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

Authors: David Teng , Hao Shen
Date: 05/08/2025

Data Description :

This study utilizes the Right Heart Catheterization (RHC) dataset, comprising 5,735 critically ill adult patients from the SUPPORT study (1989–1994). The SUPPORT study was a multicenter observational study conducted across five U.S. teaching hospitals. The dataset is publicly available through the Vanderbilt University Department of Biostatistics.

Purpose:

To assess how **sex, age group,** and **cardiovascular history** influence **hospital stay duration**.

Modeling Approaches:

ANOVA is used to evaluate differences in hospital length of stay across groups.

```
aov(length_of_stay ~ sex * age * cardiohx, data = df_clean)
```

Response Variables:

length_of_stay — A continuous variable representing the number of days from hospital admission to discharge.

Independent Variables:

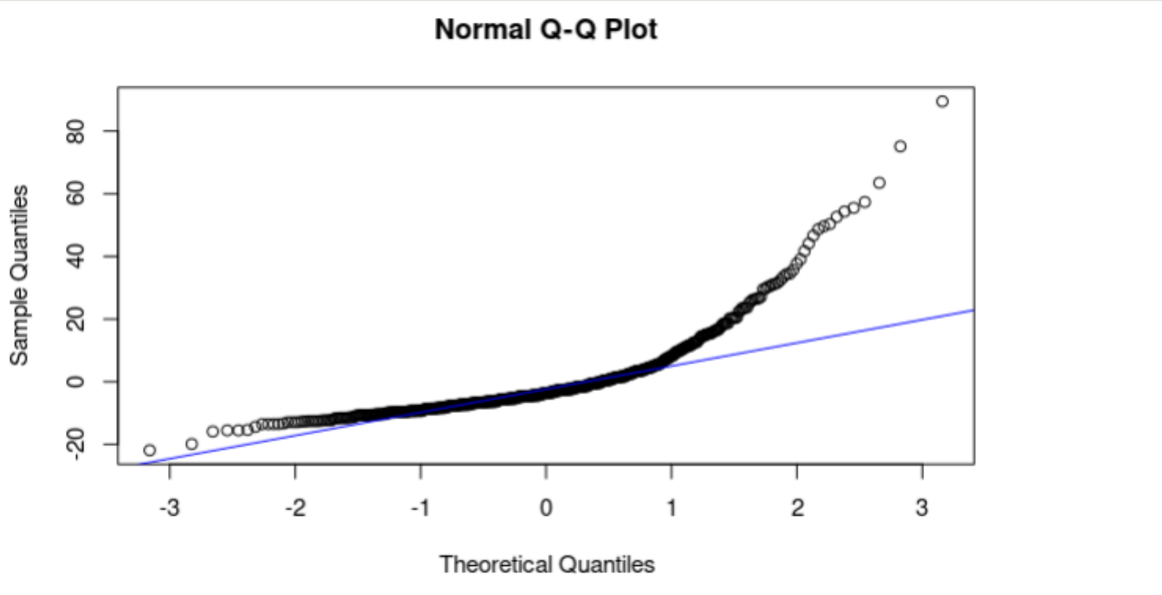
sex — Patient's biological sex (Male or Female).

age_group — Patient age categorized into four groups: <50, 50–65, 65–80, and 80+.

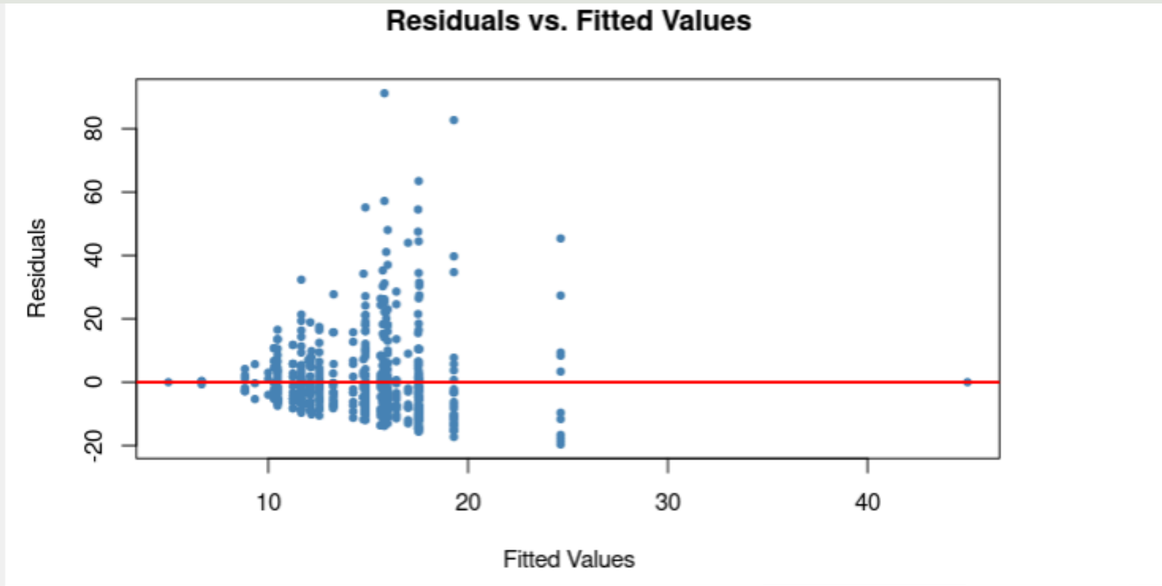
cardiohx — Indicates whether the patient has a history of cardiovascular disease (Yes or No).

ASSUMPTION CHECKING :

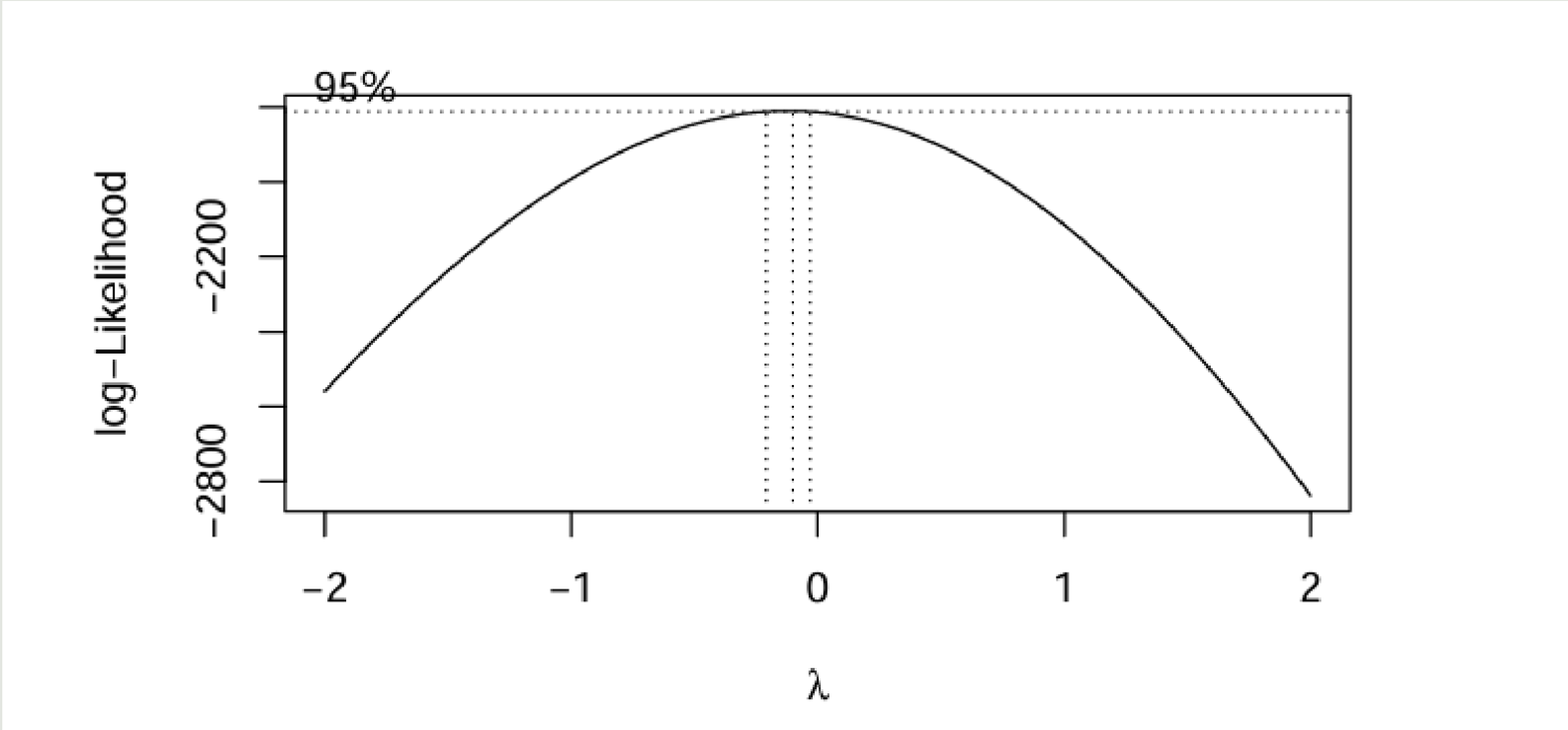
```
        Shapiro-Wilk normality test

data:  resid_anova
W = 0.76042, p-value < 2.2e-16
```




```
Levene's Test for Homogeneity of Variance
      Df F value Pr(>F)
group  30  1.0808 0.3533
      603
```

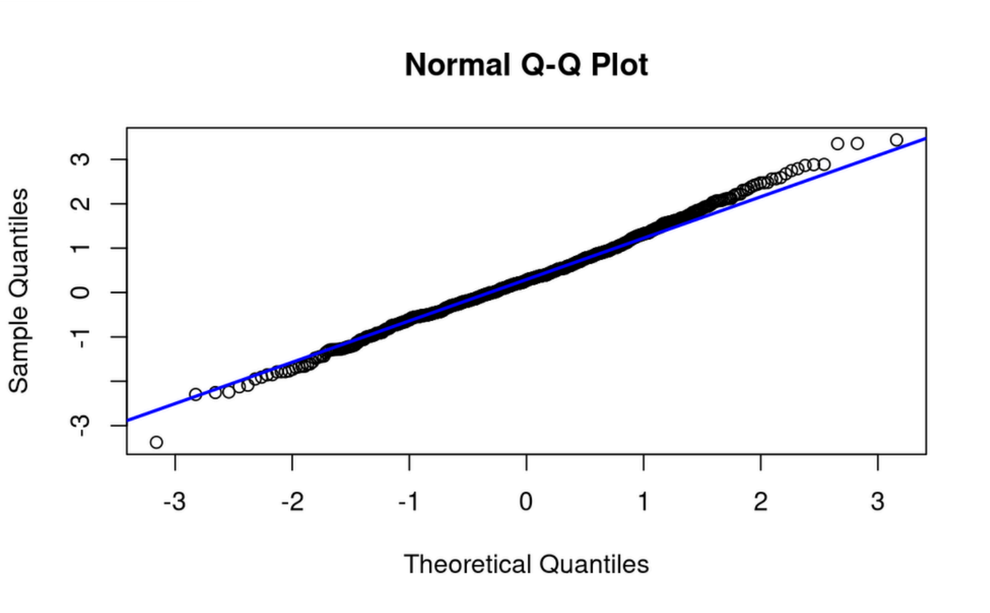


BOX-COX TRANSFORMATION

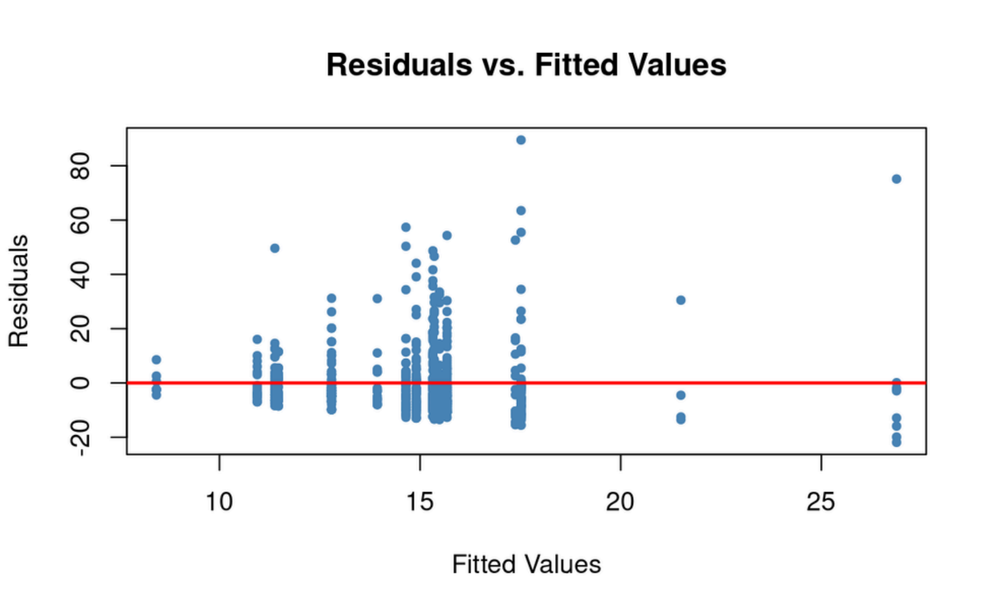


Since lambda ≈ 0, then log transformation is best.

LOG TRANSFORMATION



- Shapiro-Wilk normality test
- data: resid_log
- W = 0.99643, p-value = 0.1671



- Levene's Test for Homogeneity of Variance (center = median)
- Df F value Pr(>F)
- group 15 1.3023 0.1948
  618

THREE-WAY ANOVA

```r
df_clean$log_los <- log(df_clean$length_of_stay + 1)
anova_model_log <- aov(log_los ~ sex * age_group * cardiohx, data = df_clean)
summary(anova_model_log)
```

```
                          Df Sum Sq Mean Sq F value Pr(>F)
sex                        1   0.10  0.1044   0.254 0.6143
age_group                  3   1.21  0.4028   0.981 0.4013
cardiohx                   1   1.01  1.0135   2.468 0.1167
sex:age_group              3   0.16  0.0522   0.127 0.9440
sex:cardiohx               1   0.23  0.2288   0.557 0.4557
age_group:cardiohx         3   4.06  1.3547   3.299 0.0201 *
sex:age_group:cardiohx     3   0.72  0.2387   0.581 0.6275
Residuals                618 253.80  0.4107
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Key Finding:

Only the interaction between **age group and cardiac history** had a statistically significant effect on hospital length of stay (p = 0.0201).

No significant main effects or other interactions were observed.

Tukey HSD Post-Hoc Test

Comparison of Age Groups on Hospital Length of Stay

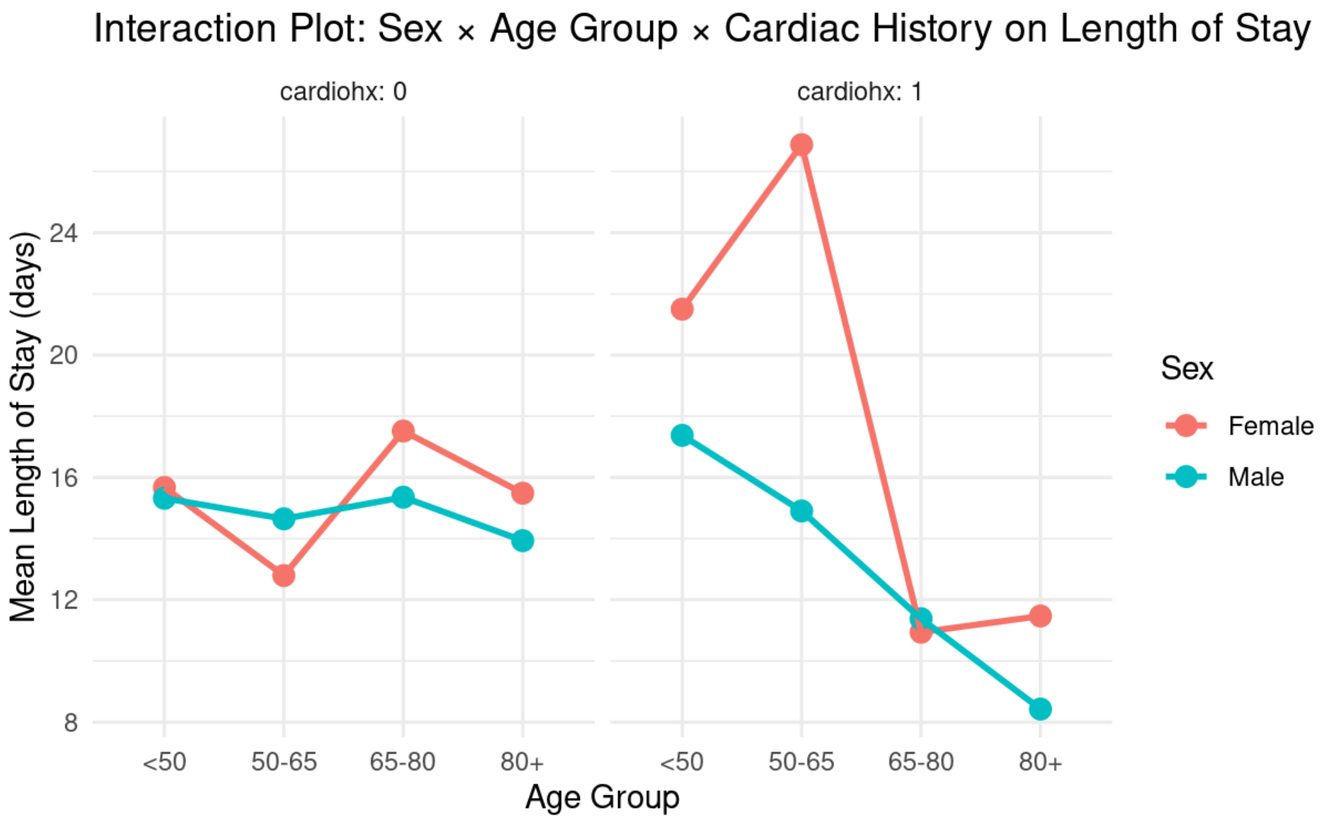
```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = log_los ~ age_group, data = df_clean)

$age_group
                    diff        lwr         upr     p adj
50-65-<50    -0.08350586 -0.2585489 0.09153719 0.6087007
65-80-<50    -0.06099560 -0.2289056 0.10691444 0.7856357
80+-<50      -0.13245221 -0.3649109 0.10000649 0.4577271
65-80-50-65   0.02251026 -0.1479293 0.19294983 0.9864422
80+-50-65    -0.04894635 -0.2832387 0.18534602 0.9496878
80+-65-80    -0.07145661 -0.3004689 0.15755568 0.8526800
```

Age Group Differences in Length of Stay Among Patients With Cardiac History

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = log_los ~ age_group, data = with_cardio_1)

$age_group
                    diff        lwr         upr     p adj
50-65-<50    -0.02513487 -0.4720526 0.42178286 0.9988855
65-80-<50    -0.27089314 -0.7187567 0.17697045 0.3984398
80+-<50      -0.28535352 -0.8094110 0.23870397 0.4926420
65-80-50-65 -0.24575827 -0.5631131 0.07159653 0.1883910
80+-50-65    -0.26021865 -0.6782719 0.15783461 0.3724913
80+-65-80    -0.01446038 -0.4335247 0.40460389 0.9997413
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = log_los ~ age_group, data = with_cardio_0)

$age_group
                    diff        lwr         upr     p adj
50-65-<50    -0.11681488 -0.31465163 0.08102187 0.4248586
65-80-<50     0.02146110 -0.16281054 0.20573275 0.9905841
80+-<50      -0.05523578 -0.32592298 0.21545142 0.9527487
65-80-50-65  0.13827598 -0.06419746 0.34074943 0.2937004
80+-50-65     0.06157910 -0.22181288 0.34497109 0.9437186
80+-65-80    -0.07669688 -0.35079117 0.19739741 0.8885284
```


Log Length of Stay by Age, Sex, and Cardiac History
0
1
Log Length of Stay
4
3
2
<50
50-65
65-80
80+
Age Group
sex
Female
Male


Interaction Plot: Sex × Age Group × Cardiac History on Length of Stay
cardiohx: 0
cardiohx: 1
Mean Length of Stay (days)
24
20
16
12
8
<50
50-65
65-80
80+
Age Group
Sex
Female
Male

Conclusion

- Only the age group × cardiac history interaction significantly affected hospital length of stay (p = 0.0201).
- No significant pairwise differences were found between age groups in post-hoc Tukey HSD tests.

Limitation

- Results are based on log-transformed length of stay to meet ANOVA assumptions, which may limit interpretability.
- Excluding observations with missing values may have introduced mild bias or reduced sample size, potentially affecting the robustness of the results.

Thank You

For your attention