Nowacek MidtermQ4

Q4

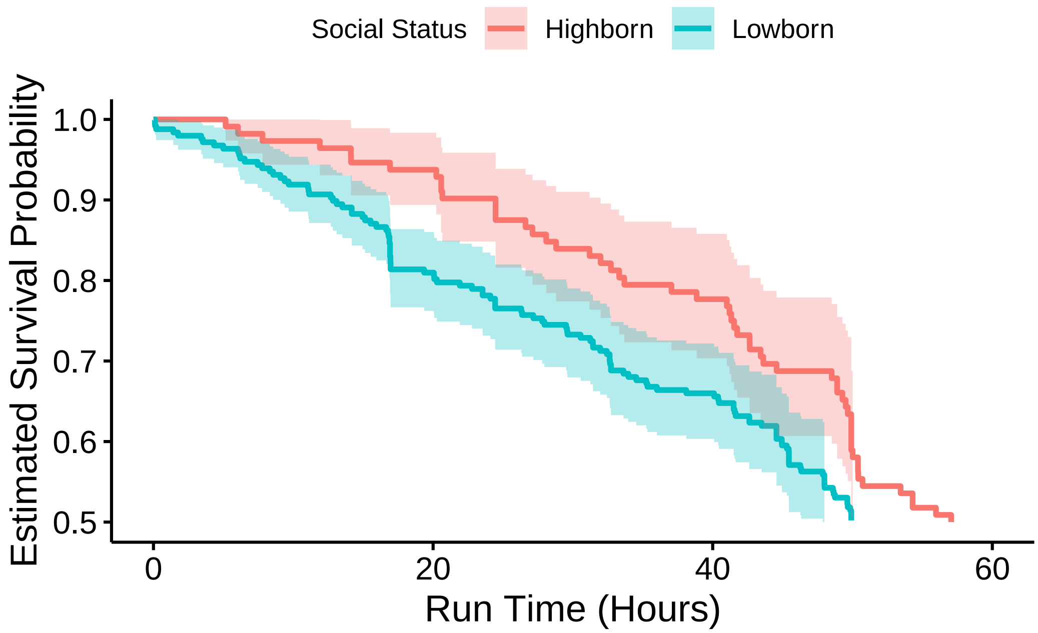
```
data <-
read.csv("https://www2.stat.duke.edu/courses/Fall24/sta490.01/data/got_dat.csv")
```

Part 1

```
kme <- survfit(Surv(event_time_hrs, dth_flag) ~ social_status, data = data)

ggsurvplot(kme, data = data,
           main = "Kaplan-Meier Estimates by Social Status",
           xlab = "Run Time (Hours)",
           ylab = "Estimated Survival Probability",
           legend.title = "Social Status",
           legend.labs = c("Highborn", "Lowborn"),
           ylim = c(0.5, 1), conf.int = TRUE, censor = FALSE)
```

```
Warning: Removed 30 rows containing missing values or values outside the scale range
(`geom_step()`).
Removed 30 rows containing missing values or values outside the scale range
(`geom_step()`).
```



Part 2

```
logrt <- survdiff(Surv(event_time_hrs, dth_flag) ~ social_status, data = data)
logrt
```

```
Call:
survdiff(formula = Surv(event_time_hrs, dth_flag) ~ social_status,
    data = data)

                  N Observed Expected (O-E)^2/E (O-E)^2/V
social_status=1 112       67     71.5     0.286     0.434
social_status=2 247      145    140.5     0.146     0.434

 Chisq= 0.4  on 1 degrees of freedom, p= 0.5
```

At the $\alpha = 0.05$ level, this test is not significant as the p-value is 0.5. This indicates that we can not reject the null hypothesis which states that there is no difference expected survival between high-born and low-born individuals. Therefore, we fail to reject the hypothesis that high-born individuals and low-born individuals have a different expectation in terms of the number of cumulative show-hours that they survive based on their social status alone.

Part 3

```r
aft <- survreg(Surv(event_time_hrs, dth_flag) ~ intro_time_hrs +
                 social_status + sex,
               data = data, dist = "weibull")
summary(aft)
```

```
Call:
survreg(formula = Surv(event_time_hrs, dth_flag) ~ intro_time_hrs +
    social_status + sex, data = data, dist = "weibull")
                  Value Std. Error     z       p
(Intercept)     3.68869    0.22506 16.39 < 2e-16
intro_time_hrs  0.01777    0.00284  6.25 4.1e-10
social_status  -0.22269    0.10283 -2.17 0.03033
sex             0.39897    0.11658  3.42 0.00062
Log(scale)     -0.39494    0.06108 -6.47 1.0e-10

Scale= 0.674

Weibull distribution
Loglik(model)= -1092.4   Loglik(intercept only)= -1126.3
    Chisq= 67.78 on 3 degrees of freedom, p= 1.3e-14
Number of Newton-Raphson Iterations: 6
n= 359
```

The question is, is there evidence for differential survival based on when a character first appears on screen.

The coefficient for when a character first appears on screen is positive and significant at the $\alpha = 0.05$ level, indicating that as the introduction time increases, that is the character was introduced later in the show, their expected survival also increases, controlling for social status and sex.

The coefficient of sex is also positive and significant at the $\alpha = 0.05$ level, indicating that women are expected to live longer than men, controlling for social status and when a character first appears on screen.

The coefficient of social status is negative and significant at the $\alpha = 0.05$ level, indicating that lowborn individuals are expected to live longer than highborn individuals, controlling for social status and when a character first appears on screen.

In conclusion, controlling for the other variables in the study, women are expected to live longer than men, lowborn individuals are expected to live longer than highborn individuals, and the later you were introduced to the show, the higher your expected survival.