Homework4

Yuki Joyama

```
# import data
df = haven::read_dta("./data/umaru.dta")
```

1. Parametric/Accelerated Failure Time Models

a.

Variables: `age`, `nonwhite`, `treat`, `site`, `ivdrug`
Models: Exponential, Weibull, Log-logistic, Log-normal, Generalized Gamma

```
# define the survival object
surv_object <- Surv(time = df$time, event = df$censor)

# fit AFT models with different distributions
aft_exponential <-
  flexsurvreg(surv_object ~ age + nonwhite + treat + site + ivdrug, data = df, dist = "exponential")
aft_weibull <-
  flexsurvreg(surv_object ~ age + nonwhite + treat + site + ivdrug, data = df, dist = "weibull")
aft_llogis <-
  flexsurvreg(surv_object ~ age + nonwhite + treat + site + ivdrug, data = df, dist = "llogis")
aft_lognormal <-
  flexsurvreg(surv_object ~ age + nonwhite + treat + site + ivdrug, data = df, dist = "lognormal")
aft_gen_gamma <-
  flexsurvreg(surv_object ~ age + nonwhite + treat + site + ivdrug, data = df, dist = "gengamma")

# check the results
# aft_exponential
# aft_weibull
# aft_llogis
# aft_lognormal
# aft_gen_gamma
```

i. Values of the -2 log L, the total number of parameters (including shape and scale for ϵ) and the AIC for each of these models

```
# extract the log-likelihood, total parameters, and AIC
results <- data.frame(
  Model = c("Exponential", "Weibull", "Log-logistic", "Log-normal", "Generalized Gamma"),
  "-2LogL" = c(
    -2 * aft_exponential$loglik,
```

```
    -2 * aft_weibull$loglik,
    -2 * aft_llogis$loglik,
    -2 * aft_lognormal$loglik,
    -2 * aft_gen_gamma$loglik
  ),
  "Total Parameters" = c(
    aft_exponential$npars,
    aft_weibull$npars,
    aft_llogis$npars,
    aft_lognormal$npars,
    aft_gen_gamma$npars
  ),
  AIC = c(
    aft_exponential$AIC,
    aft_weibull$AIC,
    aft_llogis$AIC,
    aft_lognormal$AIC,
    aft_gen_gamma$AIC
  )
)

colnames(results) <- c("Model", "-2 Log L", "Total Parameters", "AIC")

results |>
  kable()
```

Model	-2 Log L	Total Parameters	AIC
Exponential	6180.608	6	6192.608
Weibull	6179.805	7	6193.805
Log-logistic	6127.191	7	6141.191
Log-normal	6137.951	7	6151.951
Generalized Gamma	6137.219	8	6153.219

The Log-logistic model has the lowest AIC (6141.191), making it the best-fitting model based on AIC.

ii.

Given the AIC, Weibull model does not provide an improved fit compared to the exponential model. I will confirm this using likelihood ratio test.

```
# extract log-likelihoods
logL_exp <- aft_exponential$loglik
logL_weib <- aft_weibull$loglik

# compute LRT statistic
lrt_stat <- -2 * (logL_exp - logL_weib)

# df
degf <- aft_weibull$npars - aft_exponential$npars

# p-value
```

```r
p_value <- pchisq(lrt_stat, df = degf, lower.tail = FALSE)

# output results
cat("LRT Statistic:", lrt_stat)
## LRT Statistic: 0.8022232
cat("Degrees of Freedom:", degf)
## Degrees of Freedom: 1
cat("p-value:", p_value)
## p-value: 0.3704295
```

Given p-value > 0.05, we fail to reject the null hypothesis and conclude that Weibull model does not improve fit compared to the exponential model.

iii.

Exponential model and Weibull model are nested within the generalized gamma model.
Generalized gamma model vs exponential model

```r
# extract log-likelihoods
logL_exp <- aft_exponential$loglik
logL_ggamma <- aft_gen_gamma$loglik

# compute LRT statistic
lrt_stat <- -2 * (logL_exp - logL_ggamma)

# df
degf <- aft_gen_gamma$npars - aft_exponential$npars

# p-value
p_value <- pchisq(lrt_stat, df = degf, lower.tail = FALSE)

# output results
cat("LRT Statistic:", lrt_stat)
## LRT Statistic: 43.38879
cat("Degrees of Freedom:", degf)
## Degrees of Freedom: 2
cat("p-value:", p_value)
## p-value: 3.786546e-10
```

Given p-value < 0.05, we reject the null hypothesis and conclude that generalized gamma model provides a better fit compared to the exponential model.

Generalized gamma model vs Weibull model

```r
# extract log-likelihoods
logL_exp <- aft_exponential$loglik
logL_weib <- aft_weibull$loglik

# compute LRT statistic
lrt_stat <- -2 * (logL_weib - logL_ggamma)

# df
```

```
degf <- aft_gen_gamma$npars - aft_weibull$npars

# p-value
p_value <- pchisq(lrt_stat, df = degf, lower.tail = FALSE)

# output results
cat("LRT Statistic:", lrt_stat)
## LRT Statistic: 42.58657
cat("Degrees of Freedom:", degf)
## Degrees of Freedom: 1
cat("p-value:", p_value)
## p-value: 6.762209e-11
```

Given p-value < 0.05, we reject the null hypothesis and conclude that generalized gamma model provides a better fit compared to the Weibull model.

b.

c.

d.

2. Hazard Rates and Survival from Parametric Models

a.

b.

c.

d.