Problem Set 4

Philip Kruger
18328699
16/04/23

Question 1

part 1

The data was loaded and 2 models were created. One which is additive and another which has an interactive term. This was done with the following code:

```
#loading libraries
library(ggplot2)
library(tidyverse)
library("eha")
library("survival")

#this fit is literally taken from https://cran.r-project.org/web/packages/eha/eha.pdf pg 13 with
#1 covariate changed
fit <- coxreg(Surv(enter, exit, event) ~ sex + m.age, data = child, coxph = TRUE)
summary(fit)

#using interactive model
interfit <- coxreg(Surv(enter, exit, event) ~ sex * m.age, data = child, coxph = TRUE)
summary(interfit)
```

This output the following two models:

```
> summary(fit)
Covariate             Mean       Coef     Rel.Risk   S.E.    LR p
sex                                                          0.002
            male      0.510     0         1 (reference)
          female      0.490    -0.082     0.921     0.027
m.age                32.010     0.008     1.008     0.002    0.000

Events                    5616
Total time at risk        325030
Max. log. likelihood      -56503
LR test statistic         22.52
Degrees of freedom        2
Overall p-value           1.28921e-05




Model:
Surv(enter, exit, event) ~ sex * m.age
          Df    AIC   LRT Pr(>Chi)
<none>       113013
sex:m.age  1 113011 0.106     0.74

Covariate             Mean       Coef     Rel.Risk   S.E.    Wald p
sex
```

```
            male       0.510     0          1 (reference)
          female       0.490    -0.127      0.881      0.140    0.365
m.age                 32.010     0.007      1.007      0.003    0.017
sex:m.age
   female:                     0.001     1.001     0.004     0.744

Events                    5616
Total time at risk        325030
Max. log. likelihood      -56503
LR test statistic         22.62
Degrees of freedom        3
Overall p-value           4.83661e-05
```

The first major thing to notice is that the chi squared for the interactive model is .74. This means it is not statistically significant and thus we will be using the additive model (which has .002 for sex and .0003 for m.age which is statistically significant in both cases). For the additive model, we can see that female has a coefficient of -0.82, this means that if the child is female the log likelyhood of the child dying (between the ages of 0 and 15) decreases by .82 over the basline (male children). The coefficient for m.age is 0.008. This means that for a one unit increase in the age of the mother (presumably when the child is born but, documentation doesn't specify) there is a 0.008 increase in the log likelyhood of the child dying.

To illustrate these results I plotted the cumulative hazard function and the Kaplan-Meier curve with the following code:

```
    #plotting the results
plot(fit, main = "Cumulative probability of a child dying 0-15")
plot(survfit(fit),main = "Cumulative probability of a child surviving 0-15")
```

This resulted in the following graphs:

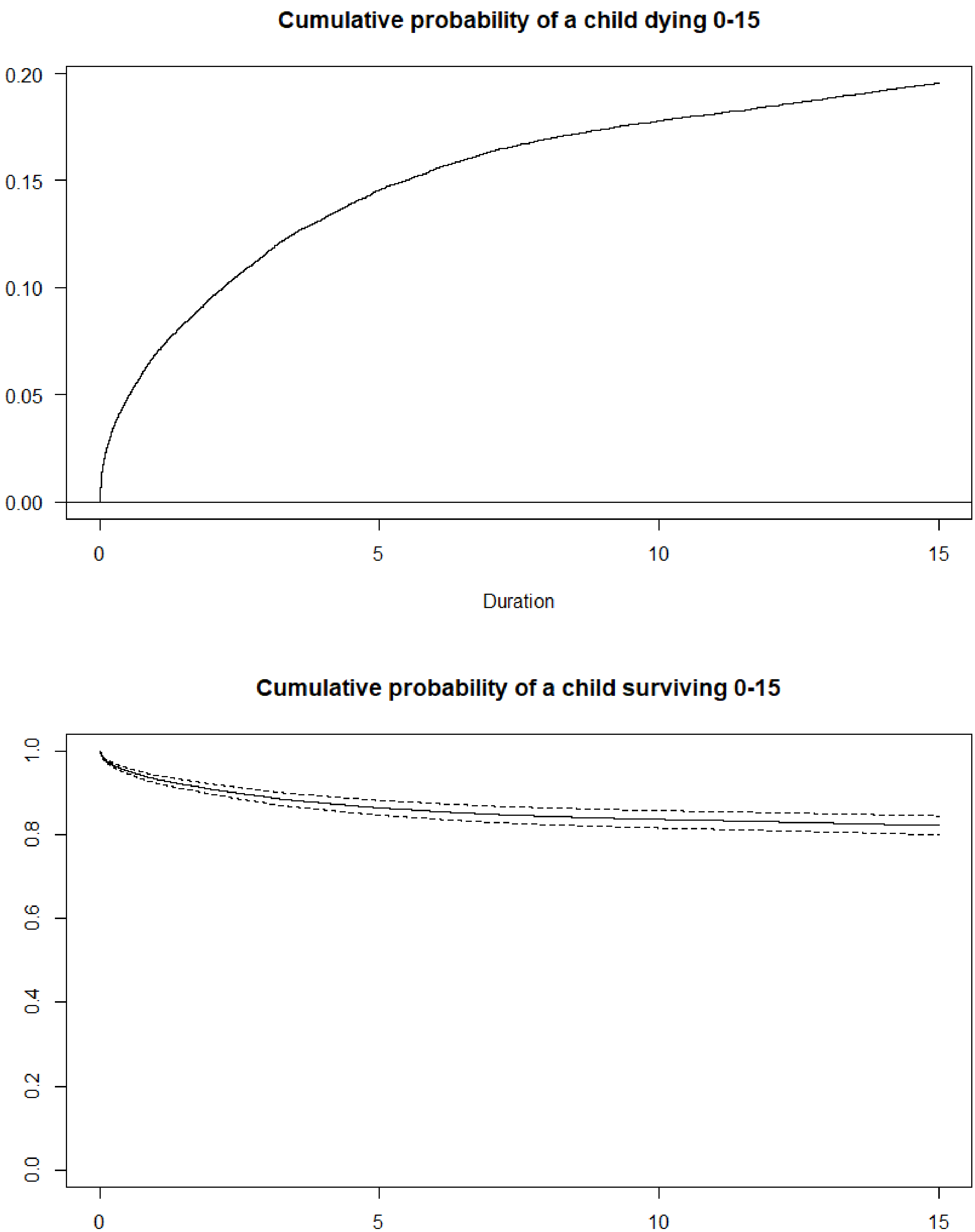

Figure 1: cumulative probility and Kaplan-Meier curve

These plots are clearly inverses of eachother. This makes sense since one is measuring mortality hazard and one is representing survival probability. We can see that the curve is logarithmic which means that there is a relatively high probablity of the child dying very young (within first year) and it becomes less probable as the child gets older. We can see that the probability of reaching the age of 1 is roughly 92 percent, reaching 5 is roughly 85 percent and reaching 15 is roughly 80 percent.