

## BIOSTAT 625 Proposal: Cost Analysis and Forecasting for Hospital Financial Performance

### Objectives

The goal of this project is to analyze and forecast hospital operating costs and revenue trends to assist administrators in strategic planning and decision-making. By leveraging advanced statistical modeling and some basic machine learning techniques to forecast, we aim to provide actionable insights into financial performance and potential future trends.

### Proposed Approach

- Data Source: Use the [CMS Hospital Provider Cost Report dataset](#)
- Normalize variables such as "Operating Costs," "Revenue," and "Bad Debt Expense" to account for variability in hospital size. Moreover, we also need to clean out other covariate variables, and we need to merge different year's data together to form a panel data so that we can make use of all statistical methods we learned.
- Aggregate financial data by Multiple Years (2011 - 2022). Calculate derived metrics:
  - Cost-to-Revenue Ratio = Operating Costs / Total Revenue.
  - Revenue per Bed = Total Revenue / Number of Beds.

other covariate data includes but is not limited to whether the hospital is located in an urban area or rural area, and the hospital zip code and the total cost and so on.

### Exploratory Analysis

- **Descriptive Statistics:** We will create a table to represent the summary statistics of the data by year and the by region.
- **Visualization:** Yearly Time series plots to show historical trends in operating costs and revenue.
- **Dashboard Integration:** *Interactive Visualization:* Line charts for predicted vs. actual costs and revenue trends. Dropdown menus to filter forecasts by hospital type, region, or size. *Scenario Analysis:* Include sliders for users to adjust key variables (e.g., charity care costs, bed count). Display real-time impacts on forecasted metrics.

### Modeling and Forecasting

**Model Selection:** We start with the linear regression model to capture the general trend of the data, and then we further use machine learning methods to predict the key attributes of the hospital (e.g., total revenue or operating costs).

**Multivariate Regression:** For identifying key predictors of costs and revenue. Include variables such as bed count, charity care costs, and uncompensated care costs as predictors. Add lagged variables to capture temporal dependencies (e.g., revenue from previous periods).

$$Y_{it} = \beta_0 + \beta_1 \text{Bed\_Count}_{it} + \beta_2 \text{Charity\_Care\_Costs}_{it} + \beta_3 \text{Uncompensated\_Care\_Costs}_{it} + \beta_4 \text{Revenue}_{it-1} + \mu_i + \tau_t + \epsilon_{it}$$

$Y_{it}$ : Target variable (e.g., total revenue or operating costs).  $\beta_j$ : Coefficients for predictors.  $\mu_i$  is the hospital fixed effect and the  $\tau_t$  is the time fixed effect (year).  $\epsilon_{it}$ : Error term. Fitting: Use Ordinary Least Squares (OLS) to estimate coefficients. Include interaction terms and polynomial features (if needed) to capture non-linear relationships.

**Machine Learning Prediction:** Possible models: There are a lot of models that we can use, including but not limited to random forest, lasso / ridge regression. The training is based on the 30% of the random sampling of the data, and 30% for the testing and the other for the validation. We can also try to fit the model by year and predict by other years.

### Outcome

The proposed analysis and forecasting framework will:

- Help hospital administrators anticipate future financial performance.
- Provide actionable insights for resource allocation and budget planning.
- Identify trends and outliers in operating costs and revenue.