

Bayesian Regression: A Breast Cancer Survival Analysis

William Qian

1 Option Selection

I prefer **Option 1: Data Analysis**.

2 Introduction

Breast cancer is the most common cancer affecting women worldwide, representing a significant public health challenge. It arises from the cells of the breast, potentially leading to a malignant tumor that can invade surrounding tissues or spread to other parts of the body. Early detection and treatment of breast cancer are essential for improving survival rates and quality of life for breast cancer patients. Survival analysis can also help researchers understand the influencers of breast cancer and develop more effective treatments as well as preventive measures. That is why survival analysis on breast cancer patients is of great importance.

Traditional regression analysis, commonly applied in medical research for survival prediction, yields fixed estimates without adequately addressing the uncertainty surrounding these predictions. This approach may overlook the variability in patient responses to treatment and the inherent unpredictability of disease progression, potentially leading to suboptimal clinical decisions.

In order to address that problem, this analysis aims to apply Bayesian regression to model the survival outcomes of breast cancer patients. Through the use of Bayesian methods, it's possible to estimate the probability distributions of survival outcomes rather than just point estimates, offering a deeper understanding of the range of possible patient prognoses, ultimately aiming to enhance survival rates and patient care by acknowledging and quantifying the uncertainty in predictions.

3 Data Description

This dataset of breast cancer patients was obtained from the 2017 November update of the SEER Program of the NCI, which provides information on population-based cancer statistics. The dataset involved female patients with infiltrating duct and lobular carcinoma breast cancer (SEER primary cites recode NOS histology codes 8522/3) diagnosed in 2006-2010. Patients with unknown tumor size examined regional LNs, positive regional LNs, and patients whose survival months were less than one month were excluded; thus, 4024 patients were ultimately included.

The access to this dataset has been confirmed.