Ansell & Phillips

Mean square error

The mean square error of an estimator $\hat{\beta}$ is the expectation of the square of the difference between $\hat{\beta}$ and the parameter β. It measures how 'far away' the estimator is from the parameter β 'on average'.

2.6.2 CLASSICAL METHODS FOR THE DETERMINATION OF ESTIMATORS

One of the oldest methods of determining estimators is the *method of moments* (associated with K. Pearson). It is necessary to give two definitions for *moments* to understand this method.

Population moments

The ith *population moment* of the random variable T, say, is defined to be $E(T^i)$. So the expectation μ is the first population moment and the variance σ^2 is defined in terms of the second population moment.

Sample moments

The ith *sample moment* for a sample $t_1, t_2, ..., t_n$ of size n is defined to be

$$\tilde{\mu}_i = \sum_{j=1}^{n} t_j^i / n.$$

(So the first sample moment $\tilde{\mu}_1 = \bar{t}$, the sample mean.)

With these two definitions it is now possible to outline the method.

Method of moments

If there are k parameters to be estimated then the first k population moments are equated to the corresponding sample moments. The solutions of the k equations give the estimators of the parameters, which will be functions of the sample moments. These estimators will generally be consistent and though their efficiency is often less than one they may be used as first approximations from which more efficient estimates may be obtained.

Example 2.6.1

For the exponential parametric distribution, given in Subsection 2.3.1, the expected value is $1/\lambda$. Equating this expectation to the first sample moment, $\bar{t}$, so that $1/\hat{\lambda} = \bar{t}$, the estimator $\hat{\lambda}$ for λ is given by

$$\hat{\lambda} = \frac{1}{\bar{t}},$$

which is the inverse of the sample mean of the observed failure times.

Using the example of failure times for the 20 pressure vessels presented in Subsection 1.4.1 the estimate of λ is $1/575.3 = 0.00174$.

□

It is often convenient for simplicity to consider only those estimators which are linear functions of the sample observations.

Best linear unbiased estimators

Among those unbiased estimators which are linear functions of the sample observations the *best linear unbiased estimator* is the estimator with minimum variance.

Example 2.6.2

For the binomial parametric distribution, given in Examples 1.6.1 and 1.7.2, the parameter p is the probability that a system fails. Let X_i be

$$X_i = \begin{cases} 1, & \text{if failure is observed,} \\ 0, & \text{otherwise.} \end{cases}$$

Then for a sample of size n a linear estimator is given by

$$\hat{p} = \sum_{j=1}^{n} c_j X_j,$$

where the c_j are suitably chosen constants. If this estimator is unbiased then $\sum_{j=1}^{n} c_j = 1$. The variance is given by $pq \sum_{j=1}^{n} c_j^2$ and it can be shown that this variance is minimized if $c_j = 1/n$, for $j = 1, \ldots, n$. Hence the best linear unbiased estimator is given by

$$\hat{p} = \bar{X}.$$

This is a point estimator.

Again using the example of failure times for the 20 pressure vessels presented in Subsection 1.4.1, consider finding the estimate of p, the probability that the failure time of a pressure vessel is less than 100 hours. Hence the point estimate is $5/20 = 0.25$.

□

Least squares estimators

The *least squares* estimator of a parameter β is the estimator which minimizes the sum of the squares of the differences between the observations of a sample and their expectations, which are a function of the parameter β. There are two reasons for using squares of the differences. Firstly, the square is a function which is non-negative and when minimized cannot be smaller than zero. Secondly, this will imply that large differences between an observation and its expectation will be penalized. Usually each observation is equally weighted, but there are often good reasons for using different weights, which will then give a *weighted least squares* estimator.

Example 2.6.3

For the binomial parametric distribution, considered in Example 2.6.2, the parameter p is the probability that a system fails. Then $E(X_i) = p$ and the least squares estimator of p will minimize

$$\sum_{i=1}^{n}(x_i - p)^2.$$

Hence, by differentiation by p, the least squares estimator is also found to be given by

$$\hat{p} = \bar{x}.$$

□

As will be shown in Subsection 2.6.3, when likelihood methods are considered, the least squares estimator is also the maximum likelihood estimator when a random sample is taken from a *normal population.*

2.6.3 LIKELIHOOD METHODS

The second method of determining estimators which is widely used is the *likelihood method*, which was propounded by R.A. Fisher. This method uses the *likelihood function*, which is the probability density function (probability function) of the joint distribution of the observations of the sample for a continuous (discrete) random variable, considering this function as a function of the parameter $\boldsymbol{\beta}$. In the case of a random sample from a random variable T with probability density function $f_T(t;\boldsymbol{\beta})$, say, where the observations are IID, the likelihood function $L(\boldsymbol{\beta})$ is given by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f_T(t_i;\boldsymbol{\beta}).$$

The principle of maximum likelihood

An estimator is obtained using the *principle of maximum likelihood* by choosing the value of $\boldsymbol{\beta}$ which maximizes the likelihood function $L(\boldsymbol{\beta})$. So for a fixed sample the *maximum likelihood estimator* maximizes the probability of observing that sample over all possible values of the parameter.

If the sample is from a normal population then using the probability density function given in Subsection 2.3.4 gives a likelihood function which is the exponential function of minus the sum of squares of the differences between the observations and their expectations. Hence the maximum likelihood estimator and the least squares estimator for any parameter used to model the expectation are equivalent in the case of a normal population.

Example 2.6.4

For the exponential parametric distribution, given in Subsection 2.3.1, the likelihood function for a sample of size n is

$$L(\lambda) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} t_i).$$

When this likelihood is maximized as a function of λ this gives the estimator $\hat{\lambda} = 1/\bar{t}$, as was obtained in Example 2.6.1, by the method of moments. So using the example of failure times for the 20 pressure vessels presented in Subsection 1.4.1 the estimate of λ is again $1/575.3 = 0.00174$.

□

The maximization of the likelihood function is often accomplished by using differential calculus to obtain appropriate equations which can then be solved. This maximization can often be more easily performed on the logarithm of the likelihood (*log likelihood*) function and this will be denoted by

$$l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}).$$

Example 2.6.4 assumed that all the observations were failure times. However, as was illustrated in Section 2.4 many practical situations result in the collection of right censored observations. Hence the likelihood function given above for a complete random sample must be modified by replacing the probability density function by the reliability function for the right censored observations. Using the log likelihood function and the censoring indicator d_i introduced in Subsection 2.4.1 gives

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} d_i \log f_T(t_i;\boldsymbol{\beta}) + \sum_{i=1}^{n}(1 - d_i)\log R_T(t_i;\boldsymbol{\beta}).$$

Example 2.6.5

Consider again the example of Newton (1991) which was used in Example 2.5.3 to illustrate the estimation of the reliability function for component failure times by non-parametric methods. Suppose a parametric model is to be used to model the component failure time distribution in order to investigate the behaviour of the hazard function. If the Weibull distribution defined in Subsection 2.3.2 is chosen as a model then the log likelihood function is given by

$$l(\kappa,\theta) = d\log\kappa + (\kappa - 1)\sum_{i=1}^{n} d_i \log(t_i) - \sum_{i=1}^{n}(t_i/\theta)^{\kappa} - d\kappa\log\theta,$$

for $d = \sum_{i=1}^{n} d_i$ failure times and $n - d$ censored times. The partial differentials of this log likelihood function are given by

$$\begin{aligned}\frac{\partial l}{\partial \theta} &= \frac{\kappa \sum_{i=1}^{n} t_i^{\kappa}}{\theta^{\kappa+1}} - \frac{d\kappa}{\theta} \\ \frac{\partial l}{\partial \kappa} &= \frac{d}{\kappa} + \sum_{i=1}^{n} d_i \log(t_i) - \frac{\sum_{i=1}^{n} t_i^{\kappa} \log(t_i/\theta)}{\theta^{\kappa}} - d\log\theta.\end{aligned}$$

The maximum likelihood estimators are the solutions of the two equations obtained by equating the two partial derivatives to zero. Then the parameter estimator $\hat{\kappa}$ is the solution of the equation in κ given by

$$\frac{\sum_{i=1}^{n} t_i^{\kappa} \log(t_i)}{\sum_{i=1}^{n} t_i^{\kappa}} - \frac{1}{\kappa} - \frac{\sum_{i=1}^{n} d_i \log(t_i)}{d} = 0.$$

Hence, by solving the above equation numerically, the shape parameter κ is estimated as $\hat{\kappa} = 1.662$ using the maximum likelihood estimation method for Example 2.5.3. The scale parameter, or characteristic life, estimator $\hat{\theta}$ is given by

$$\hat{\theta} = \left(\frac{\sum_{i=1}^{n} t_i^{\hat{\kappa}}}{d}\right)^{\frac{1}{\hat{\kappa}}},$$

and hence the characteristic life is estimated as $\hat{\theta} = 619.2$. These equations for the Weibull distribution were first given by Cohen (1965).

As the hazard function of a Weibull distribution is $\kappa t^{\kappa-1}/\theta^{\kappa}$ and κ is estimated to be greater than unity this analysis leads to the conclusion that the hazard function is an increasing function of t.

□

So far only point estimates have been considered. However, interval estimators give an interval and a confidence level for that interval which indicate the precision of the estimator. This is achieved by calculating a standard error for the estimator, which can be done using large sample results with the likelihood function. Theoretical results for the maximum likelihood estimation of the parameters $\boldsymbol{\beta}$ of a model show that in the limit as the sample size n gets large the distribution of the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ is a normal distribution. If there are r parameters then this limiting (asymptotic) distribution is a multivariate normal distribution. The multivariate normal distribution like the univariate normal distribution defined in Subsection 2.3.4 is defined using the first two population moments and this is done using a vector of expectations $\boldsymbol{\mu}$ and a variance covariance matrix $\mathbf{V}$, which contains the variances in the main diagonal and the covariances in the off-diagonal entries. (The covariance is defined

for any pair of random variables as the difference between the expectation of the product of the two random variables and the product of the expectations of these two random variables. The covariance is zero for two *independent* random variables.)

To obtain the variance covariance matrix for $\hat{\boldsymbol{\beta}}$, minus the expectation of the second partial derivatives of the log likelihood function is used to obtain a matrix, $\mathbf{I}_f$, which is known as the *Fisher information* matrix. Under mild regularity conditions the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ will have a multivariate normal distribution with expectations $\boldsymbol{\beta}$ and variance covariance matrix given by $\mathbf{I}_f^{-1}$, the inverse of the Fisher information matrix, in the limit as the sample size n increases to infinity. Hence this result can be used for large sample sizes. In practice sample sizes do not have to be very large for this result to give a good approximation providing the proportion of censored observations is small. To use this result it is necessary to evaluate the expectations of the second derivatives of the log likelihood function. Fortunately in practice it is usually sufficient to use the value of the second derivative evaluated at the values observed in the sample, replacing any unknown parameters by their maximum likelihood estimates. The matrix $\mathbf{I}_o$ containing these second derivatives is known as the *observed information* matrix. The square roots of the entries in the main diagonal of $\mathbf{I}_o^{-1}$, the inverse of the observed information matrix, can then be used to give standard errors of the estimators. Also because the distribution is multivariate normal then a covariance of zero for any pair of estimators implies that these estimators are *independently* distributed. (*Note*: This last result is only true because the distributions are normal. It is **not generally** true that zero covariance is equivalent to statistical independence.)

Example 2.6.5 (continued)
Continuing the Weibull distribution example the second derivatives of the log likelihood are

$$\begin{aligned}\frac{\partial^2 l}{\partial \theta^2} &= -\frac{\kappa(\kappa+1)\sum_{i=1}^{n} t_i^{\kappa}}{\theta^{\kappa+2}} + \frac{d\kappa}{\theta^2} \\ \frac{\partial^2 l}{\partial \theta \partial \kappa} &= \frac{\sum_{i=1}^{n} t_i^{\kappa}}{\theta^{\kappa+1}} + \frac{\kappa \sum_{i=1}^{n} t_i^{\kappa} \log(t_i/\theta)}{\theta^{\kappa+1}} - \frac{d}{\theta} \\ \frac{\partial^2 l}{\partial \kappa^2} &= -\frac{d}{\kappa^2} - \frac{\sum_{i=1}^{n} t_i^{\kappa} \log(t_i/\theta)^2}{\theta^{\kappa}}.\end{aligned}$$

Hence, on substituting the observed sample values and replacing the parameters with their maximum likelihood estimates, the observed information matrix is

$$\begin{pmatrix} 0.000115 & -0.002726 \\ -0.002726 & 9.038249 \end{pmatrix}.$$

Hence, by inverting this matrix, and taking the square roots of the entries in the main diagonal, the standard errors of $\hat{\kappa}$ and $\hat{\theta}$ are 0.3338 and 93.59, respectively.

With these standard errors it is possible to construct 95% confidence intervals. These are (1.008, 2.317) for κ and (435.8, 802.7) for θ, respectively.

□

2.6.4 BAYESIAN METHODS

The third method of estimation is the *Bayesian* method due to the followers of T. Bayes. This method is philosophically different from the previous two methods as it considers the parameter β as the realization of a random variable B, say. The distribution of this random variable is known as the *prior* distribution. Advocates of Bayesian methods combine the prior distribution with the sample observations to produce the *posterior* distribution by using a procedure which is now referred to as *Bayes' rule.*

Bayes' rule

The probability density function $f_{B|\mathbf{T}}(\beta \mid \mathbf{T})$ of the *posterior* distribution of β obtained from the sample of observations $\mathbf{T}$ and the probability density function $f_B(\beta)$ of the *prior* distribution of B is given by

$$f_{B|\mathbf{T}}(\beta \mid \mathbf{T}) = \frac{f_{\mathbf{T}|B}(\mathbf{T} \mid \beta) f_B(\beta)}{f_{\mathbf{T}}(\mathbf{T})},$$

where

$$f_{\mathbf{T}}(\mathbf{T}) = \int_{-\infty}^{\infty} f_{\mathbf{T}|B}(\mathbf{T} \mid \beta) f_B(\beta) d\beta.$$

Bayes' rule produces a distribution for the parameter β, not a point estimator. Thus probability statements can be made about β using the posterior distribution, usually by constructing Bayesian confidence *(credible)* intervals. If a single value is required to give a point estimator then the *mode* of the distribution is often used, though the expected value could be used.

Example 2.6.6

A random sample of observations $t_1, t_2, ..., t_n$ was obtained from a population of failure times with an exponential distribution with parameter λ, as defined in Subsection 2.3.1. A *prior* distribution for λ is chosen to be the gamma distribution, as defined in Subsection 2.3.1, with *shape* parameter α but with *scale* parameter a rather than λ. This prior distribution will have a mode equal to $(\alpha - 1)/a$.

Applying Bayes' rule to this problem gives a *posterior* distribution λ which is also gamma but with shape parameter $\alpha + n$ and scale parameter $a + n\bar{t}$, where $\bar{t}$ is the sample (arithmetic) mean of the observations. So the effect of the information from the observations is to change the parameters of the *prior* gamma distribution. (However, if **no** observations are taken then there is **no** change from the *prior* distribution.) If the mode of the *posterior* distribution is used to give a single value for λ then the value is

$$\frac{\alpha - 1 + n}{a + n\bar{t}}.$$

This mode takes the value $(\alpha - 1)/a$ when $n = 0$ and approaches $1/\bar{t}$ when n is large, which is the answer obtained for the estimator of λ in both Examples 2.6.1 and 2.6.4. For the exponential distribution the reciprocal of λ is the expected value. So considering the reciprocals, the analysis using Bayes' rule is consistent with the view that as a large amount of information is obtained from the sample the 'best' value for the *population* mean (expected value, $1/\lambda$) is the *sample* mean (arithmetic mean of the sample, $\bar{t}$).

Using the example of failure times for the 20 pressure vessels presented in Subsection 1.4.1 with $\alpha = 2$ and $a = 400$, this gives a mode for the prior distribution of λ equal to 0.0025, which is equivalent to an expectation of 400. The mode of the posterior is equal to 0.00176, which is equivalent to an expectation of 567.2. Hence the estimate of the expectation obtained by the use of the prior distribution, corresponding to a low expectation of 400, is slightly less than 575.3, which is the estimate which would be obtained from the estimate of λ given in Examples 2.6.1 and 2.6.4.

□

2.7 Hypothesis testing and goodness-of-fit tests

Another objective of the reliability study may be to decide, when using a parametric model with parameters $\boldsymbol{\beta}$, whether the reliability data, which are the observations collected (in a possibly censored or truncated form) from a random sample, are consistent with the parameters $\boldsymbol{\beta}$ having the value $\boldsymbol{\beta}_0$, say. *Inference* is the science which provides methods for answering such questions. The *classical* method is to use the approach of *hypothesis testing* due to J. Neyman and E.S. Pearson. However, there is also an approach based on the use of *likelihood.* The parameter $\boldsymbol{\beta}$ will be considered as a single variable for ease of explanation. The ideas may be generalized to more variables.

2.7.1 HYPOTHESIS TESTS

The simplest form of the hypothesis test is the situation where there is a choice between two simple hypotheses or two values of β, say β_0 and β_1. The preferred value of these two, usually denoted by β_0, is known as the *null* hypothesis while the other value, β_1, is known as the *alternative* hypothesis. The aim of the test is to decide whether the observed data are consistent with the null hypothesis (and so should be *accepted*) or whether the null hypothesis should be *rejected* (and hence the alternative hypothesis be accepted). In performing the test it is possible to make two types of *error*, namely rejecting the null hypothesis when it is true (known as a *Type I error*) and secondly accepting the null hypothesis when the alternative hypothesis is true (known as a *Type II error*). The probability of an error being made is known as the *size*. It would be nice to minimize the sizes of both these errors but in general it is necessary to trade one off against the other.

The quantity given by one minus the size of the Type II error is known as the *power* of the test. So if there is a choice between two tests which have the same size of the Type I error (called the *significance level of the test*) then the test with the larger power would be chosen as it will have the smaller size of the Type II error. So power is a criterion for choosing between tests and ideally the *most powerful* test would be preferable. Neyman and Pearson provided a lemma for constructing the most powerful test.

Neyman–Pearson lemma
In testing two simple hypotheses on the basis of a sample of observations $\mathbf{X}$, say, of size n, say, the test is defined as rejecting the null hypothesis for one set of observations (known as the *critical region*) and accepting the null hypothesis otherwise. The Neyman–Pearson lemma gives a method of constructing the *critical region* for the *most powerful* test for testing the *null* hypothesis against the *alternative* hypothesis amongst all tests with a given *significance level* and *sample size*.

The simple hypotheses situation is obviously a very idealized situation. However, it is possible to extend the ideas of hypothesis testing to *composite* hypotheses, which are defined by using intervals, rather than single values, for the values of the parameter β. The concept of *power* is extended to the *power function*, which is the probability that $\mathbf{X}$ belongs to the *critical region* (so that the null hypothesis is rejected), given a certain hypothesis is true. This is a function of the value of the parameter β and so a 'good' test would have low values when β takes values that define the null hypothesis and high values when β takes values that define the alternative hypothesis. The significance level of the test is defined as the maximum of the significance levels for tests with a null hypothesis defined by values of β which define

the *composite* null hypothesis. However, there are difficulties. It may not be possible to discriminate between two tests (as was done for simple hypotheses) by deciding that one test is 'uniformly more powerful' than the other.

When using a hypothesis test it is usual to quote the significance level and this is commonly referred to as the 'p-value'. It is conventional to reject the null hypothesis if this p-value is less than 0.05.

Example 2.7.1
Consider carrying out a hypothesis test of a particular value of $\lambda = \lambda_0$, say, for the exponential parametric distribution, given in Subsection 2.3.1, against an alternative hypothesis that $\lambda = \lambda_1$, where $\lambda_0 > \lambda_1$. In Example 2.6.4 it was shown that

$$\prod_{i=1}^{n} f_T(t_i;\lambda) = \lambda^n \exp(-n\lambda/\hat{\lambda}),$$

where $\hat{\lambda} = 1/\bar{t}$ is the maximum likelihood estimator of λ. The Neyman–Pearson lemma defines the critical region as

$$\frac{\prod_{i=1}^{n} f_T(t_i;\lambda_0)}{\prod_{i=1}^{n} f_T(t_i;\lambda_1)} = \frac{\lambda_0^n \exp(-n\lambda_0/\hat{\lambda})}{\lambda_1^n \exp(-n\lambda_1/\hat{\lambda})} < C,$$

which rearranges to give

$$\hat{\lambda} < \frac{n(\lambda_0 - \lambda_1)}{\log((\lambda_0/\lambda_1)^n C)},$$

for a suitable value of C. The value of C is determined by the size of the Type I error, conventionally chosen as 0.05, by using the distribution of $\hat{\lambda}$. Under the null that $\lambda = \lambda_0$ the distribution of $2n\lambda_0/\hat{\lambda}$ is chi-square with $2n$ degrees of freedom (χ^2_{2n}).

Using the example of failure times for the 20 pressure vessels presented in Subsection 1.4.1, consider testing the null hypothesis $\lambda_0 = 0.0025$. Then, using the chi-square distribution with 40 degrees of freedom (χ^2_{40}), the critical region is defined by

$$\hat{\lambda} < 0.00182.$$

Hence as the estimate of λ is 0.00174 the null hypothesis is rejected in favour of the alternative hypothesis that $\lambda < 0.0025$.

□

One popular way to test a hypothesis that $\beta = \beta_0$, say, is to use a Wald test. This simply uses the estimator $\hat{\beta}$ of β and its standard error $se(\hat{\beta})$ and the result that for large sample sizes the ratio $(\hat{\beta}-\beta_0)/se(\hat{\beta})$ will approximately have a standard normal distribution, which has an expected value of 0 and a standard deviation of 1. Hence $(\hat{\beta}-\beta_0)^2/(se(\hat{\beta}))^2$ has a chi-square distribution with 1 degree of freedom (χ_1^2). A hypothesis test based on this statistic is known as a Wald test.

Example 2.7.2
Consider carrying out a hypothesis test of a particular value of $\kappa = 1$ for the Weibull distribution in Example 2.6.5, against an alternative hypothesis that $\kappa = 1$. The null hypothesis is equivalent to assuming the exponential parametric distribution. The Wald statistic is $(1.662-1)^2/0.3338^2 = 3.93$ with a p-value of 0.047. The conventional interpretation of this result would be to reject the null hypothesis of $\kappa = 1$, which is consistent with the fact that the confidence interval obtained in Example 2.6.5 does not contain the value 1.

□

2.7.2 LIKELIHOOD RATIO TESTS

In the case of simple hypotheses the Neyman–Pearson lemma leads to considering the ratio of likelihoods to construct the most powerful test. This idea can be extended to composite hypotheses by considering the ratio of the maxima of likelihoods over the respective composite hypotheses. This procedure produces statistics for tests which are known as *likelihood ratio tests*. The procedure often enables a test to be found either easily by tractable mathematical methods or by numerical methods. The behaviour of such tests for large sample sizes has been extensively studied and is well known in many situations.

Example 2.7.3
Consider carrying out a hypothesis test of a particular value of $\lambda = \lambda_0$, say, for the exponential parametric distribution, given in Subsection 2.3.1, against an alternative hypothesis that $\lambda = \lambda_0$. In Example 2.6.4 it was shown that

$$L(\lambda) = \lambda^n \exp(-n\lambda/\hat{\lambda}),$$

where $\hat{\lambda} = 1/\bar{t}$ is the maximum likelihood estimator of λ. Hence the log likelihood function is given by

$$l(\lambda) = n\log\lambda - n\lambda/\hat{\lambda}.$$

Then the test is based on Λ, the likelihood ratio test statistic, which is minus twice the difference of the log likelihoods maximized over the null

hypothesis and the alternative composite hypothesis, respectively. Hence as $l(\lambda)$ is maximized at $\lambda = \lambda_0$ and $\lambda = \hat{\lambda}$, respectively, then

$$\Lambda = -2(l(\lambda_0) - l(\hat{\lambda})) = 2n(\log(\hat{\lambda}/\lambda_0) + \lambda_0/\hat{\lambda} - 1),$$

which under the null hypothesis that $\lambda = \lambda_0$ has the chi-square distribution with 1 degree of freedom (χ_1^2), for large sample size, n. Choosing the size of the Type I error, conventionally chosen as 0.05, and using the distribution of Λ gives a critical region for rejecting the null hypothesis defined by

$$2n(\log(\hat{\lambda}/\lambda_0) + \lambda_0/\hat{\lambda} - 1) > 3.84,$$

which is equivalent to

$$\hat{\lambda} < C_1 \text{ or } \hat{\lambda} > C_2$$

for suitable values of C_1 and C_2.

Using the example of failure times for the 20 pressure vessels presented in Subsection 1.4.1, consider testing the null hypothesis $\lambda_0 = 0.0025$. Then the critical region is defined by

$$\hat{\lambda} < 0.00166 \text{ or } \hat{\lambda} > 0.00401.$$

Hence as the estimate of λ is 0.00174 the null hypothesis is accepted as opposed to the alternative hypothesis that $\lambda = 0.0025$. Note the difference between the inference obtained in this example and that obtained in Example 2.7.1. This is because the alternative hypotheses considered are *two sided* as λ can be either less than or greater than λ_0. In Example 2.7.1 it was assumed that for the alternative hypothesis λ was less than λ_0.

□

Table 1.3. Times to failure (in hours) for pressure vessels

274	28.5	1.7	20.8	871	363	1311	1661	236	828
458	290	54.9	175	1787	970	0.75	1278	776	126

probability that the failure time of a pressure vessel is less than 100 hours, is equal to 0.5. The Pearson chi-square statistic is given by

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i},$$

where O_i is the number observed in category i and E_i the respective expectation. For the pressure vessels example $E_i = 10$ for times less than 100 hours and for times greater than 100 hours. The observed values are 5 and 15, respectively. Hence $X^2 = 5$. Under the null hypothesis this statistic X^2 has a chi-square distribution with 1 degree of freedom (χ^2_1), for large sample sizes. Hence the null hypothesis of $p = 0.5$ would be rejected at the 0.05 level of significance as the critical region corresponds to $X^2 > 3.84$.

□

2.8 Discussion

In this chapter the main models used in the analysis of failure data in reliability studies were introduced. In Section 2.3 the properties of four important parametric families of distributions, the gamma, Weibull, extreme value, and normal distributions, were described. The first two are defined for positive values, though the other two are defined for both positive and negative values of time. This means that though the first two can be used to model failure times, it is more usual to use the other two (the extreme value and normal) to model the logarithm of time.

As was demonstrated in the examples in Chapter 1, Section 1.4, data obtained from reliability studies are often incomplete owing to censoring. The various ways that these data may be obtained were outlined in Section 2.4. Though parametric models for lifetime distributions are important, the use of non-parametric methods is beneficial and these were illustrated in Section 2.5 for right censored data.

In order to use parametric models for lifetime distributions it is necessary to estimate the values of the parameters of the model. A brief account of the classical, likelihood, and Bayesian methods of estimation was presented in Section 2.6, though the various merits of the methods were not discussed, and these methods are illustrated in later chapters in various examples.

The chapter was concluded in Section 2.7 by a brief account of inferential methods for performing statistical tests to be able to assess whether a hypothesized lifetime distribution is consistent with observed failure times.

Ansell & Phillips

3
Analysis of lifetimes with covariates

3.1 Introduction

In many reliability studies as well as the information on the component's (or system's) lifetime there is also available other information about the component (system) or its environment. Typically there may be information on the design of the component or the wear the component has suffered. Hence instead of just recording a set of lifetimes there may be associated with each lifetime other variables. These variables are known as covariates. (The term is often used to cover factors, see Section 3.2, as well as measurable variables.) This extra information should yield more understanding about the performance of the component or system. The more information used the better the understanding, hopefully.

Recently there has been considerable interest in the analysis of such data sets in the reliability and related literature. Part of the interest may be due to the availability of the proportional hazards model, suggested by Cox (1972, 1975). This has focused interest in relating both variables and factors to components' lifetimes. However, the proportional hazards model is not the only model available and a number of approaches predate it, such as the use of the accelerated failure time model (Nelson, 1993) and the Weibull regression model (Smith, 1991). All the models and techniques discussed in this chapter are regression-type techniques.

Ansell (1987) suggested that there are four main reasons for the analysis of these types of data within the reliability literature. These are:
(a) to find significant factors (or variables) which affect lifetime;
(b) to remove nuisance variables which distort analysis;
(c) to increase comprehension of the failure model; and
(d) to produce a better prediction of the failure rate.
Examples of each of these uses appear in the reliability literature, see for example Bendell and Wightman (1985), Ansell and Ansell (1987), Dale (1983), Drury *et al.* (1987), and Jardine and Anderson (1984).

In using regression models the aim is to account for the variation in one variable in terms of other variables or factors. In reliability modelling this

usually means relating the variation in the lifetimes to other variables or factors. In many regression analyses the lifetimes are assumed to have a normal distribution. In lifetime studies this is rarely the case; more usually exponential or Weibull models are used. This will affect the algorithms used for estimation, though the general approach to, and concerns which arise from, regression modelling are still applicable in lifetime studies.

For illustration suppose the lifetime, T, of a component is a random variable which has an exponential distribution, with an expectation which is possibly related to a set of variables $\mathbf{z} = (z_1, z_2, ..., z_k)$. Then from Chapter 2, Subsection 2.3.1, the probability density function $f_T(t)$ of T is given by

$$f_T(t) = \begin{cases} \lambda(\mathbf{z};\boldsymbol{\beta}) \exp\left(-\lambda(\mathbf{z};\boldsymbol{\beta})t\right), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0, \end{cases}$$

where λ is a function of $\mathbf{z}$ and $\boldsymbol{\beta}$, which is a vector of unknown parameters. It is necessary to decide on the appropriate function λ and then it will be necessary to estimate the parameters $\boldsymbol{\beta}$. Usually $\log \lambda$ will be taken to be a linear function.

Historically least squares estimation was used for regression models, though it is common to use maximum likelihood estimation. If maximum likelihood estimation is used then a Newton–Raphson numerical procedure can be applied. Properties of the estimators of the parameters may be determined by exact methods or by use of the asymptotic theory approximations, see Chapter 2, Subsection 2.6.3. The maximum likelihood estimate of $\boldsymbol{\beta}$ will have asymptotically a multivariate normal distribution with mean $\boldsymbol{\beta}$ and a variance which is given by the reciprocal of the Fisher information. The appropriateness of these asymptotic results will usually depend on the sample size but may also depend on other variables such as total observation time or even some of the parameters of the model, see Sweeting (1992). Tests on both the parameters and the models can be constructed using this asymptotic theory. The tests will be discussed in the relevant sections. It is also possible to consider 'goodness-of-fit' of the models to the data. Some of the tests of the appropriateness of the model for the data are based on the residuals. A number of residuals can be defined for the lifetime regression models and these are discussed in Sections 3.3, 3.4, and 3.5.

One aspect which differentiates lifetime regression models from other regression models is censoring or truncation of the data. As stated before it is not unusual that upwards of 90% of the data will be censored, see Chapter 1, Subsection 1.4.2. The estimation procedure must be capable of taking account of this censoring.

3.2 Data and design

For the type of analyses discussed in this chapter as well as the times of failures, or times between failures, there will be data on other variables, usually referred to as covariates or concomitant variables. These extra data typically describe the construction of the component (or system) or the condition in which it functions. This is extra information and may account for the lifetime of the component. There may, of course, be other factors which affect lifetimes which have not been measured. The covariates may be controllable, in that they can be selected by an experimenter, or alternatively they may be nuisance factors outside the control of the experimenter.

Example 3.2.1

An experiment may be designed to investigate the best position of a brake on a wheel. The choice of site for the brake might be controlled by the experimenter, but the road conditions under which the experiment may be run might be beyond the control of the experimenter. So one variable would be the site of the brake and another might be the wetness of the road surface. The variables can be measures or factors. A measure is a direct reading, for example a temperature, the amount of wear, etc. A factor is a variable which is discrete, taking only a limited number of values. For example, the position of the brake could be a factor.

When modelling the effect of a factor it is common practice to use dummy variables. For example, if considering the position of the brake there are two possible sites, A and B, then there are two levels to the factor. Then if the brake was sited at position A the associated variable would take the value 0 and if the brake was sited at position B it would take the value 1.

□

Example 3.2.2

Suppose there is interest in the tensile strength of steel rods; then the carbon content might only be recorded as high, medium, or low. The variable describing the amount of carbon would be a factor with three levels, so it is possible to use three dummy variables: one for high carbon content, one for medium carbon content, and one for low carbon content. If a rod had high carbon content then it would have 1 for the first variable and 0 for the other two. These variables are related, with one out of the three being 1 whilst the other two are 0. Hence there are a range of alternative formulations which can be used. Three formulations are given in Table 3.1.

□

Table 3.1. Three possible formulations for the factor representing the content of carbon in steel rods considered in Example 3.2.2

Carbon content	Formulations		
	Three dummy variables	Two dummy variables	
	d_1 d_2 d_3	d_1 d_2	d_1 d_2
High	1 0 0	1 0	1 0
Medium	0 1 0	0 1	0 1
Low	0 0 1	−1 −1	0 0

Whilst often reliability data come from unplanned situations, there are some cases when the data do arise from laboratory trials or pre-planned experiments. There are several advantages of designed experiments; the two most important are the clarification of the aim of the experiment and the efficiency of collection of the data that can result from design. Design of experiments is a large field in statistics and has in the past been primarily associated with agricultural experiments, though it has been applied successfully in many engineering contexts, see Grove and Davis (1992). Currently the subject is attracting considerable attention through the drive for quality in production arising out of the quality movements, total quality management, and statistical process control. In the context of reliability, however, the subject has not received sufficient attention, though some authors have addressed this issue recently, see Davis (1991). There are a number of papers on design where the underlying distribution is assumed to be Weibull, see Zelen (1959) and Smith (1991).

It should be remembered, though, that unfortunately the majority of reliability data come from unplanned situations with little control on the number of variables collected, the regularity of measurement of variables, and checks on the accuracy of recording. Hence the analyst may encounter situations where there are a large number of variables to choose from to explain the variation in the lifetimes of the components. Selecting the variables to include in an analysis can be an art in itself and will be discussed in Section 3.8 of this chapter.

3.3 Weibull regression model

The material covered in this section should enable readers to consider both the Weibull regression model and other distributional regression models. The algorithms will, of course, differ for other distributions from those given in this section but the general principles will still hold.

It is usual for regression models to describe one or more of the distribution's parameters in terms of the covariates $\mathbf{z}$. The relationship is usually linear, though this is not always the case. The Weibull distribution, see Chapter 2, Subsection 2.3.2, has a reliability function which can be given by

$$R_T(t) = \exp(-\lambda t^{\kappa}), \text{ for } t > 0,$$

where $\lambda = 1/\theta^{\kappa}$ and θ is the scale parameter, and κ is the shape parameter. Each of these parameters could be described in terms of the covariates $\mathbf{z}$, though it is more usual to define either the scale or shape parameter in terms of $\mathbf{z}$. For example, if the scale parameter was chosen then a common model would be to have $\lambda(\mathbf{z};\boldsymbol{\beta}) = \exp(\boldsymbol{\beta}^T\mathbf{z})$, where the number of covariates $k = r$, the number of parameters. Then the reliability function would be

$$R_T(t \mid \mathbf{z};\boldsymbol{\beta}) = \exp(-\exp(\boldsymbol{\beta}^T\mathbf{z})t^{\kappa}), \text{ for } t > 0.$$

The probability density function is given by

$$f_T(t \mid \mathbf{z};\boldsymbol{\beta}) = \kappa\exp(\boldsymbol{\beta}^T\mathbf{z})t^{\kappa-1}\exp(-\exp(\boldsymbol{\beta}^T\mathbf{z})t^{\kappa}), \text{ for } t > 0.$$

This model is commonly referred to as the Weibull regression model but there are alternatives which have been studied, see Smith (1991), where the shape parameter is dependent also on the covariate.

There are advantages to reparameterizing this model by taking logs, so that the model takes the form of Gumbel's extreme value distribution, see Chapter 2, Subsection 2.3.3. A reason for this is to produce a model more akin to the normal regression model, but also it allows a more natural extension of the model and hence greater flexibility. Define $Y = \log T$ so that the reliability function is given by

$$R_Y(y \mid \mathbf{z};\boldsymbol{\beta}) = \exp(-\exp(\kappa y + \boldsymbol{\beta}^T\mathbf{z}))$$

so that

$$E(\log T) = -\frac{\gamma}{\kappa} - \frac{\boldsymbol{\beta}^T\mathbf{z}}{\kappa},$$

where γ is Euler's constant. It is usual to estimate the parameters by using the maximum likelihood approach. Suppose that there are n observations

some of which may be right censored. Using the d_i notation introduced in Chapter 2, Subsection 2.4.1, then the log likelihood l is given by

$$l(\kappa, \boldsymbol{\beta}; \mathbf{y}, \mathbf{Z}) = \sum_{i=1}^{n} d_i \log f_Y(y_i, \mathbf{z}_i; \boldsymbol{\beta}) + \sum_{i=1}^{n} (1 - d_i) \log R_Y(y_i, \mathbf{z}_i; \boldsymbol{\beta}),$$

where $\mathbf{y} = (y_1, y_2, ..., y_n)^T$, $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n)$, and $\mathbf{z}_i = (z_{i1}, z_{i2}, ..., z_{ir})^T$. Substituting for f_Y and R_Y gives

$$l(\kappa, \boldsymbol{\beta}; \mathbf{y}, \mathbf{Z}) = d \log \kappa + \sum_{i=1}^{n} d_i(\kappa y_i + \boldsymbol{\beta}^T \mathbf{z}_i) - \sum_{i=1}^{n} \exp(\kappa y_i + \boldsymbol{\beta}^T \mathbf{z}_i),$$

where $\sum_{i=1}^{n} d_i = d$. The equations for the maximum likelihood estimators of the β and κ are

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{n} d_i z_{ij} - \sum_{i=1}^{n} z_{ij} \exp(\kappa y_i + \boldsymbol{\beta}^T \mathbf{z}_i) = 0; \\ \frac{\partial l}{\partial \kappa} &= \frac{d}{\kappa} + \sum_{i=1}^{n} d_i y_i - \sum_{i=1}^{n} y_i \exp(\kappa y_i + \boldsymbol{\beta}^T \mathbf{z}_i) = 0. \end{aligned}$$

The second derivatives of the log likelihood are

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_j^2} &= -\sum_{i=1}^{n} z_{ij}^2 \exp(\kappa y_i + \boldsymbol{\beta}^T \mathbf{z}_i), \\ \frac{\partial^2 l}{\partial \beta_j \beta_k} &= -\sum_{i=1}^{n} z_{ij} z_{ik} \exp(\kappa y_i + \boldsymbol{\beta}^T \mathbf{z}_i), \\ \frac{\partial^2 l}{\partial \beta_j \partial \kappa} &= -\sum_{i=1}^{n} z_{ij} y_i \exp(\kappa y_i + \boldsymbol{\beta}^T \mathbf{z}_i), \\ \frac{\partial^2 l}{\partial \kappa^2} &= -\frac{d}{\kappa^2} - \sum_{i=1}^{n} y_i^2 \exp(\kappa y_i + \boldsymbol{\beta}^T \mathbf{z}_i). \end{aligned}$$

Solutions to these equations for the maximum likelihood estimators can be obtained using an iterative Newton–Raphson procedure. However, Aitkin and Clayton (1980) suggest that convergence is slow for κ using such a procedure. They suggest fitting the Weibull regression model using the statistical software package GLIM. The procedure consists of iteratively fitting a Poisson distribution until convergence is reached. The shape parameter (κ) is usually set initially to 1 and is updated by using fitted values of the Poisson mean, see Aitkin and Clayton (1980) and Roger (1985).

Table 3.2. Lifetimes (in cycles) of sodium sulphur batteries

Batch 1	164	164	218	230	263	467	538	639	669
	917	1148	1678+	1678+	1678+	1678+			
Batch 2	76	82	210	315	385	412	491	504	522
	646+	678	775	884	1131	1446	1824	1827	2248
	2385	3077							

Note: Lifetimes with + are right censored observations, not failures.

Standard errors may be obtained by use of the second derivatives to obtain the observed information matrix, $\mathbf{I}_o$, as was explained in Chapter 2, Subsection 2.6.3, and this matrix is usually calculated in the standard statistical software packages. The estimated variance covariance matrix will be $\mathbf{I}_o^{-1}$ and the standard errors will be the square roots of the diagonal elements of this inverted matrix.

Example 3.3.1

Ansell and Ansell (1987) analysed the data given in Table 3.2 in a study of the performance of sodium sulphur batteries. The data consist of lifetimes (in cycles) of two batches of batteries. The covariate vector for the ith battery is given by $\mathbf{z}_i = (z_{i1}, z_{i2})^T$, where $z_{i1} = 1$ and z_{i2} represents whether the battery comes from batch 1 or batch 2, so that

$$z_{i2} = \begin{cases} 0, & \text{if battery } i \text{ is from batch 1,} \\ 1, & \text{if battery } i \text{ is from batch 2.} \end{cases}$$

Hence β_2 represents the difference in performance between batteries from batch 2 and batch 1.

Fitting the Weibull regression model results in $\hat{\beta}_2 = 0.0156$ and $\hat{\kappa} = 1.127$. Using the observed information matrix the standard error of $\hat{\beta}_2 = 0.386$. Hence a 95% confidence interval for β_2 is $(-0.740, 0.771)$.

An obvious test to perform is to see if β_2 is non-zero. If it is non-zero this would imply there is a difference between the two batches of batteries. The hypotheses are:

$$\begin{aligned} H_0 : \beta_2 &= 0. \\ H_1 : \beta_2 &\neq 0. \end{aligned}$$

The log likelihood evaluated under H_0 is -49.7347 and under H_1 is -49.7339. Hence the likelihood ratio test statistic has a value of 0.0016. Under the null hypothesis the test statistic has a χ^2 distribution with 1 degree of freedom. Therefore the hypothesis that β_2 is zero, which is equivalent to no difference between the batches, is accepted. There are alternative tests, see Lawless (1982). Using the estimate $\hat{\beta}_2$ of β_2 and its standard error gives a Wald statistic of 0.0016, almost the same value as the likelihood ratio test statistic, and hence leads to the same inference. Both these inferences are consistent with the confidence interval, as it includes zero.

□

Examination of the goodness-of-fit and the appropriateness of the assumptions made in fitting the regression model can be based on graphical approaches using residuals, see Smith (1991). The Cox and Snell generalized residuals, see Cox and Snell (1968), are defined as

$$e_i = -\log R_T(t_i \mid \mathbf{z}_i; \hat{\boldsymbol{\beta}}),$$

where $R_T(t_i \mid \mathbf{z}_i; \hat{\boldsymbol{\beta}})$ is the reliability function evaluated at t_i and $\mathbf{z}_i$ with estimates $\hat{\boldsymbol{\beta}}$.

Cox and Snell (1968) provided a first-order correction to these residuals, though in many cases these are not used, see Smith (1991). Obviously one problem that arises is with the residuals for censored observations and authors generally, see Lawless (1982), suggest using

$$e_i = -\log R_T(t_i \mid \mathbf{z}_i; \hat{\boldsymbol{\beta}}) + 1.$$

Cox and Snell residuals should be independent and identically distributed (IID) random variables with a unit exponential distribution, i.e. with an expectation of one. From the result given in Chapter 1, Example 1.6.3, the cumulative hazard function is a linear function with a slope of one. Also the cumulative hazard function is minus the log of the reliability function. Hence a plot of minus the log of the estimated reliability function for the residuals, $-\log \hat{R}(e_i)$, against e_i should be roughly linear with a slope of one when the model is adequate. Other graphs can be informative; see Smith (1991) where plots against covariates and plots of subgroups of data are considered.

Example 3.3.1 (continued)

Using the data in Example 3.3.1 and fitting the Weibull regression model the generalized residuals have been calculated and are presented in Fig. 3.1. The plot is of minus the log of the reliability function of the generalized residuals against the generalized residuals. Since some of the points are

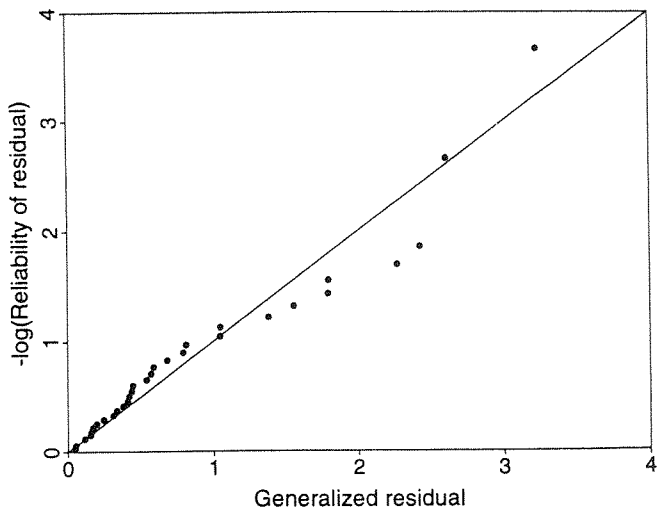

Fig. 3.1. Plot of the generalized residuals of the Weibull regression model for the sodium sulphur battery data

far from the line with slope 1 it would seem that the current model is not necessarily appropriate. Further investigation would be required to clarify if this was due to the choice of the distribution or the current model.

□

The approach described above is applicable to many distributions, though, of course, the algorithms for obtaining the estimates will vary. A special case of the Weibull regression model is the exponential regression model, when the shape parameter is taken to be 1 ($\kappa = 1$). This has been studied by M. Glasser (1967), Cox and Snell (1968), and Lawless (1982). There are a number of other lifetime regression models which have been studied: gamma, log-logistic, and lognormal, see Lawless (1982). Many of these models are covered by the general term location-scale models, or accelerated failure time models. The general model will be explored in Section 3.4.

Selection of an appropriate distribution model depends both on the physical context and on the actual fit achieved. There ought to be good physical justification for fitting a distribution using the context under study. Basing the decision about the distribution purely on fit can be very misleading, especially if there are a number of possible covariates to be chosen.

It is always possible by fitting extra variables to achieve a better fit to a set of data, though such a fit will have little predictive power.

3.4 Accelerated failure time model

The Weibull regression model discussed in the last section can be regarded as an example of an accelerated failure time model. Accelerated failure time models were originally devised to relate the performance of components put through a severe testing regime to a component's more usual lifetime. It was assumed that a variable or factor, such as temperature or number of cycles, could be used to describe the severity of the testing regime. This problem has been considered by a number of authors; see Nelson (1993) for a comprehensive account of this area.

Suppose in a study that the covariate is z, which can take the values 0 and 1, and that it is assumed that the hazard functions are

$$\lambda(t \mid z = 0) = \lambda_0$$

and

$$\lambda(t \mid z = 1) = \phi\lambda_0,$$

so that ϕ is the *relative risk* for $z = 1$ versus $z = 0$.

Then

$$R(t \mid z = 1) = R(\phi t \mid z = 0)$$

and, in particular,

$$E(T \mid z = 1) = \frac{E(T \mid z = 0)}{\phi}.$$

So the time for components with $z = 1$ is passing at a rate ϕ faster than for the components with $z = 0$. Hence the name of the model.

The model can be extended as follows. Suppose ϕ is replaced by $\phi(z)$ with $\phi(0) = 1$; then

$$R(t \mid z) = R(\phi(z)t \mid z = 0),$$

and hence

$$\lambda(t \mid z) = \phi(z)\lambda(\phi(z)t \mid z = 0)$$

and

$$E(T \mid z) = \frac{E(T \mid z = 0)}{\phi(z)}.$$

In using the model for analysis a parametric model is specified for $\phi(z)$ with β as the parameter, which will be denoted by $\phi(z;\beta)$. A typical choice would be

$$\phi(z;\beta) = \exp(\beta z).$$

This choice leads to a linear regression model for $\log T$ as $\exp(\beta z)T$ has a distribution which does not depend on z. Hence $\log T$ is given by

$$\log T = \mu_0 - \beta z + \epsilon,$$

where μ_0 is $E(\log T \mid z = 0)$ and ϵ is a random variable whose distribution does not depend on the covariate z.

To estimate β there is the need to specify the distribution. If the distribution of T is lognormal then least squares estimation may be used as $\mu_0 + \epsilon$ has a normal distribution, see Chapter 2, Subsection 2.3.4. If the distribution of T is Weibull with a shape parameter κ, see Chapter 2, Subsection 2.3.2, then $\kappa(\mu_0 + \epsilon)$ has a standard extreme value distribution, see Chapter 2, Subsection 2.3.3. Hence, as was stated at the beginning of this section, an example of the accelerated failure time model is the Weibull regression model studied in Section 3.3. Other such models have been widely applied in reliability, see Cox (1964), Fiegl and Zelen (1965), Nelson and Hahn (1972), Kalbfleisch (1974), Farewell and Prentice (1977), and Nelson (1993). However, they are not regularly applied by reliability engineers because of the perceived difficulties with estimation because the estimators vary depending on the distribution. There are a number of GLIM macros for specific distributions and the general approach of Aitkin and Clayton (1980) encompasses a number of these distributions. These models can also be fitted using the statistical software package SAS with PROC LIFEREG. Plotting techniques, such as using Cox and Snell generalized residuals as defined for the Weibull regression model, may be used for assessing the appropriateness of the model.

Example 3.4.1

Elsayed and Chan (1990) presented data collected from tests for the time-dependent dielectric breakdown of metal–oxide–semiconductor integrated circuits, which was described in Chapter 1, Subsection 1.4.4, with the data given in Table 1.6. The data consist of times to failure (in hours) for three different temperatures (170 °C, 200 °C, and 250 °C). Elsayed and Chan (1990) suggest a model where the covariate of interest is the inverse of the absolute temperature. So the covariate vector for the ith circuit is given by $\mathbf{z}_i = (z_{i1}, z_{i2})^T$, where $z_{i1} = 1$ and z_{i2} represents the inverse absolute temperature at which the test was performed, and takes the three values 0.001911, 0.002113, and 0.002256. Hence β_2 represents the coefficient of the inverse absolute temperature covariate.

Fitting the Weibull regression model results in $\hat{\beta}_2 = -7132.4$ and $\hat{\kappa} = 0.551$. Using the observed information matrix the standard error of $\hat{\beta}_2$ is 1222.0. Hence a 95% confidence interval for β_2 is $(-9527.5, -4737.1)$. This interval indicates that β_2 is non-zero and this would imply there is a difference in the performance of the circuits at the different temperatures. This can be confirmed by performing a hypothesis test to see if β_2 is non-zero. The hypotheses are

$$\begin{aligned} H_0 : \beta_2 &= 0. \\ H_1 : \beta_2 &= 0. \end{aligned}$$

The log likelihood evaluated under H_0 is -148.614 and under H_1 is -130.112. Hence the likelihood ratio test statistic has a value of 37.00. Under the null hypothesis the test statistic has a χ^2 distribution with 1 degree of freedom. Therefore the hypothesis that β_2 is zero is not accepted and this implies there is a difference between the circuits depending on the temperatures of the tests. Using the estimate $\hat{\beta}_2$ of β_2 and its standard error gives a Wald statistic of 34.06, almost the same value as the likelihood ratio test statistic, and hence leads to the same inference.

The generalized residuals are presented in Fig. 3.2. The plot is of minus the log of the reliability function of the generalized residuals against the generalized residuals. The points lie closer to a line with slope 1 than was the case in Fig. 3.1 for Example 3.3.1. So the current model may be appropriate.

□

3.5 Proportional hazards model

This model has been widely used in reliability studies by a number of authors: Bendell and Wightman (1985), Ansell and Ansell (1987), and Jardine and Anderson (1984). A diversity of uses have been found for the model.

The model Cox (1972) proposed assumed that the hazard function for a component could be decomposed into a baseline hazard function and a function dependent on the covariates. The hazard function at time t with covariates $\mathbf{z}$, $\lambda(t, \mathbf{z})$ would be expressed as

$$\lambda(t \mid \mathbf{z}) = \psi[\lambda_0(t), \phi(\mathbf{z}; \boldsymbol{\beta})],$$

where ψ would be an arbitrary function, $\lambda_0(t)$ would be the baseline hazard function, ϕ would be another arbitrary function of the covariates, $\mathbf{z}$, and

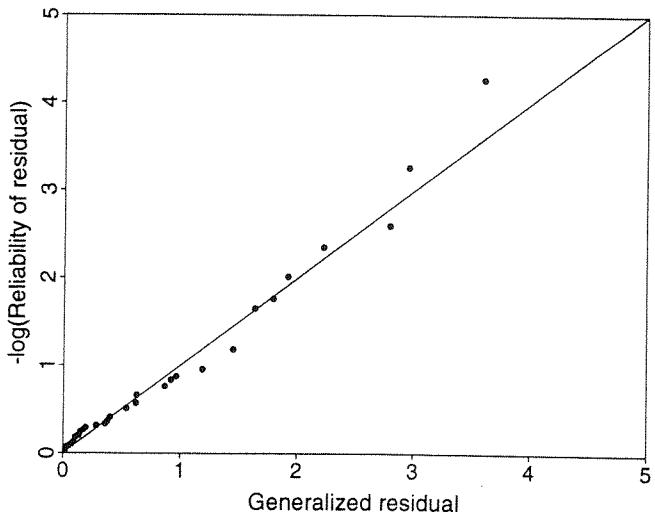

Fig. 3.2. Plot of the generalized residuals of the Weibull regression model for the semiconductor integrated circuit data

$\boldsymbol{\beta}$ the parameters of the function ϕ. Cox (1972) suggested that ψ might be a multiplicative function and that ϕ should be the exponential with a linear predictor for the argument. This *proportional hazards* model has the advantage of being well defined for most purposes, given that a hazard function must be non-negative. It yields

$$\lambda(t \mid \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}).$$

However, this is only one possible selection for ψ and ϕ. Using the multiplicative formulation it is usual to define $\phi(\mathbf{z}; \boldsymbol{\beta})$ so that $\phi(\mathbf{0}; \boldsymbol{\beta}) = 1$ so that $\phi(\mathbf{z}; \boldsymbol{\beta})$ is the relative risk for a component with covariate $\mathbf{z}$ compared with a component with covariate $\mathbf{z} = \mathbf{0}$. Thus the reliability function is given by

$$R(t \mid \mathbf{z}) = R(t \mid \mathbf{z} = \mathbf{0})^{\phi(\mathbf{z}, \boldsymbol{\beta})},$$

where $R(t \mid \mathbf{z} = \mathbf{0})$, often denoted simply as $R_0(t)$, is the baseline reliability function. Etezardi-Amoli and Ciampi (1987), amongst others, have considered an alternative additive model.

Cox (1972) considered the case in which the hazard function is a semi-parametric model; $\lambda_0(t)$ is modelled non-parametrically (or more accurately

using infinitely many parameters). It is possible to select a specific parametric form for $\lambda_0(t)$, which could be a hazard function from one of the families of distributions discussed in Chapter 2, Section 2.3.

In the case of the semi-parametric model Cox (1975) introduced the concept of *partial* likelihood to tackle the problems of statistical inference. This has been further supported by the work of Andersen and Gill (1982).

In Cox's approach the partial likelihood function is formed by considering the components at risk at each of the n_0 failure times $t_{[1]}, t_{[2]}, t_{[3]}, \ldots, t_{[n_0]}$, as defined in Chapter 2, Section 2.5. This produces a function which does not depend on the underlying distribution and can therefore be used to obtain estimates of $\boldsymbol{\beta}$. The partial likelihood function $L(\boldsymbol{\beta})$ is given by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n_0} \frac{\exp\left(\boldsymbol{\beta}^T \mathbf{z}_{[i]}\right)}{\sum_{w \epsilon R_i} \exp\left(\boldsymbol{\beta}^T \mathbf{z}_w\right)},$$

where R_i is the risk set and $\mathbf{z}_i$ is the observed covariate for the failure at time t_i. This can be maximized directly using the Newton–Raphson procedure or by use of an E–M algorithm.

Ties often occur in practice in data and adjustments should be made to the estimation procedure to account for ties. Breslow (1974) and Peto (1972) have made suggestions for replacing the partial likelihood function by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n_0} \frac{\exp\left(\boldsymbol{\beta}^T \mathbf{S}_{[i]}\right)}{\left(\sum_{w \epsilon R_i} \exp\left(\boldsymbol{\beta}^T \mathbf{z}_w\right)\right)^{m_i}},$$

where m_i is the number of failure times equal to $t_{[i]}$ and $\mathbf{S}_{[i]} = (S_{[i]1}, S_{[i]2}, \ldots, S_{[i]r})$ is the sum of the covariate vectors $\mathbf{z}_i$ for this time.

The first differentials of the log of the partial likelihood function $l(\boldsymbol{\beta})$ are

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n_0} \left(S_{[i]j} - \frac{m_i \sum_{w \epsilon R_i} z_{wj} \exp(\boldsymbol{\beta}^T \mathbf{z}_w)}{\sum_{w \epsilon R_i} \exp(\boldsymbol{\beta}^T \mathbf{z}_w)} \right),$$

for $j = 1, \ldots, r$. These derivatives can be equated to zero and solved to give estimators $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$.

The second differentials are

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n_0} m_i \left[\frac{\sum_{w \epsilon R_i} z_{wj} z_{wk} \exp(\boldsymbol{\beta}^T \mathbf{z}_w)}{\sum_{w \epsilon R_i} \exp(\boldsymbol{\beta}^T \mathbf{z}_w)} - \frac{\left(\sum_{w \epsilon R_i} z_{wj} \exp(\boldsymbol{\beta}^T \mathbf{z}_w)\right)\left(\sum_{w \epsilon R_i} z_{wk} \exp(\boldsymbol{\beta}^T \mathbf{z}_w)\right)}{\left(\sum_{w \epsilon R_i} \exp(\boldsymbol{\beta}^T \mathbf{z}_w)\right)^2} \right].$$

These second derivatives can be evaluated at $\hat{\boldsymbol{\beta}}$ to produce minus an 'information' matrix, whose inverse can be used to obtain standard errors.

Example 3.5.1

Returning to the sodium sulphur battery data used in Example 3.3.1, put $\mathbf{z} = z_2$, which was defined in Example 3.3.1 to indicate whether the battery comes from batch 1 or batch 2. (*Note*: The variable $z_1 = 1$ is not required as the arbitrary term $\lambda_0(t)$ will contain any arbitrary constant which was provided in the Weibull regression model in Example 3.3.1 by the parameter β_1.) Then $\hat{\beta} = -0.0888$. Using the information matrix, the standard error of $\hat{\beta}$ is 0.4034. Hence a 95% confidence interval for β is $(-0.879, 0.702)$.

A test can be performed to see if β is non-zero. If it is non-zero this would imply there is a difference between the two batches of batteries. The hypotheses are

$$\begin{aligned} H_0 : \beta &= 0. \\ H_1 : \beta &= 0. \end{aligned}$$

It is possible to use the partial log likelihood, which when evaluated under H_0 is -81.262 and under H_1 is -81.238. Hence the 'likelihood' ratio test statistic (twice the difference between these partial log likelihoods) has a value of 0.048. Under the null hypothesis this test statistic can be shown to have a χ^2 distribution with 1 degree of freedom, in the same way as for likelihood. Therefore the hypothesis that β is zero is accepted and this implies there is no difference between the batches.

□

An alternative non-parametric approach can be taken in the case when comparing two distributions to see if they are the same. In the case of two groups a test of whether $\beta = 0$ is equivalent to testing whether the two reliability functions, $R_1(t)$ for group 1 and $R_2(t)$ for group 1, are the same. The hypotheses are

$$\begin{aligned} H_0 : R_2(t) &= R_1(t) \\ H_1 : R_2(t) &= R_1(t)^{\exp(\beta)}, \text{ for } \beta \text{ not equal to 0.} \end{aligned}$$

Suppose $t_{[i]}, i = 1, \ldots, n_0$, are the ordered failure times and r_{1i} and r_{2i} are the number at risk at time $t_{[i]}$ for group 1 and group 2, respectively, and let m_{1i} be the number of failures of group 1 and let m_{2i} be the number of failures of group 2 at time $t_{[i]}$; then a test statistic would be U^2/I, where

$$U = \sum_{i=1}^{n_0} \left(m_{2i} - \frac{m_i r_{2i}}{r_i} \right),$$

$$I = \sum_{i=1}^{n_0} \frac{m_i(r_i - m_i)r_{1i}r_{2i}}{r_i^2(r_i - 1)},$$

$m_i = m_{1i} + m_{2i}$, and $r_i = r_{1i} + r_{2i}$. Under the null hypothesis this test statistic can be shown to have a χ^2 distribution with 1 degree of freedom. This procedure was originally proposed by Mantel (1966).

Example 3.5.2

Returning to Example 3.5.1, a test of whether $\beta = 0$ is equivalent to testing whether the reliability functions are the same for both batches of batteries. Now $U = 0.542$ and $I = 6.052$ and hence $U^2/I = 0.0485$. This statistic, though not identical, is similar to that obtained from the likelihood ratio in Example 3.5.1, and hence leads to the same conclusion. Therefore the null hypothesis that β is zero is accepted.

□

These examples have used a categorical covariate which indicates different groups of observations. The next example uses a continuous covariate.

Example 3.5.3

Returning to the semiconductor integrated circuit data used in Example 3.4.1, put $\mathbf{z} = z_2$, which was defined in Example 3.4.1 as the inverse of the absolute temperature. (*Note*: Again the variable $z_1 = 1$ is not required as the arbitrary term $\lambda_0(t)$ will contain any arbitrary constant which was provided in the Weibull regression model in Example 3.4.1 by the parameter β_1.) Then $\hat{\beta} = -7315.0$. Using the information matrix, the standard error of $\hat{\beta}$ is 1345.0. Hence a 95% confidence interval for β is $(-9951.2, -4678.8)$.

This interval indicates that β is non-zero and this would imply there is a difference in the performance of the circuits at the different temperatures. This can be confirmed by performing a test to see if β is non-zero. The hypotheses are

$$\begin{aligned} H_0 : \beta &= 0. \\ H_1 : \beta &= 0. \end{aligned}$$

It is possible to use the partial log likelihood, which when evaluated under H_0 is -190.15 and under H_1 is -173.50. Hence the 'likelihood' ratio test statistic (twice the difference between these partial log likelihoods) has a value of 33.31. Under the null hypothesis this test statistic can be shown to have a χ^2 distribution with 1 degree of freedom, in the same way as for likelihood. Therefore the hypothesis that β is zero is not accepted and this implies there is a difference between the circuits depending on the temperatures of the tests.

Using the estimate $\hat{\beta}$ of β and its standard error gives a Wald statistic of 29.60, almost the same value as the likelihood ratio test statistic, and hence leads to the same inference.

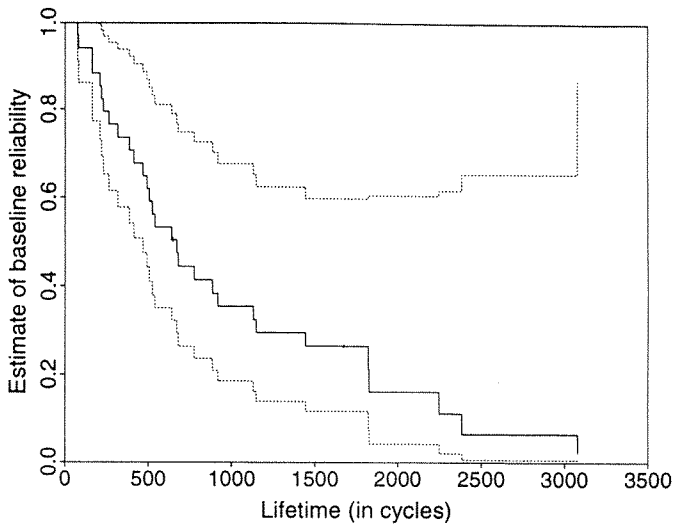

Fig. 3.3. Plot of the baseline reliability function for the proportional hazards model for the sodium sulphur battery data with 95% confidence limits

So far the estimation of β has been considered so that the relative risk can be estimated.

□

3.5.1 ESTIMATION OF THE BASELINE RELIABILITY FUNCTION

The reliability function has to be estimated and the usual approach is to estimate first the baseline reliability function, $R_0(t)$. This reliability function is first described as a product of the α so that

$$R_0(t) = \prod_{i:t_{[i]}<t} \alpha_i,$$

where $t_{[i]}$ are the ordered failure times. Hence the baseline reliability function is estimated at the failure times of the data. It is possible to derive a likelihood for the α and to obtain equations for the estimators $\hat{\alpha}$, which if there is only a single failure at $t_{[i]}$ are given by

$$\hat{\alpha}_i^{\exp(\hat{\boldsymbol{\beta}}^T \mathbf{z}_{[i]})} = 1 - \frac{\exp(\hat{\boldsymbol{\beta}}^T \mathbf{z}_{[i]})}{\sum_{w \in R_i} \exp(\hat{\boldsymbol{\beta}}^T \mathbf{z}_w)},$$

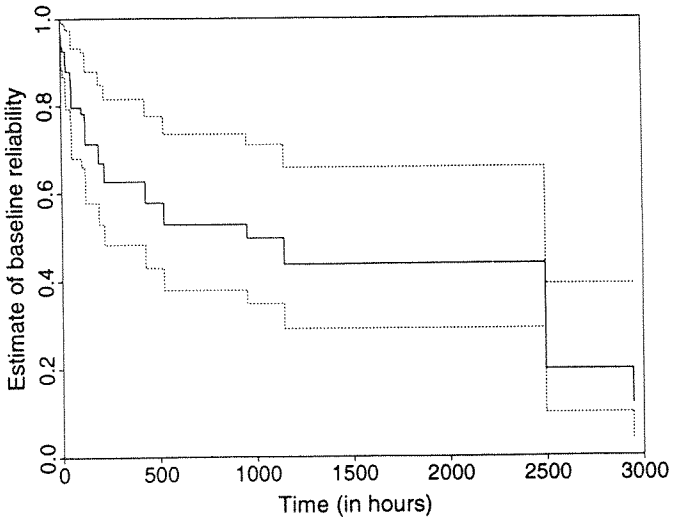

Fig. 3.4. Plot of the baseline reliability function for the proportional hazards model for the semiconductor integrated circuit data with 95% confidence limits

but otherwise $\hat{\alpha}$ has to be found by an iterative method. A suggested initial approximation, see Lawless (1982), is

$$\hat{\alpha}_i = \exp\left(\frac{-m_i}{\sum_{w \epsilon R_i} \exp(\hat{\boldsymbol{\beta}}^T \mathbf{z}_w)}\right),$$

which can be used to obtain the estimate of $R_0(t)$ and hence of $R(t)$.

Example 3.5.4
The estimate of the baseline reliability function for the data on sodium sulphur batteries introduced in Example 3.3.1 with the estimate of $\hat{\beta}$ of β as given in Example 3.5.1 are given in Fig. 3.3 with 95% confidence limits.

□

Baseline has been defined as the case with covariate $z = 0$. However, this is not always a sensible choice of the covariate. This is illustrated in the next example.

Example 3.5.5
For the semiconductor integrated circuit data used in Example 3.4.1, $\mathbf{z} = z$ was defined as the inverse of the absolute temperature. This will only be

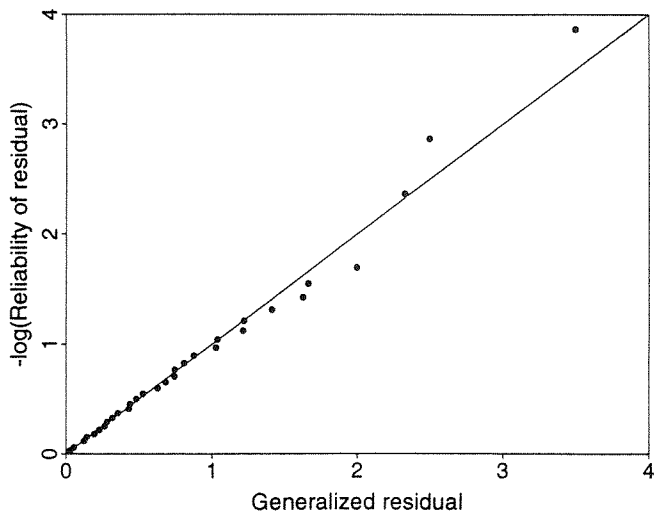

Fig. 3.5. Plot of the generalized residuals of the proportional hazards model for the sodium sulphur battery data

zero if the temperature is infinitely large. Hence it makes more sense to take the baseline value to be one of the temperatures used in the tests. The *smallest* temperature will be chosen, which corresponds to 0.002256, the *largest* value of the inverse absolute temperature. Then the estimate of the baseline reliability function for the data with the estimate of β as given in Example 3.5.3 are given in Fig. 3.4 with 95% confidence limits.

□

3.5.2 RESIDUAL PLOTS

Given the estimates of the reliability function it is then possible to obtain the Cox and Snell generalized residuals which are for an uncensored time

$$e_i = [-\log \hat{R}_0(t_i)] \exp(\hat{\boldsymbol{\beta}}^T \mathbf{z}_i)$$

and for a right censored lifetime

$$e_i = [-\log \hat{R}_0(t_i)] \exp(\hat{\boldsymbol{\beta}}^T \mathbf{z}_i) + 1.$$

These residuals have approximately a unit exponential distribution, i.e. with an expectation of one. From the result given in Chapter 1, Example

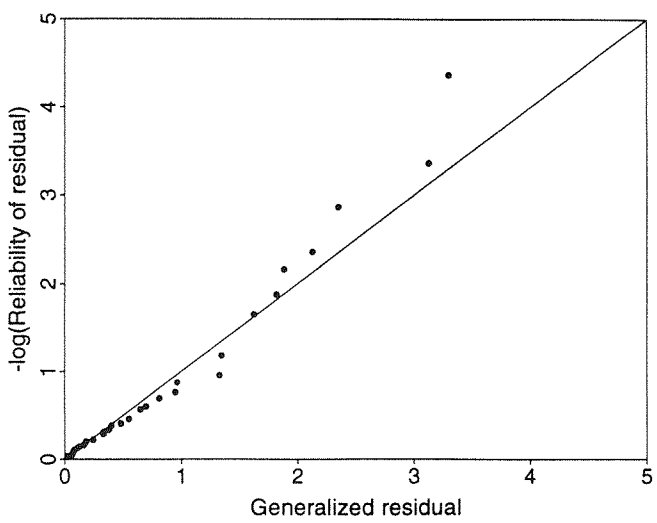

Fig. 3.6. Plot of the generalized residuals of the proportional hazards model for the semiconductor integrated circuit data

1.6.3, the cumulative hazard function is a linear function with a slope of one. Also the cumulative hazard function is minus the log of the reliability function. Hence a plot of minus the log of the reliability function of the residuals against their value should be approximately a straight line with a slope of one. Any departure from such would be a cause for concern.

Example 3.5.6

Using the sodium sulphur battery data from Example 3.3.1 the generalized residuals have been calculated and a plot of minus the log of their reliability function against the generalized residuals is given in Fig. 3.5. There seems to be a better agreement with the straight line than that obtained before in Fig. 3.1 with the Weibull regression model.

□

Example 3.5.7

Using the semiconductor integrated circuit data considered in Example 3.4.1 the generalized residuals have been calculated and a plot of minus the log of their reliability function against the generalized residuals is given in Fig. 3.6. There seems be no better agreement with the straight line than that obtained before in Fig. 3.2 with the Weibull regression model.

□

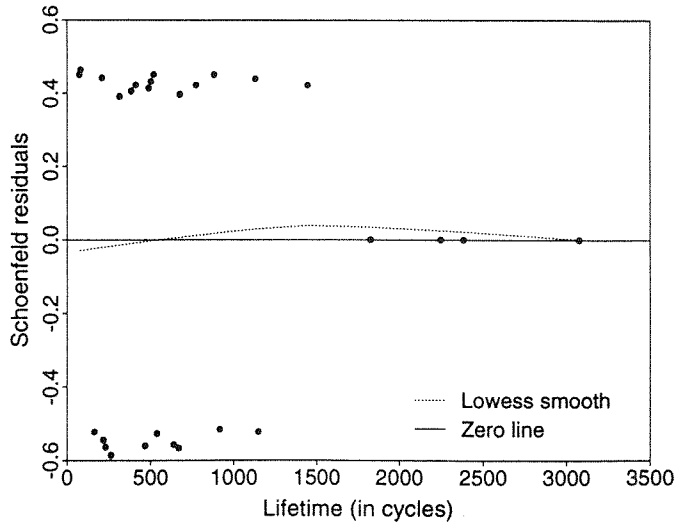

Fig. 3.7. Plot of the Schoenfeld residuals for batch of the proportional hazards model for the sodium sulphur battery data

Other residuals have been suggested for the Cox model. Schoenfeld (1982) suggested partial residuals to examine the proportionality assumption made in the Cox model. These are defined for the jth covariate as follows:

$$r_{ij} = z_{ij} - E(z_{ij} \mid R_i),$$

where $E(z_{ij} \mid R_i)$ is given by

$$E(z_{ij} \mid R_i) = \frac{\sum_{w \epsilon R_i} z_{wj} \exp(\hat{\boldsymbol{\beta}}^T \mathbf{z}_w)}{\sum_{w \epsilon R_i} \exp(\hat{\boldsymbol{\beta}}^T \mathbf{z}_w)}.$$

These residuals can then be plotted against time, and if the proportional hazards assumption holds then the residuals should be randomly scattered about zero, with no time trend.

Example 3.5.8

Using the sodium sulphur battery data from Example 3.3.1 and fitting the proportional hazards model, the Schoenfeld residuals are calculated and are presented in Fig. 3.7 as a plot of the residuals against the time. A non-parametric estimate of the regression line, the 'lowess' line, see Cleveland

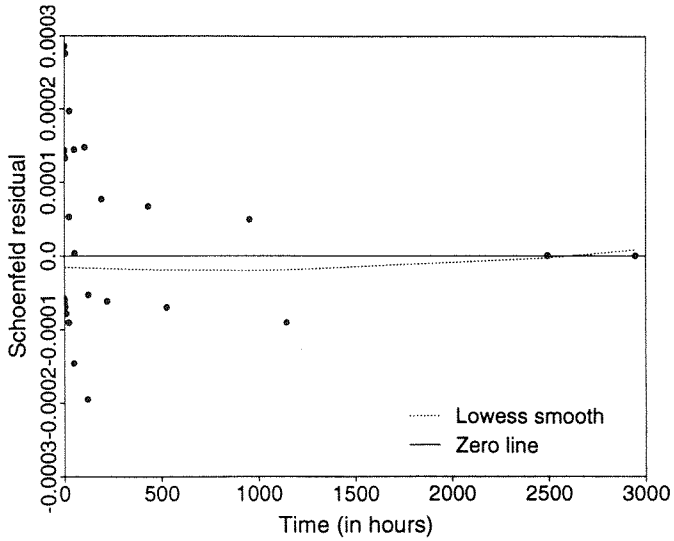

Fig. 3.8. Plot of the Schoenfeld residuals for the inverse of absolute temperature of the proportional hazards model for the semiconductor integrated circuit data

(1979), is included on the plot as well as the zero residual line. There seems to be no significant trend. A test for *linear* trend can be used. For this test the statistic is 0.0654, which would be from a χ^2 distribution with 1 degree of freedom if there was no linear trend. Hence there is no evidence of a linear trend and it is probably safe to accept the proportional hazards assumption.

□

Example 3.5.9

Using the semiconductor integrated circuit data as in Example 3.4.1 and fitting the proportional hazards model, the Schoenfeld residuals are calculated and are presented in Fig. 3.8 as a plot of the residuals against the time. A non-parametric estimate of the regression line, the 'lowess' line, is included on the plot as well as the zero residual line. For the test for *linear* trend the statistic is 1.60, which would be from a χ^2 distribution with 1 degree of freedom if there was no linear trend. Hence there is no evidence of a linear trend. However, it is not clear whether to accept the proportional hazards assumption.

□

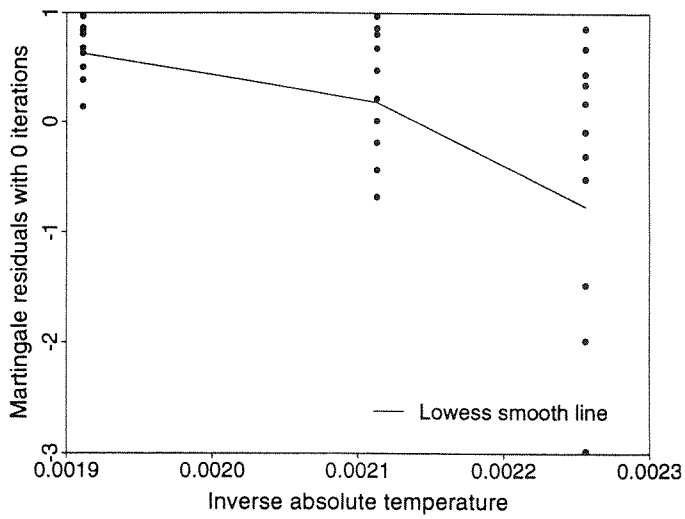

Fig. 3.9. Plot of the martingale residuals versus the inverse of the absolute temperature of the null proportional hazards model for the semiconductor integrated circuit data

A number of statistical software packages facilitate proportional hazards modelling, including SAS and S-PLUS. There are also GLIM macros for proportional hazards models, based on work by Whitehead (1980).

Therneau *et al.* (1990) suggest two alternative residuals, a martingale residual and a deviance residual. Except in the case of discrete covariates these residuals are far from simple to calculate; however, statistical software is available for their estimation, for example the SAS procedure PROC PHREG and S-PLUS functions **coxreg** and **agreg**. In the case of discrete covariates the martingale residuals are a transformation of the Cox and Snell generalized residuals. The deviance residuals are a transformation of the martingale residuals to correct for skewness. Therneau *et al.* (1990) suggest that the martingale residuals are useful in deciding about (a) appropriate functional relationships between the covariates and their survival, (b) proportionality of the hazard functions, and (c) the influence of observations. They suggest that the deviance residuals are more useful in identifying the observations which may be outliers.

Example 3.5.10

Using the semiconductor integrated circuit data as in Example 3.4.1 and

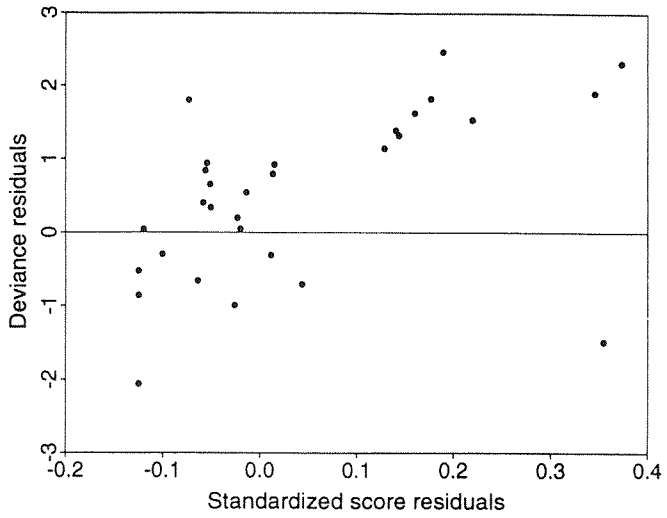

Fig. 3.10. Plot of the deviance residuals versus the standardized score residuals for the inverse of the absolute temperature of the proportional hazards model for the semiconductor integrated circuit data

fitting the proportional hazards model, the martingale residuals are calculated and are presented in Fig. 3.9 as a plot of the residuals against the covariate, the inverse of the absolute temperature. A non-parametric estimate of the regression line, the 'lowess' line, is included. There is some suggestion that a linear fit might not be best and a quadratic function might be an improvement, but this was found not to be the case.

The deviance residuals are calculated and can be used with the standardized score residuals to identify outliers. The plot of these residuals in Fig. 3.10 suggests that there are some outliers with standardized score residuals which are larger than 0.3.

□

3.6 Non-proportional hazards models

An assumption of the Cox regression model was that the hazards are proportional. This can be interpreted as the distance between the $\log(-\log)$ of the reliability functions not varying with time. Cox (1972) suggested a test for this proportionality by adding an extra covariate of log time to the

model. In the case of two groups with z_i indicating membership of a group ($z_i = 0$ if the ith individual belongs to group 1, $z_i = 1$ belongs to group 2), then the hazard functions become

$$\text{for group 1:}\lambda_0(t)$$

$$\text{for group 2:}\lambda_0(t)t^{\beta_2}\exp(\beta_1 z_i).$$

If the coefficient of log time is significantly different from zero the reliability functions are non-proportional, otherwise they are proportional. There was concern about the use of such models with time-dependent covariates, though this was justified by Andersen and Gill (1982).

Example 3.6.1
Using the sodium sulphur battery data in Example 3.3.1 and fitting Cox's non-proportional hazards model, $\beta_2 = 0.0900$ with standard error 0.523. These give a Wald statistic of 0.0296, which is not significantly different from zero, for a χ^2 distribution with 1 degree of freedom. Hence the null hypothesis of proportionality of the hazards is accepted, which agrees with the conclusion made in Example 3.5.8, when using the Schoenfeld residuals.

□

Many authors have extended this use of time-dependent variables in studies of lifetimes, for example Gore *et al.* (1984). A number of models have been proposed for the nature of dependency. Some of the models can lead to computational difficulties, see Holford (1976) and Anderson and Senthilselvan (1982). The other obvious problem is selection of an appropriate model. A plot of $\log(-\log)$ of the reliability function or a log hazard plot may reveal some suitable function or the nature of the problem may suggest some particular structure. Alternatively one might use the smoothed martingale residuals suggested by Therneau *et al.* (1990) plotted against the covariate to seek functional form. The danger is in producing too complex a model which does not increase insight. When the variables are quantitative rather than qualitative then the problem is exacerbated and greater caution is necessary.

When predicting future performance there may be difficulties with time-dependent covariates. It may then be necessary to resort to simulation.

3.7 Logistic model

Besides the models described so far there are a number of other models which are either specific to a type of application, such as growth models, see Chapter 7, or specific to the type of data available. In this section