Libraries

```r
library(survival) # For survival analysis
library(ggplot2) # For data visualization
library(ggfortify) # Visualize survival curves
library(gridExtra) # Arrange multiple plots
library(MASS) # For stepAIC model selection
library(SurvRegCensCov) # For parametric survival models
```

```
## Warning: package 'SurvRegCensCov' was built under R version 4.2.3
```

```
## Registered S3 method overwritten by 'SurvRegCensCov':
##   method     from
##   print.src dplyr
```

Data Preprocessing

```r
# Load pbc dataset
data(pbc)

# Create binary event indicator
pbc$status = ifelse(pbc$status == 2, 1, 0)

# Convert time to years
pbc$time <- pbc$time/365.25

# Convert variables to factors
pbc$edema <- as.factor(pbc$edema)
pbc$trt <- as.factor(pbc$trt)
pbc$sex <- as.factor(pbc$sex)
pbc$ascites <- as.factor(pbc$ascites)
pbc$hepato <- as.factor(pbc$hepato)
pbc$spiders <- as.factor(pbc$spiders)
pbc$stage <- as.factor(pbc$stage)
```

Kaplan-Meier Curves

```r
# Estimate Kaplan-Meier curves for variables
km_fit2 <- survfit(Surv(time, status) ~ sex, data = pbc, type = "kaplan-meier")
g1 <- autoplot(km_fit2) + labs(fill = "Sex", color = "Sex", x = "Time (Years)",
                               y = "Survival")

km_fit3 <- survfit(Surv(time, status) ~ ascites, data = pbc, type = "kaplan-meier")
g2 <- autoplot(km_fit3) + labs(fill = "Ascites", color = "Ascites", x = "Time (Years)",
                               y = "Survival")
```

```
km_fit4 <- survfit(Surv(time, status) ~ hepato, data = pbc, type = "kaplan-meier")
g3 <- autoplot(km_fit4) + labs(fill = "Hepato", color = "Hepato", x = "Time (Years)",
                               y = "Survival")

km_fit5 <- survfit(Surv(time, status) ~ spiders, data = pbc, type = "kaplan-meier")
g4 <- autoplot(km_fit5) + labs(fill = "Spiders", color = "Spiders", x = "Time (Years)",
                               y = "Survival")

km_fit6 <- survfit(Surv(time, status) ~ edema, data = pbc, type = "kaplan-meier")
g5 <- autoplot(km_fit6) + labs(fill = "Edema", color = "Edema", x = "Time (Years)",
                               y = "Survival")

km_fit16 <- survfit(Surv(time, status) ~ stage, data = pbc, type = "kaplan-meier")
g6 <- autoplot(km_fit16) + labs(fill = "Stage", color = "Stage", x = "Time (Years)",
                                y = "Survival")

km_fit17 <- survfit(Surv(time, status) ~ trt, data = pbc, type = "kaplan-meier")
g7 <- autoplot(km_fit17) + labs(fill = "Treatment", color = "Treatment", x = "Time (Years)",
                                y = "Survival")

# Arrange multiple KM curves
grid.arrange(g1, g2, g3, g4, g5, g6, g7, nrow = 4, ncol = 2)
```
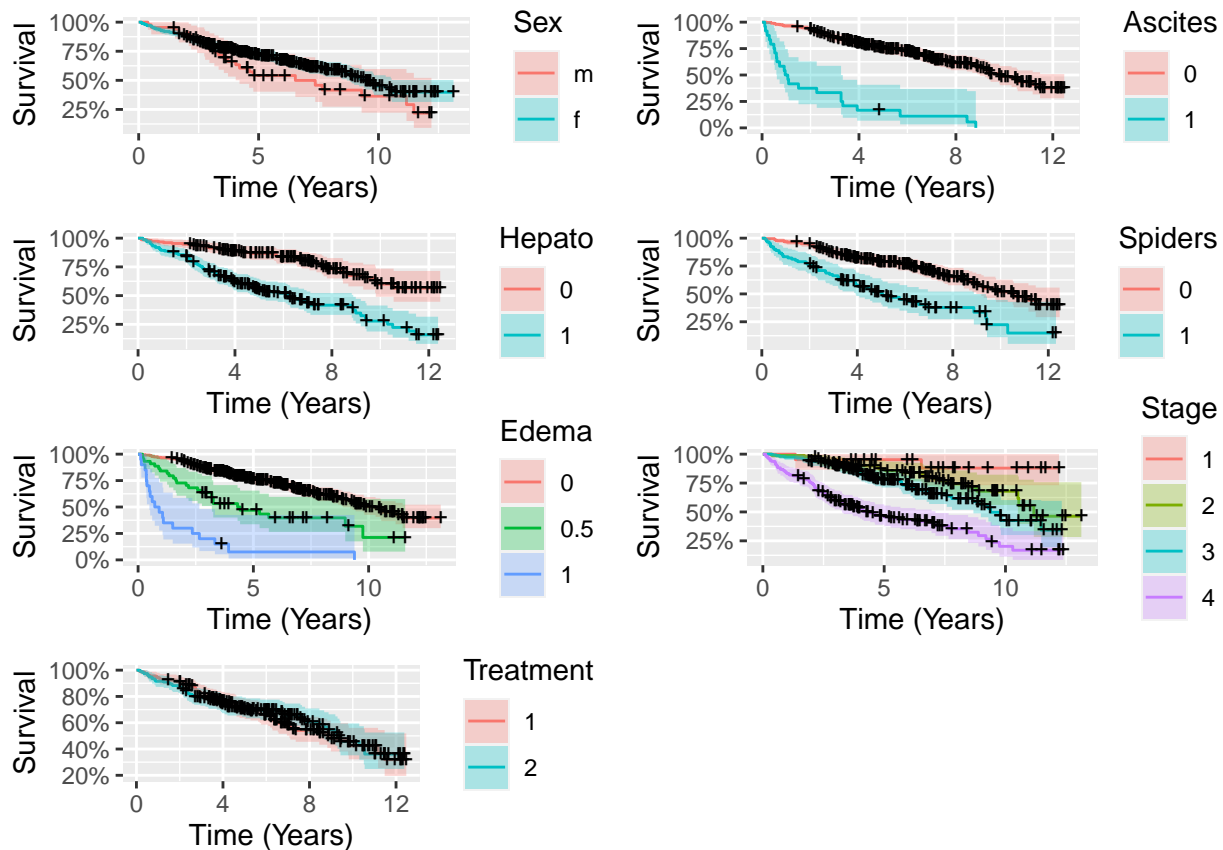
Cox Proportional Hazards Models

```r
# Fit Cox PH model
pbc2 <- na.omit(pbc) # Remove missing data
coxph(Surv(time, status) ~ age + edema + bili + albumin + copper + ast + protime + stage,
      data = pbc2)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ age + edema + bili + albumin +
##     copper + ast + protime + stage, data = pbc2)
##
##                coef  exp(coef)   se(coef)       z        p
## age       0.0313188  1.0318144  0.0102909   3.043  0.00234
## edema0.5  0.1598036  1.1732804  0.3054890   0.523  0.60090
## edema1    0.9121653  2.4897075  0.3545431   2.573  0.01009
## bili      0.0869270  1.0908171  0.0198104   4.388 1.14e-05
## albumin  -0.7387129  0.4777284  0.2792471  -2.645  0.00816
## copper    0.0027867  1.0027905  0.0009912   2.811  0.00493
## ast       0.0039562  1.0039641  0.0018415   2.148  0.03168
## protime   0.2642049  1.3023951  0.1122859   2.353  0.01862
## stage2    1.3596258  3.8947355  1.0808743   1.258  0.20843
## stage3    1.6823556  5.3782101  1.0478753   1.605  0.10839
## stage4    2.0627073  7.8672396  1.0432133   1.977  0.04801
##
## Likelihood ratio test=165.7  on 11 df, p=< 2.2e-16
## n= 276, number of events= 111
```

```r
# Fit full Cox PH model
model <- coxph(data = pbc2, Surv(time, status) ~ trt + age + sex + ascites + hepato + spiders
               + edema + bili + chol + albumin + copper + alk.phos + ast + trig + platelet
               + protime + stage + age:chol + trt:age)

# Backwards selection
step <- stepAIC(model, direction = "backward")
```

```
## Start:  AIC=973.62
## Surv(time, status) ~ trt + age + sex + ascites + hepato + spiders +
##     edema + bili + chol + albumin + copper + alk.phos + ast +
##     trig + platelet + protime + stage + age:chol + trt:age
##
##              Df    AIC
## - alk.phos    1 971.62
## - ascites     1 971.63
## - hepato      1 971.66
## - age:chol    1 971.67
## - spiders     1 971.76
## - trig        1 971.81
## - platelet    1 971.99
## - trt:age     1 972.67
## - sex         1 972.95
## <none>          973.62
## - ast         1 974.78
```

```
## - stage      3 975.60
## - copper     1 975.81
## - protime    1 976.38
## - edema      2 976.87
## - albumin    1 977.12
## - bili       1 979.78
##
## Step:  AIC=971.62
## Surv(time, status) ~ trt + age + sex + ascites + hepato + spiders +
##      edema + bili + chol + albumin + copper + ast + trig + platelet +
##      protime + stage + age:chol + trt:age
##
##              Df     AIC
## - ascites    1 969.63
## - hepato     1 969.66
## - age:chol   1 969.67
## - spiders    1 969.77
## - trig       1 969.81
## - platelet   1 970.00
## - trt:age    1 970.67
## - sex        1 970.95
## <none>         971.62
## - ast        1 972.79
## - stage      3 973.61
## - copper     1 974.17
## - protime    1 974.44
## - edema      2 974.88
## - albumin    1 975.32
## - bili       1 977.78
##
## Step:  AIC=969.63
## Surv(time, status) ~ trt + age + sex + hepato + spiders + edema +
##      bili + chol + albumin + copper + ast + trig + platelet +
##      protime + stage + age:chol + trt:age
##
##              Df     AIC
## - hepato     1 967.67
## - age:chol   1 967.69
## - spiders    1 967.78
## - trig       1 967.81
## - platelet   1 968.01
## - trt:age    1 968.67
## - sex        1 968.98
## <none>         969.63
## - ast        1 970.81
## - stage      3 971.80
## - copper     1 972.43
## - protime    1 972.59
## - edema      2 973.03
## - albumin    1 974.16
## - bili       1 975.87
##
## Step:  AIC=967.67
## Surv(time, status) ~ trt + age + sex + spiders + edema + bili +
```

4

```
##     chol + albumin + copper + ast + trig + platelet + protime +
##     stage + age:chol + trt:age
##
##              Df     AIC
## - age:chol  1 965.74
## - spiders   1 965.83
## - trig      1 965.88
## - platelet  1 966.03
## - trt:age   1 966.71
## - sex       1 967.04
## <none>        967.67
## - ast       1 968.89
## - copper    1 970.53
## - protime   1 970.63
## - edema     2 971.04
## - stage     3 972.16
## - albumin   1 972.25
## - bili      1 974.04
##
## Step:  AIC=965.74
## Surv(time, status) ~ trt + age + sex + spiders + edema + bili +
##     chol + albumin + copper + ast + trig + platelet + protime +
##     stage + trt:age
##
##              Df     AIC
## - spiders   1 963.90
## - trig      1 963.96
## - platelet  1 964.07
## - trt:age   1 964.73
## - chol      1 965.08
## - sex       1 965.11
## <none>        965.74
## - ast       1 967.10
## - copper    1 968.82
## - protime   1 968.91
## - edema     2 969.10
## - stage     3 970.21
## - albumin   1 970.25
## - bili      1 972.04
##
## Step:  AIC=963.9
## Surv(time, status) ~ trt + age + sex + edema + bili + chol +
##     albumin + copper + ast + trig + platelet + protime + stage +
##     trt:age
##
##              Df     AIC
## - trig      1 962.09
## - platelet  1 962.18
## - trt:age   1 962.82
## - sex       1 963.17
## - chol      1 963.37
## <none>        963.90
## - ast       1 965.18
## - protime   1 967.17
```

```
## - copper    1 967.37
## - edema     2 967.90
## - albumin   1 968.30
## - stage     3 969.67
## - bili      1 970.58
##
## Step:  AIC=962.09
## Surv(time, status) ~ trt + age + sex + edema + bili + chol +
##     albumin + copper + ast + platelet + protime + stage + trt:age
##
##            Df    AIC
## - platelet  1 960.30
## - trt:age   1 960.98
## - sex       1 961.29
## - chol      1 961.59
## <none>        962.09
## - ast       1 963.41
## - copper    1 965.39
## - protime   1 965.52
## - albumin   1 966.34
## - edema     2 967.13
## - stage     3 967.67
## - bili      1 969.59
##
## Step:  AIC=960.3
## Surv(time, status) ~ trt + age + sex + edema + bili + chol +
##     albumin + copper + ast + protime + stage + trt:age
##
##            Df    AIC
## - sex       1 959.33
## - trt:age   1 959.35
## - chol      1 960.13
## <none>        960.30
## - ast       1 961.41
## - protime   1 963.65
## - copper    1 963.92
## - albumin   1 964.46
## - edema     2 965.20
## - stage     3 965.67
## - bili      1 968.16
##
## Step:  AIC=959.33
## Surv(time, status) ~ trt + age + edema + bili + chol + albumin +
##     copper + ast + protime + stage + trt:age
##
##            Df    AIC
## - trt:age   1 958.39
## - chol      1 959.20
## <none>        959.33
## - ast       1 960.98
## - protime   1 962.65
## - albumin   1 962.79
## - edema     2 963.61
## - stage     3 964.57
```

```
## - copper   1 965.43
## - bili     1 966.67
##
## Step:  AIC=958.39
## Surv(time, status) ~ trt + age + edema + bili + chol + albumin +
##     copper + ast + protime + stage
##
##            Df    AIC
## - trt      1 957.00
## - chol     1 957.95
## <none>       958.39
## - ast      1 959.90
## - protime  1 961.95
## - albumin  1 961.97
## - edema    2 962.17
## - stage    3 963.60
## - copper   1 964.13
## - age      1 965.62
## - bili     1 966.07
##
## Step:  AIC=957
## Surv(time, status) ~ age + edema + bili + chol + albumin + copper +
##     ast + protime + stage
##
##            Df    AIC
## - chol     1 956.69
## <none>       957.00
## - ast      1 958.23
## - protime  1 960.27
## - edema    2 960.61
## - albumin  1 960.82
## - stage    3 961.86
## - copper   1 963.15
## - bili     1 964.38
## - age      1 965.43
##
## Step:  AIC=956.69
## Surv(time, status) ~ age + edema + bili + albumin + copper +
##     ast + protime + stage
##
##            Df    AIC
## <none>       956.69
## - ast      1 958.84
## - edema    2 958.86
## - protime  1 959.90
## - stage    3 961.14
## - albumin  1 961.39
## - copper   1 961.88
## - age      1 963.90
## - bili     1 970.89
```

Model Diagnostics

```r
# Check model residuals
mr <- residuals(step, type = "martingale")

# Plot residuals against covariates
par(mfrow = c(3,2))
plot(pbc2$age, mr, xlab = "Age (Years)", ylab = "Martingale Residuals")
plot(pbc2$bili, mr, xlab = "Bilirubin (mg/dl)", ylab = "Martingale Residuals")
plot(pbc2$albumin, mr, xlab = "Albumin (gm/dl)", ylab = "Martingale Residuals")
plot(pbc2$copper, mr, xlab = "Copper (ug/day)", ylab = "Martingale Residuals")
plot(pbc2$ast, mr, xlab = "SGOT (U/ml)", ylab = "Martingale Residuals")
plot(pbc2$protime, mr, xlab = "Prothrombin Time (seconds)", ylab = "Martingale Residuals")
```
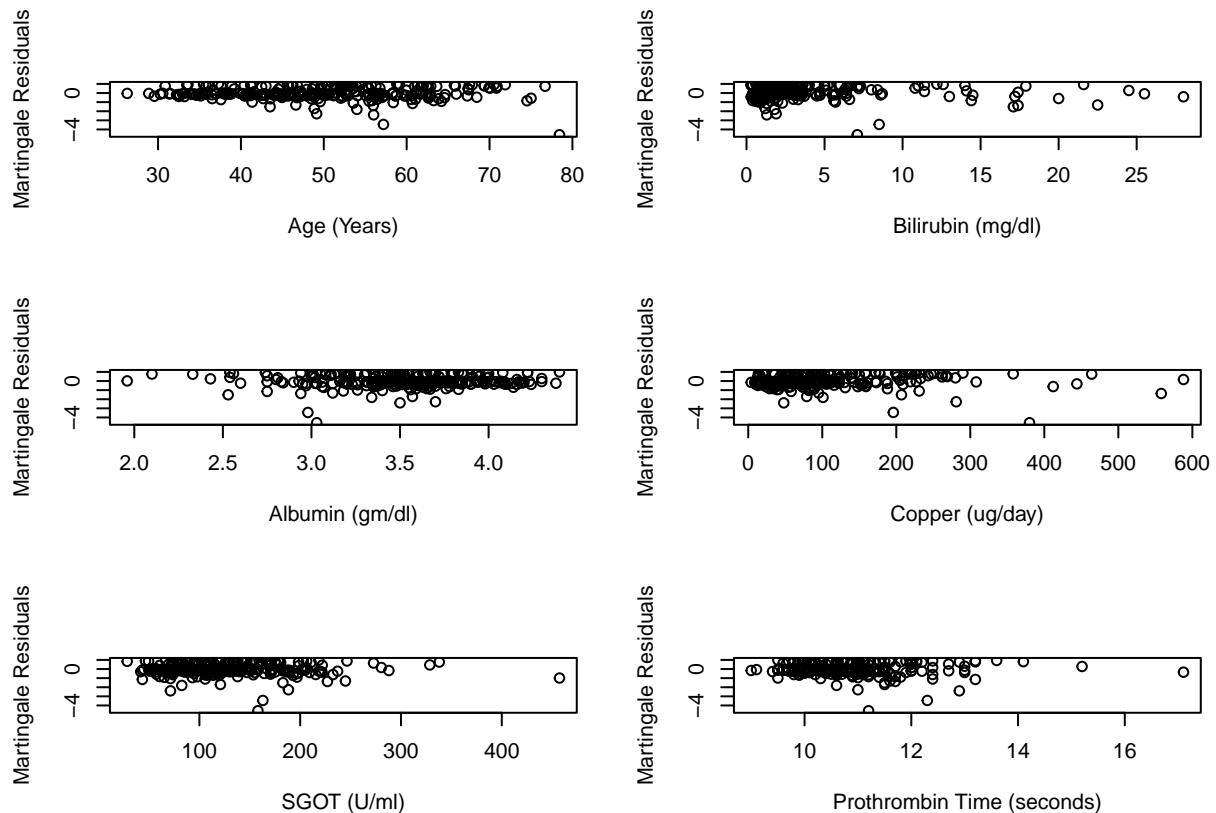

Weibull Model

```r
# Fit Weibull AFT model on pbc data
ret <- survreg(Surv(time, status) ~ trt + age + sex + ascites + hepato +
                spiders + edema + bili + chol + albumin + copper +
                alk.phos + ast + trig + platelet + protime + stage +
                age:chol + trt:age,
              dist = "weibull", data = pbc2)
```

```r
# Exponentiate coefficients to get acceleration factors
# Values > 1 indicate shorter survival times
acceleration_factors <- exp(ret$coefficients)

# Rescale coefficients by estimated shape parameter
# Allows coefficients to be interpreted as log hazard ratios
shape <- 0.606
haz_ratios <- exp(-ret$coefficients/shape)
```

Model Validation

```r
# Compute Cox-Snell residuals
CS <- pbc2$status - residuals(step)

# Check cumulative hazard assumption
plot(survfit(Surv(CS, pbc2$status) ~ 1, type = "flem"), fun = "cumhaz")
```