

STAT 5600 / 6600 – Probability and Statistics for Data Science

Dataset List and Descriptions

Dataset	Where Used	Data File	Description
aphasia	Categorical data analysis / contingency tables	aphasia.txt	Clinical aphasia study data used for probability models and categorical data analysis.
diabetes	Exploratory data analysis / regression	diabetes.csv	Patient-level diabetes measurements used for exploratory analysis and regression modeling.
gasdata	Time series / stochastic processes	gasdata.txt	Gas consumption time series data used to illustrate dependence and temporal modeling.
new_data	Simulation and probability modeling	New_data.csv	Synthetic dataset used for simulation-based probability experiments.
demographic_data	Sampling / multivariate analysis	P2-Demographic-Data.csv	Demographic survey data used for sampling, descriptive statistics, and multivariate analysis.
rainfall	Continuous distributions / modeling	rainfall.txt	Rainfall measurements used to study continuous distributions and model fitting.