Homework4

Yuki Joyama

```r
# 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

```r
# 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

```r
# 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,
```

```r
    -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.

```r
# 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
```

```r
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.

Time ratio: $\phi = e^{\beta}$

```r
# function to compute time ratio and confidence interval
tr_ci <- function(model) {
  # extract coefficient and standard error for ivdrug
  beta <- model$res["ivdrug", "est"]
  se <- model$res["ivdrug", "se"]

  # compute time ratio (phi) and 95% confidence interval
  phi <- exp(beta)
  ci_lower <- exp(beta - 1.96 * se)
  ci_upper <- exp(beta + 1.96 * se)

  # return as a named vector
  c("Time Ratio (phi)" = phi,
    "95% CI Lower" = ci_lower,
    "95% CI Upper" = ci_upper)
}

# apply the function to all models
results <- data.frame(
  Model = c("Exponential", "Weibull", "Log-logistic", "Log-normal", "Generalized Gamma"),
  rbind(
    tr_ci(aft_exponential),
    tr_ci(aft_weibull),
    tr_ci(aft_llogis),
    tr_ci(aft_lognormal),
    tr_ci(aft_gen_gamma)
  )
)

colnames(results) = c("Model", "Time Ratio (phi)", "95% CI Lower", "95% CI Upper")

results |>
  kable()
```

Model	Time Ratio (phi)	95% CI Lower	95% CI Upper
Exponential	1.5210016	1.2386592	1.8677018
Weibull	0.6527657	0.5274755	0.8078157
Log-logistic	0.6610930	0.5254482	0.8317545
Log-normal	0.6547605	0.5161506	0.8305934
Generalized Gamma	0.6520866	0.5154833	0.8248900

In the Wibull model, the estimated ϕ for the `ivdrug` covariate is 0.653 (95%CI: 0.527-0.808). On average, individuals with IV drug use have survival times that are about 65.3% of those without IV drug use, after adjusting for other covariates. This association is statistically significant.

c.

Exponential model
$\beta_{HR} = \beta_{AFT}$
$HR = e^{\beta_{HR}} = e^{\beta_{AFT}}$
95%CI $= e^{\beta_{HR} \pm 1.96 \times SE_{HR}}$

```r
# extract AFT beta and SE for IV drug use (exponential model)
beta_aft_exp <- aft_exponential$res["ivdrug", "est"]
se_aft_exp <- aft_exponential$res["ivdrug", "se"]

# log-hazard ratio for exponential
beta_hr_exp <- beta_aft_exp

# HR and 95% CI
hr_exp <- exp(beta_hr_exp)
hr_exp_ci <- exp(c(beta_hr_exp - 1.96 * se_aft_exp, beta_hr_exp + 1.96 * se_aft_exp))
```

Webull model
$\beta_{HR} = -\beta_{AFT}/\alpha$, where $\alpha =$ shape parameter
$HR = e^{\beta_{HR}} = e^{\beta_{AFT}/\alpha}$
95%CI $= e^{\beta_{HR} \pm 1.96 \times SE_{HR}}$

```r
# extract AFT beta and SE for IV drug use (Weibull model)
beta_aft_weib <- aft_weibull$res["ivdrug", "est"]
se_aft_weib <- aft_weibull$res["ivdrug", "se"]

# extract Weibull shape parameter
alpha <- aft_weibull$res["shape", "est"]

# log-hazard ratio for Weibull
beta_hr_weib <- -beta_aft_weib / alpha
se_hr_weib <- se_aft_weib / alpha

# HR and 95% CI
hr_weib <- exp(beta_hr_weib)
hr_weib_ci <- exp(c(beta_hr_weib - 1.96 * se_hr_weib, beta_hr_weib + 1.96 * se_hr_weib))
```

Results:

```r
# summarize results
results <- data.frame(
  Model = c("Exponential", "Weibull"),
  "Log-HR" = c(beta_hr_exp, beta_hr_weib),
  "HR" = c(hr_exp, hr_weib),
  "95% CI Lower" = c(hr_exp_ci[1], hr_weib_ci[1]),
  "95% CI Upper" = c(hr_exp_ci[2], hr_weib_ci[2])
)

colnames(results) = c("Model", "Log-HR", "HR", "95% CI Lower", "95% CI Upper")

results |>
  kable()
```

Model	Log-HR	HR	95% CI Lower	95% CI Upper
Exponential	0.4193691	1.521002	1.238659	1.867702
Weibull	0.4412566	1.554660	1.247057	1.938136

In the exponential model, the HR for IV drug use is 1.52 (95% CI: 1.24–1.87), indicating that individuals with IV drug use have a 52% higher hazard of the event compared to those without IV drug use, holding other variables constant. Similarly, in the Weibull model, the HR is 1.55 (95% CI: 1.25–1.94), showing consistent evidence of increased hazard associated with IV drug use.

d.

2. Hazard Rates and Survival from Parametric Models

a.

b.

c.

d.