The Effects of Sex, Age Group, and Cardiac History on Hospital Length of Stay

Authors: David Teng, Hao Shen

Date: May 12, 2025

Introduction

Hospital length of stay is a key metric in both clinical care and healthcare administration. It reflects the duration of inpatient treatment and is commonly associated with illness severity, treatment complexity, and hospital resource utilization. Prolonged hospitalizations can increase healthcare costs and risk of complications, while shorter stays may indicate more efficient care or lower acuity. Identifying factors that influence the duration of hospitalization can support better care planning and resource allocation.

This study examines whether patient age and history of cardiovascular disease are associated with variation in hospital length of stay among critically ill adults. Specifically, it evaluates whether the effect of cardiac history on length of stay varies across age groups. The hypothesis is that patient age and cardiac history interact to influence hospitalization patterns, particularly in older populations.

Methods

This analysis uses the Right Heart Catheterization (RHC) dataset, publicly available through the Vanderbilt University Department of Biostatistics. The dataset includes 5,735 critically ill adult patients from the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT), a multicenter observational study conducted between 1989 and 1994 across five U.S. teaching hospitals.

The outcome variable for this analysis is hospital length of stay, defined as the number of days between hospital admission and discharge. The primary explanatory variables considered are age group, cardiac history, and sex. Age was categorized into four groups: younger than 50, 50 to 65, 65 to 80, and 80 or older. Cardiac history was recorded as a binary variable indicating whether the patient had a known history of cardiovascular disease. Sex was initially included as a covariate in the modeling process but was ultimately excluded in the final model based on its lack of statistical significance and contribution to model fit.

All observations with missing values were excluded. The distribution of the length of stay variable was strongly right-skewed, violating the assumption of normality required for analysis of variance. A Box-Cox transformation was conducted and supported the use of a log transformation. The log-transformed length of stay variable was used in all subsequent models. Levene's test confirmed that the assumption of equal variance across groups was satisfied.

A three-way Type II ANOVA model was fitted using linear regression to examine all main effects and interactions among age group, cardiac history, and sex. Type II sums of squares were selected due to the unbalanced design of the dataset, which included unequal group sizes across combinations of predictor variables. Following the initial results, a reduced model excluding sex and its associated interactions was evaluated. Both models were compared using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and nested model ANOVA to determine the most parsimonious and statistically appropriate model.

Results

The full three-way Type II ANOVA model identified the interaction between age group and cardiac history as the only statistically significant effect influencing log-transformed hospital length of stay, with a p-value of 0.0201. Neither age group, cardiac history, nor sex showed significant main effects, and none of the other two-way or three-way interaction terms reached significance. These findings suggest that the effect of age or cardiac history on hospital length of stay cannot be fully understood in isolation and that their joint influence provides a more meaningful explanation of variability in hospitalization duration.

Following the identification of this significant interaction, a reduced two-way Type II ANOVA model was fitted, including only age group, cardiac history, and their interaction as predictors. In this reduced model, the interaction between age group and cardiac history remained statistically significant, with a p-value of 0.0388. The main effects of age group and cardiac history were not significant, reaffirming that their combined effect contributes more substantially to explaining differences in hospital length of stay than either variable alone.

To determine whether the reduced model offered a more appropriate fit than the full model, formal model comparisons were conducted using multiple criteria. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were both lower for the reduced model. Specifically, the AIC decreased from 1252.78 in the full model to 1241.21 in the reduced model, and the BIC decreased from 1328.47 to 1281.28. A nested analysis of variance comparison between the two models yielded a *p*-value of 0.8258, indicating that the full model did not significantly improve model fit over the reduced model. These results support the selection of the reduced Type II ANOVA model as more parsimonious and equally effective in capturing the relevant effects. See Table 1 in the Appendix for a summary of the results.

Although the interaction term was statistically significant, pairwise comparisons using Tukey's Honestly Significant Difference test did not identify significant differences between specific age groups, either in the full dataset or within subgroups defined by cardiac history. However, interaction plots and boxplots revealed meaningful patterns. Among patients with cardiac history, hospital length of stay tended to decrease with increasing age. In contrast, for patients without cardiac history, hospital length of stay remained relatively stable or slightly increased across age groups. These visual patterns reinforce the conclusion that the effect of age on hospital length of stay varies depending on the presence of cardiac history, even though the differences between individual group means were not statistically significant. In addition, a plot of estimated marginal means with 95% confidence intervals was generated from the reduced Type II ANOVA model. This plot revealed that among patients with cardiac history, mean length of stay decreased with age, whereas no clear trend was observed among those without cardiac history. These results align with the statistically significant interaction found in the model and provide a clearer representation of the joint effect of age and cardiac history. See Figure 1, Figure 2 and Figure 3 in the Appendix.

Discussion

This analysis shows that hospital length of stay among critically ill patients is significantly influenced by the interaction between age group and cardiac history. Neither variable had a significant main effect when considered alone, but their combination revealed differences in hospitalization patterns. Specifically, older patients with cardiac history experienced shorter hospital stays, while those without cardiac history did not exhibit the same trend. These findings underscore the value of evaluating interaction effects in clinical data, particularly when modeling complex outcomes like inpatient duration.

The reduced Type II ANOVA model, excluding sex, was preferred for both statistical and interpretive reasons. It retained the significant interaction between age group and cardiac history and yielded better fit statistics compared to the full model. Eliminating non-significant terms improved clarity without compromising model validity. Graphical summaries further supported the interaction, illustrating distinct patterns in length of stay distributions based on both age and cardiac history. The use of model-adjusted means and confidence intervals provided a clearer visualization of these patterns and complemented the statistical findings from the reduced Type II ANOVA model.

This study has several limitations. Categorizing age may have reduced sensitivity to age-related trends compared to treating age as a continuous variable. Important clinical factors such as ICU type, comorbidities, and treatment interventions were not included, which may have influenced hospitalization duration. Additionally, removing missing data may have introduced mild bias, and the dataset reflects hospital practices from the early 1990s, which may limit the applicability of findings to current clinical settings.

Nonetheless, the findings demonstrate the usefulness of factorial analysis of variance in evaluating interactions in healthcare data. The results suggest that age and cardiac history should not be considered in isolation when analyzing predictors of hospitalization duration. Joint consideration of demographic and clinical variables can provide more accurate and meaningful insights for patient care and hospital resource planning. These findings may inform risk stratification and discharge planning efforts in critical care settings by identifying patient subgroups more likely to experience prolonged hospitalization.

References

Connors, A. F., Speroff, T., Dawson, N. V., Thomas, C., Harrell, F. E., Wagner, D., ... & Vidaillet, H. (1996). The effectiveness of right heart catheterization in the initial care of critically ill patients. *JAMA*, *276* (11), 889–897. https://doi.org/10.1001/jama.1996.03540110043030

Vanderbilt University Department of Biostatistics. RHC Dataset. https://hbiostat.org/data/repo/rhc.html

Appendix

Table 1. Model Comparison Between Full and Reduced Type II ANOVA Models

Model	Residual DF	Residual SS	AIC	BIC	Nested ANOVA p-value
Full Model	618	253.80	1252.78	1328.47	0.8258
Reduced Model	626	255.58	1241.21	1281.28	—

Figure 1. Interaction Plot from the Reduced Type II ANOVA Model: Age Group × Cardiac History on Log-Transformed Length of Stay

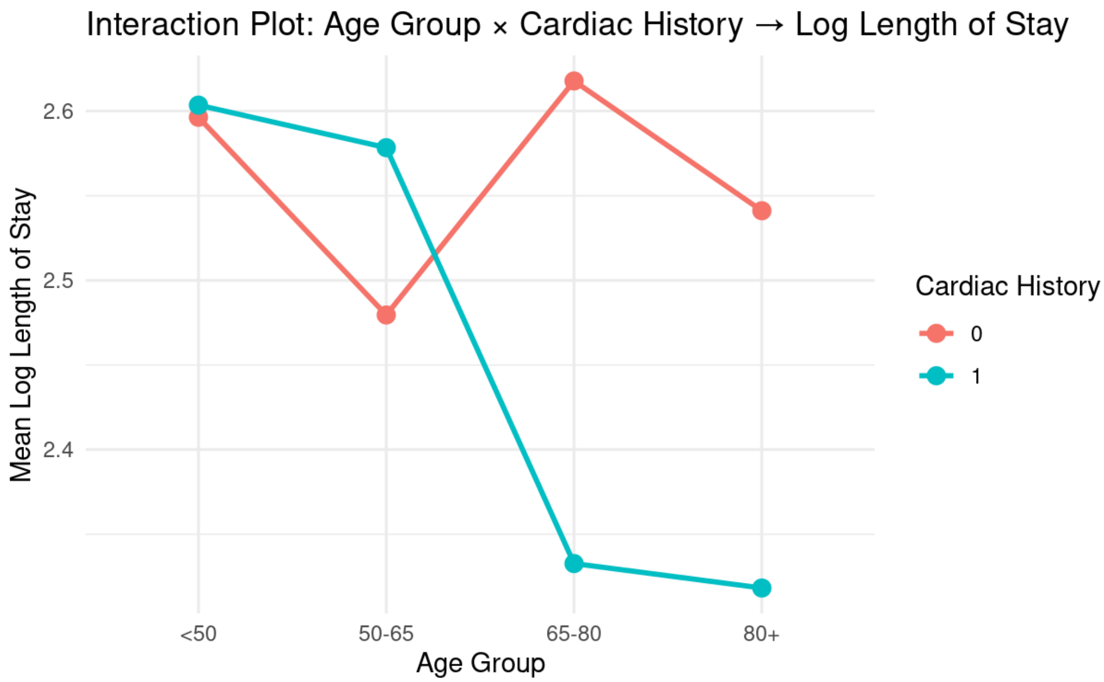


Figure 2. Boxplot from the Reduced Type II ANOVA Model: Age Group × Cardiac History on Log-Transformed Length of Stay

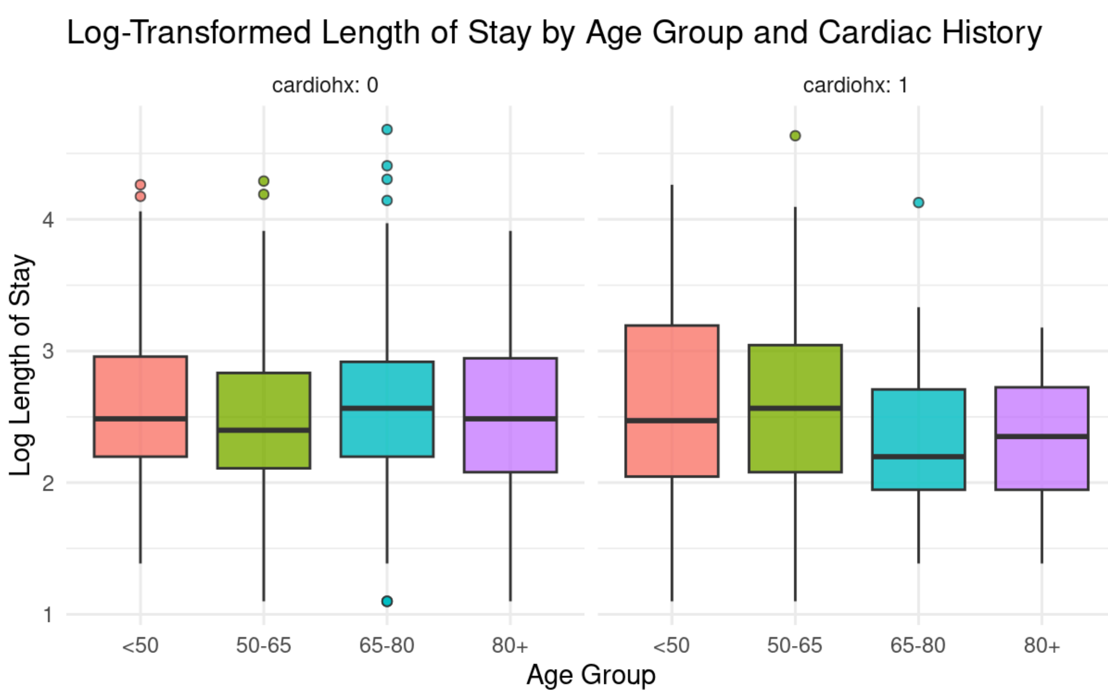
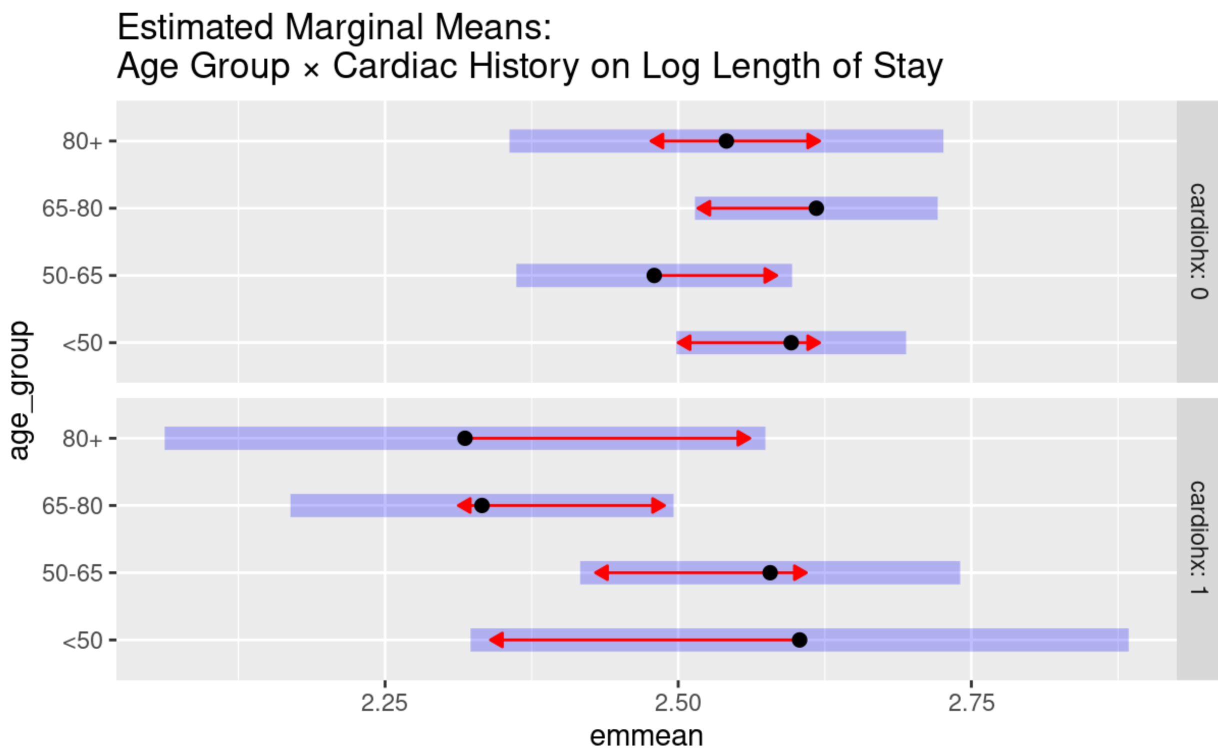


Figure 3. Estimated Marginal Means with 95% Confidence Intervals from the Reduced Type II ANOVA Model: Age Group × Cardiac History on Log-Transformed Length of Stay



All R scripts used in the data analysis and model building are publicly available at the following GitHub repository:
https://github.com/David-W-Teng/STAT-631-Analysis-of-Variance-Models

---

End of Paper