes and Digestive and Kidney Diseases. (2019, June 5). *Hemodialysis* | NIDDK. National Institute of Diabetes and Digestive and Kidney Diseases. https://www.niddk.nih.gov/health-information/kidney-disease/kidney-failure/hemodialysis

Cleveland Clinic. (n.d.). *Dialysis.* Cleveland Clinic. https://my.clevelandclinic.org/health/treatments/14618-dialysis