DataQuest Solutions

Survival Analysis with R

Lesson 1. Basics of Survival Analysis

- Definition and purpose
- Key Concepts:
    - Survival Time
    - Censoring (right, left, interval)
    - Event Occurrence

Lesson 2. Survival Data Structures

- Data Requirements: time, event indicator and covariates
- Creating Survival Objects:
  Surv(time, event) from the survival package
- Understanding censoring in Surv()

Lesson 3. Descriptive Survival Analysis

- Kaplan-Meier Estimator
- Life Tables
- Median Survival Time, Confidence interval

Lesson 4. Comparing Survival Curves

- Log-rank test (survdiff() in survival)
- Wilcoxon test (Breslow)
- Stratified analysis (sub-group comparisons)

Lesson 5. Cox Proportional Hazards Model

- Model Building: coxph(Surv(time, event) ~ covariates, data)
- Hazard ratios (HRs) and Confidence intervals
- Model Assumptions
  Proportional hazards assumption

Checking with cox.zph()
- Time-varying covariates: tt() function or survival::tmerge

Lesson 6. Accelerated Failure Time (AFT) Models

- Model: survreg(Surv(time, event) ~ covariates, dist = "Weibull")
- Parametric nodelling exponential, Weibull, log-logistic, log-normal

Lesson 7. Model Diagnostics and Validation

- Assess propotional hazards: cox.zph(), Schoenfeld, and cox-smell
- Residuals:
  Martingale, deviance, Schoenfeld, and Cox-Smell
- Goodness of fit
- ROC Curves for Survival Data: timeROC, survivalROC packages
- Concordance index (C-index)

Lesson 8. Stratified and Interaction Models

- Stratification in Cox: coxph(Surv() ~ covariates + strata(variable))
- Interaction terms: covariate1 * covariate2

Lesson 9. Time-Dependent Covariates

- Extended Cox Models
- Creating time-dependent datasets using survival::tmerge

Lesson 10. Competing Risks

- Event types: death from different causes
- Cumulative incidence functions
- Cmprsk package: cuminc(), crr() for Fine & Gray Model

Lesson 11. Recurrent events

- Counting process approach
- Andersen-Gill model (coxph() with counting process format)

Lesson 12. Visualization

- Ggsurplot() from survminer()
- Forest plots (forestmodel)
- Hazard curves, cumulative hazard (basehaz())

Lesson 13. Advanced Topics

- Multiple imputation for censored data
- Joint models (longitudinal + survival) using JM or JMbayes packages
- Flexible parametric models (flexsurv, rstpm2)
- DeepSurv and ML survival models

Lesson 14. Reporting Survival Analysis

- APA-style summaries
- HRs with 95% CI and p-values
- Kaplan-Meier plots with risk tables