Time-to-Event Analysis: Practical 2

Instructor: Graeme L. Hickey

27th October 2014

Contents

1 Introduction

The afternoon lecture covered regression modelling of survival data. During this practical we will review how to fit and interpret regression models fitted in R. It is necessary for us to utilize everything we covered during the morning practical. As usual, if you have any questions, do not hesitate to ask.

Before we begin, it is important to remember that if you are opening a new R session, you will need to re-load the `survival` library.

```
library(survival)
```

```
## Loading required package: splines
```

2 Example 1: Cox proportional hazards regression

Reload the `dialysis` data (Diggle and Chetwynd, 2011) from the morning practical class if it is not already in your R workspace.

```
dialysis <- read.table("dialysis.data", header = TRUE)
```

We fit a Cox proportional hazards regression model to the data including only one variable — `method` — using the `coxph()` function. That is, we are fitting the model

$$h(t, \texttt{method}) = h_0(t) \exp(\beta \times \texttt{method}),$$

where `method` $= 0$ if CAPD was given, and $= 1$ is APD was given.

The syntax is almost identical to the `survfit()` function used to calculate the Kaplan-Meier estimator.

```
fit0 <- coxph(Surv(days, dead) ~ method, data = dialysis)
```

We called our regression model `fit0`. If we type `fit0` into the console we get some basic information back.

```
fit0
```

```
## Call:
## coxph(formula = Surv(days, dead) ~ method, data = dialysis)
##
##
##          coef exp(coef) se(coef)     z     p
## method -0.48     0.619     0.25 -1.92 0.055
##
## Likelihood ratio test=3.68  on 1 df, p=0.055  n= 124, number of events= 65
```

This is actually sufficient for our needs. However, applying the `summary()` function to our model provides us with further details of the model.

```
summary(fit0)
```

```
## Call:
## coxph(formula = Surv(days, dead) ~ method, data = dialysis)
##
##   n= 124, number of events= 65
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## method -0.4796    0.6190   0.2502 -1.917   0.0553 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## method     0.619      1.615    0.3791     1.011
##
## Concordance= 0.608  (se = 0.032 )
## Rsquare= 0.029   (max possible= 0.99 )
## Likelihood ratio test= 3.68  on 1 df,   p=0.05495
## Wald test            = 3.67  on 1 df,   p=0.05529
## Score (logrank) test = 3.74  on 1 df,   p=0.05306
```

We estimate $\hat{\beta} = -0.48$, which gives a hazard ratio for APD (relative to CAPD) of $\exp(-0.48) = 0.62$. The Wald-test P-value for this variable is 0.055. The coefficient of method is negative with a marginally significant P-value, suggesting some evidence that the hazard is lower for patients under APD than those under CAPD, according to the proportional hazards model. Refer back to the morning practical to compare this to the log-rank test.

Now recall that as well as treatment type (CAPD or APD), we have information on the patient's age at the start of dialysis which is likely to be an important variable in describing survival. A reasonable question to ask now is "does the effect of dialysis delivery method change if we take age at first dialysis into account?" In other words, we should include age at first dialysis in our model alongside the method of dialysis delivery to address this question. The Cox model as described in the lecture takes the form

$$h(t, \texttt{method}, \texttt{age}) = h_0(t) \exp(\beta_1 \times \texttt{method} + \beta_2 \times \texttt{age})$$

To implement this in R we **augment** the formula with the additional variables separated by a '+' symbol. We can add as many variables as we like using this method (within the limits of what the data permits and what is sensible) and R will estimate a coefficient for each one.

```
fit1 <- coxph(Surv(days, dead) ~ method + age, data = dialysis)
summary(fit1)
```

```
## Call:
## coxph(formula = Surv(days, dead) ~ method + age, data = dialysis)
##
##   n= 124, number of events= 65
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## method -0.990757  0.371296  0.264176 -3.750 0.000177 ***
## age     0.056637  1.058272  0.008381  6.758  1.4e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## method    0.3713     2.6933    0.2212    0.6231
## age       1.0583     0.9449    1.0410    1.0758
##
## Concordance= 0.753  (se = 0.038 )
## Rsquare= 0.352   (max possible= 0.99 )
## Likelihood ratio test= 53.89  on 2 df,   p=1.99e-12
## Wald test            = 49.71  on 2 df,   p=1.603e-11
## Score (logrank) test = 55.8  on 2 df,   p=7.633e-13
```

We find that age appears to be associated with hazard also, and the positive coefficient of age suggests that older patients experience a greater hazard, as intuition would suggest. That is, we obtain a hazard ratio of 1.06 per each additional year of age ($P < 0.001$).

In addition the coefficient of `method` has changed, as has its associated level of significance, although the sign (negative) of the coefficient remains the same, again indicating evidence that the hazard is lower for patients under APD than those under CAPD. The hazard ratio is now 0.37 ($P < 0.001$).

3 Problem 1: Cox proportional hazards regression

3.1 The `veteran` data

The `veteran` dataset is contained in the `survival` package. It records survival times in a study trial comparing two treatments for lung cancer. See Kalbfleisch and Prentice (1980) for further details.

The dataset contains 8 variables:

- `trt`: 1 = standard, 2 = test
- `celltype`: squamous, small-cell, adeno, large
- `time`: survival time
- `status`: censoring status

- `karno`: Karnofsky performance score (100 = good)
- `diagtime`: months from diagnosis to randomization
- `age`: in years
- `prior`: prior therapy 0 = no, 1 = yes

3.2 Exercises

1. Fit a Cox proportional hazards regression model to the `veteran` data including only the treatment variable. Do the data suggest any difference in the long-term survival probability according to the method of treatment?

2. Try adding in additional covariates.[1] Are the hazard ratios (coefficients) intuitively sensible? How does the treatment coefficient change?

4 Example 2: Assessing the proportional hazards assumption

We have now found that it is quite straightforward to fit a Cox proportional hazards model. However, that does not mean the model is appropriate for the data at hand. During the lecture we noted that there are a number of tests we can apply:

1. Simple check: if Kaplan-Meier curves cross the PH violated (this is a one-way argument: non-crossing curves does not ensure PH satisfied)
2. Plot $\log\left[-\log(\hat{S}(t))\right]$ against $\log(t)$ and check they are parallel
3. Plot smoothed (scaled) Schoenfeld residuals against predictor and check slope = 0
4. Grambsch-Therneau test

The first two tests are straightforward, and we applied them in Practical 1. The second two tests are more advanced, but in the interests of completeness I give an overview (optional reading). Notice, however, that the first two tests only apply to categorical variables; for continuous variables we must use the latter two methods, or arbitrarily dichotomize / categorize the continuous variable.

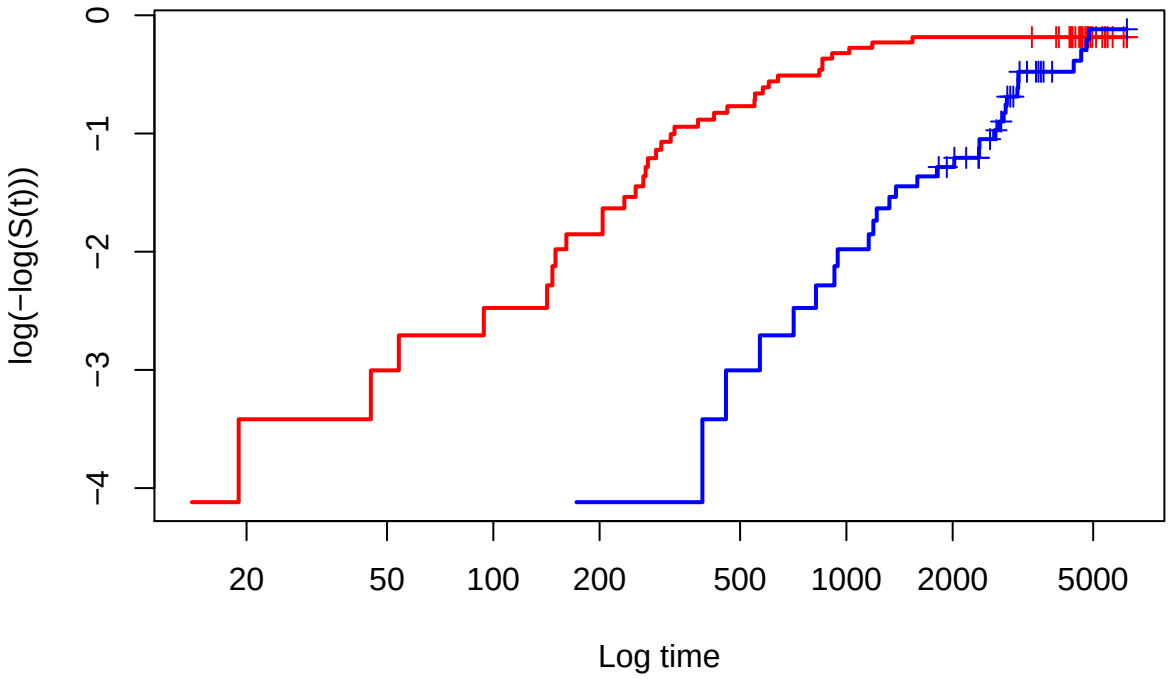
Let us apply each of these in turn to the `dialysis` data.

Check 1. The Kaplan-Meier curves do cross (see morning lecture slides and Practical 1) at around 13-years. However, the uncertainty in the survival probabilities around this time (and extending to the end of the study) are large and overlapping. So, we would not rule out the PH assumption on the basis of this plot alone.

[1] Regression model building and selection is a broad topic that crosscuts through most of statistics, therefore the focus on this exercise is not to justify model choice, but just to build a multivariable model.

Check 2. In the final exercise of Practical 1 we noted that we can plot $\log\left(\hat{H}(t)\right)$ against $\log(t)$, which is equivalent to plotting $\log\left[-\log(\hat{S}(t))\right]$ against $\log(t)$, using `plot(fit.km, fun = "cloglog")`.

```
fit.dialysis <- survfit(Surv(days, dead) ~ method, data = dialysis)
plot(fit.dialysis, fun = "cloglog",
     col = c(2, 4),
     lwd = 2,
     xlab = "Log time",
     ylab = "log(-log(S(t)))")
```



These lines are not parallel, so we would rule out the proportional hazards assumption and appeal to a parametric model that did not depend on the assumption.

Check 3 *(Optional reading).* Like all regression models, we can use residuals as diagnostic tools for assessing model fit. As is often the case, there are more than one type of residuals. Schoenfeld residuals derived from a Cox regression model play a central role in evaluating whether the proportional hazards assumption is violated.

Recall from the lecture that (if the model is correct) and a failure is to occur at time $t^\star$, then

$$P(\text{subject } i \text{ dies}) = \frac{\exp(\beta^T \mathbf{x}_i)}{\sum_{k \in R(t^\star)} \exp(\beta^T \mathbf{x}_k)}$$

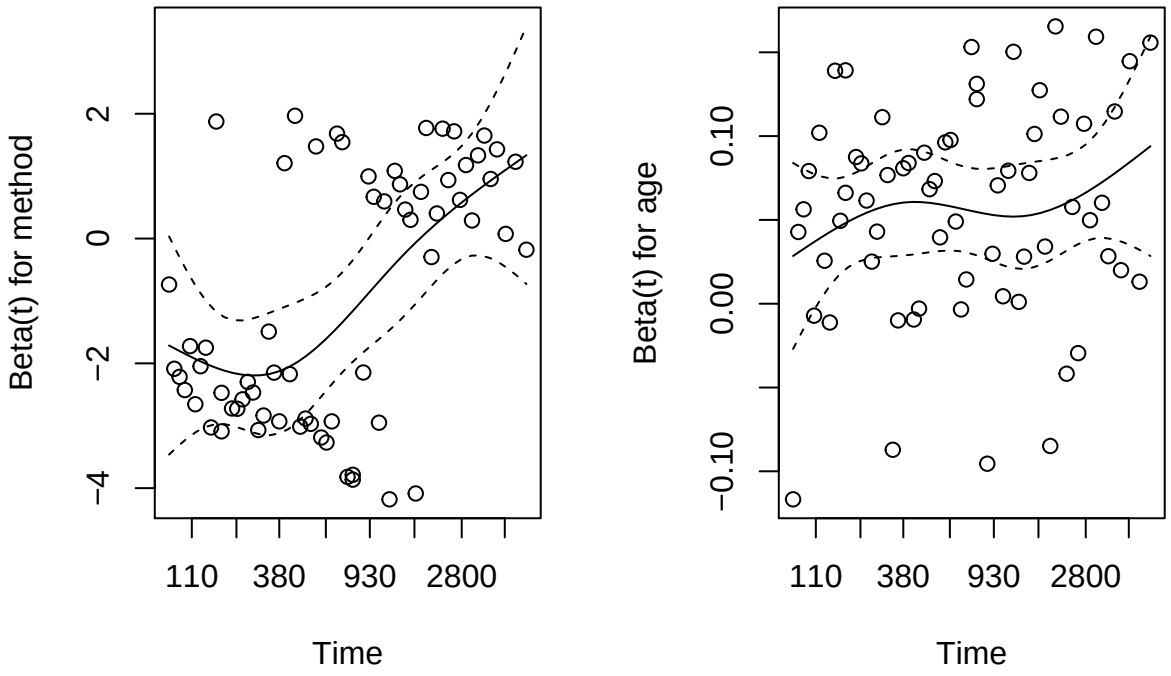
where $R(t)$ is the set of all subjects alive and uncensored at time t (called the risk set).

For any covariate in the model we can then calculate the *expectation* of this value for the subject who ultimately fails at time t_i by plugging-in the maximum partial likelihood estimates

for β.[2] The difference between the observed covariate value and the expected covariate value is called the *Schoenfeld residual*. Note that we calculate separate residuals for each covariate and subject with observed failure time.

R reports something called a *scaled* Schoenfeld residual for reasons I won't go into here. See Grambsch and Therneau (1994) for further details. We then use the `cox.zph()` function to plot these residuals against some function of time, typically log time.

```
dialysis.phtest <- cox.zph(fit1)
par(mfrow = c(1, 2)) # Allows me to plot 2 figures side-by-side
plot(dialysis.phtest)
```



If the proportional hazards assumption is satisfied then we should observe no association between the residuals and (some function of) time. In the plots above we find that the assumption appears plausible for age; however for the variable method the assumption is violated. In fact, we see that there is a positive linear trend, suggesting that the *relative* hazard for APD increases with time.

Check 4 *(Optional reading).* Whilst a graphical inspection of the smoothed (scaled) Schoenfeld residuals is useful, we would like a formal test statistic. The Grambsch and Therneau test can be subdivided into two tests.

1. An individual test for each covariate, such that the test statistic is the Pearson sample correlation coefficient ρ between the (scaled) Schoenfeld residuals and log time. A P-value is also provided under the null hypothesis of $\rho = 0$.
2. A 'global test' that takes into account all covariates.

[2]Recall: the expectation of a discrete random variable X is $E(X) = \sum_{x \in \mathcal{X}} xP(X = x)$.

Applying the `print()` function to our saved object `dialysis.phtest` gives the Grambsch and Therneau statistics.

```
print(dialysis.phtest)
```

```
##          rho chisq        p
## method 0.538 16.15 5.84e-05
## age    0.151  1.32 2.51e-01
## GLOBAL    NA 21.49 2.16e-05
```

The results indicate that age does not provide sufficient evidence to reject the null hypothesis of proportional hazards under this model. However, the method variable is strongly significant, and this is also driving the global test to be significant. Therefore we would conclude that the proportional hazards assumption is not satisfied in this model.

5 Problem 2: Assessing the proportional hazards assumption

1. For the two Cox PH models fitted in Problem 1 (one with treatment only and one with treatment + other relevant covariates), evaluate the proportional hazards assumption for the treatment variable included using any appropriate method(s). Does the inclusion of additional covariates in the model affect inferences about the proportional hazards assumption for the treatment variable?

2. Evaluate whether there is evidence to suggest that the variable `celltype` violates the proportional hazards assumption?

3. As mentioned in the lecture, one method to resolve non-proportional hazards is to *stratify* on the offending variable. This method includes a separate baseline hazard function in the Cox model for each stratum. Fit two models: one with `celltype` + other relevant variables, and one with `celltype` as a stratifying variable (use `strata(celltype)` in the model formula). Did this affect any of the other coefficients? What is the immediately obvious drawback of using a stratified model? (Hint: what is the effect size for `celltype`?)

6 Example 3: Parametric regression

We next turn to fitting a Weibull regression model. If you haven't already completed Problem 3 from Practical 1, then it is advised you do so before reading on.

In Problem 3 in Practical 1, we made use of the `survreg()` function. In particular, we included formulae of the sort `Surv(time, event) ~ 1`. If we replace the right hand-side

(~ 1) with our covariates, then we can implement a regression model. (Note that the ~1 notation tells R to only estimate an intercept, or what we called μ in the AFT model.)

We now fit two Weibull regression models to the `dialysis` data: the first including the treatment method only, the second including treatment method and age.

```
fit.weib0 <- survreg(Surv(days, dead) ~ method, data = dialysis,
                     dist = "weibull")
summary(fit.weib0)
```

```
##
## Call:
## survreg(formula = Surv(days, dead) ~ method, data = dialysis,
##     dist = "weibull")
##             Value Std. Error     z         p
## (Intercept) 8.450      0.276 30.56 3.60e-205
## method      0.712      0.406  1.76  7.92e-02
## Log(scale)  0.476      0.110  4.35  1.39e-05
##
## Scale= 1.61
##
## Weibull distribution
## Loglik(model)= -609.2   Loglik(intercept only)= -610.8
##  Chisq= 3.17 on 1 degrees of freedom, p= 0.075
## Number of Newton-Raphson Iterations: 5
## n= 124
```

```
fit.weib1 <- survreg(Surv(days, dead) ~ method + age, data = dialysis,
                     dist = "weibull")
summary(fit.weib1)
```

```
##
## Call:
## survreg(formula = Surv(days, dead) ~ method + age, data = dialysis,
##     dist = "weibull")
##                Value Std. Error     z        p
## (Intercept) 12.1053     0.6991 17.32 3.62e-67
## method       1.3185     0.3354  3.93 8.45e-05
## age         -0.0774     0.0115 -6.73 1.69e-11
## Log(scale)   0.2571     0.1045  2.46 1.39e-02
##
## Scale= 1.29
##
## Weibull distribution
```

```
## Loglik(model)= -581.3   Loglik(intercept only)= -610.8
##  Chisq= 58.98 on 2 degrees of freedom, p= 1.6e-13
## Number of Newton-Raphson Iterations: 5
## n= 124
```

Comparing these models side-by-side we find:

1. The coefficient for method increases for 0.71 to 1.32. These correspond to prolonging the time to death by factors of $\exp(0.71) = 2.03$ and $\exp(1.32) = 3.74$ respectively; thus indicating that APD is associated with better prognosis. As per the Cox regression model, the P-value for method goes from being marginally significant ($P = 0.079$) to highly significant ($P < 0.001$).

2. The intercepts are markedly different: 8.45 and 12.11 respectively. However, this is **not** unexpected. The intercept for the second model is based on a patient aged 0-years old! Therefore it is not directly comparable to the intercept in the first model. If we multiply the coefficient by the mean patient age, we obtain a more comparable intercept of 8.13. One way of getting around this would have been to replace `age` with `I(age - mean(age))` in the model formula;[3] this uses the mean centered age $\texttt{age}_i - \overline{\texttt{age}}$ for each subject so that the intercept is interpretable for a subject with mean age.

3. The `scale` parameters were estimated as 1.6 and 1.3 respectively. These correspond to shape parameters p of 0.6 and 0.8 respectively, both suggesting a decreasing hazard function. Each model gave strong evidence to reject the special case of the exponential model (i.e. $p = 1$).

4. The coefficient for age in the second model is -0.077, which corresponds to prolonging the time to death by a factor of $\exp(-0.077) = 0.93$ for each additional year of age. Or to phrase it better, time to death accelerates by a factor of $\exp(+0.077) = 1.08$ for each additional year of age.

5. We can compare the Weibull model *hazard ratios* to the Cox PH model hazard ratios using the relationship: $\beta = -p\alpha$. For example, the log hazard ratio for treatment method as estimate by the Weibull model is $-1.3185/1.29 = -1.02$, which is comparable to the Cox model estimate of -0.99.

We should note, however, that the proportional hazards assumption has also been shown to be unsupported, and therefore the standard Weibull model is also inappropriate. One option we could use would be to fit a *stratified* Weibull model. In this case we would estimate separate scale parameters for each stratum.

[3] The `I()` function, or *inhibit interpretation function* is used to tell R to apply the transformation **before** trying to fit the model.

7 Problem 3: Parametric regression

Fit a Weibull survival model to the `veteran` data and interpret the output.

8 Problem 4: A case study

8.1 The `pneumon` data

Data[4] gathered from 3,470 annual personal interviews conducted for the National Longitudinal Survey of Youth (NLSY, 1995) from 1979 through 1986 was used to study whether the mother's feeding choice (breast feeding vs. never breast fed) protected the infant against hospitalized pneumonia in the first year of life. Information obtained about the child included whether it had a normal birth-weight, as defined by weighing at least 5.5 pounds (36%), race (56% white, 28% black, and 16% other), number of siblings (range 0–6), and age at which the child was hospitalized for pneumonia, along with an indicator variable as to whether the child was hospitalized. Demographic characteristics of the mother, such as age (average is 21.64 years with a range of 14–29 years), years of schooling (average of 11.4 years), region of the country (15% Northeast, 25% North-central, 40% South, and 20% West), poverty (92%), and urban environment (76%). Health behavior measures during pregnancy, such as alcohol use (36%) and cigarette use (34%), were also recorded.

This dataset was presented by Klein and Moeschberger (1997; §1.13); see book for further details. The text above has been copied verbatim. The file is saved as `pneumon.data` on the USB flash drive, but can also be accessed (if you have internet access) using

```
install.packages("KMsurv")
data(pneumon, package = "KMsurv") # Now saved in R workspace
```

8.2 Exercises

1. An investigator is interested in assessing the association between race and poverty status on time to hospitalization of pneumonia.

    a. Use non-parametric survival estimators and suitable tests to contrast the data for each variable of interest.
    b. Fit a Cox PH model to the data, reporting the hazard ratios and 95% confidence intervals. Interpret the estimates.

2. Ages of young children at the time they were hospitalized with pneumonia were recorded as well as the observed ages of those infants that were not hospitalized with pneumonia during the study period.

[4] This description is copied verbatim from Klein and Moeschberger (1997; §1.13)

a. Test the hypothesis that the survival functions for the two types of breast feeding are equal using the log-rank test.
b. From question 1, estimate an *adjusted* hazard ratio for breast feeding taking into account race and poverty status. Are the *unadjusted* and *adjusted* inferences consistent?
c. Also available in the data set is information on other factors that may be associated with the timing of hospitalized pneumonia. These factors are age of the mother at the infant's birth, rural-urban environment of the mother, use of alcohol by the mother (no drinks, less than one drink, 1–2 drinks, 3–4 drinks, or more than 4 drinks per month), mother's cigarette use (none, less than 1 pack/day, 1 or more pack/day), region of country (northeast, north central, south, or west), birth-weight of infant (less the 5.5 lbs or 5.5 lbs or more), poverty status of mother (yes/no), race of mother (white, black, or other), or number of siblings of infant. Test the hypothesis that the times to hospitalized pneumonia are the same for the two feeding groups adjusting for these factors in a separate model.

3. Evaluate the proportional hazards assumption in the model fitted above. In particular, examine whether proportional hazards is satisfied for the feeding type.

4. Fit a Weibull regression model to the data.

   a. Interpret the coefficient for feeding type.
   b. Is a simpler exponential model supported?

9 References

- Diggle PJ and Chetwynd AG (2011). Statistics and Scientific Method: an Introduction for Students and Researchers. Oxford: *Oxford University Press*.
- D Kalbfleisch and RL Prentice (1980), The Statistical Analysis of Failure Time Data. Wiley, New York.
- Grambsch PM, Therneau TM (1994), Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81, 515-26.
- National Longitudinal Survey of Youth. NLS Handbook. Center for Human Resource Research. The Ohio State University, Columbus, Ohio, 1995.