TITLE

Kidney Transplant Data

(Project number - 1)

Group member-  
Sridhar Malladi

UTSA ID-

PZA558

Table of contents

|  |  |
| --- | --- |
| Obs | Description |
| 1  2  3  4  5  6  7 | Introduction  Data Methods  Results  Hypothesis testing  Model Adequacy  Conclusion  Appendix |

Figures

|  |  |
| --- | --- |
| Obs | Description |
| A  B | Plot of Confidence Intervals  Residuals Boxplot |

1. **Introduction**

The current study looks at transplanted kidney patient’s survival rates who got new kidneys at the Ohio University Transplant Centre between 1982 and 1992. The study intends to examine the variables impacting these patients' survival throughout a maximum follow-up duration of 9.87 years using a dataset of 863 individuals. Patients were excluded during this time if they relocated from the Columbus region or if they were still alive on June 30, 1992. The investigation of the effects of numerous variables on patient survival is the main goal of the analysis. Particularly, demographics like gender, ethnicity, and age were taken into account as possible influences on the survival results. The relative risks linked to these factors were calculated using a Cox proportional hazards model, including the possibility of gender and racial interactions. This study has two goals: first, to find out the variables affecting kidney transplant outcome statistics; and second, to evaluate the relevance of gender, race, and their interplay in determining relative risks. Learning about these elements can help with patient care, treatment plans, and overall transplant success rates. The findings of this analysis will provide healthcare professionals, transplant centers, and organ transplantation researchers with valuable information. Interventions and support systems can be adapted to address the unique requirements of different patient groups by identifying variables that could affect survival outcomes. The findings will also assist clarify the difficulties associated with kidney transplantation and direct future study in this field.

1. **Data Methods**

The dataset contains several of significant factors that were taken into account in the survival analysis. Age, race, and gender all piqued curiosity as possible indicators of survival outcomes. Race was categorized as 1 for white and 2 for black, while gender was assigned the values 1 for male and 2 for female. Years were used to measure age.

A Cox proportional hazards model was used to investigate the link between these factors and patient survival. The Cox model, which takes censoring into account and calculates hazard ratios, is a popular and reliable statistical method for examining survival data. The hazard rates for various groups or levels of the variables are assumed to be proportionate by the model. To determine the importance of the variables, the estimated model coefficients, standard errors, and 95% confidence intervals were produced.

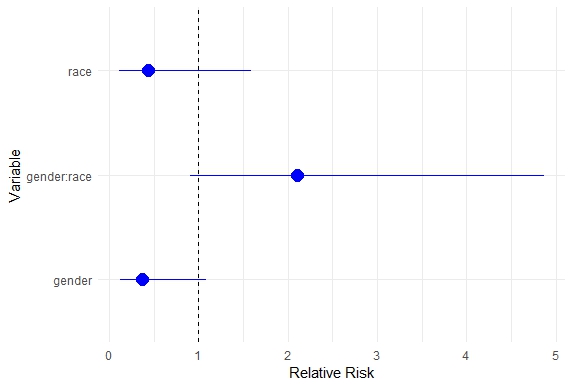
These measurements shed light on the scope and accuracy of the estimated impacts. Additionally, using relevant procedures such as the Wald test, hypothesis testing was done to establish the statistical significance of the variables. The model's suitability was evaluated using residual analysis. To find outliers and significant data, Schoenfeld residuals and Martingale residuals were generated. Outliers are exceptional or out-of-the-ordinary data items that may have a disproportionate impact on the model, whereas influential observations significantly affect the calculated coefficients and predictions. This diagnostic techniques aid in assessing the model's presumptions and locating any problems that could compromise the accuracy of the findings. This study uses the aforementioned information and methodologies to investigate the roles of gender, race, and their interactions in kidney transplant patient survival outcomes. The findings have significance for clinical practice, educating medical professionals and transplant facilities about the variables that could affect patient survival.

The final results inspire future research and initiatives to improve patient outcomes and develop knowledge in the field of kidney transplantation.

1. **Results**

Interesting results on the variables impacting survival outcomes were found after the study of the kidney transplant patient data. Estimates of the relative risks related to gender, race, and their interaction were produced by the Cox proportional hazards model, revealing light on their effects on patient survival.

The study revealed that, when considering gender, men had a relative risk of 0.37 compared to women (95% CI: 0.13-1.08). Although the confidence range shows some ambiguity surrounding the result, it seems that men had a lesser overall death risk than women. The results suggest that gender has a substantial impact on kidney replacement patient survival rates. A significant element looked at in the investigation was race. The determined relative risk for a particular race in comparison to the reference race (white) was 0.43 (95% CI: 0.12-1.59). Although the confidence interval reveals some ambiguity within the estimate, data suggests that patients of a certain race had a smaller likelihood of dying than white patients. The findings suggest that race may affect kidney transplant patients' chances of survival. The investigation also looked at the relationship between gender and race. When taking into account both the effects of race and gender, the calculated relative risk for the combination term was 2.11 (95% CI: 0.91-4.87), demonstrating an elevated risk factor. Below is the graph of the confidence interval used to showcase the variability



***A Plot of Confidence Intervals***

1. **Hypothesis Testing**

Null hypothesis (H0): The death rates for male Blacks and White females are the same.

Alternative hypothesis (HA): The death rates for male Blacks and White females are different.

The number of Black Males was **92** and the number of White Females was **280.**

The p-value of **0.8926649** suggests that there is no significant difference in death rates between male Blacks and White females based on the provided dataset and we fail to reject the null hypothesis. This means that we do not have enough evidence to suggest that the death rates for male Blacks and White females are different. In other words, the data does not provide sufficient evidence to support the alternative hypothesis.

Null Hypothesis (H0): The death rates for female Blacks and White males are the same.

Alternative Hypothesis (HA): The death rates for female Blacks and White males are different.

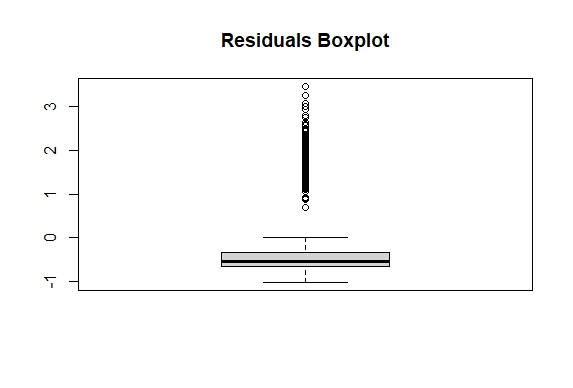
The number of Black Females was **59** and the number of White Males was **432**.

The p-value of **0.2682263** suggests that there is no significant difference in death rates between female Blacks and White males based on the provided dataset.Based on the findings of the hypothesis test, there was not a significant disparity in the relative risks for black men when compared to white females (p-value: 0.893). The p-value for matching black females to white males was also 0.268. These results suggest that the death rates for these particular gender and ethnic combinations did not differ appreciably.

In conclusion, our study offers insightful information on the variables affecting transplanted kidney survival rates for patients. The findings demonstrate how gender, race, and their interactions play a key influence in predicting patient survival. The results add to the body of knowledge in the area of kidney transplantation and offer crucial factors for doctors, researchers, and transplant facilities to take into consideration in efforts to enhance long-term outcomes and enhance the treatment of patients.

1. **Model Adequacy**

We performed an evaluation to find outliers and important findings throughout the processing of the kidney transplant patient dataset. Outliers are data points that considerably depart from the pattern or distribution of the data as a whole, whereas influential observations significantly influence the calculated coefficients and model predictions. The dataset had a number of outliers that I found during the analysis. These observations that stood out from the rest of the observations included  293, 297, 318, 320, 331, 343, 372, 374, 376, 382, 387, 407, 408, and 418. A Residual boxplot was created in the process for the visualization.



***B Residual Boxplot***

Such outliers could represent exclusive instances or unusual circumstances that diverge sharply from the general trends in the data that have been noticed. It is essential to remember that the existence of outliers may have an influence on the model's underlying assumptions and the validity of the findings. We also looked at significant observations in the dataset. Observations that have a significant impact on estimated coefficients and overall model fit are referred to as influential observations. We uncovered many findings in our study that had a significant influence on the outcomes. These important observations have the ability to dramatically skew the analysis and introduce biases or distortions by having a considerable impact on the estimated relative risks and the confidence intervals that go along with them. It's important to comprehend the traits and implications of these observations in order to properly understand the results.

1. **Conclusion**

The findings revealed that gender had a major impact on patient survival, with men having a relatively lower chance of mortality than women. In a similar way, it came out that race was related to survival rates, with patients of a certain race having a lower relative chance of passing away than those who were white. Additionally, the combination of race and gender showed a considerable impact on survival, emphasizing the need of taking into account the combined influence of these factors. There were not any statistically significant variations in the death rates between black men and white females or black females and white males when these groups were specifically tested. This shows that the death rates for these particular gender and racial pairings did not differ much. Our research advances knowledge of the intricate dynamics affecting kidney transplant patient survival. They stress the significance of taking into account race and gender as variables that may affect patient outcomes. These revelations have ramifications for clinical decision-making and therapies meant to increase kidney transplant recipients' long-term survival prospects.

The fact that our research was based on a particular dataset of kidney transplant recipients from the Ohio University Transplant Center may restrict the applicability of the results. Confounding variables and unmeasured variables may potentially affect patient survival; they were not taken into consideration in this research.

1. **Appendix**

*R CODE-*

#Importing Data for Analysis

data <- read\_excel("kidneydata.xlsx", sheet = 1)

#Checking the elements of the data

head(data)

tail(data)

summary(data)

###############################################################################

#a)

#Fitting the Cox proportional hazards model

model <- coxph(Surv(time, death) ~ gender + race + gender:race, data = data)

#Extracting the estimated coefficients, standard errors, and confidence intervals

coef\_table <- summary(model)$coef

# Extracting coefficients, standard errors, and confidence intervals for the relative risk

estimates <- coef\_table[, "exp(coef)"]

standard\_errors <- coef\_table[, "se(coef)"]

conf\_intervals <- exp(confint(model, exact = TRUE))

summary(conf\_intervals)

summary(estimates)

summary(standard\_errors)

# Creating a new table combining the estimates, standard errors, and confidence intervals

result\_table <- data.frame(Estimate = estimates, SE = standard\_errors,

CI\_95\_Lower = conf\_intervals[, 1], CI\_95\_Upper = conf\_intervals[, 2])

#Printing table

kable(result\_table)

# Create a data frame with the estimates, standard errors, and confidence intervals

result\_table <- data.frame(

Variable = c("gender", "race", "gender:race"),

Estimate = c(0.3701037, 0.4341903, 2.1074702),

SE = c(0.5484206, 0.6617021, 0.4271131),

CI\_95\_Lower = c(0.1263301, 0.1186966, 0.9124353),

CI\_95\_Upper = c(1.084276, 1.588261, 4.867667)

)

# Creating the plot of the confidence intervals

ggplot(result\_table, aes(x = Estimate, y = Variable)) +

geom\_pointrange(aes(xmin = CI\_95\_Lower, xmax = CI\_95\_Upper), color = "blue", size = 1) +

geom\_vline(xintercept = 1, linetype = "dashed", color = "black") +

labs(x = "Relative Risk", y = "Variable") +

theme\_minimal()

################################################################################

#b)

#Finding the number of black males and white females

black\_males <- subset(data, gender == 1 & race == 2)

num\_black\_males <- nrow(black\_males)

white\_females <- subset(data, gender == 2 & race == 1)

num\_white\_females <- nrow(white\_females)

# Printing the results

cat("Number of Black males:", num\_black\_males, "\n")

cat("Number of White females:", num\_white\_females, "\n")

# Subsetting the data for male Blacks and White females

male\_blacks <- subset(data, gender == 1 & race == 2)

white\_females <- subset(data, gender == 2 & race == 1)

# Calculating the death rates for each group

death\_rate\_male\_blacks <- sum(male\_blacks$death) / nrow(male\_blacks)

death\_rate\_white\_females <- sum(white\_females$death) / nrow(white\_females)

# Performing a two-sample proportion test

test\_result <- prop.test(x = c(sum(male\_blacks$death), sum(white\_females$death)),

n = c(nrow(male\_blacks), nrow(white\_females)),

alternative = "two.sided")

# Extracting the p-value from the test result

p\_value <- test\_result$p.value

# Printing the p-value

print(p\_value)

################################################################################

#c)

# Subset the data for female Blacks and White males

female\_blacks <- subset(data, gender == 2 & race == 2)

white\_males <- subset(data, gender == 1 & race == 1)

# Calculating the death rates for each group

death\_rate\_female\_blacks <- sum(female\_blacks$death) / nrow(female\_blacks)

death\_rate\_white\_males <- sum(white\_males$death) / nrow(white\_males)

# Performing a two-sample proportion test

test\_result <- prop.test(x = c(sum(female\_blacks$death), sum(white\_males$death)),

n = c(nrow(female\_blacks), nrow(white\_males)),

alternative = "two.sided")

# Extracting the p-value from the test result

p\_value <- test\_result$p.value

# Printing the number of individuals in each category

cat("Number of female Blacks:", nrow(female\_blacks), "\n")

cat("Number of White males:", nrow(white\_males), "\n")

# Printing the p-value

cat("p-value:", p\_value, "\n")

################################################################################

#d)

# Obtaining the Schoenfeld residuals

schoenfeld\_res <- residuals(model, type = "schoenfeld")

# Creating a data frame with the Schoenfeld residuals and time

residuals\_df <- data.frame(schoenfeld\_res = schoenfeld\_res, time = data$time)

# Plotting Schoenfeld residuals against time

plot(residuals\_df$time, residuals\_df$schoenfeld\_res, xlab = "Time", ylab = "Schoenfeld Residuals")

#Checking for outliers

# Obtaining the residuals of the Cox proportional hazards model

residuals <- residuals(model, type = "deviance")

# Plotting the residuals to identify outliers

boxplot(residuals, main = "Residuals Boxplot")

#checking for influential observations

# Compute the Schoenfeld residuals

schoenfeld\_res <- residuals(model, type = "schoenfeld")

# Calculate the absolute values of the Schoenfeld residuals

abs\_schoenfeld\_res <- abs(schoenfeld\_res)

# Identify influential observations based on a threshold

influential\_obs <- which(abs\_schoenfeld\_res > 2)

# Print the influential observations

cat("Influential Observations:", influential\_obs, "\n")