```
%weibdiag(SURVDATA=ovdata, TIME=survtime, STATUS=status,
    VSTATUS=0, XVARS=age treat, NXVARS=2, RSCORE=eta,
    SIGMA=0.5489, OUTVALS=results);

proc print data=results;
run;
```

Again, plots constructed from the diagnostics produced by this macro can easily be obtained.

12.4 Further reading

Allison (1995) provides a comprehensive guide to the SAS software for survival analysis. Der and Everitt (2001) also include material on survival analysis in their text on the use of SAS in data analysis. Therneau and Grambsch (2000) give a detailed account of how SAS and S-PLUS are used to fit the Cox regression model, and extensions to it. This book includes a description of a number of SAS macros and S-PLUS functions that supplement the standard facilities available in these packages. The use of S-PLUS in survival analysis is described in Everitt and Rabe-Hesketh (2001). Venables and Ripley (2002) show how to use the S environment as a powerful and graphical data analysis system, implemented in S-PLUS, and include a chapter on survival analysis. The use of Stata is illustrated in Rabe-Hesketh and Everitt (2000).

David Collett Modelling Survival Data in Med. Res. 2nd Edition 2003

APPENDIX A

Maximum likelihood estimation

This appendix gives a summary of results on maximum likelihood estimation that are relevant to survival analysis. The results presented apply equally to inferences based on a partial likelihood function, and so can be used in conjunction with the Cox regression model described in Chapter 3, and the fully parametric models introduced in Chapters 5 and 6. A full treatment of the theory of maximum likelihood estimation and likelihood ratio testing is given by Cox and Hinkley (1974).

A.1 Inference about a single unknown parameter

Suppose that the likelihood of n observed survival times, $t_1, t_2, \ldots, t_n$, is a function of a single unknown parameter β, and denoted $L(\beta)$. The *maximum likelihood estimate* of β is then the value $\hat{\beta}$ for which this function is a maximum. In almost all applications, it is more convenient to work with the natural logarithm of the likelihood function, $\log L(\beta)$. The value $\hat{\beta}$, which maximises the log-likelihood, is the same value that maximises the likelihood function itself, and is generally found using differential calculus.

Specifically, $\hat{\beta}$ is the value of β for which the derivative of $\log L(\beta)$, with respect to β, is equal to zero. In other words, $\hat{\beta}$ is such that

$$\left.\frac{\mathrm{d}\log L(\beta)}{\mathrm{d}\beta}\right|_{\hat{\beta}} = 0.$$

The first derivative of $\log L(\beta)$ with respect to β is known as the *efficient score* for β, and is denoted $u(\beta)$. Therefore,

$$u(\beta) = \frac{\mathrm{d}\log L(\beta)}{\mathrm{d}\beta},$$

and so the maximum likelihood estimate of β, $\hat{\beta}$, satisfies the equation

$$u(\hat{\beta}) = 0.$$

The asymptotic variance of the maximum likelihood estimate of β can be found from

$$\left(-\mathrm{E}\left\{\frac{\mathrm{d}^2\log L(\beta)}{\mathrm{d}\beta^2}\right\}\right)^{-1} \tag{A.1}$$

or from the equivalent formula,

$$\left(\mathrm{E}\left\{\frac{\mathrm{d}\log L(\beta)}{\mathrm{d}\beta}\right\}^2\right)^{-1}.$$

The variance calculated from either of these expressions can be regarded as the approximate variance of $\hat{\beta}$, although it is usually more straightforward to use expression (A.1). When the expected value of the derivative in expression (A.1) is difficult to obtain, a further approximation to the variance of $\hat{\beta}$ is found by evaluating the derivative at $\hat{\beta}$. The approximate variance of $\hat{\beta}$ is then given by

$$\operatorname{var}(\hat{\beta}) \approx -\left(\frac{\mathrm{d}^2\log L(\beta)}{\mathrm{d}\beta^2}\right)^{-1}\Bigg|_{\hat{\beta}}. \tag{A.2}$$

The second derivative of the log-likelihood function is sometimes known as the *Hessian*, and the quantity

$$-\mathrm{E}\left\{\frac{\mathrm{d}^2\log L(\beta)}{\mathrm{d}\beta^2}\right\}$$

is called the *information function*. Since the information function is formed from the expected value of the second derivative of $\log L(\beta)$, it is sometimes called the *expected information function*. In contrast, the negative second derivative of the log-likelihood function itself is called the *observed information function*. This latter quantity will be denoted $i(\beta)$, so that

$$i(\beta) = -\left\{\frac{\mathrm{d}^2\log L(\beta)}{\mathrm{d}\beta^2}\right\}.$$

The reciprocal of this function, evaluated at $\hat{\beta}$, is then the approximate variance of $\hat{\beta}$ given in equation (A.2), that is,

$$\operatorname{var}(\hat{\beta}) \approx \frac{1}{i(\hat{\beta})}.$$

The standard error of $\hat{\beta}$, that is, the square root of the estimated variance of $\hat{\beta}$, is found from

$$\operatorname{se}(\hat{\beta}) = \frac{1}{\sqrt{\{i(\hat{\beta})\}}}.$$

This standard error can be used to construct confidence intervals for β.

In order to test the null hypothesis that $\beta = 0$, three alternative test statistics can be used. The *likelihood ratio test statistic* is the difference between the values of $-2\log L(\hat{\beta})$ and $-2\log L(0)$. The *Wald test* is based on the statistic

$$\hat{\beta}^2 i(\hat{\beta}),$$

and the *score test statistic* is

$$\frac{\{u(0)\}^2}{i(0)}.$$

Each of these statistics has an asymptotic chi-squared distribution on 1 d.f.,

under the null hypothesis that $\beta = 0$. Note that the Wald statistic is equivalent to the statistic

$$\frac{\hat{\beta}}{\operatorname{se}(\hat{\beta})},$$

which has an asymptotic standard normal distribution.

A.2 Inference about a vector of unknown parameters

The results in Section A.1 can be extended to the situation where n observations are used to estimate the values of p unknown parameters, $\beta_1, \beta_2, \ldots, \beta_p$. These parameters can be assembled into a p-component vector, $\boldsymbol{\beta}$, and the corresponding likelihood function is $L(\boldsymbol{\beta})$. The maximum likelihood estimates of the p unknown parameters are the values $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$, which maximise $L(\boldsymbol{\beta})$. They are therefore found by solving the p equations

$$\frac{\mathrm{d}\log L(\boldsymbol{\beta})}{\mathrm{d}\beta_j}\bigg|_{\hat{\boldsymbol{\beta}}} = 0,$$

for $j = 1, 2, \ldots, p$, simultaneously.

The vector formed from $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ is denoted $\hat{\boldsymbol{\beta}}$, and so the maximised likelihood is $L(\hat{\boldsymbol{\beta}})$. The efficient score for β_j, $j = 1, 2, \ldots, p$, is

$$u(\beta_j) = \frac{\mathrm{d}\log L(\boldsymbol{\beta})}{\mathrm{d}\beta_j},$$

and these quantities can be assembled to give a p-component vector of efficient scores, denoted $\boldsymbol{u}(\boldsymbol{\beta})$. The vector of maximum likelihood estimates is therefore such that

$$\boldsymbol{u}(\hat{\boldsymbol{\beta}}) = \mathbf{0},$$

where $\mathbf{0}$ is the $p \times 1$ vector of zeroes.

Now let the matrix $\boldsymbol{H}(\boldsymbol{\beta})$ be the $p \times p$ matrix of second partial derivatives of the log-likelihood function, $\log L(\boldsymbol{\beta})$. The (j, k)th element of $\boldsymbol{H}(\boldsymbol{\beta})$ is then

$$\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial\beta_j \partial\beta_k},$$

for $j = 1, 2, \ldots, p$, $k = 1, 2, \ldots, p$, and $\boldsymbol{H}(\boldsymbol{\beta})$ is called the *Hessian matrix*. The matrix

$$\boldsymbol{I}(\boldsymbol{\beta}) = -\boldsymbol{H}(\boldsymbol{\beta})$$

is called the *observed information matrix*. The (j, k)th element of the corresponding *expected information matrix* is

$$-\mathrm{E}\left(\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial\beta_j \partial\beta_k}\right).$$

The variance-covariance matrix of the p maximum likelihood estimates, $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$, written $\operatorname{var}(\hat{\boldsymbol{\beta}})$, can then be approximated by the inverse of the observed information matrix, evaluated at $\hat{\boldsymbol{\beta}}$, so that

$$\operatorname{var}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{I}^{-1}(\hat{\boldsymbol{\beta}}).$$

The square root of the (j, j)th element of this matrix can be taken to be the standard error of $\hat{\beta}_j$, for $j = 1, 2, \ldots, p$.

The test statistics given in Section A.1 can be generalised to the multi-parameter situation. Consider the test of the null hypothesis that all the β-parameters in a fitted model are equal to zero. The likelihood ratio test statistic is the value of

$$2\left\{\log L(\hat{\boldsymbol{\beta}}) - \log L(\mathbf{0})\right\},$$

the Wald test is based on

$$\hat{\boldsymbol{\beta}}' \boldsymbol{I}(\hat{\boldsymbol{\beta}})\, \hat{\boldsymbol{\beta}},$$

and the score test statistic is

$$\boldsymbol{u}'(\mathbf{0})\boldsymbol{I}^{-1}(\mathbf{0})\boldsymbol{u}(\mathbf{0}).$$

Each of these statistics has a chi-squared distribution on p d.f. under the null hypothesis that $\boldsymbol{\beta} = \mathbf{0}$.

In comparing alternative models, interest centres on the hypothesis that some of the β-parameters in a model are equal to zero. To test this hypothesis, the likelihood ratio test is the most suitable, and so we only consider this procedure here. Suppose that a model that contains $p + q$ parameters, $\beta_1, \beta_2, \ldots, \beta_p, \beta_{p+1}, \ldots, \beta_{p+q}$, is to be compared with a model that only contains the p parameters $\beta_1, \beta_2, \ldots, \beta_p$. This amounts to testing the null hypothesis that the q parameters $\beta_{p+1}, \beta_{p+2}, \ldots, \beta_{p+q}$ in the model with $p + q$ unknown parameters are all equal to zero. Let $\hat{\boldsymbol{\beta}}_1$ denote the vector of estimates under the model with $p + q$ parameters and $\hat{\boldsymbol{\beta}}_2$ that for the model with just p parameters. The likelihood ratio test of the null hypothesis that $\beta_{p+1} = \beta_{p+2} = \cdots = \beta_{p+q} = 0$ in the model with $p + q$ parameters is then based on the statistic

$$2\left\{\log L(\hat{\boldsymbol{\beta}}_1) - \log L(\hat{\boldsymbol{\beta}}_2)\right\},$$

which has a chi-squared distribution on q d.f., under the null hypothesis. This test forms the basis for comparing alternative models, and was described in greater detail in Section 3.5 of Chapter 3.

APPENDIX B

Likelihood function for randomly censored data

Suppose that lifetime data for a sample of n individuals is a mixture of event times and right-censored observations. Denote the observed time for the ith individual by t_i, and let δ_i be the corresponding event indicator, $i = 1, 2, \ldots, n$, so that $\delta_i = 1$ if t_i is an event time, and $\delta_i = 0$ if the time is censored.

The random variable associated with the event time of the ith individual will be denoted by T_i. The censoring times will be assumed to be random, and C_i will denote the random variable associated with the time to censoring. The value t_i is then an observation on the random variable $\tau_i = \min(T_i, C_i)$. The density and survivor functions of T_i will be denoted by $f_{T_i}(t)$ and $S_{T_i}(t)$, respectively. Also, $f_{C_i}(t)$ and $S_{C_i}(t)$ will be used to denote the density and survivor functions of the random variable associated with the censoring time, C_i.

We now consider the probability distribution of the pair (τ_i, δ_i) for censored and uncensored observations, respectively. Consider first the case of a censored observation, so that $\delta_i = 0$. The joint distribution of τ_i and δ_i is described by

$$\mathrm{P}(\tau_i = t, \delta_i = 0) = \mathrm{P}(C_i = t, T_i > t).$$

This joint probability is a mixture of continuous and discrete components, but to simplify the presentation, $\mathrm{P}(T_i = t)$, for example, will be understood to be the probability density function of T_i. The distribution of the event time, T_i, is now assumed to be independent of that of the censoring time, C_i. Then,

$$\begin{aligned}\mathrm{P}(C_i = t, T_i > t) &= \mathrm{P}(C_i = t)\,\mathrm{P}(T_i > t),\\ &= f_{C_i}(t) S_{T_i}(t),\end{aligned}$$

so that

$$\mathrm{P}(\tau_i = t, \delta_i = 0) = f_{C_i}(t) S_{T_i}(t).$$

Similarly, for an uncensored observation,

$$\begin{aligned}\mathrm{P}(\tau_i = t, \delta_i = 1) &= \mathrm{P}(T_i = t, C_i > t),\\ &= \mathrm{P}(T_i = t)\,\mathrm{P}(C_i > t),\\ &= f_{T_i}(t) S_{C_i}(t),\end{aligned}$$

again assuming that the distributions of C_i and T_i are independent. Putting these two results together, the joint probability, or likelihood, of the n observations, $t_1, t_2, \ldots, t_n$, is therefore

$$\prod_{i=1}^{n} \{f_{T_i}(t_i) S_{C_i}(t_i)\}^{\delta_i} \{f_{C_i}(t_i) S_{T_i}(t_i)\}^{1-\delta_i},$$

which can be written as

$$\prod_{i=1}^{n} f_{C_i}(t_i)^{1-\delta_i} S_{C_i}(t_i)^{\delta_i} \times \prod_{i=1}^{n} f_{T_i}(t_i)^{\delta_i} S_{T_i}(t_i)^{1-\delta_i}.$$

On the assumption of non-informative censoring, the first product in this expression will not involve any parameters that are relevant to the distribution of the survival times, and so can be regarded as a constant. The likelihood of the observed data is therefore proportional to

$$\prod_{i=1}^{n} f_{T_i}(t_i)^{\delta_i} S_{T_i}(t_i)^{1-\delta_i},$$

which was given in expression (5.12) of Chapter 5.

It can also be shown that when the study has a fixed duration, so that individuals who have not experienced an event by the end of the study are censored, the same likelihood function is obtained. Details are not given here, but see Klein and Moeschberger (1997) or Lawless (2002), for example.

APPENDIX C

Standard error of percentiles

In this appendix, an expression for the standard error of the percentiles of the Weibull distribution is derived. More general results are given for the standard error of the pth percentile in the Weibull proportional hazards model and the general accelerated failure time model.

C.1 Standard error of a percentile of the Weibull distribution

In equation (5.25) of Chapter 5, the estimated pth percentile of the Weibull distribution with scale parameter λ and shape parameter γ was shown to be given by

$$\hat{t}(p) = \left[\frac{1}{\hat{\lambda}} \log\left(\frac{100}{100-p}\right)\right]^{1/\hat{\gamma}}.$$

The variance of $\hat{t}(p)$ is most easily found by first obtaining the variance of $\log \hat{t}(p)$. Now,

$$\log \hat{t}(p) = \frac{1}{\hat{\gamma}} \log\left\{\hat{\lambda}^{-1} \log\left(\frac{100}{100-p}\right)\right\},$$

and so

$$\log \hat{t}(p) = \frac{1}{\hat{\gamma}}\left\{c_p - \log \hat{\lambda}\right\},$$

where

$$c_p = \log\log\left(\frac{100}{100-p}\right).$$

This is a function of two parameter estimates, $\hat{\lambda}$ and $\hat{\gamma}$, and so we can use the result in equation (5.43) of Chapter 5 to obtain the approximate variance of $\log \hat{t}(p)$. From equation (5.43),

$$\operatorname{var}\{\log \hat{t}(p)\} \approx \left(\frac{\partial \log \hat{t}(p)}{\partial \hat{\lambda}}\right)^2 \operatorname{var}(\hat{\lambda}) + \left(\frac{\partial \log \hat{t}(p)}{\partial \hat{\gamma}}\right)^2 \operatorname{var}(\hat{\gamma}) + 2\frac{\partial \log \hat{t}(p)}{\partial \hat{\lambda}} \frac{\partial \log \hat{t}(p)}{\partial \hat{\gamma}} \operatorname{cov}(\hat{\lambda}, \hat{\gamma}).$$

Now, the derivatives of $\log \hat{t}(p)$ with respect to $\hat{\lambda}$ and $\hat{\gamma}$ are given by

$$\frac{\partial \log \hat{t}(p)}{\partial \hat{\lambda}} = -\frac{1}{\hat{\lambda}\hat{\gamma}},$$

$$\frac{\partial \log \hat{t}(p)}{\partial \hat{\gamma}} = -\frac{c_p - \log \hat{\lambda}}{\hat{\gamma}^2},$$

and so the approximate variance of $\log \hat{t}(p)$ is

$$\frac{1}{\hat{\lambda}^2\hat{\gamma}^2}\,\text{var}\,(\hat{\lambda}) + \frac{\left(c_p - \log\hat{\lambda}\right)^2}{\hat{\gamma}^4}\,\text{var}\,(\hat{\gamma}) + \frac{2\left(c_p - \log\hat{\lambda}\right)}{\hat{\lambda}\hat{\gamma}^3}\,\text{cov}\,(\hat{\lambda},\hat{\gamma}). \qquad (\text{C.1})$$

The variance of $\hat{t}(p)$ itself is found from the result in equation (2.9) of Chapter 2, from which

$$\text{var}\,\{\hat{t}(p)\} \approx \hat{t}(p)^2\,\text{var}\,\{\log\hat{t}(p)\}.$$

Therefore, from expression (C.1),

$$\text{var}\,\{\hat{t}(p)\} \approx \frac{\hat{t}(p)^2}{\hat{\lambda}^2\hat{\gamma}^4}\left\{\hat{\gamma}^2\,\text{var}\,(\hat{\lambda}) + \hat{\lambda}^2\left(c_p - \log\hat{\lambda}\right)^2\,\text{var}\,(\hat{\gamma}) + 2\hat{\lambda}\hat{\gamma}\left(c_p - \log\hat{\lambda}\right)\,\text{cov}\,(\hat{\lambda},\hat{\gamma})\right\},$$

and so the standard error of $\hat{t}(p)$ is the square root of this expression, that is,

$$\text{se}\,\{\hat{t}(p)\} = \frac{\hat{t}(p)}{\hat{\lambda}\hat{\gamma}^2}\left\{\hat{\gamma}^2\,\text{var}\,(\hat{\lambda}) + \hat{\lambda}^2\left(c_p - \log\hat{\lambda}\right)^2\,\text{var}\,(\hat{\gamma}) + 2\hat{\lambda}\hat{\gamma}\left(c_p - \log\hat{\lambda}\right)\,\text{cov}\,(\hat{\lambda},\hat{\gamma})\right\}^{\frac{1}{2}}. \qquad (\text{C.2})$$

A $100(1-\alpha)\%$ confidence interval for the true pth percentile is found from exponentiating the corresponding confidence limits for $\log t(p)$. These limits are

$$\log\hat{t}(p) \pm z_{\alpha/2}\,\text{se}\,\{\log\hat{t}(p)\},$$

where $\text{se}\,\{\log\hat{t}(p)\}$ is the square root of expression (C.1), and $z_{\alpha/2}$ is the upper $\alpha/2$-point of the standard normal distribution.

Note that for the special case of the exponential distribution, where the shape parameter, γ, is equal to unity, the standard error of the estimated pth percentile is

$$\frac{\hat{t}(p)}{\hat{\lambda}}\,\text{se}\,(\hat{\lambda}).$$

From equation (5.15) of Chapter 5,

$$\text{se}\,(\hat{\lambda}) = \hat{\lambda}/\sqrt{r},$$

where r is the number of death times in the data set, and so

$$\text{se}\,\{\hat{t}(p)\} = \hat{t}(p)/\sqrt{r},$$

as in equation (5.19).

C.2 Standard error of a percentile in the Weibull model

Using the parameterisation of the general Weibull proportional hazards model adopted in Section 5.5 and elsewhere, the estimated pth percentile for an individual for whom the values of the explanatory variables form the vector $\boldsymbol{x}$ is

$$\hat{t}(p) = \left\{\frac{1}{\hat{\lambda}\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x})}\log\left(\frac{100}{100-p}\right)\right\}^{1/\hat{\gamma}}.$$

Again, the variance of $\log\hat{t}(p)$ is found first, and writing

$$c_p = \log\log\left(\frac{100}{100-p}\right),$$

we get

$$\log\hat{t}(p) = \frac{1}{\hat{\gamma}}\left(c_p - \log\hat{\lambda} - \hat{\boldsymbol{\beta}}'\boldsymbol{x}\right). \qquad (\text{C.3})$$

This is a function of the $p+2$ parameter estimates $\hat{\lambda}$, $\hat{\gamma}$ and $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$, and the approximate variance of this function can be found using a further generalisation of the result in equation (5.43). This generalisation is best expressed in matrix form.

Suppose that $\hat{\boldsymbol{\theta}}$ is a vector formed from k parameter estimates, $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$, and that $g(\hat{\boldsymbol{\theta}})$ is a function of these k estimates. The approximate variance of $g(\hat{\boldsymbol{\theta}})$ is then given by

$$\text{var}\,\{g(\hat{\boldsymbol{\theta}})\} \approx \boldsymbol{d}(\hat{\boldsymbol{\theta}})'\,\text{var}\,(\hat{\boldsymbol{\theta}})\,\boldsymbol{d}(\hat{\boldsymbol{\theta}}), \qquad (\text{C.4})$$

where $\text{var}\,(\hat{\boldsymbol{\theta}})$ is the $k \times k$ variance-covariance matrix of the estimates in $\hat{\boldsymbol{\theta}}$, and $\boldsymbol{d}(\hat{\boldsymbol{\theta}})$ is the k-component vector whose ith element is

$$\left.\frac{\partial g(\boldsymbol{\theta})}{\partial\theta_i}\right|_{\hat{\boldsymbol{\theta}}},$$

for $i = 1, 2, \ldots, k$.

We now write $\boldsymbol{V}$ for the $(p+2)\times(p+2)$ variance-covariance matrix of $\hat{\lambda}$, $\hat{\gamma}$ and $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$. Next, from equation (C.3), the derivatives of $\log\hat{t}(p)$ with respect to these parameter estimates are

$$\frac{\partial\log\hat{t}(p)}{\partial\hat{\lambda}} = -\frac{1}{\hat{\lambda}\hat{\gamma}},$$

$$\frac{\partial\log\hat{t}(p)}{\partial\hat{\gamma}} = -\frac{c_p - \log\hat{\lambda} - \hat{\boldsymbol{\beta}}'\boldsymbol{x}}{\hat{\gamma}^2},$$

$$\frac{\partial\log\hat{t}(p)}{\partial\hat{\beta}_j} = -\frac{x_j}{\hat{\gamma}},$$

for $j = 1, 2, \ldots, p$, where x_j is the jth component of $\boldsymbol{x}$. The vector $\boldsymbol{d}(\hat{\boldsymbol{\theta}})$ in equation (C.4) can be expressed as $-\hat{\gamma}^{-1}\boldsymbol{d}_0$, where $\boldsymbol{d}_0$ is a vector with components $\hat{\lambda}^{-1}$, $\hat{\gamma}^{-1}\{c_p - \log\hat{\lambda} - \hat{\boldsymbol{\beta}}'\boldsymbol{x}\}$, and $x_1, x_2, \ldots, x_p$. Then, the standard error of $\log\hat{t}(p)$ is given by

$$\text{se}\,\{\log\hat{t}(p)\} = \hat{\gamma}^{-1}\sqrt{(\boldsymbol{d}_0'\boldsymbol{V}\boldsymbol{d}_0)}, \qquad (\text{C.5})$$

from which the standard error of $\hat{t}(p)$ itself is obtained using

$$\text{se}\{\hat{t}(p)\} = \hat{t}(p)\,\text{se}\{\log \hat{t}(p)\}. \qquad \text{(C.6)}$$

Notice that for the null model, that contains no explanatory variables, the standard error of $\log \hat{t}(p)$ in equation (C.5) reduces to the result in equation (C.2).

C.3 Standard error of a percentile in the AFT model

The general accelerated failure time model for the random variable associated with the survival time of the ith of n individuals, T_i, is such that

$$\log T_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_p x_{pi} + \sigma\epsilon_i,$$

where $x_{1i}, x_{2i}, \ldots, x_{pi}$ are the values of p explanatory variables in the model, μ is the intercept parameter, σ is the scale parameter, and ϵ_i has a particular probability distribution. This model was considered in detail in Chapter 6. On fitting this model, the estimated pth percentile can be expressed as

$$\hat{t}_i(p) = \exp\{\hat{\sigma}\epsilon_i(p) + \hat{\mu} + \hat{\alpha}_1 x_{1i} + \hat{\alpha}_2 x_{2i} + \cdots + \hat{\alpha}_p x_{pi}\},$$

where $\epsilon_i(p)$ is the pth percentile of the distribution of ϵ_i and $\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p, \hat{\sigma}$ are the estimated values of the parameters in the accelerated failure time model. See Table 6.2 of Chapter 6 for a summary of the possible forms of $\epsilon_i(p)$.

The estimated percentiles of the distribution of T_i are functions of the parameter estimates in the log-linear accelerated failure time model, and so the standard error of these estimates can be found using the general result in equation (C.4). Specifically, the vector $\hat{\boldsymbol{\theta}}$ now has $p+2$ components, namely $\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p, \hat{\sigma}$, and var $(\hat{\boldsymbol{\theta}})$ is the variance-covariance matrix of these parameter estimates. Equation (C.4) shows how the variance of a function of the parameter estimates can be obtained from the vector of derivatives, $\boldsymbol{d}(\hat{\boldsymbol{\theta}})$, of the estimated percentiles. However, it is much more straightforward to first obtain the variance of $\log \hat{t}_i(p)$, since the derivatives of $\log \hat{t}_i(p)$, with respect to $\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p, \hat{\sigma}$, are $1, x_{1i}, x_{2i}, \ldots, x_{pi}, \epsilon_i(p)$, respectively. Equation (C.6) is then used to obtain the standard error of the estimated percentile. Confidence intervals for a percentile will usually be formed from an interval estimate for $\log t_i(p)$. Note that most computer software for survival analysis can provide the standard error of specified percentiles.

APPENDIX D

Additional data sets

This appendix contains a number of data sets, together with some suggestions for analyses that could be carried out. These data sets may be downloaded from the publishers's or author's web site, at the location given in the preface.

D.1 Chronic active hepatitis

In a clinical trial described by Kirk *et al.* (1980), 44 patients with chronic active hepatitis were randomised to the drug prednisolone, or an untreated control group. The survival time of the patients, in months, following admission to the trial, was the response variable of interest. The data, which were given in Pocock (1983), are shown in Table D.1, in which an asterisk denotes a censored survival time.

Table D.1 *Survival times of patients suffering from chronic active hepatitis.*

Prednisolone		Control	
2	131*	2	41
6	140*	3	54
12	141*	4	61
54	143	7	63
56*	145*	10	71
68	146	22	127*
89	148*	28	140*
96	162*	29	146*
96	168	32	158*
125*	173*	37	167*
128*	181*	40	182*

Summarise the data in terms of the estimated survivor function for each treatment group. Compare the groups using the log-rank and Wilcoxon tests. Fit Cox and Weibull proportional hazards model to determine the significance of the treatment effect. Compare the results from these different analyses in terms of the significance of the treatment effect, and, for the model-based analyses, the estimated hazard ratio and corresponding 95% confidence limits. Obtain a log-cumulative hazard plot of the data, and comment on which method of analysis is the most appropriate.