

# **Associations of Healthcare Costs and Early Detection with Colorectal Cancer Mortality in the USA**

Yulin Yuan, Ruohan Sun, Tony Lee

## **Introduction**

### **Study Objective**

Clearly define the research questions or objectives.

### **Dataset Overview**

Describe:

- Where the data comes from
- What each variable represents
- How the data was collected (if known)

### **Motivation**

- Why is this question or data important/interesting?
- Need to explain why we remove country-level variable and Patient\_ID, reasonable is fine.

## Analysis

### Exploratory Data Analysis (EDA)

#### Balance of Response Variable

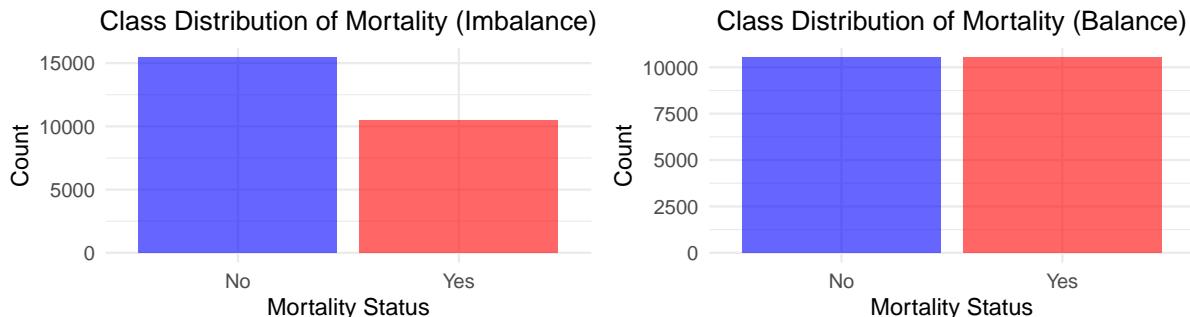


Figure 1: Class Distribution of the Mortality Variable Before and After Balancing

#### Continuous Variables

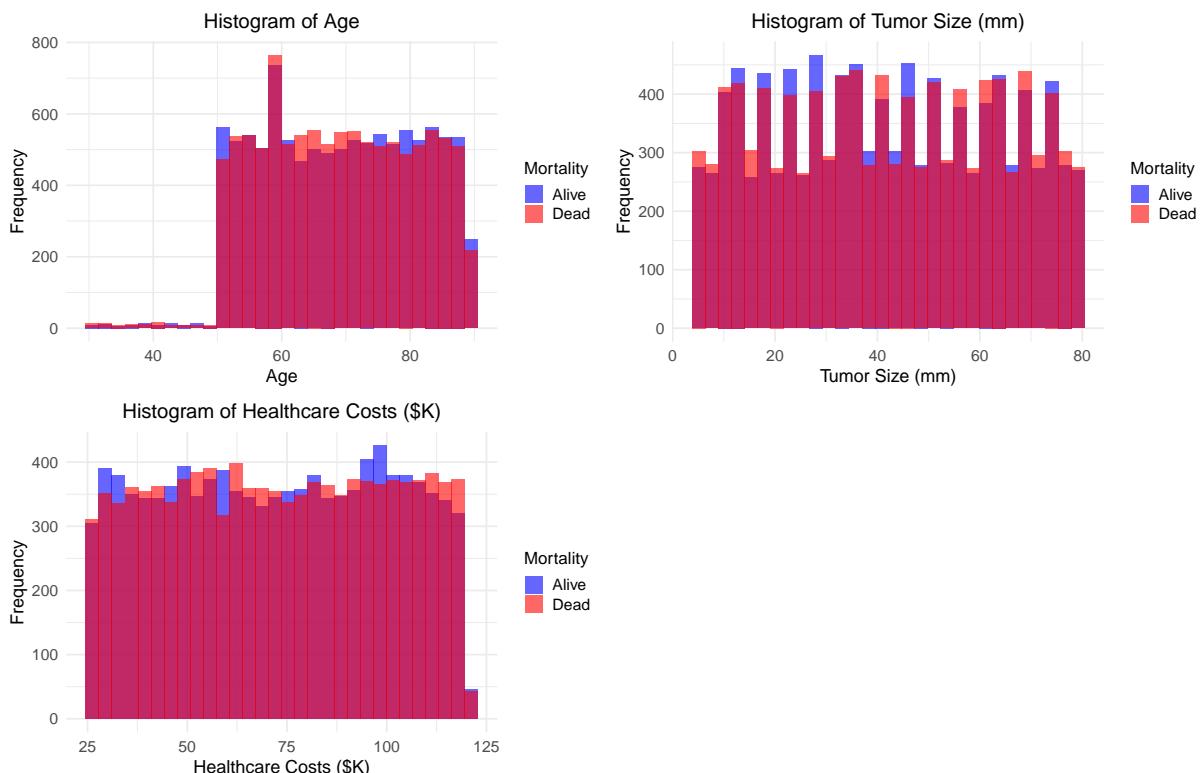
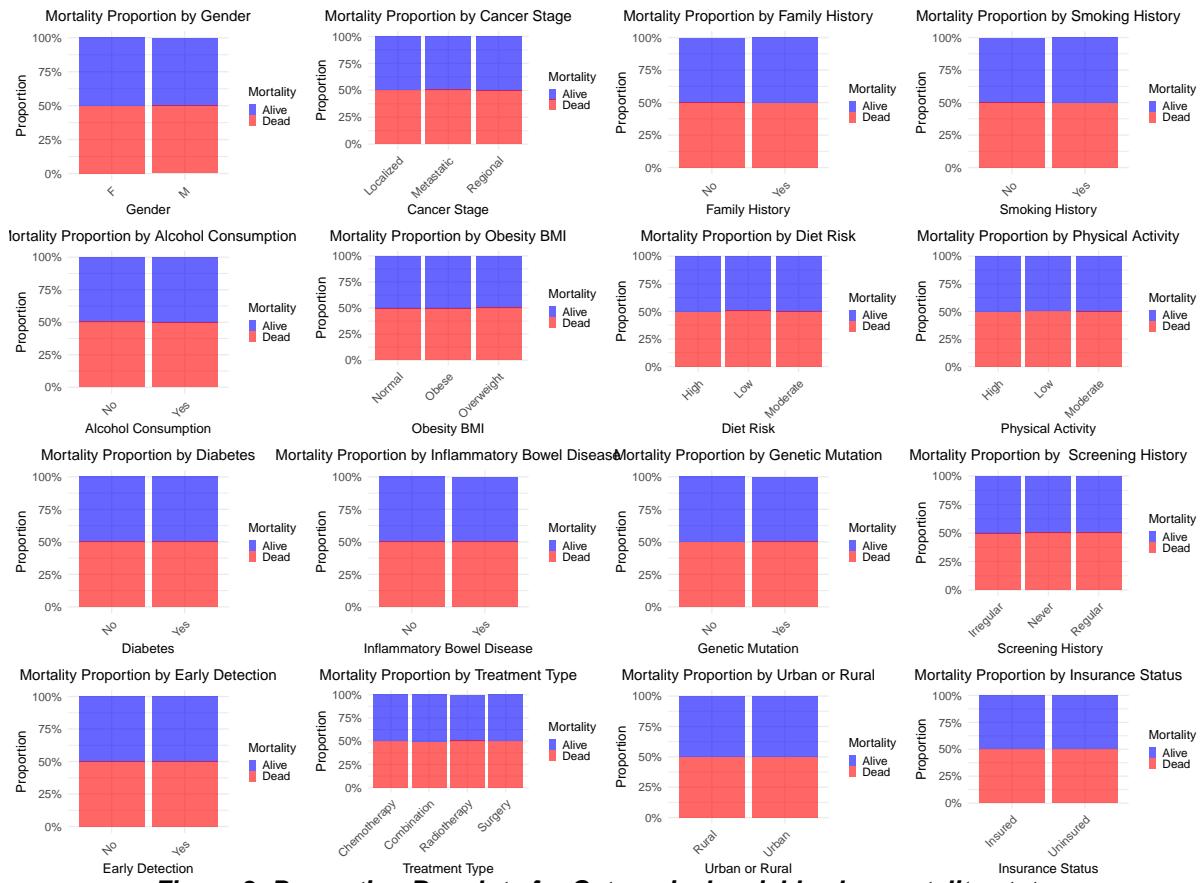


Figure 2: Distribution of continuous variables by mortality status

## Categorical Variables



**Figure 3: Proportion Bar plots for Categorical variables by mortality status**

## Interpretation of Findings

Pattern, trends, suggested operations

## Model Choice and Reasoning

### Logistic Regression

explain why choose this model based on EDA and Data description

## Assumption Check

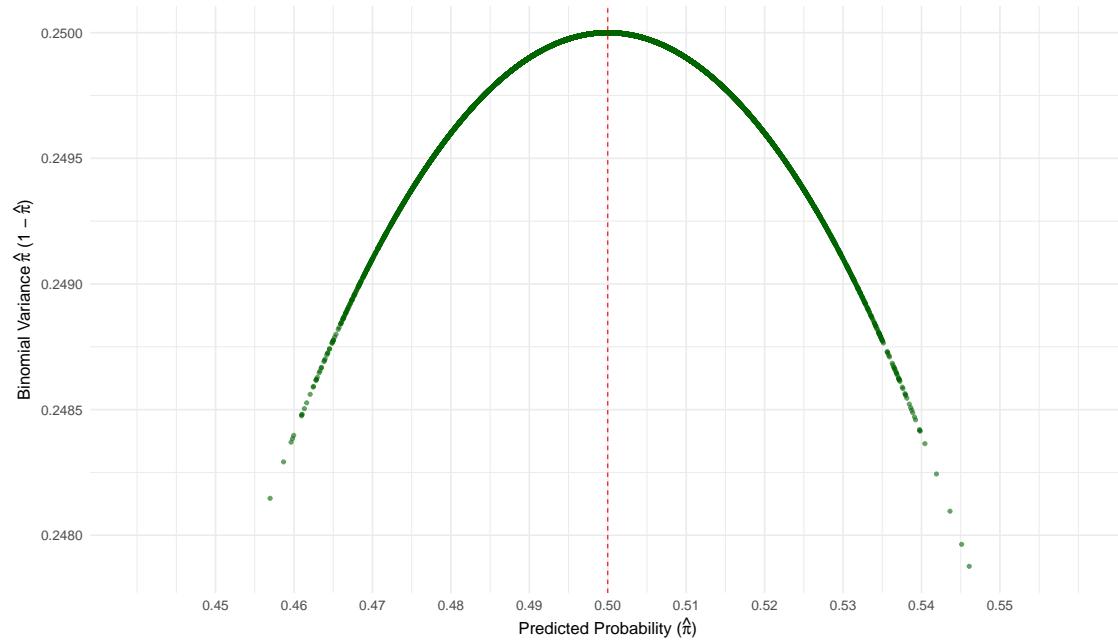
### 1. Binary Response

Based on *Figure 1*, the response variable is binary

## 2. Independence

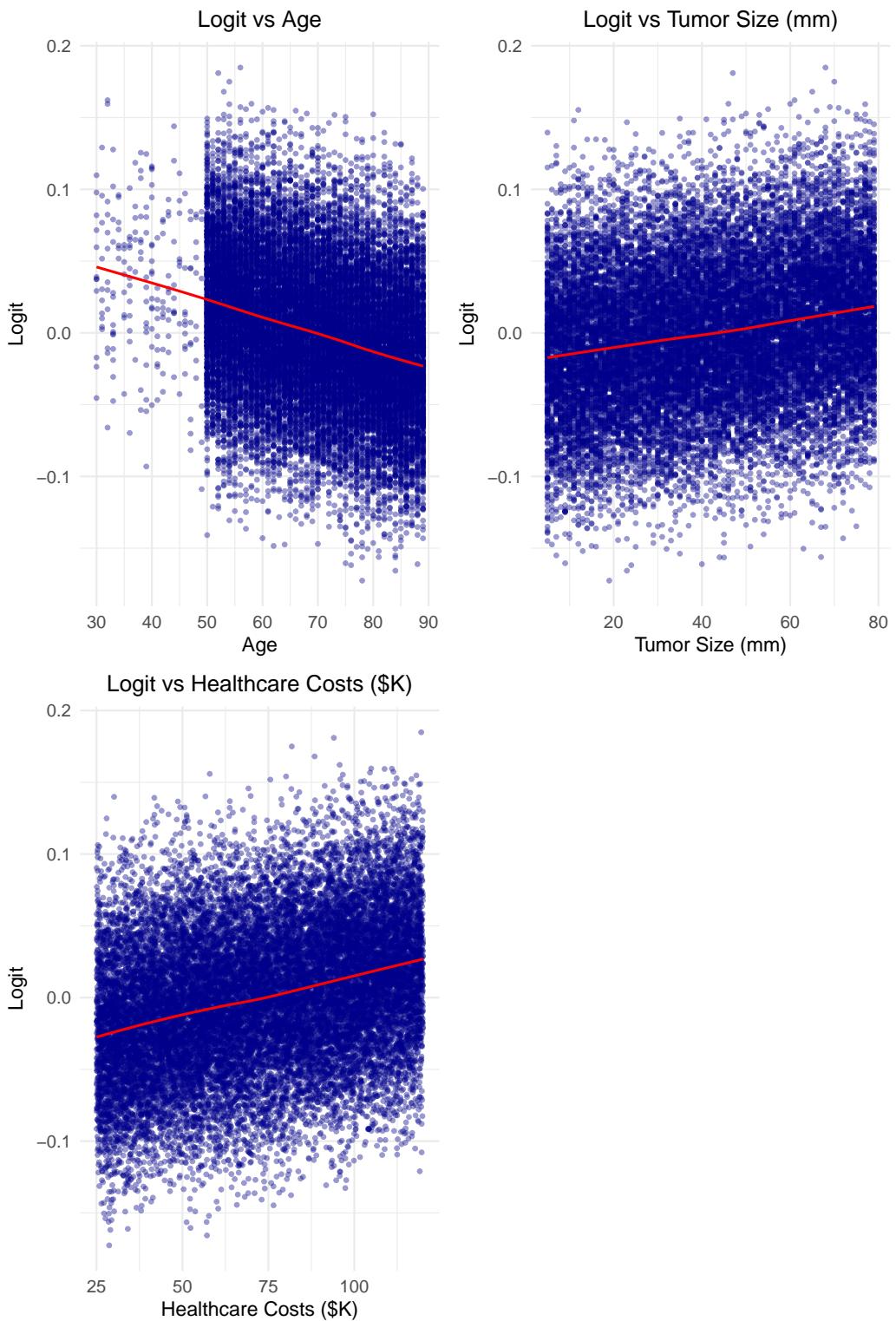
No duplicate rows, independence hold

## 3. Variance Structure



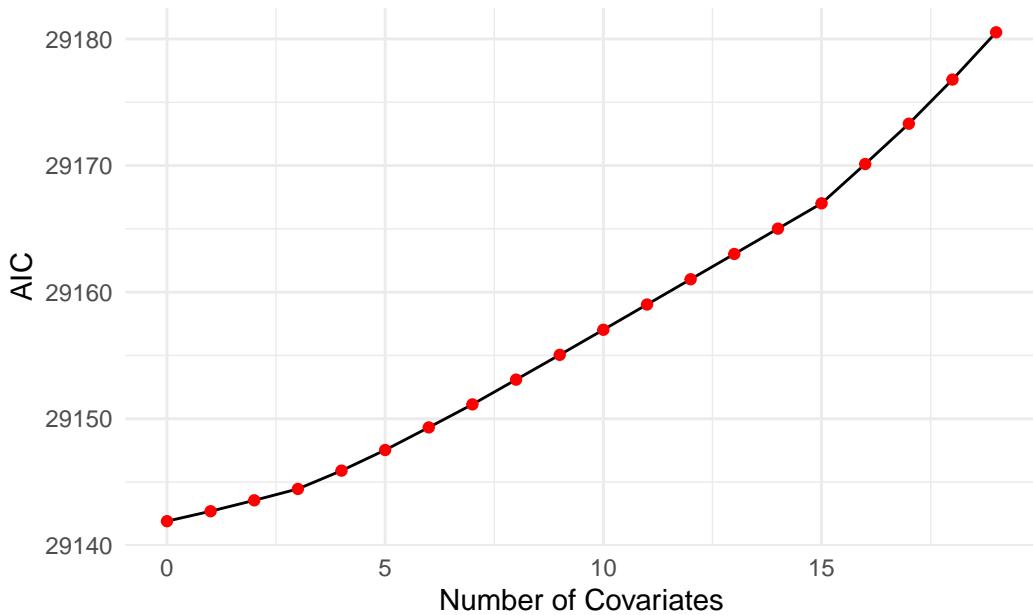
**Figure 4: Variance peaks at predicted probability = 0.5.**

## 4. Linearity



**Figure 5: Linearity check – logit of mortality plotted against continuous variables**

## Feature Selection



*Figure 6: AIC vs Number of Covariates – demonstrating backward feature selection.*

## Statistical Analysis

### Results Interpretation

inference results

## Conclusion

### Main Findings

Interpreting result in real-world context, careful about causality

### Limitations

- Discuss possible sources of bias, limitations in data, model assumptions
- Suggest improvements or next steps

## **Potential Further research**

Mention anything interesting you found that doesn't fit elsewhere

## **Appendix**

- Full regression output
- Extra plots or tables not essential to the main body
- Model selection steps