Stat 525 Homework 11

Kenny Flagg

December 7, 2016

A study was conducted on the effects of ploidy on the prognosis of patients with cancers of the mouth. Patients were selected who had a paraffin-embedded sample of the cancerous tissue taken at the time of surgery. Follow-up survival data were obtained on each patient. The tissue samples were examined to determine if the tumor had a aneuploid (abnormal) or diploid (normal) DNA profile. The data set contains `type` *with a 1 indicating aneuploid and a 2 indicating diploid tumor types,* `time` *is survival in weeks post-surgery, and* `delta` *with a 1 indicating death and 0 indicating censoring. The question of interest is whether or not survival is associated with tumor type. We will examine this question by fitting the Cox proportional hazards model and a Weibull parametric model.*

1. *Fit a Cox proportional hazards model to these data. Give summary results and interpret the results in terms of the problem. Give both a point estimate and an approximate 95% CI. Assess the adequacy of the proportional hazards assumption. Note that aneuploidy is the reference type.*

```
library(survival)
library(KMsurv)
library(SurvRegCensCov)
data(tongue)

fit.cox <- coxph(Surv(time, delta) ~ factor(type), tongue)
summary(fit.cox)

Call:
coxph(formula = Surv(time, delta) ~ factor(type), data = tongue)

  n= 80, number of events= 53

                coef exp(coef) se(coef)      z Pr(>|z|)
factor(type)2 0.4664    1.5942   0.2804 1.663   0.0963

              exp(coef) exp(-coef) lower .95 upper .95
factor(type)2     1.594     0.6273    0.9201     2.762

Concordance= 0.564  (se = 0.036 )
Rsquare= 0.033   (max possible= 0.993 )
Likelihood ratio test= 2.67  on 1 df,   p=0.102
Wald test            = 2.77  on 1 df,   p=0.09632
Score (logrank) test = 2.81  on 1 df,   p=0.09343
```

We estimate that the hazard of death for patients with diploid tumors is 1.594 times larger than the hazard of death for patients with aneuploid tumors, and we are 95% confident that the true hazard ratio is between 0.9201 and 2.762.

There is no evidence ($\chi_1^2 = 0.473$, p-value $= 0.491$) that the proportional hazards assumption is violated. This is confirmed by the lack of trends in the plot of Schoedfeld residuals over time.

```
fit.zph <- cox.zph(fit.cox)
fit.zph

                 rho chisq     p
factor(type)2 -0.0958 0.473 0.491

plot(fit.zph)
```
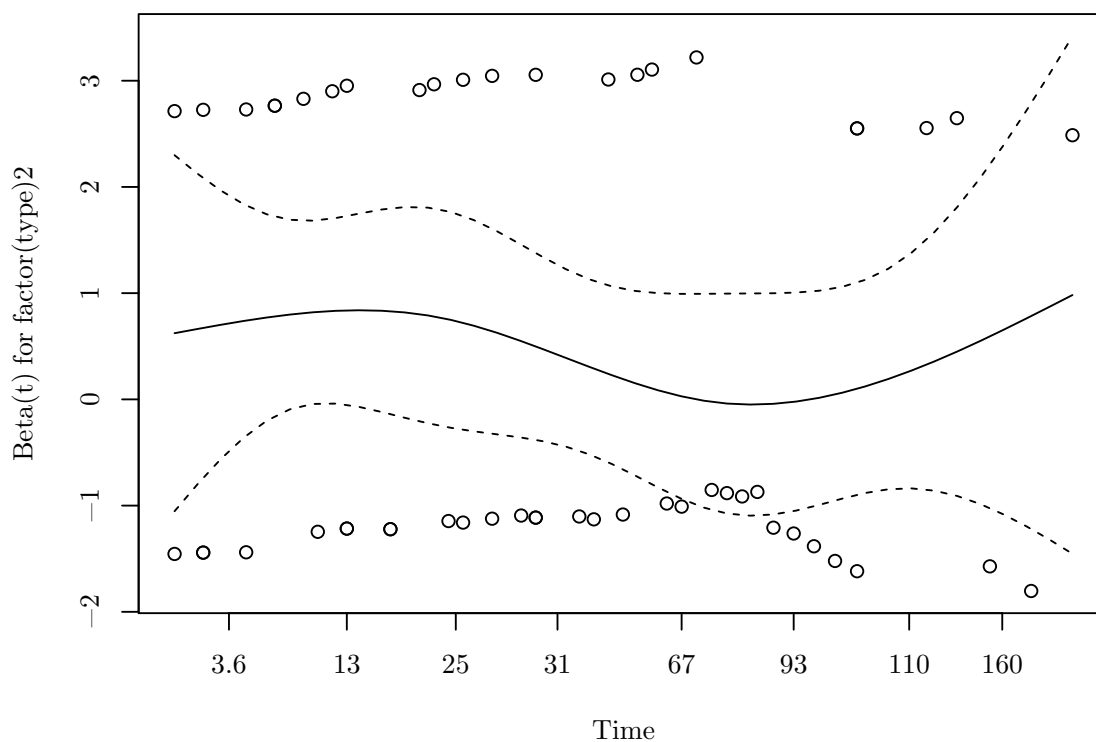
2. *Fit a parametric Weibull model. Again give summary results, interpret the results in terms of the problem giving a point estimate and an approximate 95% CI.*

```
fit.wei <- WeibullReg(Surv(time, delta) ~ factor(type), tongue)
fit.wei

$formula
Surv(time, delta) ~ factor(type)

$coef
                Estimate          SE
lambda        0.01819544 0.008475542
gamma         0.80590167 0.093217809
factor(type)2 0.53911791 0.279160048

$HR
                    HR        LB       UB
factor(type)2 1.714494 0.9920064 2.963176

$ETR
                   ETR        LB       UB
factor(type)2 0.5122398 0.2574967 1.019002

$summary

Call:
survival::survreg(formula = formula, data = data, dist = "weibull")
              Value Std. Error     z         p
(Intercept)   4.972      0.227 21.93 1.46e-106
factor(type)2 -0.669      0.351 -1.91  5.66e-02
Log(scale)    0.216      0.116  1.87  6.21e-02

Scale= 1.24

Weibull distribution
Loglik(model)= -298.9   Loglik(intercept only)= -300.7
Chisq= 3.58 on 1 degrees of freedom, p= 0.059
Number of Newton-Raphson Iterations: 5
n= 80
```

We estimate that the hazard of death for patients with diploid tumors is 1.714 times larger than the hazard of death for patients with aneuploid tumors, and we are 95% confident that the true hazard ratio is between 0.9920 and 2.963.

We continue with the tongue cancer survival data.

1. *For the Weibull model give an interpretation in terms of the accelerated failure-time model including both the point estimate and an approximate 95% CI. I want you to compare the median survival times in the accelerated failure time representation.*

   We estimate the median survival time for patients with diploid tumors to be 0.512 times the median survival time for patients with aneuploid tumors, with a 95% confidence interval of 0.257 to 1.019.

2. *Fit a Log-Logistic parameteric model and give an interpretation in terms of the accelerated failure-time model including both the point estimate and an approximate 95% CI.*

```
fit.llogis <- survreg(Surv(time, delta) ~ factor(type), dist = 'loglogistic', tongue)
summary(fit.llogis)


Call:
survreg(formula = Surv(time, delta) ~ factor(type), data = tongue,
    dist = "loglogistic")
                Value Std. Error      z       p
(Intercept)    4.4695      0.242 18.437 6.61e-76
factor(type)2 -0.7906      0.404 -1.956 5.05e-02
Log(scale)    -0.0413      0.116 -0.356 7.22e-01

Scale= 0.96

Log logistic distribution
Loglik(model)= -298.7   Loglik(intercept only)= -300.6
Chisq= 3.79 on 1 degrees of freedom, p= 0.051
Number of Newton-Raphson Iterations: 4
n= 80
```

The median surival time for patients with diploid tumors is estimated to be $\exp(0.7906) = 2.205$ times the median survival time for patients with aneuploid tumors, with a 95% confidence interval of $\exp(0.7906 - 1.96 \times 0.404) = 0.998$ to $\exp(0.7906 + 1.96 \times 0.404) = 4.869$.

Based on the above does it appear that ploidy type is associated with survival time? Justify your answer.

All of the hazard ratio and acceleration factor confidence intervals above include 1, suggesting that the hazard ratio and acceleration factor do not differ by ploidy type, so the does not appear to be an association between ploidy type and survival time.