LENGTH OF STAY PREDICTION USING MULTIPLE LINEAR REGRESSION IN R FOR A U.S. MEDICAL CENTER

1. Background

A general hospital in Ohio struggled with bed management due to unpredictable patient discharge timelines. Administrative delays and variability in care duration impacted ER throughput and ICU readiness.

To streamline operations, the hospital's data analytics team partnered with us to develop a statistical model in R to predict expected length of stay (LOS) for inpatients. The goal was to support early discharge planning and optimize ward-level staffing and equipment allocation.

2. Objective

- To model how patient-level characteristics predict LOS using multiple linear regression in R
- To identify clinical and demographic factors most associated with prolonged hospitalization
- To support early-stage discharge decision-making with statistically validated insights

3. Data Used

Source: Hospital EHR system, anonymized inpatient data (2022)

Dataset Details:

- 5,420 unique adult admissions
- Key fields:
    - LOS_Days (continuous)
    - Age, Gender, Admission_Type (Emergency, Scheduled)
    - Primary_Diagnosis_Code, Comorbidity_Score
    - Insurance_Type, ICU_Flag, Discharge_Disposition

4. Methodology

4.1 Data Preprocessing

- Imported via readr; cleaned using dplyr
- Created dummy variables for categorical predictors: Admission_Type, Insurance_Type, ICU_Flag
- Handled skewness of LOS with log transformation: Log_LOS = log1p(LOS_Days)

4.2 Model Building

- Built linear regression with lm() in R using Log_LOS as dependent variable
- Included interaction terms between ICU_Flag and Comorbidity_Score
- Checked diagnostics:
  - VIF (from car): all values < 3
  - Residual plots: no major heteroscedasticity
  - Shapiro-Wilk test for normality (p = 0.12)

5. Model Output and Diagnostics

Predictor Variable	**Coefficient (β)**	**p-value**	**Interpretation**
Age	+0.009	< 0.001	Older patients tend to stay longer
Admission_Type: Emergency	+0.224	< 0.001	Emergency admissions have longer LOS
ICU_Flag (Yes)	+0.341	< 0.001	ICU admission increases LOS
Comorbidity_Score	+0.118	< 0.001	Each additional condition adds to LOS
ICU × Comorbidity (Interaction)	+0.052	0.031	Combined effect significantly amplifies LOS
Insurance_Type: Uninsured	+0.097	0.044	Uninsured patients stay slightly longer

- **Adjusted R^2** = 0.49
- MAE (transformed back): 1.84 days
- Residuals: no visible pattern or bias in diagnostic plots

6. Interpretation and Recommendations

- Age, ICU stay, and comorbidity are the strongest LOS drivers
- Interaction term confirmed that **complex ICU patients require extended planning**
- Emergency admissions should trigger an **automatic early discharge assessment**
- Suggested flagging LOS > 5 days during admission for case management review
- Recommended integrating model into **admission dashboard** for nurse coordinators

7. Reporting Output

- **R Markdown Report (PDF, 19 pages)**:
  - Full regression summary
  - Diagnostic visualizations: residuals, Q-Q plots, VIF table
  - Recommendations for operational policy
- **Excel File**:
  - Patient-level predictions
  - LOS deviation flags
  - Regression coefficients (log-scale and original-scale interpretation)
- **Slide Deck (Optional)**:
  - Executive summary of predictors
  - Visuals for operational planning teams

8. Business Impact

- Within 2 months:
  - **Average LOS dropped by 0.7 days** through early planning
  - **ICU turnover improved by 12%** in Q2

- Model now used in **daily morning rounds** by clinical ops teams
- Incorporated into the hospital's **discharge forecast system**