PREDICTING RISK OF HOSPITAL READMISSION WITHIN 30 DAYS USING PATIENT DEMOGRAPHIC AND CLINICAL VARIABLES WITH LOGISTIC REGRESSION IN MINITAB.

1. Background & Objectives

Hospital readmissions pose financial and operational challenges, especially under healthcare quality regulations. The goal of this project is to leverage Minitab to identify the most critical predictors of 30-day readmissions based on historical patient data. This helps both hospitals and researchers understand where intervention is needed to reduce readmission rates.

2. Methodology

Data Source:

Synthetic hospital patient data, including 3,000 discharges over one year, with fields like:

- Age, gender
- Comorbidities (hypertension, diabetes)
- Length of stay
- Type of discharge
- Readmission within 30 days (binary target variable)

Software Used:

Minitab 21

Steps Followed in Minitab:

1. **Data Import & Preparation:**
    - Imported .csv into Minitab.
    - Checked for missing values using *Data > Missing Value Pattern*.
    - Encoded categorical variables for analysis.
2. **Descriptive Statistics:**
    - Used *Stat > Basic Statistics > Display Descriptive Statistics*

- Summarized means, frequencies, and standard deviations for all variables.

3. **Chi-Square Test for Associations:**
   - *Stat > Tables > Chi-Square Test for Association* to examine initial associations between categorical variables (e.g., comorbidities vs. readmission).
4. **Binary Logistic Regression:**
   - *Stat > Regression > Binary Logistic Regression*
   - Modeled the probability of readmission as a function of patient attributes.
5. **Model Evaluation:**
   - Assessed p-values and odds ratios.
   - Used Hosmer-Lemeshow test for goodness of fit.
   - Plotted ROC curve to evaluate classification accuracy.

3. Results & Interpretation

- **Significant Predictors ($p < 0.05$):**
  - Length of stay
  - Discharge to nursing home
  - Presence of diabetes
- **Odds Ratios:**
  - Patients with diabetes were 2.4 times more likely to be readmitted.
  - Each additional hospital day increased readmission risk by 8%.
- **Model Accuracy:**
  - AUC (ROC Curve) = 0.79
  - Classification accuracy = 74%

4. Visual Outputs (Created in Minitab)

- Bar chart: Readmission rates by discharge type
- ROC Curve: Predictive performance of logistic model
- Probability plot: Predicted vs. observed probabilities

(These would be included in the final report as .jpg or .png exports.)

5. Recommendations

- Prioritize care coordination for diabetic patients before discharge.
- Monitor patients with longer hospital stays more closely post-discharge.
- Consider transitional care programs for patients discharged to nursing homes.

6. Future Work

- Include lab results and medication adherence data for improved accuracy.
- Apply machine learning algorithms in Python/R for model benchmarking.
- Validate model with data from multiple hospitals to ensure generalizability.

7. Stakeholder Relevance

Academic Use:

- Demonstrates logistic regression in applied healthcare research.
- Suitable for coursework in epidemiology, public health, or biostatistics.

Corporate Use (Hospitals/Consulting):

- Helps quality control teams reduce penalties for avoidable readmissions.
- Provides actionable insights for discharge planning and resource allocation.