Project Report – Survival Analysis of Heart Failure Clinical Data

This project looked at survival patterns in 299 heart failure patients to understand what factors most influence the risk of death. I worked through the entire process in SAS — from cleaning and preparing the data, to running survival analyses, and then testing how standardizing the variables affected the results.

About the Data

The dataset contained a mix of demographic information (like age and sex), clinical measurements (ejection fraction, serum creatinine, serum sodium, creatinine phosphokinase), and health conditions (anaemia, diabetes, high blood pressure, smoking). Each record also included how long the patient survived after diagnosis and whether or not they died during the follow-up period.

Step 1 – Getting the Data Ready

I started by importing and checking the dataset in SAS to make sure all variables were in the right format. I explored key variables like age, ejection fraction, and serum creatinine to get a sense of the patient population.

Step 2 – Unadjusted Survival Analysis

Using Kaplan–Meier curves, I compared survival between patients with and without high blood pressure. The difference was clear — patients with high blood pressure tended to have shorter survival times, and the log-rank test confirmed this was statistically significant.

Step 3 – Adjusted Survival Modeling (Unstandardized Variables)

Next, I built a Cox regression model to see how high blood pressure affected survival when accounting for other factors. In this model, high blood pressure, older age, lower ejection fraction, and higher serum creatinine all increased the risk of death. Anaemia, creatinine phosphokinase, and serum sodium were also significant predictors.

Step 4 – Adjusted Survival Modeling (Standardized Variables)

I then standardized all the continuous variables so they were on the same scale (mean = 0, standard deviation = 1). This means the hazard ratios now represent the change in risk per **one standard deviation** instead of per raw unit. After standardization, two predictors really stood out:

- **Ejection fraction** showed a much stronger protective effect.
- **Serum creatinine** emerged as a major risk factor.

Interestingly, age was no longer statistically significant, suggesting its importance in the unstandardized model was partly due to its larger numerical range.

Improvements After Standardization

- It became much easier to compare which predictors had the biggest impact, since they were all on the same scale.
- Ejection fraction and serum creatinine emerged as the most critical factors.
- Variables like age lost their inflated importance.
- Predictors with large value ranges, like creatinine phosphokinase, had more stable coefficient estimates.

Final Conclusion

By standardizing the continuous variables, the model shifted from giving unit-based results to producing insights that are directly comparable across predictors. This not only made the findings easier to interpret but also made it clear which factors truly matter most for patient survival. If missing data is addressed before standardization, the model could keep its interpretability benefits without losing statistical power — making it a stronger and more reliable tool for guiding clinical decisions in heart failure care.