Chapter Ten

Ryan Honea

Exercise One

Question

The data in the table below (Table 10.4) are survival times, in weeks, for leukemia patients. There is no censoring. There are two covariates, white blood cell count (WBC) and the result of a test (AG positive and AG negative). The data set is from Feigl and Zelen (1965) and the data for the 17 patients with AG positivbe test results are described in Exercise 4.2.

AG positive		AG negative	
Survival time	White Blood cell count	Survival time	White Blood cell count
65	2.30	56	4.40
156	0.75	65	3.00
100	4.30	17	4.00
134	2.60	7	1.50
16	6.00	16	9.00
108	10.50	22	5.30
121	10.00	3	10.00
4	17.00	4	19.00
39	5.40	2	27.00
143	7.00	3	28.00
56	9.40	8	31.00
26	32.00	4	26.00
22	35.00	3	21.00
1	100.00	30	79.00
1	100.00	4	100.00
5	52.00	43	100.00
65	100.00		

Solution

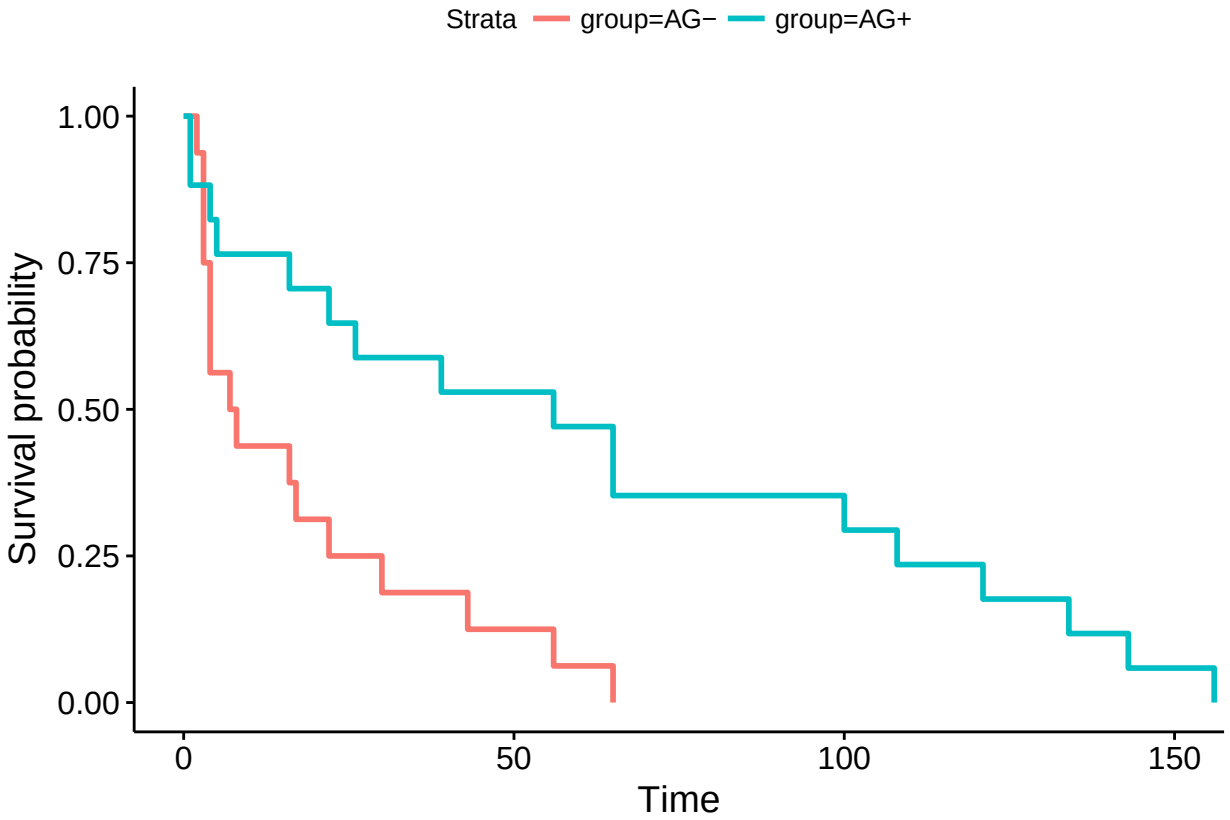
Below is just data entry.

```
df <- data.frame(
  survival = c(65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65,
               56, 65, 17, 7, 16, 22, 3, 4, 2, 3, 8, 4, 3, 30, 4, 43),
  whiteblood = c(2.3, .75, 4.3, 2.6, 6.0, 10.5, 10, 17, 5.4, 7, 9.4, 32, 35, 100, 100, 52, 100,
                 4, 3, 4, 1.5, 9, 5.3, 10, 19, 27, 28, 31, 26, 21, 79, 100, 100),
  group = c(rep("AG+",17),rep("AG-",16))
)
```

(a): Obtain the empirical survivor functions $\hat{S}(y)$ for each group (AG positive and AG negative) ignoring WBC.

Solution: We can obtain both empirical survival functions below.

```
lsurv.km <- survfit(Surv(survival) ~ group, df, type='kaplan-meier')
ggsurvplot(lsurv.km)
```



(b): Use suitable plots of the estimates $\hat{S}(y)$ to select an approriate probability distribution to model the data.

Solution: I would argue that the simple exponential would work, but the Weibull distribution would also describe it. We can test the true fits which is done below.

(c): Use a parametric model to compare the survival times for the two groups, after adjustment for the covariate WBC, which is best transformed to log(WBC).

Solution: We can create two models for this data set, one exponential and the other Weibull. I will use their Akaike Information Criterion to determine which model is best to use.

```
out.weibull <- survreg(Surv(survival) ~ log(whiteblood) + group, dist = "weibull", data=df)
out.exponential <- survreg(Surv(survival) ~ log(whiteblood) + group,
                           dist = "exponential", data=df)
cat("Weibull AIC: ", AIC(out.weibull), "\n")
```

```
## Weibull AIC:  300.936
```

```
cat("Exponential AIC: ", AIC(out.exponential))
```

```
## Exponential AIC:  299.0161
```

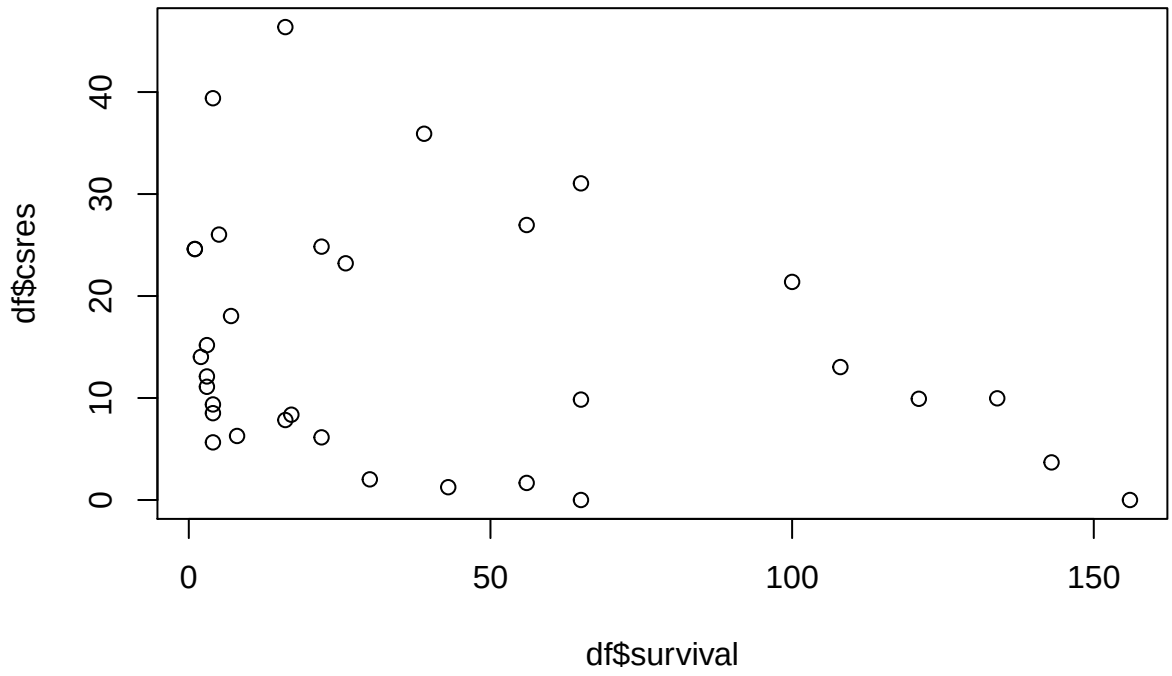
With such a small difference in AIC, the Exponential Model likely provides as good a fit as Weibull and thus I use it (due to its increased simplicity.)

(d): Check the adequacy of the model using residuals and other diagnostic tests.

Solution: I will use the Cox-Snell residuals in order to determine the adequacy of the model.

```
df$fits <- predict(out.exponential,df)
agneg <- cbind(summary(lsurv.km)$surv[1:12], summary(lsurv.km)$time[1:12])
agpos <- cbind(summary(lsurv.km)$surv[13:27], summary(lsurv.km)$time[13:27])
```

```r
h_0 <- c()
for (i in 1:17) {
  h_0[i] <- agpos[,1][agpos[,2] == df[i,1]]
}
for (i in 18:33) {
  h_0[i] <- agneg[,1][agneg[,2] == df[i,1]]
}
df$csres <- df$fits*h_0
plot(df$survival, df$csres)
```



(e): Based on this analysis, is AG a useful prognostic indicator?

Based solely on the graphs, I would say that AG is a useful indicator where positive AG usually is associated with longer survival, but the residuals associated with the added white blood cell counts seem to be widely spread out and so there is little reason to use it.

Exercise Two

Question

The **log-logistic distribution** with the probability desntiy function

$$f(y) = \frac{e^{\theta}\lambda y^{\lambda-1}}{(1+e^{\theta}y^{\lambda})^2}$$

Solution

(a): Find the survivor function $S(y)$, the hazard function $h(y)$ and the cumulative hazard function $H(y)$.

Solution:

$$
\begin{aligned}
S(y) &= 1 - F(y) \\
&= 1 - \int_0^t \frac{e^\theta \lambda y^{\lambda-1}}{(1+e^\theta y^\lambda)^2} dy \\
&= 1 - (1 - \frac{1}{1+e^\theta y^\lambda}) \\
&= \frac{1}{1+e^\theta y^\lambda} \\
h(y) &= f(y)/S(y) \\
&= \frac{e^\theta \lambda y^{\lambda-1}}{(1+e^\theta y^\lambda)^2} * 1 + e^\theta y^\lambda \\
&= \frac{e^\theta \lambda y^{\lambda-1}}{1+e^\theta y^\lambda} \\
H(y) &= -\log\left(\frac{1}{1+e^\theta y^\lambda}\right) \\
&= -\log(1) - -\log(1+e^\theta y^\lambda) \\
&= \log(1+e^\theta y^\lambda)
\end{aligned}
$$

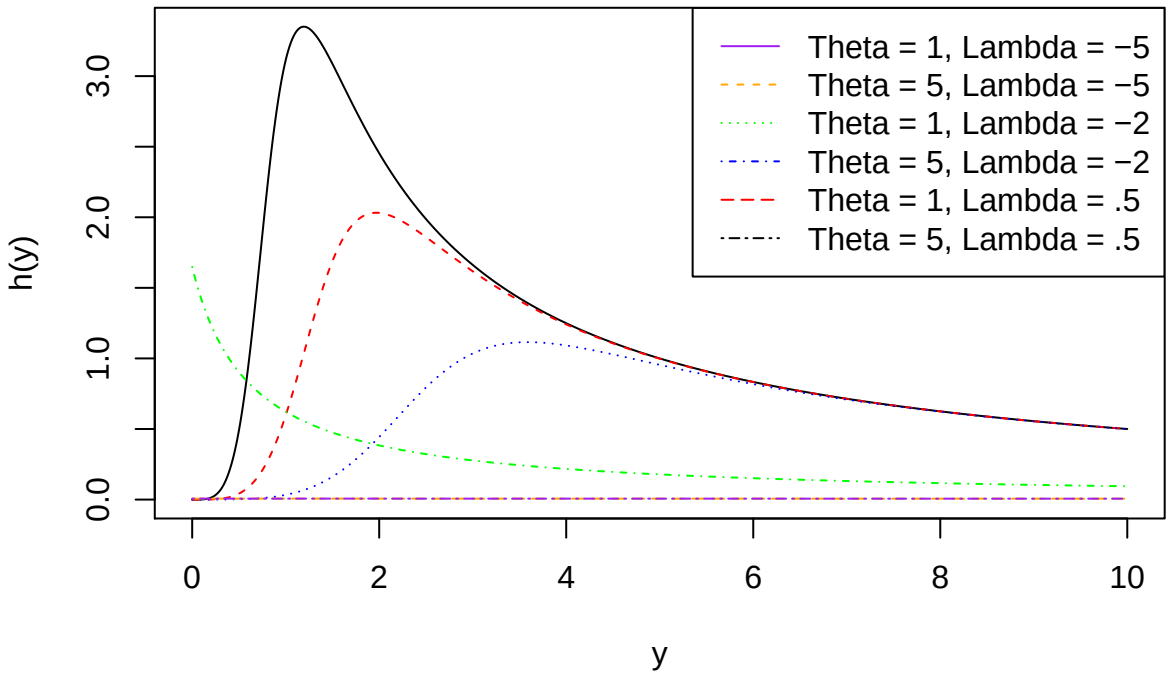
(b): Show that the median survival time is $\exp(\theta/\lambda)$.

Solution:

(c): Plot the hazard function for $\lambda = 1$ and $\lambda = 5$ with $\theta = -5$, $\theta = -2$, and $\theta = \frac{1}{2}$.

Solution:

```
y <- seq(0,10,by=.01)
lambda = c(1,5)
theta = c(-5,-2,.5)
h <- function(y, lambda, theta) {
  numerator <- exp(theta)*lambda*(y^(lambda-1))
  denominator <- 1 + exp(theta)*y^lambda
  return(numerator/denominator)
}
plot(y, h(y, lambda[2], theta[3]), type = "l", ylab = "h(y)", lty = 1)
lines(y, h(y, lambda[2], theta[2]), type = "l", col = "red", lty = 2)
lines(y, h(y, lambda[2], theta[1]), type = "l", col = "blue", lty = 3)
lines(y, h(y, lambda[1], theta[3]), type = "l", col = "green", lty = 4)
lines(y, h(y, lambda[1], theta[1]), type = "l", col = "orange", lty = 5)
lines(y, h(y, lambda[1], theta[1]), type = "l", col = "purple", lty = 6)
labels = c("Theta = 1, Lambda = -5", "Theta = 5, Lambda = -5",
           "Theta = 1, Lambda = -2", "Theta = 5, Lambda = -2",
           "Theta = 1, Lambda = .5", "Theta = 5, Lambda = .5")
cols = c("purple", "orange", "green", "blue", "red", "black")
legend("topright", labels, lty = seq(1:6), col = cols)
```


Theta = 1, Lambda = −5
Theta = 5, Lambda = −5
Theta = 1, Lambda = −2
Theta = 5, Lambda = −2
Theta = 1, Lambda = .5
Theta = 5, Lambda = .5
h(y)
0.0
1.0
2.0
3.0
0
2
4
6
8
10
y