Fatima Jammoul BSc

Modelling Breast Implant Survival Times

MASTER'S THESIS

to achieve the university degree of
Diplom-Ingenieurin

Master's degree programme: Operations Research and Statistics

submitted to

Graz University of Technology

Supervisor
Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Herwig Friedl

Institute for Statistics

Graz, January 2018

Kurzfassung

Die Lebensdaueranalyse ist ein Teilbereich der Statistik der sich mit der genauen Aufbereitung von Beobachtungen aus dem Bereich der Lebensdauer beschäftigt. Diese können allein oder in Relation zu gegebenen Prädiktoren analysiert werden. In dieser Arbeit geben wir einen Überblick über einen parametischen und einen nicht-parametrischen Ansatz der Lebensdaueranalyse. Wir konzentrieren uns hierbei nicht nur auf den theoretischen Hintergrund, sondern auch auf die praktische Anwendung in `R`. Weiters werden wir das neu erworbene Wissen in der Analyse von Implantatsdaten für Brustimplantate anwenden, welche aus dem österreichischen Silikonregister entnommen wurden.

Abstract

Survival analysis (also referred to as failure time analysis) focuses on the analysis of survival time observations, which can be considered by themselves or in relation to given predictors. In this thesis we give an overview of two major approaches in survival analysis (a parametric and a non-parametric approach), including theoretical background and practical application in `R`. Furthermore, we will use the described approaches to fit survival time models for survival times of breast implants, which are observations taken from the Austrian Silicone Registry.

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz,am ....
(Unterschrift)

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

....
date (signature)

Contents

Chapter 1

Introduction

Survival time analysis (or failure time analysis) refers to the practice of analyzing survival time observations either by themselves or in relation to given predictors. The need for such models is clear as day; not only is it imperative to be able to tell whether the survival time of a subject is related to any given set of predictors, it is also not uncommon to be interested in the actual shape of the survival distribution which indicates the probability of survival for a test subject after a certain amount of time has passed. Clearly, such models need to be approached with a delicate hand, taking into account the actual nature of observations of this type.

On a more practical note, the data that inspired this thesis actually fits the theory of survival times perfectly. In the Austrian Silicone Registry, several thousand surgeries all across the country are documented, all of which in some way relate to implants on the human body. This thesis was created in collaboration with Prof. S. Spendel and Dr. Paul Wurzer from the Department of Surgery at the Univ.Klinikum in Graz, specifically the Department for Plastic, Aesthetic and Reconstructive Surgery. The question we asked ourselves in the very beginning was simple enough: How long do breast implants last in the human body and can we find out which predictors influence the length of the survival in which way? This question inspired lengthy research into the world of survival time analysis, the results of which we describe in this thesis.

In Chapter 2, we give a detailed introduction into the nature of survival times before we move on to describe the arguably most popular approach in survival time analysis: the non-parametric models and the Cox proportional hazards model. By using very few assumptions on the actual shape of the distribution of the survival times, we deduce simple methods in order to analyze such observations both on their own as well as in relation to given predictors. Furthermore, we give an overview on diagnostic methods commonly used for such models.

In Chapter 3, we use the theoretical results of Chapter 2 in order to fit Cox models for the breast implant data taken from the Austrian Silicone Registry. We start off by describing our data and noting preliminary conjectures on the influence of the chosen predictors, before we diligently fit several models for the implant survival times. Furthermore, we take care to specify the influence of each predictor in a detailed manner.

Chapter 4 describes the parametric approach of survival time analysis, in which we describe commonly used distributions for the survival times as well as the process of fitting parametric models. Taking a more detailed look into the so-called accelerated failure time models, we decide that those actually fit our data quite nicely. Furthermore, we give a short overview on the diagnostic aspect of parametric models as well as the practical application in `R`.

In Chapter 5 we apply the theory of accelerated failure time models to the breast implant data in a more detailed manner. We choose a proper distribution for the survival times and move on to fit several models in that fashion. We also compare said models with their counterparts found in Chapter 3 and point out the differences in the results. Finally, in Chapter 6 we note our conclusions and try to summarize our findings.

Chapter 2

Non-Parametric Survival Time Models

2.1 Introduction to Survival Data

The concept of survival data in itself isn't particularly complex. One can easily visualize survival data by considering a fixed time frame, wherein a fixed set of test subjects is continuously monitored to see whether a previously agreed upon event transpires for each subject. For some of these subjects this event may happen within the time frame and the amount of time from the onset of the monitored time frame to the event is denoted by t_i, where i ranging from 1 to n refers to the i-th subject.

In a more concrete example, let us consider the breast implant data we're going to analyze further in this thesis. Every patient undergoes surgery to get an initial breast implant, which indicates the beginning of our time frame. From thereon, we count the days it takes for this implant to be either removed indefinitely or replaced by another implant. In this case, t_i refers to that amount of time for the i-th implant.

Survival data also often deals with the concept of censored data. An observation is called censored if the study ends before the event had the chance to transpire. In our breast implant example, one might say that an observation is censored if a patient passes away before the implant needs to be replaced. We may never know whether the implant will have needed replacement or not, all we know is that it "survived" from the initial surgery to the time of death of the patient. Handling censored observations is common practice when dealing with real life datasets.

2.2 Kaplan-Meier Estimators

In this section, we want to define *survival functions* and *hazard functions*. Furthermore we will move on to provide options to estimate both by introducing well-known the Kaplan-Meier approach.

From now on, we want to assume that $T_i \overset{iid}{\sim} f(t)$, where T_i refer to an iid random sample of

survival times. Subsequently, let t_i for $i = 1, \ldots, n$ be a realization of the above mentioned random sample. We define our survival function $S(t)$ as the following:

$$S(t) = P(T > t) = 1 - F(t),$$

where $F(t) = P(T \leq t) = \int_0^t f(x)dx$ as usual. To remain loyal to the example of breast implants, $S(t)$ describes the probability that an implant doesn't need to be replaced for more than t days. In the theory of survival data, it is often more convenient to consider the hazard function $\lambda(t)$, which is defined as the following:

$$\lambda(t) = \lim_{h \to 0} \frac{P(t \leq T < t + h | T \geq t)}{h}.$$

One can easily visualize $\lambda(t)$ as the instantaneous amount of events at time t. Thus, the alternative form

$$\lambda(t) = -\frac{1}{S(t)} \frac{dS(t)}{dt} = \frac{f(t)}{S(t)}$$

follows immediately. Rewriting the hazard function in this manner

$$\lambda(t) = -\frac{d \log S(t)}{dt}, \tag{2.1}$$

gives immediate insight into the behavior of a hazard function. By its nature, a survival function is a non-negative monotonically decreasing function that maps on $[0, 1]$. Thus, $\lambda(t)$ is also non-negative. It is important to note that a random variable is uniquely defined by its hazard function $\lambda(t)$ or its survival function $S(t)$. This follows simply from the definition of the survival function. Since any random variable is uniquely defined by its distribution function $F(t)$ or its density $f(t)$, the same can be concluded for $\lambda(t)$ and $S(t)$.

While these expressions are all based on the continuousness of the distribution of T, one can easily imagine how $S(t)$ and $\lambda(t)$ would change if T followed a discrete distribution instead. In that case, the time axis can be easily split up into a finite number of time intervals t_j of the same length. For example, one would have

$$S(t) = \sum_{j: t_j > t} f(t_j) \text{ with } f(t_j) = P(T \in t_j).$$

We are now interested in finding an estimate for $S(t)$. From result (2.1) it follows directly that

$$S(t) = \exp\left(- \int_0^t \lambda(x)dx \right) = e^{-\Lambda(t)}. \tag{2.2}$$

Thus, finding an estimate for $S(t)$ directly results in finding an estimate for $\Lambda(t)$, which is often referred to as the cumulative or integral hazard.

In order to introduce the Kaplan-Meier estimate for $S(t)$, consider for a sample of n subjects the times $t_1 \leq t_2 \leq \cdots \leq t_n$, which refer to the times at which the previously agreed

upon event transpires (e.g. removal of the breast implant). Notice that these times may also be censored and that two subjects can have the same survival times. Dealing with censored times is particularly simple in the Kaplan-Meier approach, which justifies its immense popularity.

Let d_i denote the number of events at t_i and let n_i the number of survivors at the time t_i. Then the Kaplan-Meier estimate for $S(t)$ is given as:

$$\hat{S}(t) = \prod_{i:t_i \leq t} \frac{n_i - d_i}{n_i}. \tag{2.3}$$

Formula (2.3) is highly flexible with regards to taking tied and censored observations into consideration. An observation t_i is referred to as *tied* with other observations $t_j, \ldots, t_k$ if $t_i = t_j = \cdots = t_k$, which of course means that the associated test subjects all experienced their respective events at the same time. So, if say two other events t_j and t_k are tied with t_i, $d_i = 3$. If the i-th observation is censored (e.g. the i-th test subjects drops out of the trial and one can no longer monitor them), then one can still take the observation into account. In that case, $d_i = 0$ since no event has transpired at time i and $n_i = n_{i-1} - 1$ in order to ensure that the conditional probability of survival to the time t_i is 1.

Given an estimate for $S(t)$ we can immediately find an estimate for $\Lambda(t)$, which is given as

$$\hat{\Lambda}(t) = -\log\left(\prod_{i:t_i \leq t} \frac{n_i - d_i}{n_i}\right) = -\sum_{i:t_i \leq t} \log\left(1 - \frac{d_i}{n_i}\right). \tag{2.4}$$

One must remind oneself that for a very small x, the approximation of $\log(1 + x) = x$ holds. Thus, equation (2.4) transforms to:

$$\hat{\Lambda}(t) \approx \sum_{i:t_i \leq t} \frac{d_i}{n_i}. \tag{2.5}$$

Turning this back onto our estimate $\hat{S}(t)$, it follows that

$$\hat{S}(t) = \exp(-\hat{\Lambda}(t)) \approx \exp\left(-\sum_{i:t_i \leq t} \frac{d_i}{n_i}\right). \tag{2.6}$$

There are a lot of packages in `R` that deal with survival functions and their visual representation, yet we will be focusing on the `survival` library as introduced by Therneau (2015). In this instance `t` is again the vector of survival times for our breast implant data. By using the `survfit` function and fitting a null-model, we get $\hat{S}(t)$ as previously discussed. Using the option `type="kaplan-meier"` ensures that the right estimate is used– other options are `"flemming-harrington"` or `"fh2"`. The `summary` function provides a deeper insight into the estimate. Take $t_1 = 0$ for example. According to the estimate, $S(0) = \frac{n_1 - d_1}{n_1}$. We see that `n.risk`, which reflects n_i and `n.event` which reflects d_i are given in the output. Thus, $S(0) = \frac{907-20}{907} = 0.97795$, just as seen in the `summary`.

```
> library(survival)
> kme <- survfit(Surv(t) ~ 1, type="kaplan-meier", conf.type="plain")
> summary(kme)

Call: survfit(formula = Surv(t) ~ 1, type = "kaplan-meier",
              conf.type = "plain")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   0    907       20  0.97795 0.00488     0.968392      0.98751
   1    887        6  0.97133 0.00554     0.960475      0.98219
   3    881        1  0.97023 0.00564     0.959171      0.98129
   4    880        1  0.96913 0.00574     0.957872      0.98039
   5    879        2  0.96692 0.00594     0.955285      0.97856
    ...
```

Plotting the resulting survival function and cumulative hazard can be easily achieved by using `plot`, just as shown below.

```
> par(mfrow=c(1,2))
> plot(kme,xlab="t", ylab="Survival", conf.int = FALSE)
> plot(kme,xlab="t", ylab="cumulative hazard", fun="cumhaz",
+      conf.int = FALSE)
```

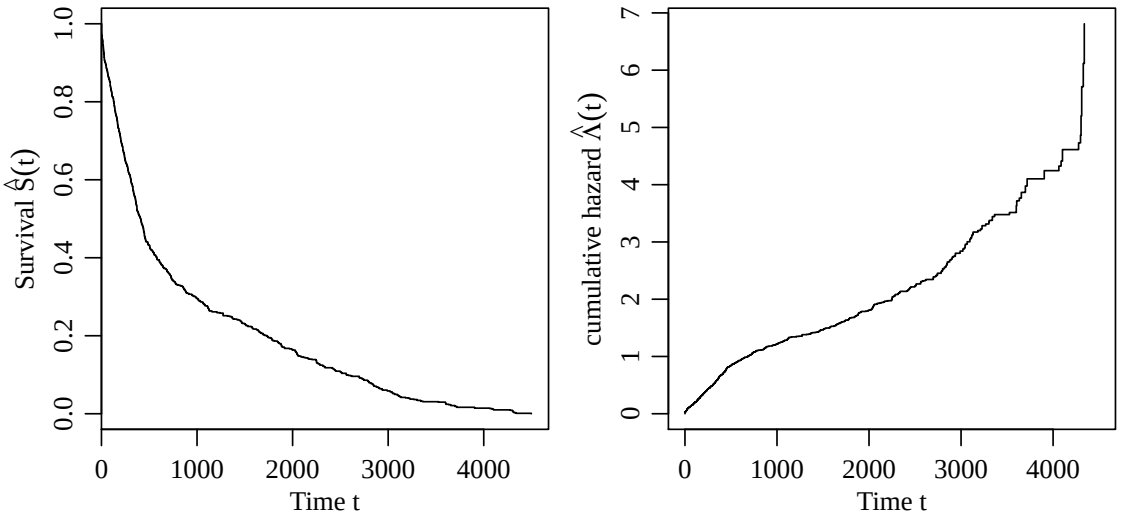

Figure 2.1: The resulting Kaplan-Meier estimates $\hat{S}(t)$ and $\hat{\Lambda}(t)$ using the implant data.

2.2.1 Variance Estimates and Confidence Intervals

Naturally, we are now interested in finding a way to estimate the variance of $\hat{S}(t)$. For a large data sample it is known that the Kaplan-Meier estimate is consistent and its distri-

bution converges in each point to a normal one (see Lawless 2003). While this seems to be good news in terms of wanting to estimate the variance, we are forced to take a detour in order to get a proper estimate. Kaplan and Meier (1958) argued that under certain circumstances $\hat{S}(t)$ is no longer defined for large t. Let t^* be the largest observation and assume that t^* is censored. Then, by definition of n_i, n_{t^*+1} has to be 0, yet $\hat{S}(t^*) > 0$. As a result, $\hat{S}(t)$ is undefined for any $t > t^*$, which proves to be problematic when attempting to find a nonparametric estimate of the variance of $\hat{S}(t)$.

By employing the popular delta method, one can compute a good estimate for the variance anyway. As a quick reminder of the delta method, let us consider a series of random variables Y_n that satisfy the property

$$\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2).$$

Furthermore consider a real function g such that $g'(\theta) \neq 0$. Then it follows that

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2[g'(\theta)]^2).$$

By performing a series of log-transformations on the individual conditional survival probabilities in each time segment (t_{i-1}, t_i) (as seen in Liu 2012), we are going to obtain an estimate for the variance of $\hat{S}(t)$. To that end, let $\hat{s}(t_i) := \frac{n_i - d_i}{n_i}$ and consider

$$\log \hat{S}(t) = \sum_{i:t_i \leq t} \log \hat{s}(t_i).$$

A variance estimate for $\hat{s}(t_i)$ is easily found– $\hat{s}(t_i)$ can be interpreted as the estimate for the conditional probability of survival in the time interval (t_{i-1}, t_i), given that the subject has survived up until t_{i-1}. Due to the fact that it is also an estimate of a proportion, more explicitly the surviving fraction out of the n_i test subjects still at risk at time t_i, one can estimate its variance by

$$\hat{V}[\hat{s}(t_i)] = \frac{\hat{s}(t_i)(1 - \hat{s}(t_i))}{n_i}.$$

We can then apply the delta method, which yields

$$\begin{aligned}\hat{V}[\log \hat{s}(t_i)] &= \frac{1}{\hat{s}(t_i)^2} \cdot \frac{\hat{s}(t_i)(1 - \hat{s}(t_i))}{n_i} = \frac{1 - \hat{s}(t_i)}{\hat{s}(t_i) n_i} \cdot \frac{n_i}{n_i} \\ &= \frac{d_i}{(n_i - d_i) n_i}.\end{aligned}$$

Thus, we obtain

$$\hat{V}[\log \hat{S}(t)] = \sum_{i:t_i \leq t} \frac{d_i}{(n_i - d_i) n_i}.$$

By yet again applying the delta method using $g(\log \hat{S}(t)) = \exp \log \hat{S}(t) = \hat{S}(t)$ we finally yield the desired variance estimate

$$\hat{V}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{i:t_i \leq t} \frac{d_i}{(n_i - d_i) n_i}. \tag{2.7}$$

The above estimate (2.7) is referred to as the Greenwood formula. Following similar steps, one can derive a similar variance estimate for $\hat{\Lambda}(t)$, which is

$$\hat{V}[\hat{\Lambda}(t)] = \sum_{i:t_i \leq t} \frac{d_i}{n_i^2}. \tag{2.8}$$

Let us return our focus back to `R` for a moment and see whether our algebraic calculations match the output.

```
> library(survival)
> kme <- survfit(Surv(t) ~ 1, type="kaplan-meier", conf.type="plain")
> summary(kme)

Call: survfit(formula = Surv(t) ~ 1, type = "kaplan-meier",
              conf.type = "plain")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0    907      20  0.97795 0.00488     0.968392      0.98751
    1    887       6  0.97133 0.00554     0.960475      0.98219
    ...
```

Following result (2.7) in order to obtain $\hat{V}[\hat{S}(t_1)]$, one would have to calculate $[\hat{S}(t_1)]^2 \frac{d_1}{(n_1-d_1)n_1}$. Plugging in our obtained estimates that means we obtain $[0.97795]^2 \frac{20}{(907-20)907}$, the square root of which is about 0.004876, just as calculated by `R`.

Regarding the computation of confidence intervals for $\hat{S}(t)$, there are plenty of ways to obtain one. The simplest method is to look at $\hat{S}(t) \pm z_{1-\alpha/2}\sqrt{\hat{V}[\hat{S}(t)]}$. In order to choose this option, one must add `conf.type = "plain"` into the initiation of the model like so.

```
> library(survival)
> survfit(Surv(t) ~ 1, type="kaplan-meier", conf.type="plain")
```

There is however a price to pay for the simplicity of the plain method. Sometimes it gives boundaries that are below 0 or above 1. Popular options to bypass that issue include using transformations like `log` (which is the default setting for `survfit`) or `log-log`.

```
> library(survival)
> kmeplain <- survfit(Surv(t) ~ 1, type="kaplan-meier",
+                     conf.type="plain")
> kmeloglog <- survfit(Surv(t) ~ 1, type="kaplan-meier",
+                      conf.type="log-log")
> par(mfrow=c(1,2))
> plot(kmeplain, xlab="t", ylab="Survival")
> plot(kmeloglog,xlab="t", ylab="Survival")
```

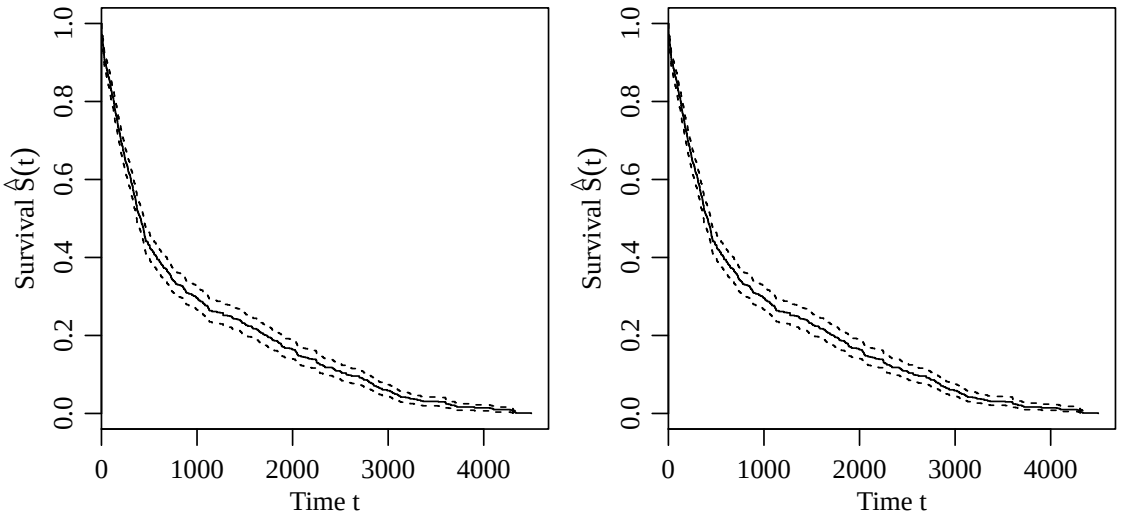

Figure 2.2: Kaplan-Meier estimates for $S(t)$ with plain and log-log 95% confidence intervals.

2.3 The Cox Proportional Hazard Model

Up until now we've done nothing but try to draw conclusions by analyzing the survival times t by themselves. Now we're interested in trying to interrelate our survival times with predictors to see whether the predictors have any significant influence on t. By introducing his take on the proportional hazard model in his 1972 paper, Cox provides the foundation of the following approach.

To that end, let us consider $X_i(t)$ to be the i-th row of design matrix. As it is common in regression approaches, $X_i(t)$ corresponds to the predictors for subject i. Thus, $X(t)$ is a matrix of dimenson $n \times p$, where n is the number of subjects and p refers to the number of predictors. For the Cox model, we will also need an arbitrary and unspecified, yet nonnegative function $\lambda_0(t)$, which we shall refer to as the *baseline hazard*. Then, the Cox model states that the hazard function for subject i can be written as

$$\lambda_i(t) = \lambda_0(t)e^{X_i(t)\beta}. \tag{2.9}$$

The fact that $\lambda_0(t)$ remains unspecified throughout the analysis is a key property of the Cox model. The lack of a proper depiction for the baseline hazard proves to be less cumbersome than initially expected, when we consider that the hazard ratio between two subjects can be written as

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)e^{X_i(t)\beta}}{\lambda_0(t)e^{X_j(t)\beta}} = e^{(X_i(t)-X_j(t))\beta}.$$

Evidently, the hazard ratio is independent of the baseline hazard. Furthermore, if the design vector $X_i(t)$ for each subject remains unchanged over the course of time, as it is

the case with the dataset we're considering, the ratio will be constant over time for two subjects i and j. Thus, the Cox model is also referred to as the *proportional hazards model*.

The Cox model puts its focus on the question whether the predictors X_i, now for simplicity's sake assuming that they remain constant over time for each subject, show significant influence on the behavior of the survival time t_i, while parametric models prefer to focus on the shape of the distribution of the survival times. Since we don't specify $\lambda_0(t)$, we are merely interested in how the predictors factor into the behavior of t_i.

To further analyze the model, let's take a look at the likelihood, for which we need to know what the density of T looks like in this model. As previously discussed

$$\lambda(t) = \frac{f(t)}{S(t)}$$

holds. Consequently, $f(t) = \lambda(t)S(t)$ follows. As we already have a description for the hazard function, we now just need to take a look at $S(t)$ within the Cox model. We remind ourselves that

$$S(t) = e^{-\Lambda(t)}$$

holds, where $\Lambda(t)$ is the integral of the hazard. Thus

$$\Lambda(t) = \int_0^t \lambda(s)ds = \int_0^t \lambda_0(s)e^{X\beta}ds = \Lambda_0(t)e^{X\beta}$$

follows for a given survival time t and the associated vector of predictors X for time t. As a result, we get

$$S(t) = e^{-\Lambda(t)} = e^{-\Lambda_0(t)e^{X\beta}}$$

and finally

$$f(t) = \lambda(t)S(t) = \lambda_0(t)e^{X\beta}e^{-\Lambda_0(t)e^{X\beta}}.$$

For a given realization of a random sample of n variables $t_1, \ldots, t_n$ from the population of survival times, we get the likelihood

$$L(\beta|t) = \prod_{i=1}^{n} \lambda_0(t_i)e^{X_i\beta}e^{-\Lambda_0(t_i)e^{X_i\beta}},$$

where as per usual X_i refers to the vector of predictors for the i-th survival time t_i. Even considering the log-likelihood, we realize that there is no way we can merely ignore the unknown baseline hazard or its integral $\Lambda_0(t)$ when searching for a MLE for β. Cox (1972) proposed a different approach– by maximizing a partial likelihood L_p with respect to β instead of the full likelihood as mentioned above, he managed to cleverly circumvent the problem of the missing and unspecified baseline hazard.

For a more specific description of the partial likelihood, please refer to Appendix A. For now, let it be known that partial likelihoods just like the one we will use to estimate β behave similarly to full likelihoods under some general assumptions (Cox, 1975).

2.3.1 Estimating $\boldsymbol{\beta}$

A logical next step would be to try to find a good estimate $\hat{\beta}$ for the coefficients. Cox (1972) utilizes the above mentioned partial likelihood approach when it comes to estimating β. To that end, let us consider that there are no tied observations, i.e. the survival times have the format $t_1 < \cdots < t_n$. Furthermore, let $R(t_i)$ denote the risk set at time t_i, which consists of all subjects that are still at risk of experiencing the event at the time t_i; those would be any subjects with a survival or censored time larger or equal to t_i. Let us also assume that we have no censored observations for the time being. Cox then suggests the following partial likelihood function in order to estimate β.

$$L_p(\beta|t) = \prod_{i=1}^{n} \frac{\lambda_i(t)}{\sum_{j \in R(t_i)} \lambda_j(t)} = \prod_{i=1}^{n} \frac{\exp\{X_i\beta\}}{\sum_{j \in R(t_i)} \exp\{X_j\beta\}}.$$

This follows from asking the following question: *Given that a subject i is at risk for the particular event at time t_i, what is the probability for i to actually experience the event at time t_i?* Indeed, this probability is expressed by the quotient of the hazard rate of subject i and the sum of the hazard rates of all subjects at risk at the time t_i, namely all subjects in $R(t_i)$ (see Liu, 2012 for further details). Calculating the partial log-likelihood l_p and differentiating with respect to β yields the score function $U(\beta)$ as follows:

$$U(\beta) = \frac{\partial l_p(\beta|t)}{\partial \beta} = \frac{\partial}{\partial \beta}\Big[\sum_{i=1}^{n} X_i\beta - \sum_{i=1}^{n} \log\big[\sum_{j \in R(t_i)} \exp\{X_j\beta\}\big]\Big].$$

Naturally, the maximum likelihood estimate $\hat{\beta}$ is found by solving the equation $U(\hat{\beta}) = 0$. Traditionally this is done by using the Newton-Raphson method, which starts off with an initiative point $\hat{\beta}_0$. In the k-th iteration, one computes

$$\hat{\beta}_{k+1} = \hat{\beta}_k + I^{-1}(\hat{\beta}_k)U(\hat{\beta}_k).$$

Hereby I denotes the negative second derivative of the log-likelihood. $\hat{\beta}$ seems to have the properties one would hope for; it is not only consistent, but also asymptotically normally distributed. As a matter of fact, it can be shown that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, I^{-1}(\hat{\beta}))$$

holds. Thus, a variance estimate for β is easily obtained by regarding the negative second derivative of the log-likelihood, which seems to be reminiscent of the properties found in estimates for generalized linear models. Continuing that thought, one is tempted to construct hypothesis tests for the resulting MLE $\hat{\beta}$. Indeed, a classic Wald Test seems appropriate when examining the hypothesis $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$ for any $j = 1, \ldots, p$. By the above mentioned asymptotic distribution of the MLE, it follows that

$$\frac{\hat{\beta}_j^2}{\widehat{var}(\hat{\beta}_j)} \overset{H_0}{\sim} \chi_1^2$$

holds asymptotically for each j. Similarly, one can consider $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$. Keeping the above mentioned property in mind, one can directly conclude that

$$\hat{\beta}^T I^{-1}(\hat{\beta})\hat{\beta} \overset{H_0}{\sim} \chi_p^2$$

holds. In the corresponding R output for such models, one will find the Wald tests executed for each coefficient of the model.

Naturally, there are also other ways to test for $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$. One classic test that comes to mind would be the Likelihood-Ratio Test. As we have estimated our β using a partial likelihood approach, it is only fitting to review

$$\Lambda_p(y) = \frac{\sup\limits_{\beta=0} L_p(y|\beta)}{\sup\limits_{\beta\in\mathbb{R}} L_p(y|\beta)} = \frac{L_p(y|0)}{L_p(y|\hat{\beta})},$$

whereas L_p refers to the partial likelihood function. As mentioned above, the commonly known property of $-2\log\Lambda_p(y) \xrightarrow{D} \chi^2_p$ holds for the partial likelihood as well.

When attempting to fit proportional hazard models in R, the library `survival` is a good way to start. Let us consider yet again our breast implant data– we now want to investigate whether the survival time of an implant (that is, the time between the surgery in which the implant is implanted and the surgery in which the implant is removed) has got anything to do with the volume of the implant itself. The function `coxph` fits a proportional hazard model in the following manner.

```
> library(survival)
> volmod <- coxph(Surv(t) ~ vol)
> volmod
Call:
coxph(formula = Surv(t) ~ vol)

        coef exp(coef) se(coef)    z       p
vol 0.002557  1.002560 0.000366 6.99 2.8e-12

Likelihood ratio test=45.1  on 1 df, p=1.88e-11
n= 903, number of events= 903
```

One yet again has the option of including censored data. If `t` were censored and if there existed a factor `status` indicating which observation is censored and which isn't, one could include this information by using `Surv(t, status)` in the above formula.

Now let us consider the `summary` of our above fitted model, which will give us more details.

```
> summary(volmod)
Call:
coxph(formula = Surv(t) ~ vol)

  n= 903, number of events= 903

        coef exp(coef) se(coef)     z Pr(>|z|)
vol 0.002557  1.002560 0.000366 6.987 2.82e-12 ***
```

```
---

    exp(coef) exp(-coef) lower .95 upper .95
vol     1.003     0.9974     1.002     1.003

Concordance= 0.573  (se = 0.011 )
Rsquare= 0.049   (max possible= 1 )
Likelihood ratio test= 45.09  on 1 df,   p=1.883e-11
Wald test            = 48.81  on 1 df,   p=2.817e-12
Score (logrank) test = 48.48  on 1 df,   p=3.344e-12
```

As we can see here, the above `summary` gives a lot more information. We conclude that the estimate $\hat{\beta}$ for volume is around 0.002557. That means, if the volume is raised by one unit, the hazard is multiplied by the exponent of $\hat{\beta}$, which indicates a growth of around 0.256%. The hypothesis $H_0 : \beta = 0$ is tested multiple ways in this output, especially since there is only one predictor. Due to that, the Wald Test at the end of the output and the z-statistic regarding `vol` are identical; the z-statistic tests for the significance of the coefficient of `vol` only, meanwhile the Wald Test tests the overall significance of all β together. We can conclude that `vol` truly seems to be a significant predictor when it comes to investigating the survival time of our breast implants. We are also shown the results of the Likelihood Ratio Test as well as a Score Test, which both corroborate the results of the Wald Test.

2.3.2 Diagnostics

One is now left to wonder just how a proportional hazard model can be assessed. Can we find visual and statistical tools to determine whether a Cox model is any good and whether the assumptions taken when fitting such a model are fulfilled in the first place?

Indeed, such tools exist and there are many different variants of evaluating a model of this type– as one might expect, the concept of residuals takes center stage in diagnostics. There are plenty of different types of residuals that can be computed and each one of them can be used to answer a specific question. By using `residuals(fit)` on a model `fit` of the type `coxph`, one can compute many of these residuals. We will focus here on four of those options to give a little insight in the diagnostic tools that are provided.

Cox-Snell Residuals: Testing Goodness of Fit

Cox-Snell residuals are perhaps the easiest type to determine– their entire theory is based on the following theorem.

Theorem 1 *Let T be a continuous non-negative random variable with associated distribution F_T and density f_T. Let Λ be its cumulative hazard and consider $\Lambda(T)$, which is a random variable as well. Then it follows that $\Lambda(T) \sim exp(1)$.*

Proof This proof will be done in two steps. First, we shall prove that $Z = F_T(T) \sim U(0, 1)$, from which we shall go ahead and prove the theorem's statement.

In order to determine the density of Z, we must utilize the transformation theorem, which states that the density of a transformed random variable can be determined by multiplying the density of the un-transformed variable with the determinant of the Jacobian of the inverse of the transformation. More specifically, this yields for the density of Z as

$$\begin{aligned} Z &= F_T(T) \\ F_T^{-1}(Z) &= T \\ \frac{\partial F_T^{-1}(Z)}{\partial Z} &= f_T^{-1}(Z) = f_T^{-1}(F(T)) \\ \Rightarrow f_Z &= f_T(F(T)) f_T^{-1}(F(T)) = 1. \end{aligned}$$

Thus, we have proven that $Z = F(T) \sim U(0,1)$. Furthermore by the definition of the survival function $S(t)$, we get $S(T) = 1 - F(T) \sim U(0,1)$. Finally, to obtain the distribution of $\Lambda(T) = -\log S(T)$, consider:

$$\begin{aligned} F_\Lambda(t) = Pr(\Lambda(T) \leq t) &= Pr(-\log S(T) \leq t) \\ &= Pr(S(T) \geq \exp(-t)) \\ &= 1 - Pr(S(T) \leq \exp(-t)) = 1 - \exp(-t). \end{aligned}$$

Thus, $\Lambda(T)$ indeed follows an exponential distribution with parameter 1. □

Armed with this knowledge, one now wants to compare $\hat{\Lambda}(t_i)$, the so-called *generalized Cox-Snell residuals*, with the exp(1)-distribution in a manner that is reminiscent of QQ-Plots in linear regression. Indeed, $\hat{\Lambda}(t_i)$ is rather easy to compute given the fitted model as shown below.

$$\hat{\Lambda}(t_i) = \int_0^{t_i} \hat{\lambda}(t)dt = \int_0^{t_i} \hat{\lambda}_0(t) e^{X_i\hat{\beta}} dt = \int_0^{t_i} \hat{\lambda}_0(t)dt \cdot e^{X_i\hat{\beta}} = \hat{\Lambda}_0(t_i) e^{X_i\hat{\beta}}.$$

Alternatively, one can also compute the Cox-Snell residuals by

$$r_i^{CS} = \hat{\Lambda}(t_i) = -\log \hat{S}(t_i).$$

In order to use the Cox-Snell residuals, one should plot $\hat{\Lambda}(r_i^{CS})$ against r_i^{CS}. If the model is proper, the plot should approximate a straight line through the origin with slope one. Cox-Snell residuals are generally used to assess the goodness of fit for a model– these types of residuals have to be enjoyed with caution, due to the fact that the correlation structure between $\hat{\beta}$ and the estimated cumulated hazard tends to mess with the results. That being said, looking at the Cox-Snell residuals in itself isn't a problem as long as one makes sure to use other diagnostic tools as well.

```
> volmod <- coxph(Surv(t) ~ vol)
> mvres <- residuals(volmod, type="martingale")
> csres <- rep(1, times=length(mvres)) - mvres
> csfit <- survfit(Surv(csres) ~ 1, type="kaplan-meier")
> plot(csfit$time, -log(csfit$surv),
+      xlab=expression(r[i]^CS), ylab=expression(hat(Lambda)(r[i]^CS))
> abline(a=0, b=1, col="red")
```

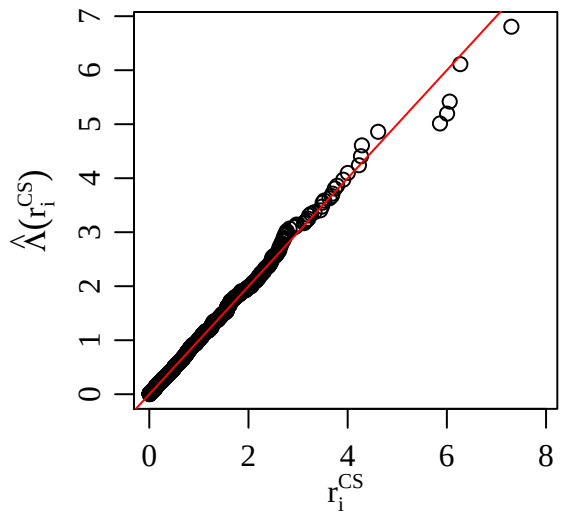

Figure 2.3: The Cox-Snell residuals for the volume model, calculated by using the martingale residuals.

Because Cox-Snell residuals aren't given as an option in `R`, one of the swiftest ways to get them is to use the martingale residuals which are readily available with the `residuals` function– we'll elaborate on the details of their relationship in the section on martingale residuals. As we can see above for our `volmod`, it appears as though the Cox-Snell residuals approximate a straight line with slope 1 quite nicely. This speaks for the goodness of fit.

Schoenfeld Residuals: Testing the Proportional Hazard Assumption

In order to test the assumption that the hazards are proportional in the first place, Schoenfeld defined these types of residuals in his 1982 paper. His idea was to construct residuals that don't depend on the change in time. Let us remind ourselves of the partial likelihood function we maximize to get $\hat{\beta}$, which is defined as

$$L_p(\beta|t) = \prod_{i=1}^{n} \frac{\exp\{X_i\beta\}}{\sum_{j\in R(t_i)} \exp\{X_j\beta\}}.$$

In order to estimate β, we proceeded by differentiating the partial log-likelihood in order to obtain the score.

$$\begin{aligned}\frac{\partial l_p(\beta|t)}{\partial\beta} &= \frac{\partial}{\partial\beta}\left[\sum_{i=1}^{n} X_i\beta - \sum_{i=1}^{n} \log\Big[\sum_{j\in R(t_i)} \exp\{X_j\beta\}\Big]\right] \\ &= \sum_{i=1}^{n}\left[X_i - \frac{\sum_{j\in R(t_i)} X_j \exp\{X_j\beta\}}{\sum_{j\in R(t_i)} \exp\{X_j\beta\}}\right].\end{aligned}$$

Then the Schoenfeld residuals $r_i^S = (r_{i1}^S, \ldots, r_{ip}^S)$ are defined as the corresponding summands in the above result for $\beta = \hat{\beta}$, indicating that the r_i^S actually measure the contri-

bution of the i-th individual to the score. Indeed, for the MLE $\hat{\beta}$, the r_i^S all sum up to zero, since for the MLE

$$\frac{\partial l_p(\beta|t)}{\partial \beta}\Big|_{\beta=\hat{\beta}} = \sum_{i=1}^{n}\left[X_i - \frac{\sum_{j\in R(t_i)} X_j \exp\{X_j\hat{\beta}\}}{\sum_{j\in R(t_i)} \exp\{X_j\hat{\beta}\}}\right] = \sum_{i=1}^{n} r_i^S = 0$$

holds. This seems to be a desirable property for any type of residual. The Schoenfeld residuals represent the amount of weight each specific subject contributes to the overall score. Now, if the assumption of proportional hazards is fulfilled, the expected value of r_i^S should be around the zero vector. Thus, a plot of each of the r_{ij}^S with $i = 1, \ldots, n$ and $j = 1, \ldots, p$ against the time t_i should yield a plot centered around zero. If the resulting residuals show no clear structure and are relatively randomly arranged around zero, we can assume that the proportional hazard assumption isn't violated for that specific predictor.

```
> schres <- residuals(volmod, type="schoenfeld")
> plot(yvalues, schres,
+      xlab="Time", ylab="Schoenfeld Residuals")
> abline(h=0, col="red")
```

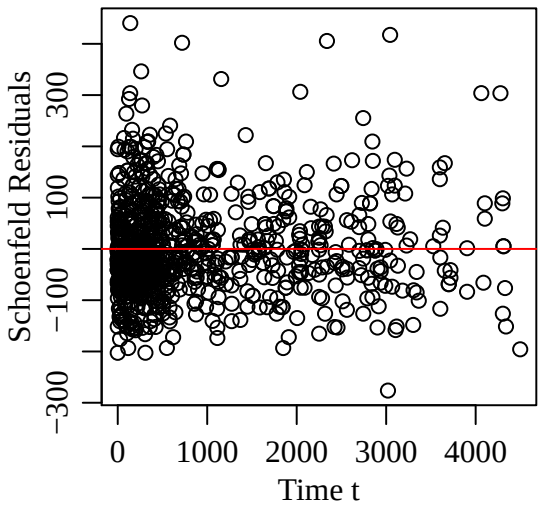

Figure 2.4: Schoenfeld residuals for the volume model.

Looking at the Schoenfeld residuals in Figure 2.4, we can observe that they seem to be rather randomly distributed around zero– while there are a few larger residuals around the 400 mark especially within the 0 to 3000 day interval, it appears to be nothing to worry about. Thus, we are reaffirmed in our beliefs that the hazards are proportional.

The `survival` package makes testing for the proportional hazards assumption even easier– by applying the function `cox.zph` on a Cox model, one performs Chi-squared tests for independence for each covariate as well as the entire model, which means that the correlation

between a scaled version of the Schoenfeld residuals and a transformed version of t is tested. The respective significance levels are then displayed alongside the estimated correlation coefficient in a compact table. One also has the option to plot the above mentioned scaled version of the Schoenfeld residuals for each predictor against a transformation of t using this function. Again one is supposed to look for any and all forms of structure within these plots– if the scaled residuals appear to follow no specific structure, one can assume that the proportional hazard assumption is not violated.

```
> lumenvolmod <- coxph(Surv(t) ~ vol + lumen)
> summary(lumenvolmod)
Call:
coxph(formula = Surv(t) ~ vol + lumen)

  n= 903, number of events= 903

                coef exp(coef)  se(coef)      z Pr(>|z|)
vol          0.002451  1.002454  0.000372  6.589 4.44e-11 ***
lumensingle -0.323640  0.723510  0.152374 -2.124   0.0337 *
> anova(lumenvolmod)
Analysis of Deviance Table
 Cox model: response is Surv(t)
Terms added sequentially (first to last)

       loglik   Chisq Df Pr(>|Chi|)
NULL  -5246.9
vol   -5224.3 45.0891  1  1.883e-11 ***
lumen -5222.3  4.1263  1    0.04222 *
```

Consider the model `lumenvolmod`, which fits the survival times t against the two predictors `vol` and `lumen`. The latter is a factor with two levels `single` and `double`, which specify a certain type of implant. It is especially interesting to see the Schoenfeld plots for these types of predictors.

```
> test.ph <- cox.zph(lumenvolmod)
> test.ph
                rho chisq      p
vol         -0.0612 3.563 0.0591
lumensingle -0.0122 0.134 0.7141
GLOBAL           NA 3.569 0.1679
```

As one can observe in the above output, none of the predictors show significant levels– thus, the null-hypothesis, which claims that the proportional hazard assumption is fulfilled, cannot be rejected. The same can be said for the test of the overall model. In fact, the model `lumenvolmod` seems to fulfill the proportional hazard assumption quite nicely. Furthermore, one can obtain plots for the Schoenfeld residuals by using the function `ggcoxzph` on the actual result of `cox.zph`.

```
> ggcoxzph(test.ph)
```

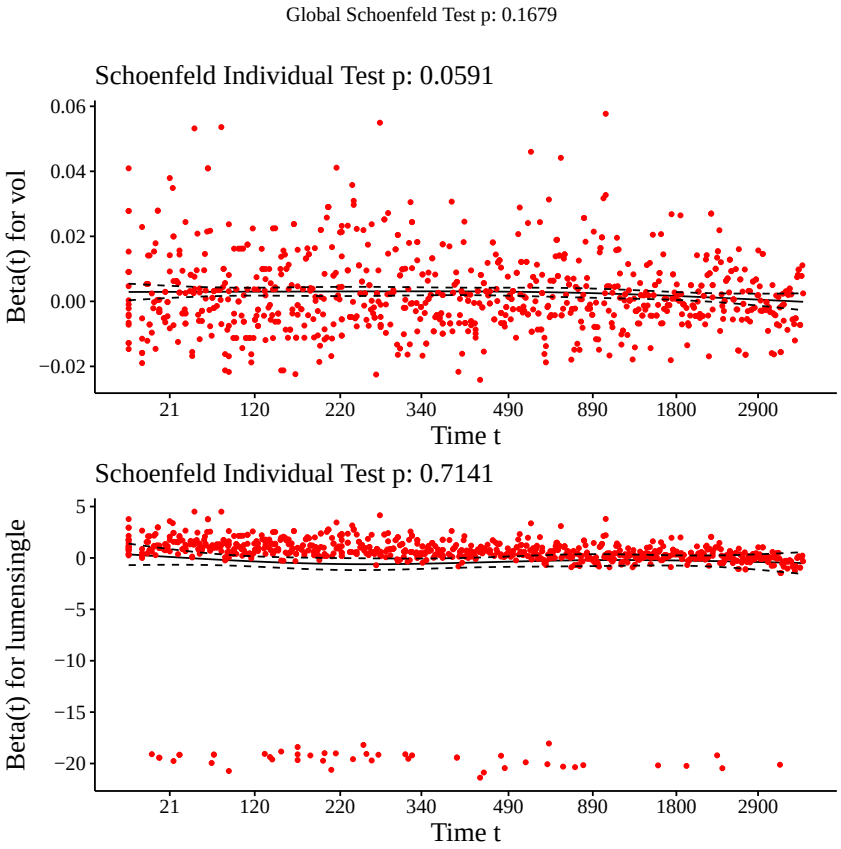

Figure 2.5: Schoenfeld residuals for the model including lumen and volume.

Martingale Residuals: Testing for Linearity

Martingale residuals stem from a different formulation for survival functions which uses counting processes. Defined by

$$M_i = \delta_i - \hat{\Lambda}_i(t_i),$$

the martingale residual can be understood as a measure for the difference between the actually observed number of events for individual i minus the estimated number of events following the fit, where δ_i is a binary indicator stating whether t_i is a censored observation or not. The name *martingale residuals* stems from the fact that M_i indeed define a martingale if the Cox model is specified correctly. A stochastic process is called a martingale if the following two properties hold:

- $E[|M_i|] < \infty$
- $E[M_i|M_1, \ldots, M_{i-1}] = M_i$.

How does one go about utilizing the martingale residuals? Now, if one is interested in finding out whether a specific predictor X_j is included in a linear form (say $X_j^T\beta$), one should go ahead and fit a `coxph` model without X_j and build the martingale residuals for

this model. Then, by laying through any kind of smoothing function (for example, `lowess` is a popular choice), one can find out what kind of a form X_j should have in the model– if the smoothed line appears to be somewhat close to linearity, X_j should be included in linear form.

```
> nullmod <- coxph(Surv(t) ~ 1)
> mres <- residuals(nullmod, type="martingale")
> plot(vol, mres,
+      xlab="Volume", ylab="Martingale Residuals")
> lines(lowess(mres~vol), col="red")
```

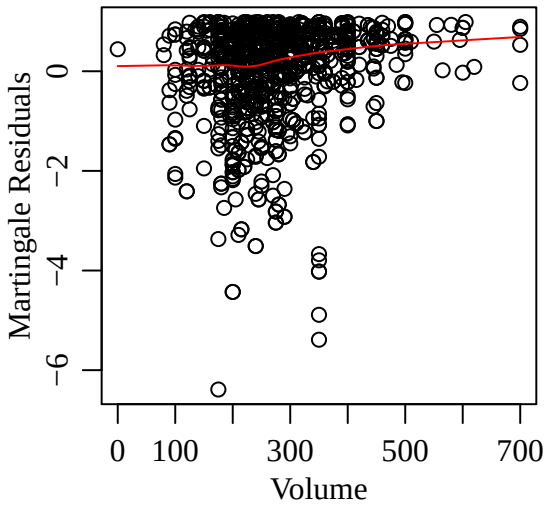

Figure 2.6: Martingale residuals for the predictor volume.

The martingale residuals as shown in Figure 2.6 seem to reaffirm our plans to include `vol` as a linear predictor– the `lowess` smooth gives an almost linear function, which indicates that `vol` should be included in a linear manner.

Deviance Residuals: Testing for Outliers

One monumental drawback of martingale residuals is their asymmetry. Deviance residuals seem to fix that particular problem. In generalized linear models, the (unscaled) deviance is defined by

$$D = 2(l(y|y) - l(\hat{\beta}|y)),$$

where $l(y|y)$ is the log-likelihood of the saturated model and $l(\hat{\beta}|y)$ is the log-likelihood of our model. One particularly useful property of the deviance is the asymptotic χ^2 distribution, which we're going to make use of in this instance since the square-root of the deviance thus approximates a normal distribution. The deviance residual for a Cox model can be expressed via the martingale residuals as

$$r_i^D = \text{sign}(M_i)\sqrt{-2[M_i + \delta_i \log(\delta_i - M_i)]}.$$

As a result, r_i^D has a mean of about zero and a variance of about one while also having a more symmetric distribution than M_i. Thus, outliers can be detected much more easily, if one plots r_i^D against the linear predictor.

```
> dres <- residuals(volmod, type="deviance")
> plot(vol, dres,
+      xlab="Volume", ylab="Deviance Residuals")
> abline(h=0, col="red")
```

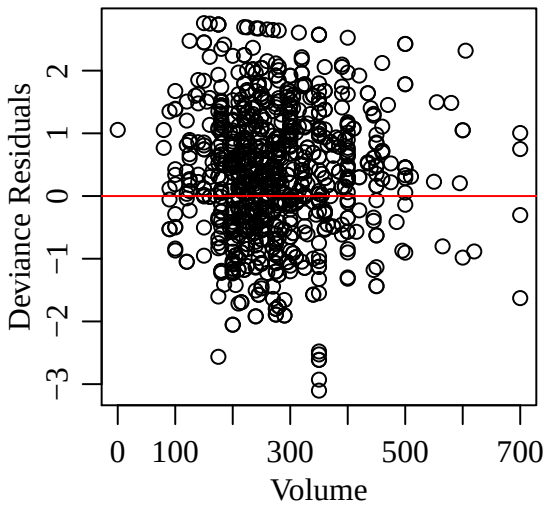

Figure 2.7: Deviance residuals for the predictor volume.

And lastly, the deviance residuals seem to be entirely inconspicuous as well– aside from maybe one or two residuals around the 310 volume mark, we can't particularly detect any drastic outliers that could distort the fit.

Chapter 3

Application of the Cox Model

3.1 Austrian Silicone Registry Implant Data

The breast implant dataset we will be analyzing further within this thesis is a subset of the Austrian silicone registry, which has been established by surgeons all over the country in an attempt to elevate the standard of care for their patients. As of 2017, there are more than 22500 entries in the registry. Each one of these entries in the registry consists of 68 columns which relate to a surgery in which an implant has either been implanted into a patient or removed. The registry has data on plenty of different types of implants, yet for the sake of consistency we will be focusing on breast implants on female patients only.

The main focus will be trying to analyze the duration with which these implants stay in the body. To that end, we have constructed 903 pairs of entries out of the registry, where one part of the pair refers to the primary surgery in which the implant was inserted into the body and the second half refers to the surgery in which the implant was removed from the body for whatever reason. The time passing between those two observations, counted as days, will be the survival time `t` for that particular implant. We intend to use `t` as response. The task in this instance is to fit models for `t` by using predictors that come from the row in the dataset associated to the primary surgery in order to figure out whether assumptions about the survival time `t` of a certain implant can be made judging by the data from its primary surgery.

All in all, we will be focusing on the following predictors in order to fit models for our response `t`:

- `vol`, which refers to the volume of each implant in cm^3.
- `surface`, which refers to the surface of the implant. `surface` is a factor with the levels `smooth`, `textured` or `polyurethan`.
- `lumen`, which specifies a certain type of implant. `lumen` is a factor with the levels `single` or `double`.
- `fill`, which refers to the substance used to fill the implant. `fill` is a factor with the levels `hydrogel`, `siliconegel`, `saline solution`, `mixed` and `other`.

- `drainage`, which indicates whether drainage was required during the primary surgery. Thus, `drainage` is a factor as well with the levels `yes` and `no`.

Response `t`

```
> summary(t)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0   167.5   412.0   881.5  1351.0  4498.0
```

Taking a closer look at our response `t`, we see that the median of the survival times is around 412 days, which is a little over a year. We thus have a large mass of short survival times at the very beginning. Contrary to that, the third quantile lies at around 1351 days, which equals a little over 3 and a half years. The maximum of 4498 days, which equals about 12 years, seems to be rare a occasion as the following plots in Figure 3.1 show.

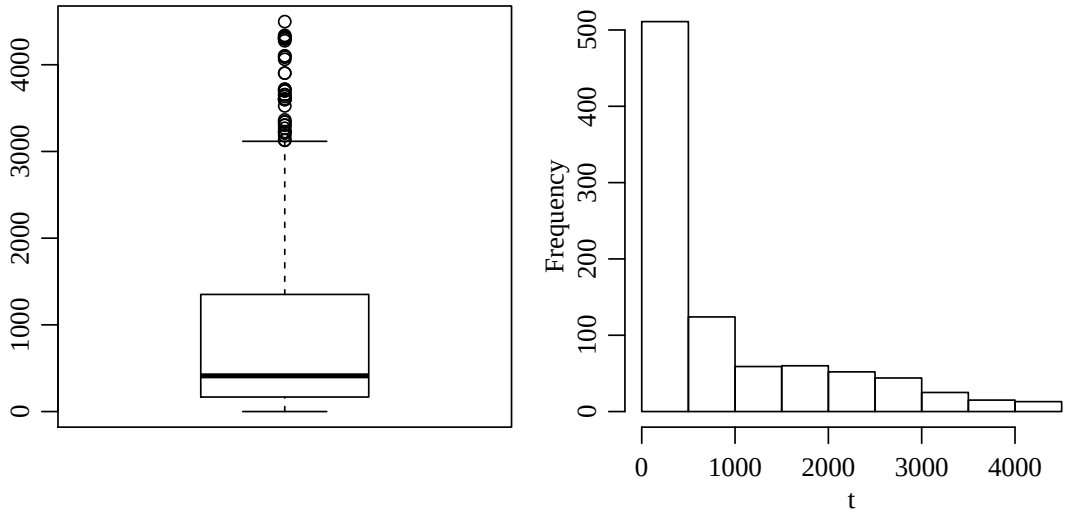

Figure 3.1: Boxplot and histogram of the survival times t.

Predictor `vol`

As for the predictor `vol`, the volume of each breast implant measured in cm^3, we obtain the following structure.

```
> summary(vol)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   80.0   215.0   260.0   276.4   325.0   700.0
```

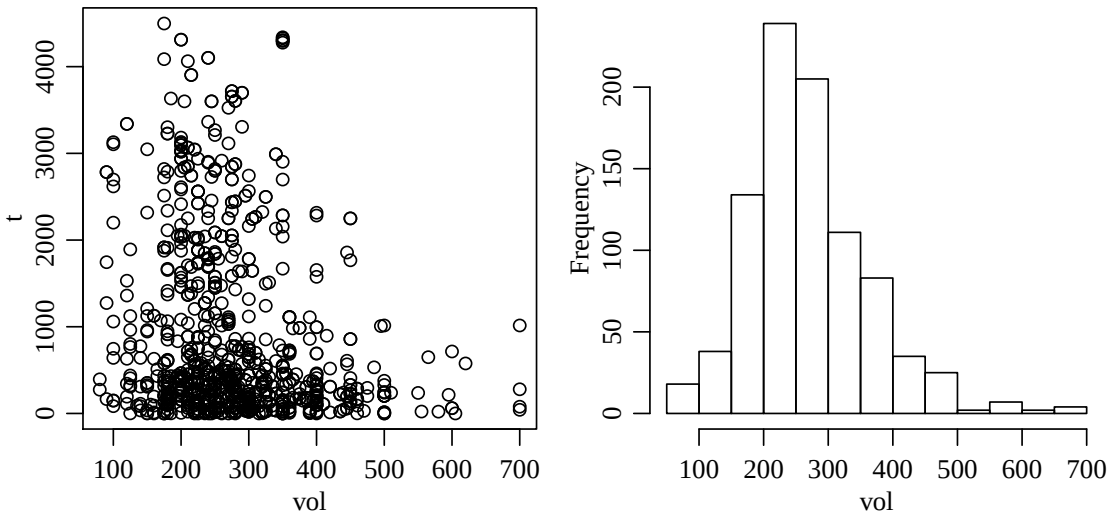

Figure 3.2: Survival times t in relation to volume and histogram of volume.

A clear first impression shows that implants with larger volumes tend to not last as long as implants with smaller volumes. As we can see in Figure 3.2, implants with volume over about 500 cm^3 tend to not survive until the 3 year mark. The longest survival times are achieved by implants with volumes between 200 and 350 cm^3.

Predictor `surface`

With regards to `surface`, we obtain the following frequencies:

```
> summary(surface)
     smooth polyurethan    textured
         10          36         857
```

Obviously, an overwhelming amount of implants appear to be of the `textured` variety, with only ten implants having a `smooth` surface. Thus, one should be careful not to interpret too much into the differences between the different surfaces. Still, one can observe that within Figure 3.3, the medians for the survival times of every level appear to be in the same ballpark.

```
> median(t[surface == "smooth"])
[1] 318.5
> median(t[surface == "polyurethan"])
[1] 244.5
> median(t[surface == "textured"])
[1] 416
```

Still, as it is expected with the massive difference in quantity between `textured` implants and the other two types of implants, the former shows more potential outliers. It is

interesting to note that for implants with a `smooth` surface or a `polyurethan` surface, the survival times tend to be on the shorter side compared to the third level. From that point of view alone, one could say that implants with a `textured` surface are to be preferred.

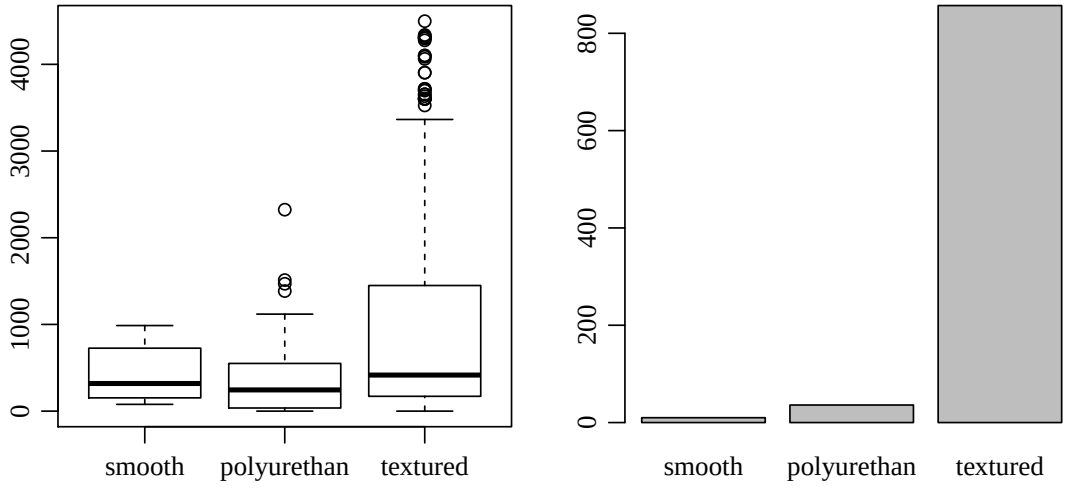

Figure 3.3: Survival times t in relation to surface and histogram of surface.

Predictor `lumen`

The predictor `lumen` indicates a special type of implant. As we can observe from the summary of this predictor, the overwhelming majority of considered implants are of the `single` variety.

```
> summary(lumen)
double single
    47    856
```

From a strictly visual point of view, not much can be said about `lumen`. All in all, we know that the interquartile range for the survival times of implants of the `double` variety seems to be much shorter than its counterpart, but we cannot say for sure that these differences cannot be lead back to the fact that there are more than 18 times as many implants of the type `single` than `double` in our dataset. Further inquiries into the significance of this predictor will have to be made later, even though judging by the plots alone it seems that implants of the `single` variety are preferable to their counterpart.

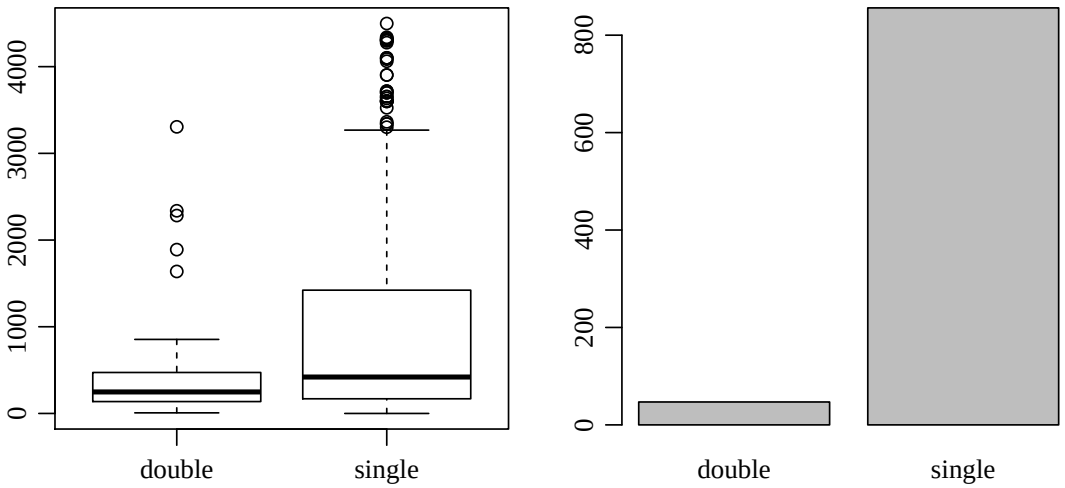

Figure 3.4: Survival times t in relation to lumen and histogram of lumen.

Predictor `fill`

The predictor `fill`, which specifies the type of filling for the associated implant, seems to be rather diverse as well. Again, we can observe that the vast majority of implants are filled with `siliconegel`.

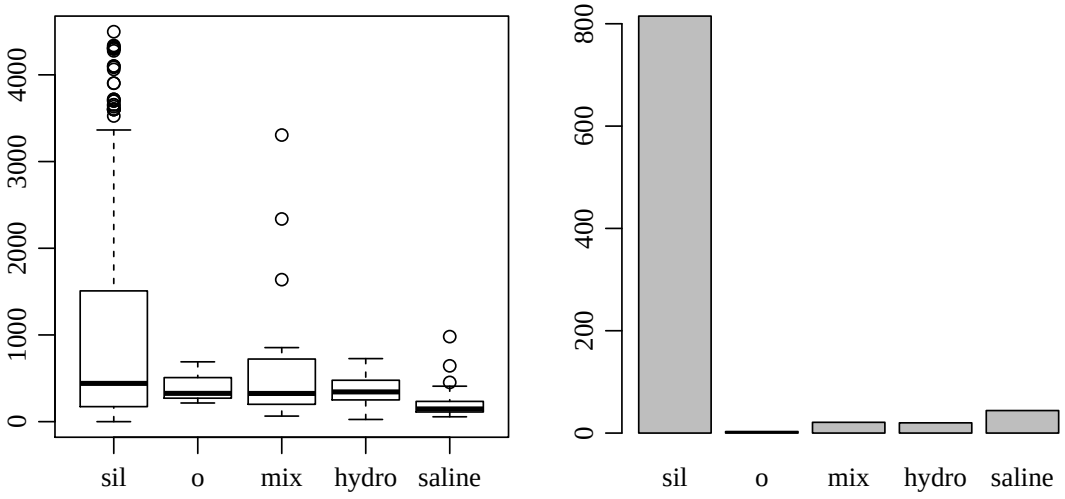

Figure 3.5: Survival times t in relation to fill and histogram of fill.

```
> summary(fill)
    siliconegel           other           mixed      hydrogel saline solution
            815               3              21              20              44
```

In Figure 3.5 we observe that implants with `saline solution` as a filling seem to tend to have shorter survival times than all the others. All in all, there seems to be a tendency of implants filled with `siliconegel` having the largest survival times. Again, such claims are to be taken with a grain of salt due to the large difference in amounts for each of the levels.

Predictor `drainage`

For the predictor `drainage`, which indicates whether during the initial surgery there was any drainage or not, we can see that there are far more observations with drainage than there are observations without.

```
> summary(drainage)
yes  no
743 160
```

In Figure 3.6, we see that the medians of both of these levels are in about the same range, whereas `no` proves to have a much shorter interquartile range than `yes`. Furthermore, we see that the bulk of observations without drainage lean towards shorter survival times with a few outliers– that leads us to believe that drainage is in fact beneficial to the survival time of an implant.

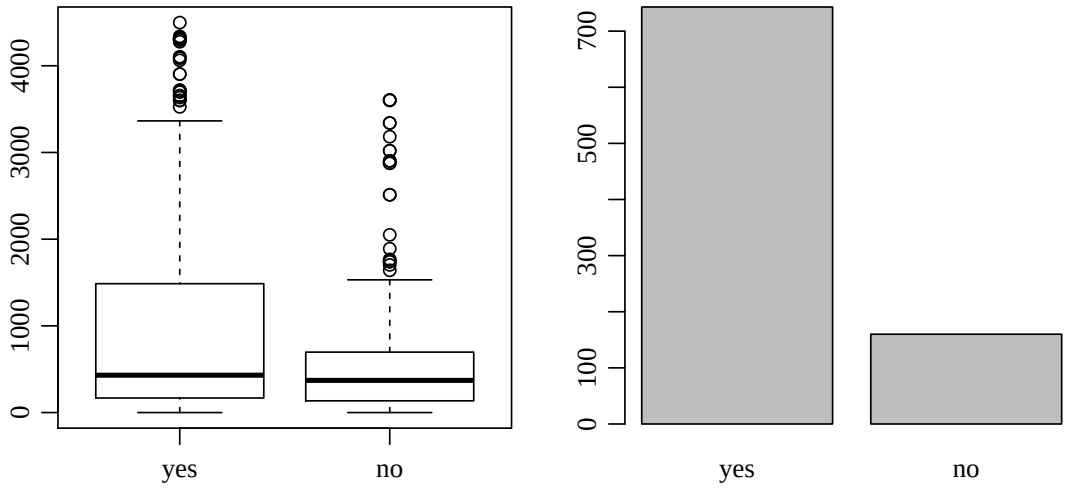

Figure 3.6: Survival times t in relation to drainage and histogram of drainage.

3.2 Cox Models

In this particular section, we will describe and analyze several models for the survival times `t`. Each one of these models includes the predictor `vol`, due to the high significance of volume as established in Section 2.3.1.

3.2.1 Volume and Surface

Recalling our first observations regarding the influence of `vol` and `surface` on the survival times `t`, we believe that the hazard will increase with increasing implant volume. Furthermore, we believe that `textured` implants will prove to be superior to implants of the `smooth` or `polyurethan` variety. In order to corroborate these assumptions, we consider a proportional Cox model that includes both of these predictors.

```
> surfacevolmod <- coxph(Surv(t) ~ vol + surface)
> anova(surfacevolmod)
Analysis of Deviance Table
 Cox model: response is Surv(t)
Terms added sequentially (first to last)

         loglik  Chisq Df Pr(>|Chi|)
NULL    -5246.9
vol     -5224.3 45.089  1  1.883e-11 ***
surface -5220.0  8.633  2    0.01335 *
```

The analysis of deviance table clearly states that both `surface` as well as `vol` are significant predictors. Their interaction however proves to be insignificant with a p-value of about 0.8 and thus we decide not to include it in further analysis.

Before we go into too much detail with regards to the results of the model, we first want to ensure that the model's assumptions are satisfied. Namely, we want to test the goodness of fit and the proportional hazards assumption. Furthermore, we will try to identify any outliers. In order to assure the goodness of fit, we consider the Cox-Snell residuals of the model. As one can see in Figure 3.7, the residuals follow the required straight line with unit slope quite nicely. We have also not detected any outliers, as indicated by the plot of the deviance residuals, shown in the right part of Figure 3.7 for the predictor `vol`. In total, the plots don't appear to be suspicious. As for the test of the proportional hazards assumption, we yet again use the function `cox.zph` to gain insight into the behavior of the Schoenfeld residuals.

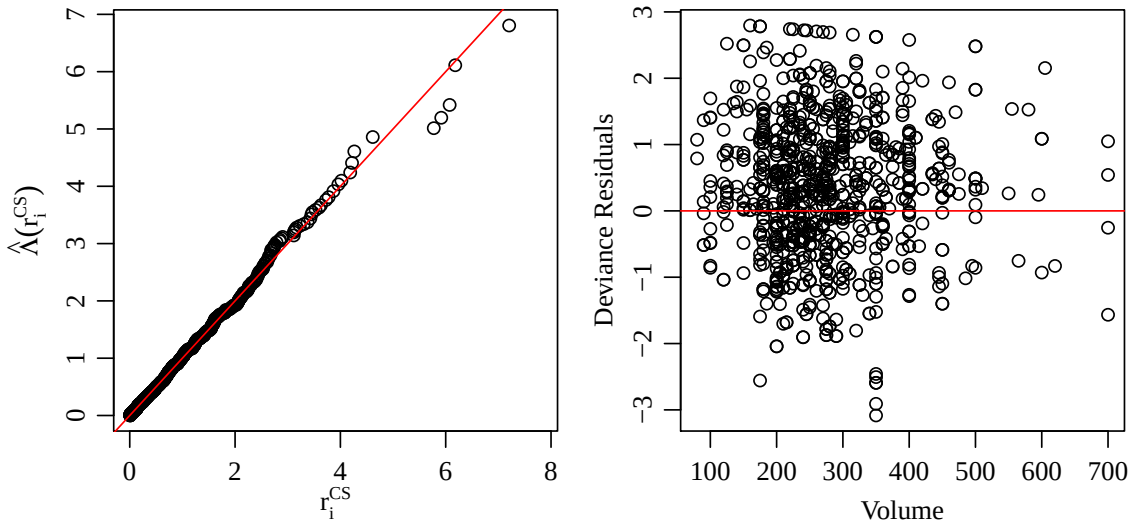

Figure 3.7: Cox-Snell residuals for surfacevolmod (left) and deviance residuals (right).

```
> cox.zph(surfacevolmod)
                         rho chisq      p
vol                  -0.0638  3.83 0.0505
surfacepolyurethan -0.0355  1.14 0.2856
surfacetextured    -0.0449  1.83 0.1759
GLOBAL                   NA  5.23 0.1557
```

None of the predictors show any suspicious significance levels in the Schoenfeld test, thus we cannot reject the assumption of proportional hazards with level $\alpha = 5\%$. Overall the model appears to adequately represent the survival times `t` and we can now go into further detail with regards to the results.

From the summary, we can conclude that under this particular model for `t` an increase in `vol` results in an increase in the hazard, just as previously assumed. As a matter of fact, an increase of one cm^3 in volume implies an increase of 0.2% in the hazard. An increase of 100 cm^3 in volume in turn translates to a 27.9% increase in hazard, which can be easily computed by considering $\exp(\hat{\beta}_{\texttt{vol}})^{100}$. The reference level for `surface` in this case is the level `smooth`. From the output we can conclude that an implant with a surface made out of `polyurethan` increases the hazard for said patient by about 22.7% in comparison to a `smooth` implant. Compared to a `textured` implant however, `smooth` is clearly inferior; in the case of a `textured` implant, the hazard reduces around 27%, just as expected.

```
> summary(surfacevolmod)
Call:
coxph(formula = Surv(t) ~ vol + surface)

  n= 903, number of events= 903
```

```
                          coef  exp(coef)   se(coef)      z Pr(>|z|)
vol                  0.0024645  1.0024675  0.0003691  6.677 2.44e-11 ***
surfacepolyurethan   0.2046081  1.2270441  0.3580238  0.571    0.568
surfacetextured     -0.3151719  0.7296634  0.3203187 -0.984    0.325
---
```

As seen in Figure 3.8, the survival function of `t` decreases much more quickly for implants with `smooth` and `polyurethan` surface in comparison to `textured`. One should note that Figure 3.8 shows the influence of `surface` on the survival times `t` without including `vol`. The probability of survival for an implant with a `textured` surface falls to about 50% after just about 500 days. After 1000 days however, only about a third of the subjects are left. In turn, patients with `smooth` implants suffer a much more difficult fate; none of those implants survive past the first 1000 days. Still, the massive difference in quantities in each level of `surface` makes it difficult to fault `smooth` or `polyurethan` implants too much. Still, we find that in conclusion implants of the `textured` variety are to be preferred.

```
> ggsurvplot(survfit(Surv(t) ~ surface, data=df), risk.table = TRUE)
```

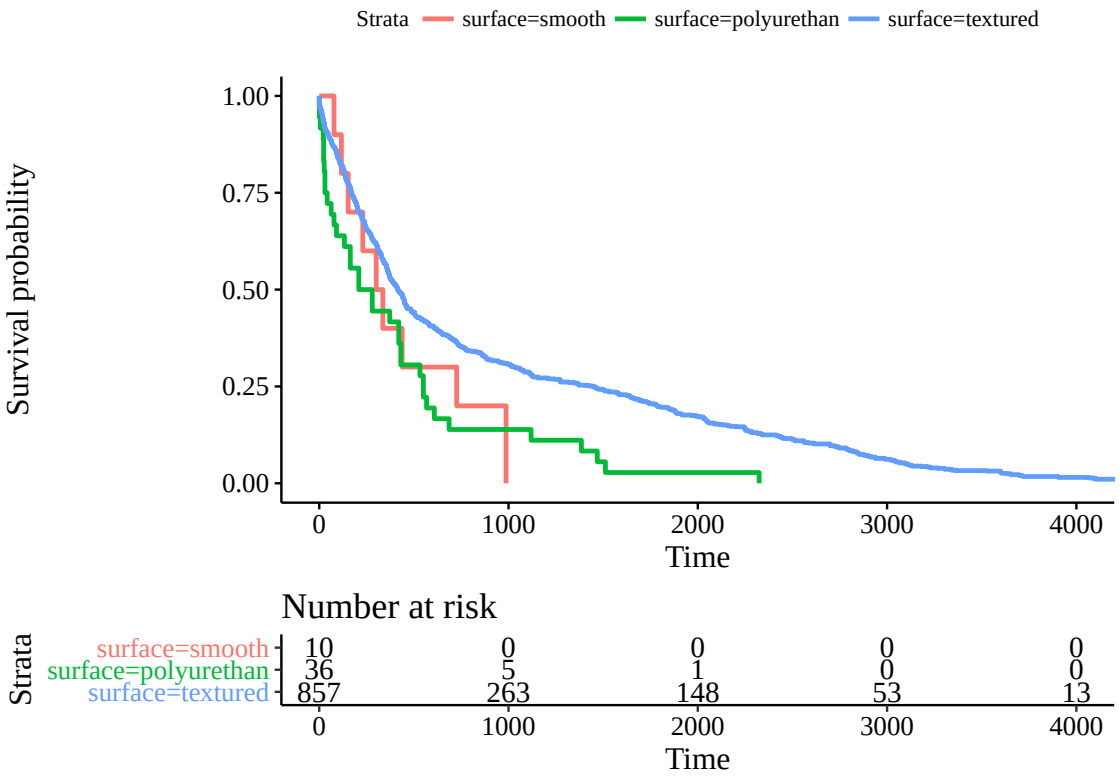

Figure 3.8: Survival function for t as a function of surface only.

3.2.2 Volume and Lumen

Judging by the discoveries of Section 3.1, the influence of `lumen` on the survival times `t` should favor implants of the `single` variety over their counterparts. We shall consider yet another model including the predictor `vol`, now adding the predictor of `lumen` into the model. Again, the interaction of these two predictors proves to be insignificant, which is why we stick to the simplified model.

```
> lumenvolmod <- coxph(Surv(t) ~ vol + lumen)
> anova(lumenvolmod)
Analysis of Deviance Table
 Cox model: response is Surv(t)
Terms added sequentially (first to last)

       loglik   Chisq Df Pr(>|Chi|)
NULL  -5246.9
vol   -5224.3 45.0891  1  1.883e-11 ***
lumen -5222.3  4.1263  1    0.04222 *
```

Before we want to continue our analysis of the model, we should check the assumptions. Both the deviance residuals as well as the Cox-Snell residuals are entirely inconspicuous, as one can see in Figure 3.9. With regards to the proportional hazards assumption, one can yet again refer to the output of the Schoenfeld tests, which show that for none of the predictors the hypothesis of proportional hazards can be rejected with level $\alpha = 5\%$. Thus, we feel confident in pursuing with an in-depth analysis of this model.

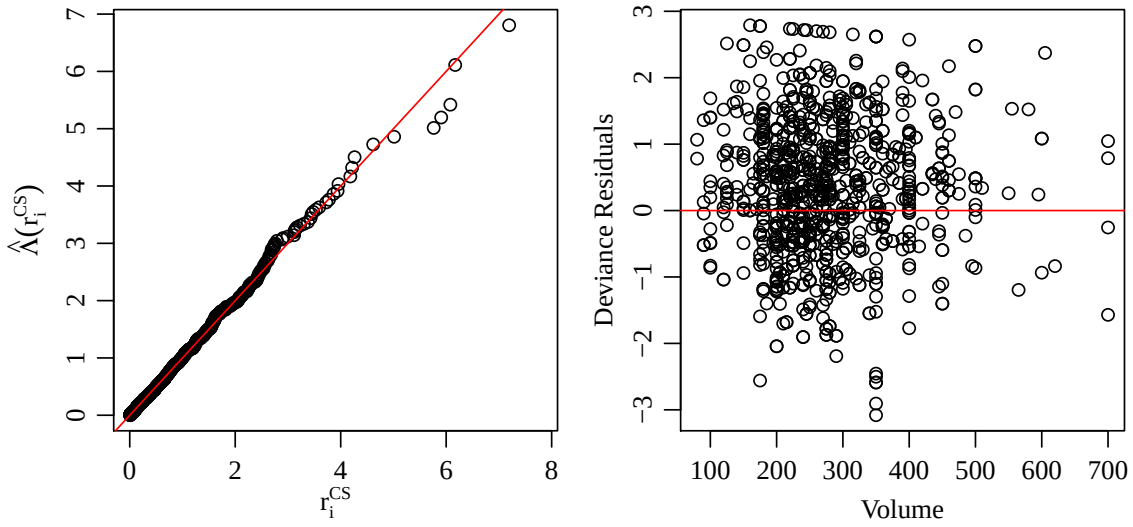

Figure 3.9: Cox-Snell residuals for lumenvolmod (left) and deviance residuals (right).

```
> cox.zph(lumenvolmod)
               rho chisq      p
```

```
vol         -0.0612 3.563 0.0591
lumensingle -0.0122 0.134 0.7141
GLOBAL           NA 3.569 0.1679
```

For this model, the reference level of `lumen` is `double`. With regards to the predictor `vol`, not much seems to change in comparison to the previously discussed model– yet again we can find that per increase of 100 cm^3, the hazard of the implant grows by about 27.7%, which is a subtle decrease compared to the model including both `vol` and `surface`. As expected, we find that implants of the form `single` prove to be the significantly better choice ($p = 0.0337$)– compared to a `double` implant, a `single` shows a 27.6% decrease in the hazard.

```
> summary(lumenvolmod)
Call:
coxph(formula = Surv(t) ~ vol + lumen)

  n= 903, number of events= 903

                coef exp(coef)  se(coef)      z Pr(>|z|)
vol          0.002451  1.002454  0.000372  6.589 4.44e-11 ***
lumensingle -0.323640  0.723510  0.152374 -2.124   0.0337 *
---

Concordance= 0.575  (se = 0.011 )
Rsquare= 0.053   (max possible= 1 )
Likelihood ratio test= 49.22  on 2 df,   p=2.056e-11
Wald test            = 53.55  on 2 df,   p=2.359e-12
Score (logrank) test = 53.54  on 2 df,   p=2.362e-12
```

Taking a closer look into the survival function of `t` dependent on the predictor `lumen` in Figure 3.10, which shows the behavior of the survival function of `t` in relation to `lumen` only, we find that the survival function for `double` dips much lower than the survival function of `single` within the first 1000 days. This fact can be lead back to the fact that out of 47 implants only 5 remain intact the first 1000 days, which is barely more than 10%. In the case of `single` implants, about 30% of implants remain intact beyond the first 1000 days. Yet again the large difference in observations for these levels suggests that `double` might not be as bad as the model makes it out to be, but still implants of the type `single` and with lower volume are to be preferred to their counterparts.

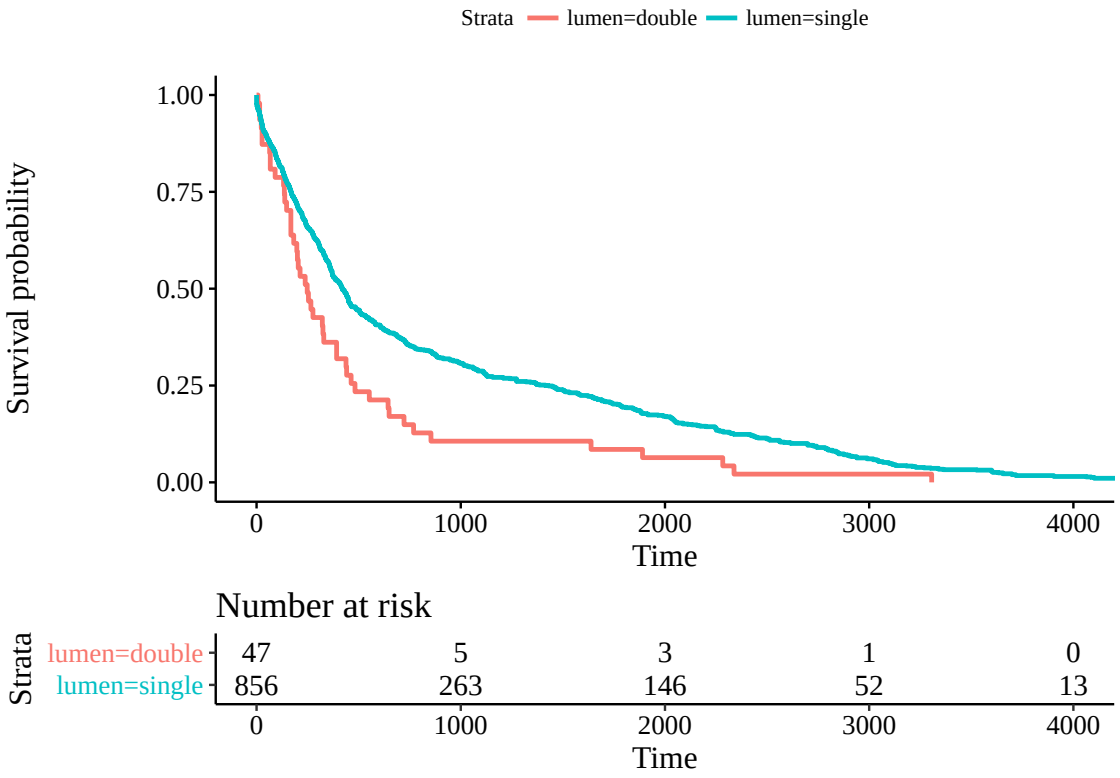

Figure 3.10: Survival function for t as a function of lumen only.

3.2.3 Volume and Filling

The predictor `fill` which describes the filling of its associated implant, has already shown in Section 3.1 that there are massive differences in the quantities of each level. Keeping that in mind, we continue to consider the model `fillvolmod`, which includes both `vol` as well as `fill` as predictors. We also test the signficance of the interaction of those two predictors. Despite the p-value of 0.09 (not shown), we decide to not include it in further analysis for simplicity's sake.

```
> fillvolmod <- coxph(Surv(t) ~ vol + fill)
> anova(fillvolmod)
Analysis of Deviance Table
 Cox model: response is Surv(t)
Terms added sequentially (first to last)

      loglik  Chisq Df Pr(>|Chi|)
NULL -5246.9
vol  -5224.3 45.089  1  1.883e-11 ***
fill -5208.0 32.757  4  1.340e-06 ***
---
```

When checking the model's assumptions we run into a few problems. While neither the Cox-Snell residuals (compare Figure 3.11) nor the deviance residuals appear to be too suspicious, we realize that the proportional hazards assumption isn't fulfilled for each level of `fill`.

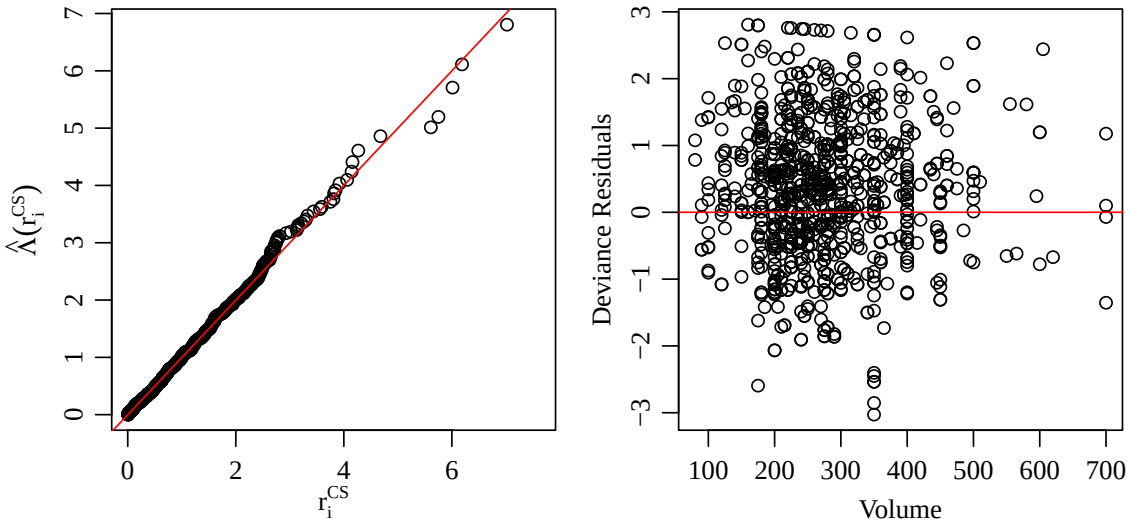

Figure 3.11: Cox-Snell residuals for fillvolmod (left) and deviance residuals (right).

```
> cox.zph(fillvolmod)
                        rho chisq      p
vol                 -0.0640  4.01 0.0452
fillother            0.0435  1.71 0.1913
fillmixed            0.0388  1.35 0.2452
fillhydrogel         0.0803  5.91 0.0151
fillsaline solution  0.0706  4.84 0.0278
GLOBAL                   NA 13.95 0.0159
```

Considering the output of `cox.zph` for the model `fillvolmod`, we realize that not all levels of the predictor fulfill the proportional hazards assumption. In fact we reject the assumption for the predictors `vol`, `hydrogel`, `saline solution` as well as for the overall model with $\alpha = 5\%$. Thus we cannot confidently state that the model is adequate. Still, we want to take a closer look at the associated estimates $\hat{\beta}$ as well as the influence of `fill` on the survival function of `t`, always keeping in mind that these observations are to be taken with a grain of salt.

```
> summary(fillvolmod)
Call:
coxph(formula = Surv(t) ~ vol + fill)

  n= 903, number of events= 903
```

```
                         coef exp(coef)  se(coef)     z Pr(>|z|)
vol                 0.0021935 1.0021959 0.0003817 5.746 9.14e-09 ***
fillother           0.1541564 1.1666733 0.5829298 0.264   0.7914
fillmixed           0.1281474 1.1367206 0.2225255 0.576   0.5647
fillhydrogel        0.4869177 1.6272927 0.2286492 2.130   0.0332 *
fillsaline solution 1.0043824 2.7302205 0.1637388 6.134 8.57e-10 ***
---

Concordance= 0.585  (se = 0.011 )
Rsquare= 0.083   (max possible= 1 )
Likelihood ratio test= 77.85  on 5 df,   p=2.331e-15
Wald test            = 94  on 5 df,   p=0
Score (logrank) test = 100.6  on 5 df,   p=0
```

Yet again we find that a 100 cm^3 increase in volume results in an increase of the hazard. This time we have a lower increase of only about 24.52%. As for the different levels of `fill` whose reference level is `siliconegel`, we see that indeed all other levels appear to be inferior. Compared to `siliconegel`, fillings of the type `mixed` and `other` only show a small increase of the hazard of 13.67% and 16.67% respectively. In that case however, one should not take the increase for `other` too seriously, as there are only 3 observations in total. `Hydrogel` shows a significant increase of about 62.73% in the hazard in comparison to `siliconegel`– still, the clear worst choice in this model is a filling made out of `saline solution`. Those actually show a 173% increase in the hazard.

Considering Figure 3.12, which shows the behavior of the survival function of `t` in relation to `fill` without considering `vol`, we see why `saline solution` seems to be so much worse than the other fillings. While for `saline solution`, `other` as well as `hydrogel` no implant remains intact beyond the first 1000 days, `saline solution` has almost double the amount of observations than the other two combined. We can yet again see why `other` should not be taken too seriously as a level within this model, due to the fact that there are only 3 observations in that category. Contrary to that, `mixed` implants tend not to last very long either– after the first 1000 days only 3 out of 21 implants are left intact, only one of which survives beyond the 3000 day mark. Furthermore, we notice that implants filled with `siliconegel` are by far the sturdiest, since even after 4000 days there are still 13 intact, despite the amount of intact implants being reduced to a third after 1000 days.

In conclusion, we need to remind ourselves that the analysis of `fillvolmod` is to be taken with a grain of salt. With the proportional hazards assumption not fulfilled for each level of the predictor `fill`, we cannot confidently stand behind the calculated percentages; still, the overall conclusion of `siliconegel` implants being the clear superior type of implants in this instance seems to be true regardless.

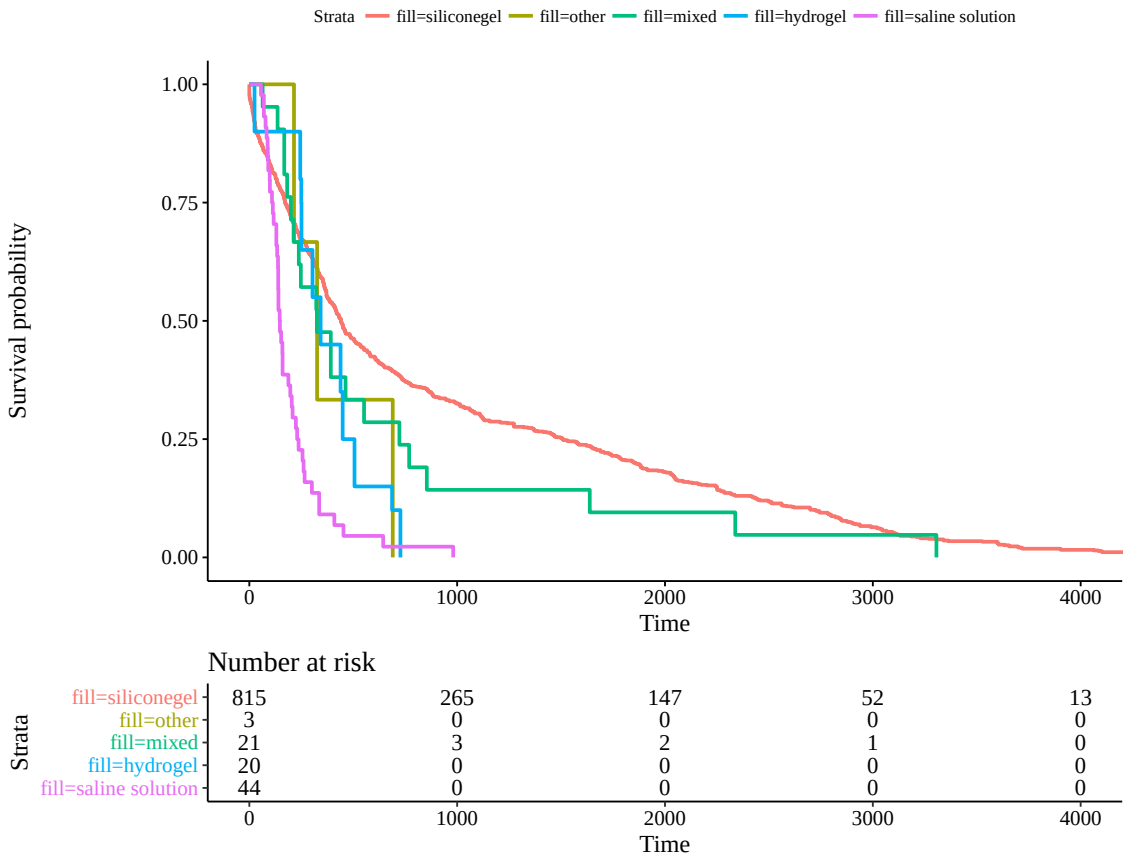

Figure 3.12: Survival function for t as a function of the filling only.

3.2.4 Volume and Drainage

Initial observations regarding the predictor `drainage` in Section 3.1 lead us to believe that it is beneficial to the survival time `t` of a certain implant to have had a `drainage` during the primary surgery. Now we want to corroborate these observations by fitting a model for `t` using the predictors `vol` and `drainage`. Including the interaction of the predictors proves to be insignificant, which is why we stick to the simple model below.

```
> drainagevolmod <- coxph(Surv(t) ~ vol + drainage)
> anova(drainagevolmod)
Analysis of Deviance Table
 Cox model: response is Surv(t)
Terms added sequentially (first to last)

          loglik   Chisq Df Pr(>|Chi|)
NULL     -5246.9
vol      -5224.3 45.0891  1  1.883e-11 ***
drainage -5221.2  6.3105  1      0.012 *
```

The model appears to be adequate when tested for Goodness of Fit as well as outliers using the usual methods described in Section 2.3.2. Furthermore, we find that the Schoenfeld residuals appear to be independent of `t` for each predictor as shown in the output of `cox.zph`. The p-values for each predictor as well as the p-value for the overall model show that we cannot reject the proportional hazard assumption with $\alpha = 5\%$. In conclusion, `drainagevolmod` represents the survival times `t` adequately.

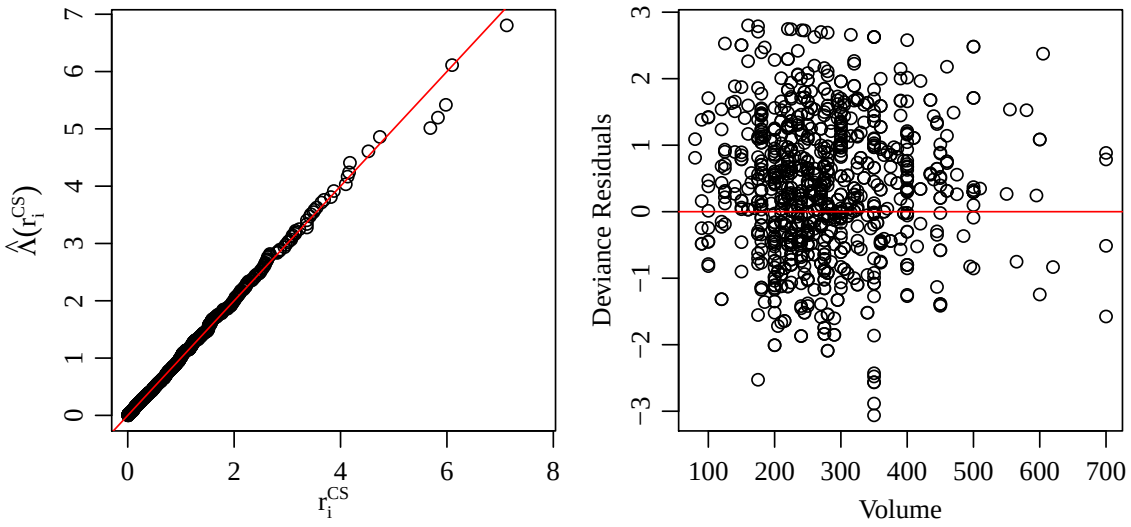

Figure 3.13: Cox-Snell residuals for drainagevolmod (left) and deviance residuals (right).

```
> cox.zph(drainagevolmod)
               rho chisq      p
vol        -0.0592 3.273 0.0704
drainageno  0.0114 0.118 0.7312
GLOBAL          NA 3.339 0.1884
```

Again we find that an increase in `vol` is counterproductive to the survival times `t`. In this case however, the hazard grows slightly stronger than it did in the other models– for an increase of 100 cm^3 in volume, the hazard increases by 28.58%. As previously expected, having no `drainage` is not beneficial to the hazard. In this case, the reference level for `drainage` is `yes`, which means that having `no` drainage increases the hazard immediately by 25.36%. Indeed, this change of the hazard is highly significant with $p = 0.0101$.

```
> summary(drainagevolmod)
Call:
coxph(formula = Surv(t) ~ vol + drainage)

  n= 903, number of events= 903

                coef exp(coef)  se(coef)     z Pr(>|z|)
```

```
vol         0.0025138 1.0025170 0.0003664 6.861 6.84e-12 ***
drainageno 0.2260197 1.2536003 0.0878577 2.573   0.0101 *
---

Concordance= 0.578  (se = 0.011 )
Rsquare= 0.055   (max possible= 1 )
Likelihood ratio test= 51.4  on 2 df,   p=6.898e-12
Wald test            = 55.55  on 2 df,   p=8.668e-13
Score (logrank) test = 55.27  on 2 df,   p=9.955e-13
```

It is interesting to note that in Figure 3.14, which shows the survival function of `t` in relation to `drainage` by itself, the survival function for `drainage` as `no` runs similarly to the function for `yes` at the very beginning and at the very end. After about a year, the survival function for `no` dips below its counterpart, yet they meet again around the 3000 day mark. After 1000 days, the number of intact implants with `drainage yes` reduces to about a third, meanwhile only about 19% are left for `drainage no` at that time. At about 3000 days, the both survival functions have a similar amount of intact implants left with about 6% for `yes` and 5% for `no`. In conclusion we find that having `drainage` in the initial surgery is to be preferred.

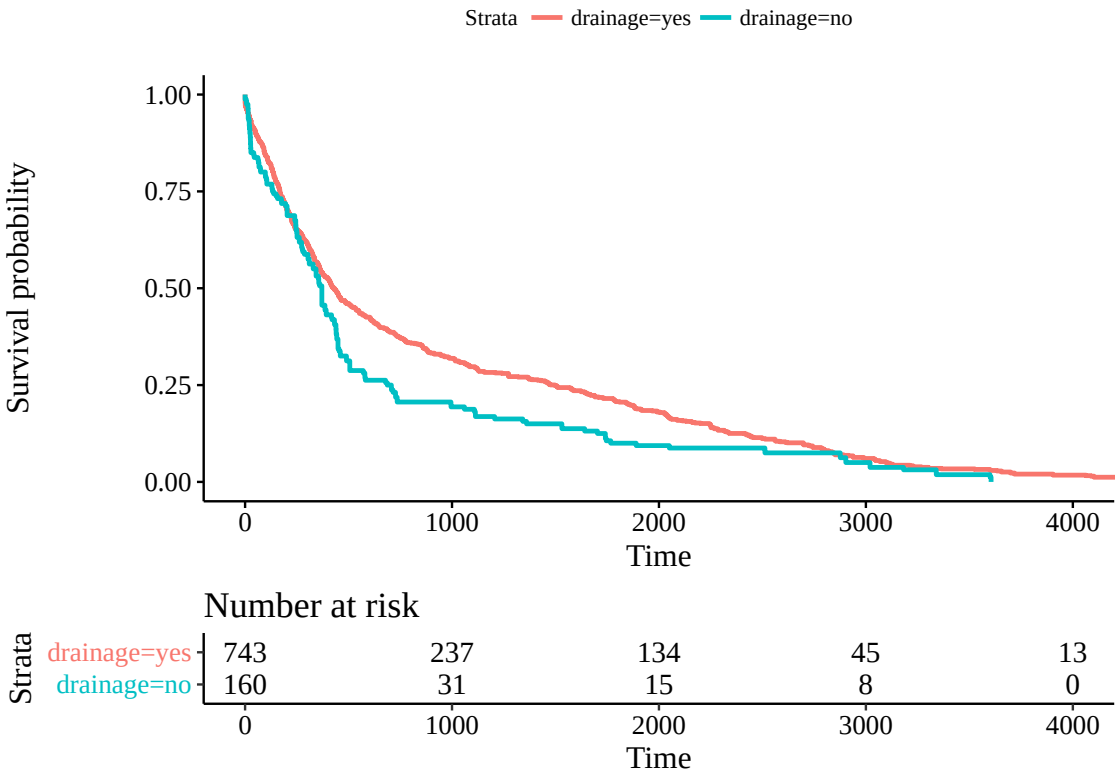

Figure 3.14: Survival function for t as a function of the drainage only.

3.2.5 Full Model

After including all of the different predictors individually, one is tempted to consider a model that includes all predictors at once. Indeed, such a model proves to be highly significant– every mentioned predictor (`vol`, `surface`, `lumen`, `fill` and `drainage`) is highly significant to the model, as has been checked diligently by utilizing the `step` function provided by `R` (not shown). The resulting model is given below.

```
> finalmod <- coxph(Surv(t) ~ vol + surface + lumen + fill + drainage)
> anova(finalmod)
Analysis of Deviance Table
 Cox model: response is Surv(t)
Terms added sequentially (first to last)

          loglik   Chisq Df Pr(>|Chi|)
NULL     -5246.9
vol      -5224.3 45.0891  1  1.883e-11 ***
surface  -5220.0  8.6330  2   0.013346 *
lumen    -5217.6  4.7681  1   0.028992 *
fill     -5200.7 33.8048  4  8.171e-07 ***
drainage -5197.3  6.7995  1   0.009118 **

> summary(finalmod)
Call:
coxph(formula = Surv(t) ~ vol + surface + lumen + fill + drainage)

  n= 903, number of events= 903

                          coef  exp(coef)   se(coef)      z Pr(>|z|)
vol                  0.0020165  1.0020186  0.0003838  5.254 1.49e-07 ***
surfacepolyurethan   0.6564823  1.9279983  0.3690523  1.779  0.07527 .
surfacetextured      0.0157394  1.0158639  0.3299402  0.048  0.96195
lumensingle         -0.4356794  0.6468251  0.2061193 -2.114  0.03454 *
fillother            0.2683088  1.3077509  0.5835382  0.460  0.64566
fillmixed           -0.2472961  0.7809094  0.2991898 -0.827  0.40849
fillhydrogel         0.3266258  1.3862827  0.2408827  1.356  0.17511
fillsaline solution  1.0236867  2.7834374  0.1721053  5.948 2.71e-09 ***
drainageno           0.2501705  1.2842444  0.0933700  2.679  0.00738 **
```

Due to the nature of the `anova` function, we are not given the right p-values in its output. Indeed, when interpreting the output of `anova`, one can only look at the very last line, whose p-value indicates the significance of the associated predictor given all above mentioned predictors are already in the model. In order to get the actual signficance of each predictor, we can utilize the function `drop1`, which performs a series of likelihood ratio tests in order to test the relevance of each predictor by comparing the likelihood of the model including said predictor and the model without that predictor. Thus, the actual significance levels of each predictor in this model can be found in the output below.

```
> drop1(finalmod, test="Chisq")
Single term deletions

Model:
Surv(t) ~ vol + surface + lumen + fill + drainage
         Df   AIC    LRT  Pr(>Chi)
<none>      10413
vol       1 10437 26.076 3.282e-07 ***
surface   2 10420 11.481  0.003213 **
lumen     1 10415  3.971  0.046300 *
fill      4 10437 32.746 1.346e-06 ***
drainage  1 10418  6.799  0.009118 **
```

The model itself doesn't appear all too conspicuous, but when it comes to checking the model assumptions, we realize that the proportional hazards assumption is not fulfilled for the entire model– indeed, it is fulfilled for only one level of one predictor, which of course means that we cannot truly use this model to express the influence of each predictor on the hazard.

```
> cox.zph(mod)
                        rho   chisq      p
vol                 -0.06237  3.7709 0.0522
surfacepolyurethan  -0.03871  1.3376 0.2475
surfacetextured     -0.04895  2.1514 0.1424
lumensingle          0.00966  0.0873 0.7676
fillother            0.04265  1.6402 0.2003
fillmixed            0.03460  1.0959 0.2952
fillhydrogel         0.07989  5.8088 0.0159
fillsaline solution  0.05483  2.9329 0.0868
drainageno          -0.01072  0.1035 0.7477
GLOBAL                   NA 15.7257 0.0728
```

The `cox.zph` output shows that the proportional hazards assumption is violated for almost every level of every predictor. In an effort to find a different method to fit a survival model for the breast implant data, we abandon the non-parametric world and attempt parametric models instead. These prove to be a bit more restrictive in their assumptions, given that they assume a specific distribution for the survival times T– whether the restrictions pay off in the end remains to be seen in Chapter 5.

Chapter 4

Parametric Survival Time Models

As the name suggests these types of regression models fit survival time data by assuming that the survival times follow a certain distribution. Contrary to nonparametric survival models, these models focus on estimating the full distribution of the survival times as a byproduct of the estimates for the influence of predictor variables, whereas nonparametric survival models are not interested in the actual shape of the hazard function. We will largely follow Kalbfleisch and Prentice (1980) in this chapter, as they provide the theoretical foundation for the associated `R` functions we will be using later on. Also, Lawless (2003) and Liu (2012) provide most of the necessary theory regarding accelerated failure time models.

4.1 Common Life Time Distributions

We are yet again going to focus our attention on survival times, where by T we denote the random variable associated with the survival times and t is a realization of T. Logic dictates that $T \geq 0$ has to hold. In the following sections, we will take a look at a few classic distributions used to estimate survival times. Furthermore, we will attempt to find out whether any of these distributions can be used to fit our breast implant data. For that purpose, let $f(t)$ denote the density function of T, $S(t)$ the survival function of T and $\lambda(t)$ the hazard function. Again, the hazard has to be a non-negative function mapping onto $[0, 1]$ as already described in Chapter 2.

It is common practice to consider a logarithmic transformation of the survival times. Indeed, all models we will discuss utilize this transformation in order to find a good fit for the survival times.

4.1.1 Extreme Value Distribution

Before we focus on distributions that will help us describe the behavior of T itself, we want to introduce the extreme value distribution, which will be useful in terms of modeling $\log T$.

For a random variable Y that follows an extreme value distribution with parameters $u \in \mathbb{R}$

for location and $b > 0$ for scale

$$f(y) = \frac{1}{b}\exp\left(\frac{y-u}{b} - e^{\frac{y-u}{b}}\right)$$
$$S(y) = \exp(-e^{\frac{y-u}{b}})$$
$$\lambda(y) = \frac{1}{b}e^{\frac{y-u}{b}}$$

holds. Also commonly referred to as the Gumpel distribution, the extreme value distribution holds great value when it comes to parametrically modeling survival times, as will be explained later on. It is worth noting that for $u = 0$ and $b = 1$ the standard extreme value distribution follows, for which

$$f(y) = e^{y-e^y}$$
$$S(y) = e^{-e^y}$$
$$\lambda(y) = e^y$$

hold for obvious reasons.

4.1.2 Exponential Distribution

As a classic distribution used for survival times, the exponential distribution is known to be a good way to start. With the parameter $\lambda > 0$, which is not to be confused with the actual hazard function $\lambda(t)$, it is well-known that for T following an exponential distribution, the density $f(t)$, the survival function $S(t)$ and the hazard function $\lambda(t)$ are given by

$$f(t) = \lambda e^{-\lambda t}$$
$$S(t) = e^{-\lambda t}$$
$$\lambda(t) = \lambda.$$

The fact that the hazard function does not depend on the time t shows why the exponential distribution is referred to as a "memoryless" distribution– indeed, the hazard for a subject that has been in the trial for a longer period of time is equal to the hazard for a subject that just entered surveillance.

In order to empirically check whether a set of observed survival times t follows an exponential distribution, one could plot the logarithm of an estimate of the survival function (for example the Kaplan-Meier Estimate explained in Section 2.2) against t. This follows directly from the fact that

$$S(t) = e^{-\lambda t} \Rightarrow \log(S(t)) = -\lambda t$$

holds. Thus, if the associated random variable T follows an exponential distribution, the resulting plot will be a straight line through the origin.

Armed with that knowledge, we want to now see whether the survival times t from our breast implant data is exponentially distributed. It is advisable to make sure to check the survival times beforehand, since there is the possibility of an accidental $\log 0$ happening; thus, we have excluded such entries in the `lowess` function we use to smooth.

```
> kme <- survfit(Surv(t) ~ 1)
> plot(kme$time, -log(kme$surv), xlab = "time", ylab = "-log(S(t))")
> lines(lowess(kme$time[1:length(kme$surv)-1],
+              -log(kme$surv)[1:length(kme$surv)-1]), col="red")
```

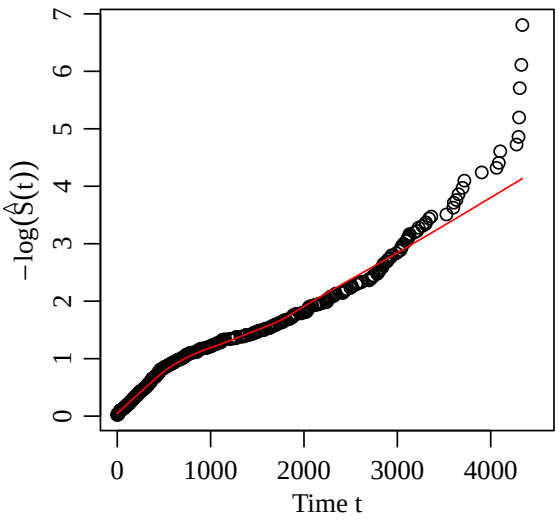

Figure 4.1: Plot of $-\log(\hat{S}(t))$ against t to check for exponential distribution.

As we can see in Figure 4.1, the breast implant survival times t could very well follow an exponential distribution, as they resemble a straight line through the origin. The plot is far from perfect though, since one can also see that towards the end of the plot there seems to be a curve in the points, leading away from the reference line. Other distributions might fit better, but the exponential distribution is a good place to start.

Since in parametric survival models a logarithmic transformation of T is often used as response, let us now consider $Y = \log T$ and its distribution. By utilizing the transformation theorem, given that $T \sim Exp(\lambda)$, we obtain the following density for Y:

$$\begin{aligned} f_Y(y) &= \lambda e^{-\lambda e^y} e^y \\ &= \exp(\log\lambda - \lambda e^y + y) \\ &= \exp(y + \log\lambda - e^{y+\log\lambda}) \\ &= \exp((y - (-\log\lambda)) - e^{y-(-\log\lambda)}). \end{aligned}$$

This however is the density of an extreme value distribution with location parameter $-\log\lambda$ and scale 1. Thus, we conclude that for a standard extreme value distributed random variable W,

$$Y = (-\log\lambda) + W$$

follows.

4.1.3 Weibull Distribution

A generalization of the exponential distribution that allows for more flexibility in terms of the hazard function is the Weibull distribution, characterized by two parameters $\lambda, p > 0$. For Weibull distributed T, we know that

$$\begin{aligned} f(t) &= \lambda^p p t^{p-1} e^{-(\lambda t)^p} \\ S(t) &= e^{-(\lambda t)^p} \\ \lambda(t) &= \lambda^p p t^{p-1} \end{aligned}$$

holds. Thus, we see that the hazard in this case is influenced by t in a polynomial manner. One can easily see that the exponential distribution is a special case of the Weibull distribution for $p = 1$. In order to check whether a set of observed survival times t follows a Weibull distribution, we consider $S(t)$, for which we know that

$$S(t) = e^{-(\lambda t)^p} \Rightarrow \log(S(t)) = -(\lambda t)^p \Rightarrow \log\left(-\log(S(t))\right) = p \log \lambda + p \log t$$

holds. Thus, we realize that for a Weibull distributed random variable, $\log\left(-\log(S(t))\right)$ should be a linear function in $\log t$. Thus, we can easily use the Kaplan Meier estimate for $S(t)$ and plot $\log\left(-\log(S(t))\right)$ against $\log t$. If the plot resembles a straight line, a Weibull distribution for T seems justified.

Since the breast implant survival times t already appear to follow an exponential distribution quite nicely, we don't expect them to misbehave when it comes to the comparison to a Weibull distribution, since it is a generalization for the exponential distribution.

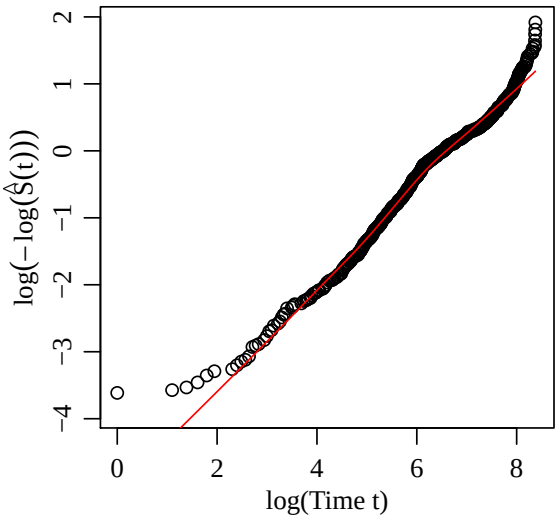

Figure 4.2: Plot of $\log(-\log(\hat{S}(t)))$ against $\log t$ to check for Weibull distribution.

```
> plot(log(kme$time), log(-log(kme$surv)),
```

```
+      xlab = "log(time)", ylab = "log(-log(S(t)))")
> lines(lowess(log(kme$time)[2:(length(kme$surv)-1)],
+              log(-log(kme$surv))[2:(length(kme$surv)-1)]), col="red")
```

As expected, we see that the survival times t follow a Weibull distribution very well.

Again we would like to classify the density of $Y = \log T$ for $T \sim Weibull(\lambda, p)$. Utilizing the transformation theorem yet again, we obtain

$$\begin{aligned} f_Y(y) &= p\lambda(e^y\lambda)^{p-1}e^{-(\lambda e^y)^p}e^y \\ &= p\exp(\log\lambda + (p-1)y + (p-1)\log\lambda - e^{p(y+\log\lambda)} + y) \\ &= p\exp(p(y+\log\lambda) - e^{p(y+\log\lambda)}). \end{aligned}$$

Again, we are faced with an extreme value distribution with location parameter $-\log\lambda$ and scale parameter p^{-1}. Thus, we conclude that for a standard extreme value distributed W, we have

$$Y = (-\log\lambda) + p^{-1}W.$$

4.2 Accelerated Failure Time Models

While there are plenty of possibilities to utilize a parametric approach when it comes to modeling survival times, we want to put specific focus on accelerated failure time models, from now on referred to as AFT models. Such models tend to be rather flexible when it comes to picking a proper underlying distribution for T and the process of fitting them is rather straight-forward. In the following theoretical exploration of AFT models, we follow the example of Liu (2012) and Lawless (2003) closely.

Suppose for a survival time T_i there exists a set of $p-1$ predictor variables X_i and we're interested in the relationship between T_i and X_i for $i = 1, \ldots, n$. Let X be the $p \times n$ design matrix, namely the matrix resulting with X_i as i-th column of said matrix. Let T be the random variable vector formed by all n individual survival times. For simplicity's sake, let the first row of X consist of a 1-vector of size $1 \times n$ in order to include an intercept for the model automatically. Then suppose that the following equation holds

$$\log T = Y = X^T\beta + \sigma W,$$

where σ is an unknown scaling parameter, W follows a known parametric distribution entirely independent of X and β refers to the vector of the regression coefficients, which will provide insight into how the predictors X factor into the behavior of Y and thus also T.

We will now provide two alternative yet equivalent depictions for the hazard function $\lambda(t|x)$, the first of which will give proper insight into why these models are referred to as *accelerated* failure time models. To that end, observe that

$$T = e^{X^T\beta}e^{\sigma W} = e^{X^T\beta}\widetilde{W}$$

holds for obvious reasons, where $\widetilde{W} = e^{\sigma W}$. Then the survival function $S(t|X)$ results by

$$\begin{aligned} S(t|X) &= P(T \geq t) \\ &= P(\exp(X^T\beta + \sigma W) \geq t) \\ &= P(\exp(\sigma W) \geq t \exp(-X^T\beta)) \\ &= P(\widetilde{W} \geq t \exp(-X^T\beta)) \\ &= S_{\widetilde{W}}(t \exp(-X^T\beta)), \end{aligned}$$

where $S_{\widetilde{W}}(w)$ refers to the survival function for the random variable $\widetilde{W}$, whose distribution is known. As a result, due to the close relationship between hazard functions and survival functions, we obtain

$$\lambda(t|X) = \lambda_{\widetilde{W}}(t \exp(-X^T\beta)) \exp(-X^T\beta)$$

and consequently

$$S(t|X) = \exp(-\Lambda_{\widetilde{W}}(t \exp(-X^T\beta))).$$

Here, $\lambda_{\widetilde{W}}(w)$ and $\Lambda_{\widetilde{W}}(w)$ refer to the hazard function and the cumulated hazard of $\widetilde{W}$, respectively. The important bit of information here is seen in the way $\lambda(t|X)$ is depicted. For $\exp(-X^T\beta) < 1$, X decelerates the failure time, whereas for $\exp(-X^T\beta) > 1$ we see an acceleration of the failure, which is where AFT models get their name from.

Equivalently, we characterize $S(t|X)$ by considering

$$\begin{aligned} S(t|X) &= P(\log T \geq \log t) \\ &= P(X^T\beta + \sigma W \geq \log t) \\ &= P\Big(W \geq \frac{\log t - X^T\beta}{\sigma}\Big) \\ &= S_W\Big(\frac{\log t - X^T\beta}{\sigma}\Big). \end{aligned}$$

Yet again, $S_W(w)$ refers to the survival function associated with the random variable W, whose parametric distribution we assume to know; for Weibull distributed T of course, W would follow a standard extreme value distribution for example. By the definition of the cumulative hazard $\Lambda(t)$, we obtain

$$\Lambda(t|X) = -\log S(t|X) = -\log S_W\Big(\frac{\log t - X^T\beta}{\sigma}\Big) = \Lambda_W\Big(\frac{\log t - X^T\beta}{\sigma}\Big) \tag{4.1}$$

and finally by differentiating the cumulative hazard, we get

$$\lambda(t|X) = \frac{1}{t\sigma}\lambda_W\Big(\frac{\log t - X^T\beta}{\sigma}\Big). \tag{4.2}$$

Furthermore, by considering the relationship between hazard function and density function as well as remembering that Y is defined as $Y = \log T$, we obtain

$$f(t|x) = \frac{1}{t\sigma} f_W\Big(\frac{\log t - X^T\beta}{\sigma}\Big)$$

as well as

$$f(y|x) = \frac{1}{\sigma} f_W\Big(\frac{y - X^T\beta}{\sigma}\Big).$$

The latter follows from simply repeating the above mentioned steps for $S(y|x)$ as well as $\lambda(y|x)$ respectively. These functions tie heavily into the process of estimation for both the regression coefficients β as well as the scale σ.

In order to estimate the unknown parameters, let $y_i = \log t_i$ and define

$$z_i = \frac{y_i - x_i^T\beta}{\sigma}.$$

Let δ_i be an indicator for the survival time t_i in the sense that $\delta_i = 1$ if t_i is not censored and $\delta_i = 0$ if it is. Furthermore, let $r = \sum_{i=1}^n \delta_i$ be the number of uncensored survival times. We will give a very short overview of the estimation of the unknown parameters here, relying heavily on the description found in Lawless (2003), whose work we refer to for any further details.

As can be found in Lawless (2003), the likelihood for a potentially censored sample $y = (y_1, \ldots, y_n)$ of log-survival times can be given by

$$L(\beta, \sigma|y) = \prod_{i=1}^n f(y_i|x)^{\delta_i} S(y_i|x)^{(1-\delta_i)} = \prod_{i=1}^n \frac{1}{\sigma^{\delta_i}} f_W(z_i)^{\delta_i} S_W(z_i)^{1-\delta_i}.$$

Basically, one can see that for an uncensored survival time t_i (or the uncensored log-survival time y_i) the contribution to the likelihood is given by the actual density function for y under the model. For a censored survival time, the contribution is given by the survival function at this point. For completely uncensored survival times, as it is the case for the breast implant data we will be analyzing later on, the likelihood consists of the product of the value of the density function at each log-survival time, as one would expect. When it comes to estimating the parameters, it is of course beneficial to consider the log-likelihood given by

$$l(\beta, \sigma|y) = -r \log \sigma + \sum_{i=1}^n [\delta_i \log f_W(z_i) + (1 - \delta_i) \log S_W(z_i)].$$

As a next step, one needs the derivatives of the log-likelihood with respect to each individual β_j with $j = 0, \ldots, p-1$ as well as σ. These are given by

$$\frac{\partial l}{\partial \beta_j} = -\frac{1}{\sigma} \sum_{i=1}^n \Big[\delta_i \frac{\partial \log f_W(z_i)}{\partial z_i} + (1 - \delta_i) \frac{\partial \log S_W(z_i)}{\partial z_i}\Big] X_{ji}$$

$$\frac{\partial l}{\partial \sigma} = -\frac{r}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^n \Big[\delta_i \frac{\partial \log f_W(z_i)}{\partial z_i} + (1 - \delta_i) \frac{\partial \log S_W(z_i)}{\partial z_i}\Big] z_i.$$

These follow from applying the chain rule and remembering that $\frac{\partial z_i}{\partial \beta_j} = -\frac{X_{ji}}{\sigma}$ and $\frac{\partial z_i}{\partial \sigma} = -\frac{z_i}{\sigma}$ holds. The above described results appear highly theoretical at a first glance, but

one must keep in mind that the very first step when applying AFT models is assuming a distribution for the survival times T. For an assumed distribution of T, for example a Weibull distribution, the density for W as well as the associated survival function can be explicitly given; in case of a Weibull distributed T, W follows a standard extreme value distribution, whose density and survival function are of course known. In order to compute the maximum likelihood estimate for both β and σ, one must solve

$$\frac{\partial l}{\partial \beta} = 0$$
$$\frac{\partial l}{\partial \sigma} = 0.$$

The solutions of these non-linear equations are the MLEs $\hat{\beta}$ and $\hat{\sigma}$, which can be computed by the Newton-Raphson method similarly to the way we explained in Section 2.3.1. As a reminder, in the k-th iteration, we compute

$$(\hat{\beta}_{k+t}, \hat{\sigma}_{k+1}) = (\hat{\beta}_k, \hat{\sigma}_k) + I^{-1}(\hat{\beta}_k, \hat{\sigma}_k) U(\hat{\beta}_k, \hat{\sigma}_k),$$

where

$$I(\hat{\beta}_k, \hat{\sigma}_k) = - \begin{bmatrix} \frac{\partial^2 l(\hat{\beta}_k, \hat{\sigma}_k | y)}{\partial \beta \partial \beta^T} & \frac{\partial^2 l(\hat{\beta}_k, \hat{\sigma}_k | y)}{\partial \beta \partial \sigma} \\ \frac{\partial^2 l(\hat{\beta}_k, \hat{\sigma}_k | y)}{\partial \sigma \partial \beta^T} & \frac{\partial^2 l(\hat{\beta}_k, \hat{\sigma}_k | y)}{\partial \sigma^2} \end{bmatrix}$$

and

$$U(\hat{\beta}_k, \hat{\sigma}_k) = \begin{bmatrix} \frac{\partial l(\hat{\beta}_k, \hat{\sigma}_k | y)}{\partial \beta} \\ \frac{\partial l(\hat{\beta}_k, \hat{\sigma}_k | y)}{\partial \sigma} \end{bmatrix}.$$

For the full computation of the second derivatives, refer to Lawless (2003). As one would expect for a larger data sample, it can be shown that for $\theta = (\beta, \sigma)$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, I^{-1}(\hat{\theta}))$$

holds. Thus, the asymptotic behavior of the estimates proves to be somewhat easy to handle.

As a logical next step, one is interested in performing hypothesis tests on the parameters using the computed estimates. For each individual β_j, one could yet again use the Wald test. So, for $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$, one could use the test statistic

$$\frac{\hat{\beta_j}^2}{\widehat{var}(\beta_j)} \overset{H_0}{\sim} \chi^2_1,$$

which is the most common way of testing H_0. A more generalized approach would be stating a hypothesis for a certain subset of coefficients of β. To that end, let $\beta = (\beta_1, \beta_2)$, where we want to test $H_0 : \beta_1 = \beta_1^0$ for β_1 a vector of length k (see Lawless, 2003). A

good first instinct would be to use the likelihood-ratio-test, whose test statistic is based on

$$\Lambda(y) = \frac{\sup\limits_{\beta_1=\beta_1^0} L(y|\beta,\sigma)}{\sup\limits_{\beta_1\in\mathbb{R}} L(y|\beta,\sigma)}.$$

As it is common with this test, $-2\log\Lambda(y) \xrightarrow{D} \chi^2_k$, which one can use to either reject or accept the hypothesis H_0.

When it comes to the implementation of AFT models in `R`, the function `survreg` proves to be an excellent option. As described briefly in Venables & Ripley (2002) the function `survreg`, which is included in the library `survival`, can fit AFT models of with the distributions `exponential`, `weibull` (which is the default), `lognormal`, `gaussian`, `logistic` and `loglogistic`. It is important to note that not all types of survival times work for all distributions– indeed, our survival times `t` need to be adapted, since the `weibull` and `exponential` options do not allow for survival times of zero. As already shown in Figure 4.2, we believe the survival times `t` follow a Weibull distribution quite nicely. Thus, we want to fit them accordingly, using `vol` as predictor.

```
> t0 <- t[which(t>0)]
> vol0 <- vol[which(t>0)]
> volparam <- survreg(Surv(t0) ~ vol0, dist="weibull")
> summary(volparam)

Call:
survreg(formula = Surv(t0) ~ vol0, dist = "weibull")
              Value Std. Error     z        p
(Intercept)  7.5551   0.126260 59.84 0.00e+00
vol0        -0.0032   0.000433 -7.39 1.51e-13
Log(scale)   0.1725   0.026293  6.56 5.33e-11

Scale= 1.19

Weibull distribution
Loglik(model)= -6848.7   Loglik(intercept only)= -6873
Chisq= 48.65 on 1 degrees of freedom, p= 3.1e-12
Number of Newton-Raphson Iterations: 6
n= 885
```

As seen in the above output, `survreg` provides plenty of information as well as coefficient estimates and separate Wald tests for each predictor and an estimate for the scale σ. Consider the coefficient $\hat{\beta}_{vol}$, which is given as -0.0032. As described before, we now know that the influence of `volume` on the behavior of t is given by $\exp(-\hat{\beta}_{vol}) = 1.003205$. That corresponds to an acceleration of the failure time with a factor of 1.003205 per one cm^3 increase in volume.

In order to give the estimated hazard function $\lambda(t|x)$, consider again that in this instance, W is standard extreme value distributed, which means that $\lambda_W(w) = e^w$. Thus, following Equation (4.2), we obtain

$$\begin{aligned}\hat{\lambda}(t|x) &= \frac{1}{t\hat{\sigma}}\lambda_W\Big(\frac{\log t - \hat{\beta}_0 - \hat{\beta}_{vol}vol}{\hat{\sigma}}\Big) \\ &= \frac{1}{1.19t}\exp\Big(\frac{\log t - 7.56 + 0.0032vol}{1.19}\Big) \\ &= \frac{0.0015}{t}\exp\Big(\frac{\log t}{1.19}\Big)1.0027^{vol}.\end{aligned}$$

That means that an increase of one cm^3 in `vol` corresponds to a 0.27% increase in the hazard. Consequently, an increase in 100 cm^3 corresponds to a 30.9% increase in the hazard.

4.2.1 Diagnostics

When evaluating an AFT model there are generally two aspects that should be considered. First, one should make sure that the basic assumptions of the model are fulfilled– that means that the previously chosen distribution matches the observed survival times $t_1, \ldots, t_n$ in the first place. This of course depends on the chosen distribution itself, but generally such assumptions can be checked by a plot of a transformation of the survival times against a transformation of an estimate of the survival function. For the details on that refer to Section 4.1, where we have discussed common distributions as well as the methods used to check whether they fit a given sample of observed survival times.

Another important aspect to be considered is the so-called Goodness of Fit, which is a general measure of how well the chosen model fits the observations that are being modeled. In Section 2.3.2, we already discussed the theory of Cox-Snell residuals. Indeed, they also prove to be an appropriate way to check the Goodness of Fit of an AFT model, as their theory uses no assumptions on the way a model is fit.

As a reminder, the Cox Snell residuals are based on the fact that the random variable $\Lambda(T)$ follows an exponential distribution with parameter 1. Since

$$\Lambda(T) = -\log S(T)$$

holds, the Cox-Snell Residuals are defined as

$$r_i^{CS} = \hat{\Lambda}(t_i) = -\log \hat{S}(t_i).$$

In order to check whether they indeed follow an exponential distribution with parameter 1, one should plot r_i^{CS} against $\hat{\Lambda}(r_i^{CS})$. If the plot resembles a straight line with unit slope it speaks for the adequacy of the model.

When implementing the Cox-Snell residuals in `R`, one has to yet again go about it the long way, since there are no specific functions that calculate these residuals for `survreg`. We will demonstrate the Cox-Snell residuals for the model `volparam`, fitted with a Weibull

distribution. As described in Section 4.1, for Weibull distributed T the $\log T$ follows an extreme value distribution. By Equation (4.1) and the fact that for a standard extreme value distribution

$$\Lambda_W(w) = -\log S_W(w) = -\log \exp(-e^w) = e^w$$

holds, we obtain that the Cox-Snell Residuals r_i^{CS} are given by

$$r_i^{CS} = \exp\left(\frac{\log t_i - X_i^T \hat{\beta}}{\hat{\sigma}}\right).$$

```
> sigmahat <- volparam$scale
> predhat <- volparam$linear.predictors
> csresv <- exp((log(t0) - predhat)/sigmahat)
> csresvfit <- survfit(Surv(csresv) ~ 1, type="kaplan-meier")
> plot(csresvfit$time, -log(csresvfit$surv),
+      xlab=expression(r[i]^CS), ylab=expression(hat(Lambda)(r[i]^CS)))
> abline(a=0, b=1, col="red")
```

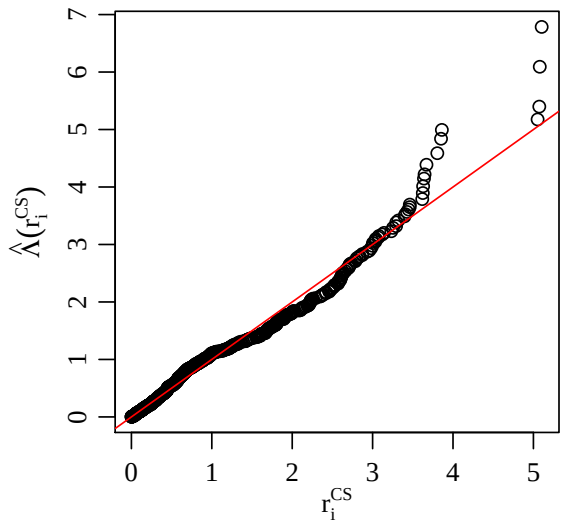

Figure 4.3: Cox-Snell Residuals for the parametric Weibull model with predictor volume.

As one can see in Figure 4.3, the Cox-Snell residuals for `volparam` follow the straight line with unit slope quite well until the very end, where a few observations seem to stray from the reference line. Overall, the model appears to be adequate, since we believe that the outliers, few as they are, do not indicate a drastic problem with the model fit.

Chapter 5

Application of AFT models

5.1 Preliminary Observations

The first step when fitting parametric models is always to find a suitable distribution for T. As already seen in Section 4.1, the exponential and Weibull distributions seem to be acceptable choices for T. We also took the time to try other distributions, for example the log-normal or log-logistic distribution, both of which can also be fit using `survreg`. As it turns out however, neither one of the two is an acceptable choice for our given data– we will not go into great detail as to how we have arrived at this conclusion, but we'll give a short overview.

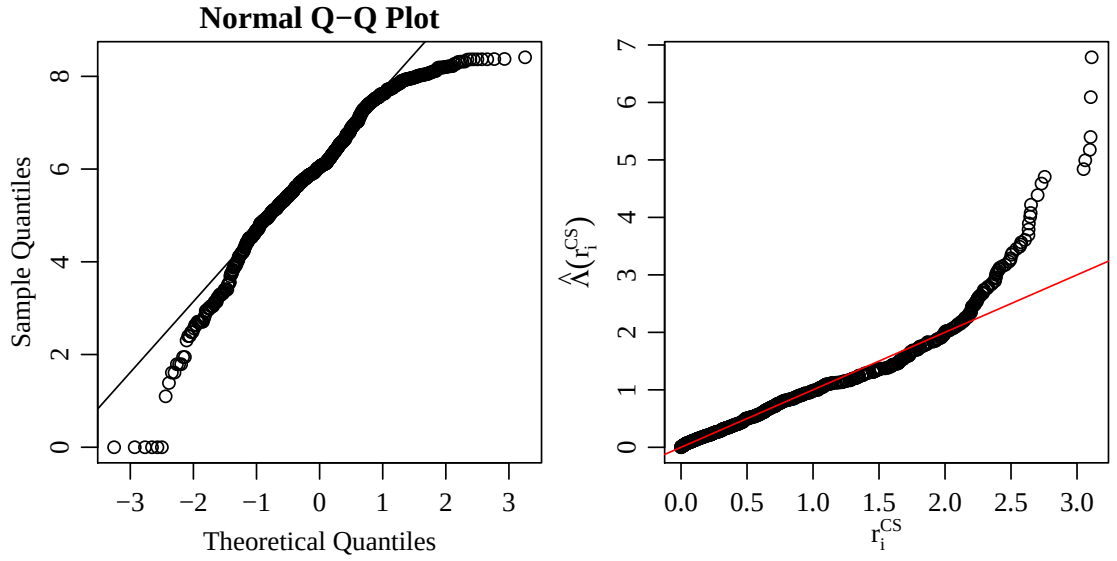

Figure 5.1: QQ-Plot of $\log t$ (left) and Cox-Snell residuals for the log-logistic model (right).

If T was log-normal distributed, $\log T$ would follow a normal distribution. As can be seen in Figure 5.1 (left), the QQ-Plot for $\log t$ doesn't seem to support that fact. Indeed, the

curved structure of the quantiles doesn't follow the required reference line of the normal distribution quantiles, which means that we reject the idea of a log-normal distribution for T. As for the log-logistic distribution, we realize that while this particular distribution for T isn't far-fetched (plot not shown here), the Cox-Snell residuals for even the simplest models don't make us feel confident in the use of a log-logistic distribution for T (see Figure 5.1 on the right).

Thus, we believe that our choice is between the exponential distribution and the Weibull distribution for T. Due to the fact that the exponential distribution is a special case of the Weibull, we ultimately chose the latter for our models since the Weibull distribution is more flexible. Furthermore, the exponential distribution is memoryless and thus the hazard for such a model would be constant over time, which is a rather restrictive assumption. Thus, we choose to move on with a Weibull distribution for T. One has to keep in mind that when using `survreg` with `dist="weibull"`, none of the survival times can be zero– thus, we from now on exclude such survival times. In total, we have to remove 18 such observations from the dataset and reduce to a total of 885 observations.

5.2 Weibull AFT Models

In this section, we fit and analyze AFT models by using a Weibull distribution for the distribution of T. Due to the high significance of the predictor `vol` as already established in Section 4.2, we decide to include it as a predictor in every model and now analyze what happens when we include further predictors. For a short data analysis on the predictors `surface`, `lumen`, `fill` and `drainage` which will be used in the models, refer to Section 3.1.

5.2.1 Volume and Surface

As already established in Section 3.2, we assume that an increase in `vol` will result in an increase of the hazard– furthermore, we assume that a `textured` surface is to be preferred compared to its counterparts. We fit an AFT model using the two predictors `vol` and `surface`, where again `surface` is a factor with baseline setting `smooth`.

```
> volsurfaceparam <- survreg(Surv(t0) ~ vol0 + surface0, dist="weibull")
> anova(volsurfaceparam)
         Df  Deviance Resid. Df    -2*LL     Pr(>Chi)
NULL     NA        NA       883 13745.98           NA
vol0      1 48.646555       882 13697.33 3.065063e-12
surface0  2  9.318328       880 13688.01 9.474378e-03
```

We see that both `vol` and `surface` appear to be significant for the model. The interaction of the two is insignificant with $p = 0.84$ (not shown) and thus can be omitted from the model. As per usual, we want to check the Goodness of Fit before we move to a more in depth analysis of the model. To that end, we analyze the Cox-Snell Residuals as described in Section 4.2.1. As observed in Figure 5.2, the resulting model `volsurfaceparam` appears to be adequate enough; the rare few outliers towards the right in the residual plot are to be expected and, while inconvenient, do not lead us to believe that the model is inadequate.

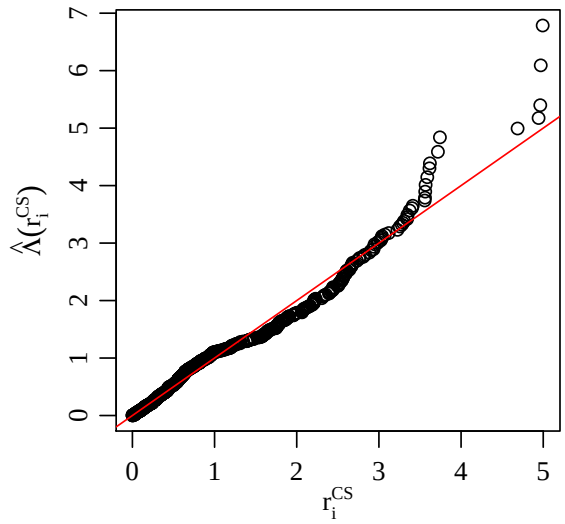

Figure 5.2: Cox-Snell Residuals for the parametric Weibull model with volume and surface.

In order to further analyze the model, let us take a closer look at the summary.

```
> summary(volsurfaceparam)

Call:
survreg(formula = Surv(t0) ~ vol0 + surface0, dist = "weibull")
                       Value Std. Error      z        p
(Intercept)          7.10839   0.406369 17.492 1.64e-68
vol0                -0.00306   0.000435 -7.029 2.08e-12
surface0polyurethan -0.20341   0.424251 -0.479 6.32e-01
surface0textured     0.43563   0.378056  1.152 2.49e-01
Log(scale)           0.16649   0.026333  6.322 2.58e-10

Scale= 1.18

Weibull distribution
Loglik(model)= -6844   Loglik(intercept only)= -6873
Chisq= 57.96 on 3 degrees of freedom, p= 1.6e-12
Number of Newton-Raphson Iterations: 6
n= 885
```

At the first glance, we notice that the Intercept is estimated to be around 7.108 and the estimate $\hat{\sigma} = 1.18$. The estimated coefficient $\hat{\beta}_{vol}$ is given by -0.00306, which indicates that with $\exp(-(-0.00306)) = 1.003065$ for each cm^3 increase in `vol`, the time `t` is accelerated with a factor of approximately 1.003. Similarly, the acceleration factor of the level `polyurethan` of the factor `surface` is given by 1.2256 in comparison to `smooth`, thus clearly a deterioration from the baseline. As expected, `textured` is to be preferred, since

the associated factor is given by 0.6469, which of course corresponds to an deceleration of the failure time.

In order to understand the influence of each predictor on the hazard, we explicitly compute the estimate $\hat{\lambda}(t|x)$, using Equation (4.2), and get

$$\begin{aligned}\hat{\lambda}(t|x) &= \frac{1}{t\hat{\sigma}}\lambda_W\Big(\frac{\log t - \hat{\beta}_0 - \hat{\beta}_{vol}vol - \hat{\beta}_{poly}I_{poly} - \hat{\beta}_{text}I_{text}}{\hat{\sigma}}\Big)\\ &= \frac{1}{1.18t}\exp\Big(\frac{\log t - 7.10839 + 0.00306vol + 0.20341I_{poly} - 0.43563I_{text}}{1.18}\Big)\\ &= \frac{0.00205}{t}\exp\Big(\frac{\log t}{1.18}\Big)1.00259^{vol}1.18813^{I_{poly}}0.691302^{I_{text}},\end{aligned}$$

where I_{poly} and I_{text} are indicator variables defined as

$$I_{poly} = \begin{cases}1 & \text{surface is } \texttt{polyurethan}\\ 0 & \text{otherwise}\end{cases} \quad I_{text} = \begin{cases}1 & \text{surface is } \texttt{textured}\\ 0 & \text{otherwise.}\end{cases}$$

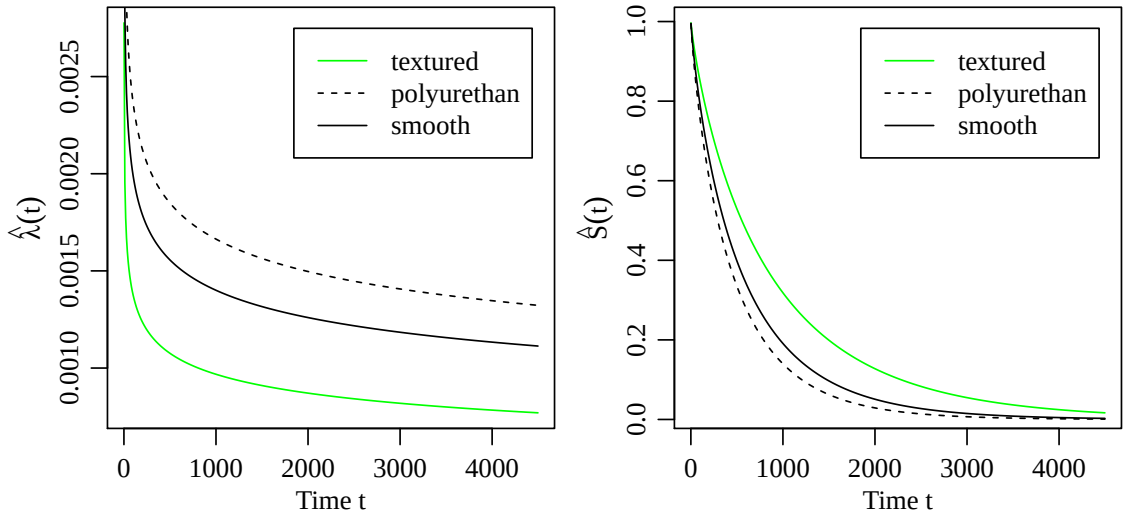

Figure 5.3: The estimated hazard function (right) and survival function (left).

Here we see the actual effect of each predictor on the hazard. As it seems, for each cm^3 increase of `vol`, the hazard increases by 0.26%, which corresponds to an increase of 29.52% in the hazard per 100 cm^3. For an implant whose `surface` is made out of `polyurethan` we see an increase of about 18.81% in the hazard. Clearly, an implant fo the `textured` variety is to be preferred, since in that case the hazard is actually reduced by 30.87% compared to its `smooth` counterpart. Below we compare the results of the corresponding Cox model described in Section 3.2 and the AFT model.

Predictor	Change in $\hat{\lambda}(t)$ for Cox mod	Change in $\hat{\lambda}(t)$ for AFT mod
`volume` per 100 cm^3	+27.9%	+29.52%
`polyurethan`	+22.7%	+18.81%
`textured`	−27.03%	−30.87%

Unsurprisingly, the two models tend to have similar results. In general, the in- and decrease of the hazard for each individual predictor is in the same ballpark for both models, as it should be. There are subtle differences in just how much the hazard is affected by each predictor, but we see that the tendency is the same.

5.2.2 Volume and Lumen

When it comes to including the predictor `lumen`, which is a factor with two levels `single` and `double`, we yet again expect an increase in `vol` of the implants to have a negative effect on the survival times. We also expect to see a preference for `single` implants in comparison to the `double`, which is the baseline.

```
> vollumenparam <- survreg(Surv(t0) ~ vol0 + lumen0, dist="weibull")
> anova(vollumenparam)
       Df  Deviance Resid. Df    -2*LL     Pr(>Chi)
NULL   NA        NA       883 13745.98           NA
vol0    1 48.646555       882 13697.33 3.065063e-12
lumen0  1  5.213548       881 13692.12 2.241156e-02
```

The interaction of `vol` and `lumen` proves to be insignificant with a p-value of 0.58 (not shown) and thus is of course excluded in the resulting model `vollumenparam`. Moving forward with `vol` and `lumen`, both of which are significant, we want to first check the Cox-Snell residuals before moving ahead and analyzing the model in a more in depth manner. As seen in Figure 5.4, the structure of the residuals appears to be rather similar to Figure 5.2, which corresponds to the model including `vol` and `surface`. We will see later that this is a common theme with these models. In general, the residuals are far from perfect given the little cluster of outliers towards the very right of the plot, but they aren't cause for too much concern.

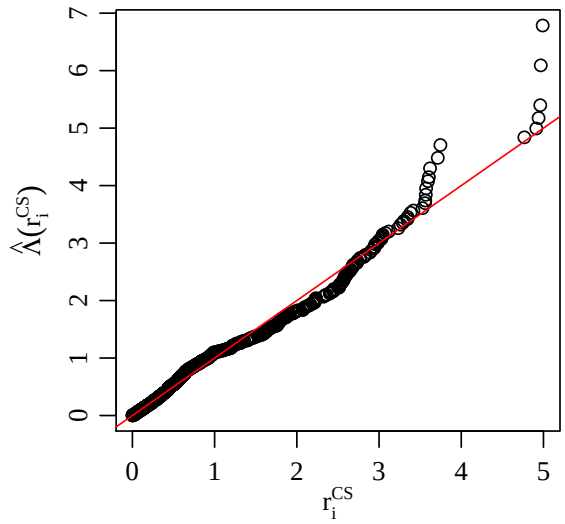

Figure 5.4: Cox-Snell Residuals for the parametric Weibull model with volume and lumen.

```
> summary(vollumenparam)

Call:
survreg(formula = Surv(t0) ~ vol0 + lumen0, dist = "weibull")
                Value Std. Error     z         p
(Intercept)   7.10062    0.22917 30.98 8.71e-211
vol0         -0.00304    0.00044 -6.92  4.51e-12
lumen0single  0.43336    0.18016  2.41  1.62e-02
Log(scale)    0.16935    0.02629  6.44  1.19e-10

Scale= 1.18

Weibull distribution
Loglik(model)= -6846.1   Loglik(intercept only)= -6873
Chisq= 53.86 on 2 degrees of freedom, p= 2e-12
Number of Newton-Raphson Iterations: 6
n= 885
```

As expected, we are faced with a significant difference for the influence of the level `single` ($p = 0.0162$) compared to `double`. Thus, `single` appears to be the preferred option in comparison to `double`. Indeed, the acceleration factor for `volume` is given by 1.0031, whereas the factor 0.6483 for `single` actually indicates a deceleration of the failure time. A closer look at the estimated hazard function $\hat{\lambda}(t|x)$ provides more insight, where I_{single} is defined as

$$I_{single} = \begin{cases} 1 & \text{lumen is } \texttt{single} \\ 0 & \text{otherwise.} \end{cases}$$

Then it follows that $\hat{\lambda}(t|x)$ is given by

$$\begin{aligned}\hat{\lambda}(t|x) &= \frac{1}{t\hat{\sigma}}\lambda_W\Big(\frac{\log t - \hat{\beta}_0 - \hat{\beta}_{vol}vol - \hat{\beta}_{single}I_{single}}{\hat{\sigma}}\Big) \\ &= \frac{1}{1.18t}\exp\Big(\frac{\log t - 7.10062 + 0.00304vol - 0.43336I_{single}}{1.18}\Big) \\ &= \frac{0.00206}{t}\exp\Big(\frac{\log t}{1.18}\Big)1.00258^{vol}0.69263^{I_{single}}.\end{aligned}$$

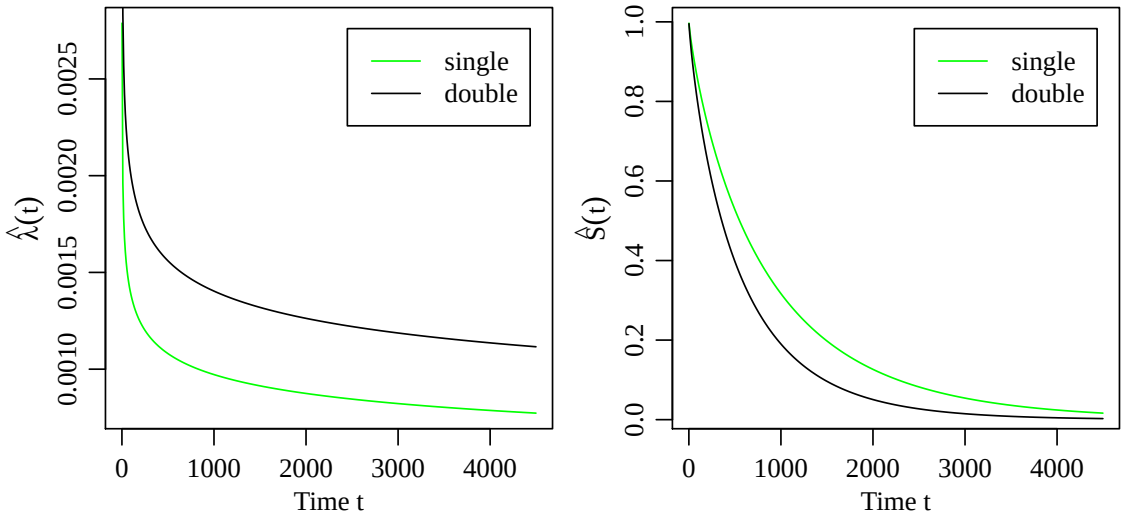

Figure 5.5: The estimated hazard function (right) and survival function (left).

As a result, we see that a 100 cm^3 increase in `vol` results in a 29.39% increase in the hazard. Furthermore, for implants of the same volume, the hazard of an implant of the type `single` is decreased by 30.74% compared to an implant of the type `double`. Thus, we clearly prefer a `single` lumen. The comparison of the results obtained by the Cox model and the AFT model yields the following table.

Predictor	Change in $\hat{\lambda}(t)$ for Cox mod	Change in $\hat{\lambda}(t)$ for AFT mod
`volume` per 100 cm^3	+27.7%	+29.39%
`single`	−27.6%	−30.74%

The differences between the two are rather small this time, showing that both models have the same tendencies. Still, for the predictor `vol` the AFT model has a slightly stronger increase in the hazard, whereas for the level `single`, the model shows a stronger decrease in the hazard compared to the Cox model.

5.2.3 Volume and Filling

The predictor `fill`, which has the baseline `siliconegel` and describes the type of filling for each implant, should be added next to the model. Again, we test the interaction for `vol`

against `fill`, but find it to be insignificant with $p = 0.07$ (not shown). Thus, proceeding with the simplified model, we see that both individual predictors are significant on their own.

```
> volfillparam <- survreg(Surv(t0) ~ vol0 + fill0, dist="weibull")
> anova(volfillparam)
      Df Deviance Resid. Df    -2*LL     Pr(>Chi)
NULL  NA       NA       883 13745.98           NA
vol0   1 48.64655       882 13697.33 3.065063e-12
fill0  4 39.63321       878 13657.70 5.154288e-08
```

A quick check of the Cox-Snell residuals (as seen in Figure 5.6) shows that the model itself appears to be more or less adequate. Again we deal with a cluster of outliers to the very right of the model that deviate from the reference line with unit slope, but we don't consider this to be a violation of the goodness of fit.

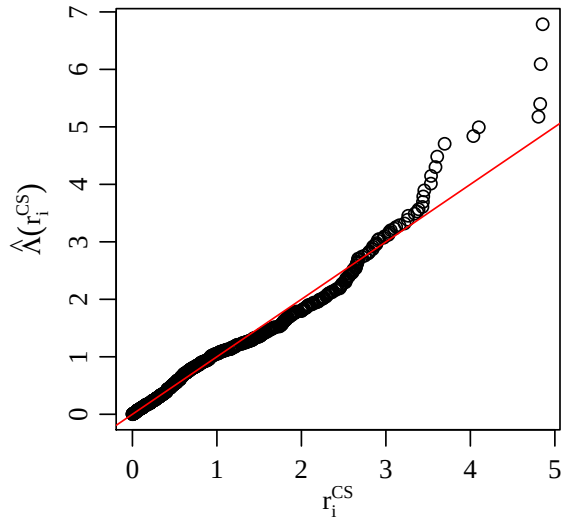

Figure 5.6: Cox-Snell Residuals for the parametric Weibull model with volume and filling.

The `summary` of `volfillparam` provides much needed insight into the detailed effects of each predictor. As expected, the failure time for an increase in `vol` is accelerated by a factor of 1.00262. We can also immediately tell that in this model, `siliconegel` is the preferable choice, as all other levels of `fill` accelerate the failure time; for `other` the failure time is accelerated by a factor of 1.35412, for `mixed` by a factor of 1.23823, for `hydrogel` by factor of 2.11446 and finally for `saline solution` by 3.49076, which is the largest factor by far.

```
> summary(volfillparam)

Call:
```

```
survreg(formula = Surv(t0) ~ vol0 + fill0, dist = "weibull")
                         Value Std. Error      z        p
(Intercept)            7.47125   0.125383 59.588 0.00e+00
vol0                  -0.00262   0.000442 -5.936 2.92e-09
fill0other            -0.30315   0.671176 -0.452 6.52e-01
fill0mixed            -0.21368   0.255906 -0.835 4.04e-01
fill0hydrogel         -0.74880   0.261582 -2.863 4.20e-03
fill0saline solution -1.25012   0.184178 -6.788 1.14e-11
Log(scale)             0.14098   0.026582  5.303 1.14e-07

Scale= 1.15

Weibull distribution
Loglik(model)= -6828.8   Loglik(intercept only)= -6873
Chisq= 88.28 on 5 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 885
```

Now let I_{other}, I_{mixed}, I_{hydro} and I_{saline} be indicator variables for each level. Then, an estimate for the hazard function $\hat{\lambda}(t|x)$ is given by

$$\hat{\lambda}(t|x) = \frac{0.0013}{t} \exp\left(\frac{\log t}{1.15}\right) 1.00228^{vol} 1.301619^{I_{other}} 1.20419^{I_{mixed}} 1.91771^{I_{hydro}} 2.96555^{I_{saline}}.$$

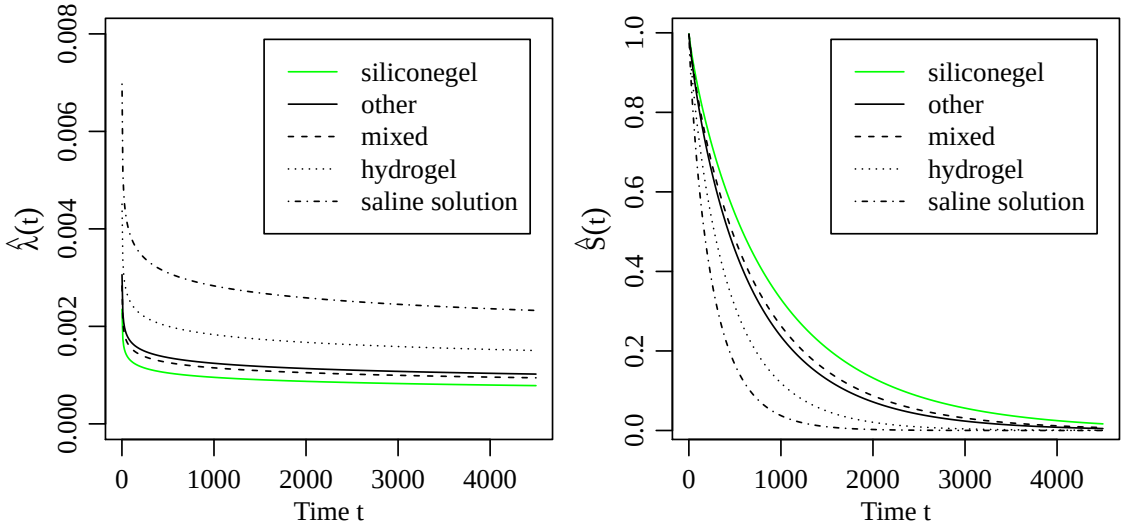

Figure 5.7: The estimated hazard function (right) and survival function (left).

In this model, we see that an increase of 100 cm^3 of `vol` corresponds to a direct 25.58% increase in the hazard function. Furthermore, our suspicions about the superiority of `siliconegel` are proven to be correct– `other` results in an increase in the hazard function

of 30.16%, `mixed` causes an increase of 20.42%, `hydro` increases the hazard by 91.77% and finally `saline solution` filling increases the hazard by a rather shocking 196.56%. When comparing these results with the results of the Cox model, it is important to note that the corresponding model described in Section 3.2.3 did not necessarily fulfill the proportional hazards assumption required for a Cox model. Thus, the following table is to be taken with a grain of salt.

Predictor	Change in $\hat{\lambda}(t)$ for Cox mod	Change in $\hat{\lambda}(t)$ for AFT mod
`volume` per 100 cm^3	+24.52%	+25.58%
`other`	+16.67%	+30.16%
`mixed`	+13.67%	+20.42%
`hydro`	+62.73%	+91.77%
`saline solution`	+173.02%	+196.56%

We observe that the increase in the hazard for `volume` is rather similar in both models– when it comes to the levels of `fill` however, there are massive differences in the individual percentages. This speaks to the previously discussed inadequacy of the Cox model; while of course the tendency remains the same with `siliconegel` being the clear favourite, the AFT model collectively has higher percentage increases of the hazard than the Cox model does.

5.2.4 Volume and Drainage

The predictor `drainage`, which is a factor with the two levels `yes` (the baseline) and `no`, indicates whether during the initial surgery drainage was required or not. We expect implants whose initial surgeries included drainage to be preferable to their counterpart. To show this, we fit an AFT model, excluding the interaction of the two since it proves to be insignificant with $p = 0.554$ (not shown).

```
> voldrainageparam <- survreg(Surv(t0) ~ vol0 + drainage0, dist="weibull")
> anova(voldrainageparam)
          Df  Deviance Resid. Df    -2*LL     Pr(>Chi)
NULL      NA        NA       883 13745.98           NA
vol0       1 48.646555       882 13697.33 3.065063e-12
drainage0  1  8.461859       881 13688.87 3.626711e-03
```

A check of the Cox-Snell residuals seen in Figure 5.8 shows that the model appears to be adequate. We see yet again the expected little cluster of outliers, but don't consider those to be too much of a bother.

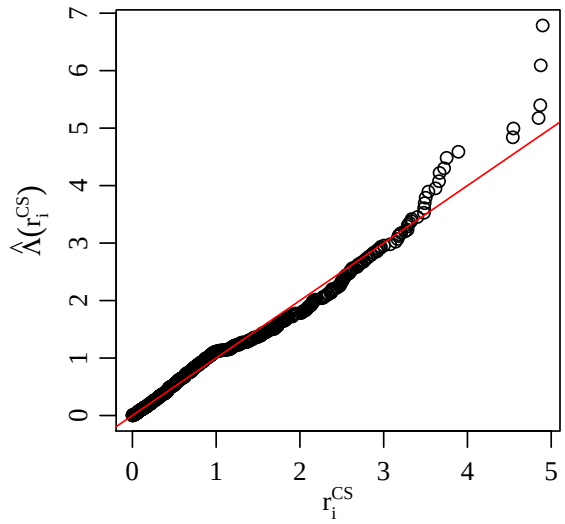

Figure 5.8: Cox-Snell Residuals for the parametric Weibull model with volume and drainage.

The `summary` shows a significant difference between the levels `no` and `yes` of `drainage` ($p = 0.0027$). Indeed, the failure time for an implant with `drainage no` is accelerated by a factor of 1.36599 compared to its counterpart. Furthermore, for a one cm^3 increase in `vol`, we obtain an acceleration of the failure time in this model by a factor of 1.00313.

```
> summary(voldrainageparam)

Call:
survreg(formula = Surv(t0) ~ vol0 + drainage0, dist = "weibull")
               Value Std. Error     z        p
(Intercept)  7.58679   0.125997 60.21 0.00e+00
vol0        -0.00312   0.000431 -7.24 4.43e-13
drainage0no -0.31188   0.103973 -3.00 2.70e-03
Log(scale)   0.16738   0.026297  6.37 1.95e-10

Scale= 1.18

Weibull distribution
Loglik(model)= -6844.4   Loglik(intercept only)= -6873
Chisq= 57.11 on 2 degrees of freedom, p= 4e-13
Number of Newton-Raphson Iterations: 6
n= 885
```

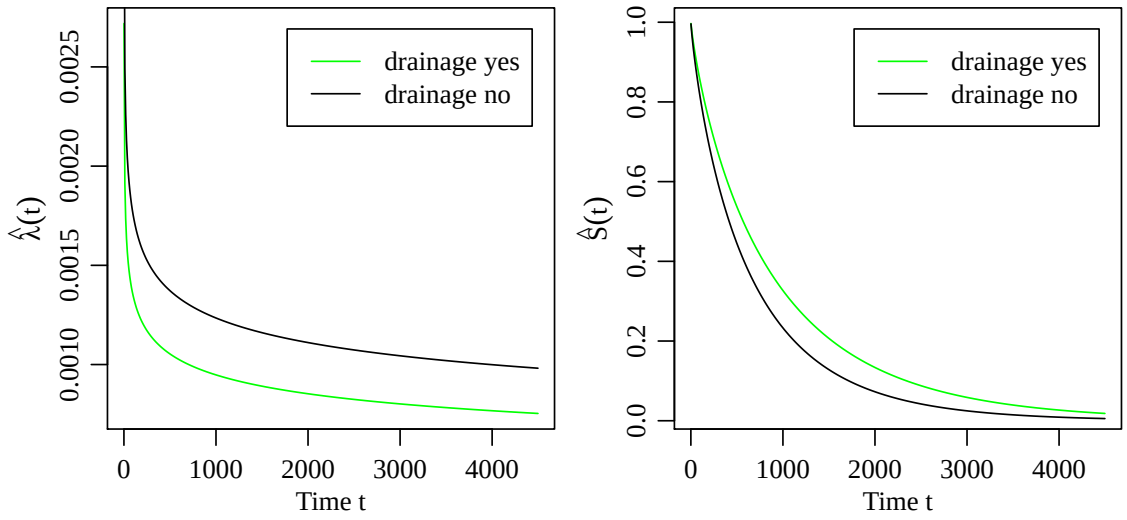

Figure 5.9: The estimated hazard function (right) and survival function (left).

By utilizing Equation (4.2) again to get an estimate for the hazard, we obtain

$$\hat{\lambda}(t|x) = \frac{0.0014}{t} \exp\left(\frac{\log t}{1.18}\right) 1.00265^{vol} 1.30252^{I_{no}},$$

where I_{no} is an indicator variable for the level `no` of the factor `drainage`. Clearly, this corresponds to an increase of 30.30% in the hazard for an increase of 100 cm^3 in volume. Similarly, we see that in comparison to `drainage yes`, `no` increases the hazard by 30.25%. Thus, we are inclined to prefer `drainage yes` over its counterpart. In comparison with the associated Cox model, we see that again the values for `vol` are rather close. The percentages for `drainage no` are a little further apart, but still there are no surprisingly large differences.

Predictor	Change in $\hat{\lambda}(t)$ for Cox mod	Change in $\hat{\lambda}(t)$ for AFT mod
`volume` per 100 cm^3	+28.58%	+30.30%
`no`	+25.36%	+30.25%

5.2.5 Full Model

Finally, we are interested in a model that includes all above mentioned predictors at once. Indeed, analysis shows that every predictor (`vol`, `surface`, `lumen`, `fill` and `drainage`) is significant in the resulting model `finalparam`, yet their interactions are not (not shown). It is imperative to note that one wants to check the significance of each predictor in this model and such a thing can't be done using `anova`, since the `anova` only gives information on whether the last predictor in the table is relevant given that all the others are already in the model. We used the method `step` provided in `R` (code not shown) to obtain `finalmod`, in which all predictors are significant. To show the significance levels of each predictor in

the model, we use the function `drop1`, which performs a series of likelihood ratio tests in order to compare the likelihood of a model including a certain predictor and the likelihood of a model without said predictor. The results are given below.

```
> finalparam <- survreg(Surv(t0) ~ vol0 + surface0
+                         + lumen0 + fill0 + drainage0, dist="weibull")
> drop1(finalparam, test="Chisq")
Single term deletions

Model:
Surv(t0) ~ vol0 + surface0 + lumen0 + fill0 + drainage0
           Df   AIC    LRT  Pr(>Chi)
<none>        13656
vol0        1 13681 27.144 1.888e-07 ***
surface0    2 13664 12.355  0.002075 **
lumen0      1 13658  4.328  0.037495 *
fill0       4 13686 37.981 1.131e-07 ***
drainage0   1 13662  8.219  0.004144 **
```

A quick look at the Cox-Snell residuals shows that `finalparam` doesn't seem to have any problems when it comes to adequacy. While the already expected outliers to the right are still there, we find the plot of the Cox-Snell residuals shown in Figure 5.10 to be even better than Figure 5.8 or Figure 5.4, for example. Overall, `finalparam` seems to be a good fit.

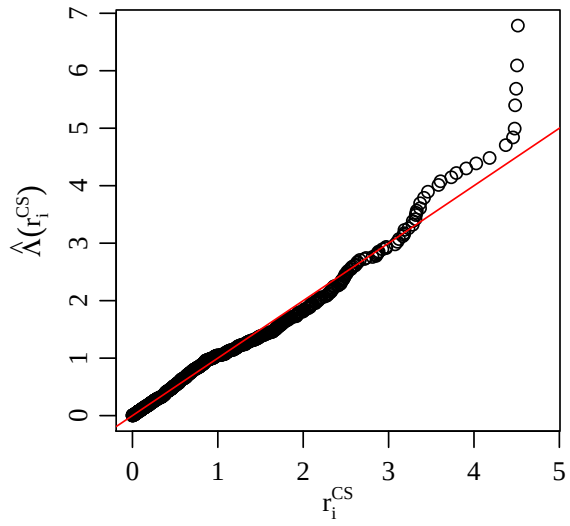

Figure 5.10: Cox-Snell Residuals for finalmod.

We will not go into too much detail regarding the factors for acceleration and deceleration time for each predictor in this particular case. Instead, we would like to focus more heavily on the resulting estimate for the hazard function, as we believe that to be a bit informative.

```
> summary(finalparam)

Call:
survreg(formula = Surv(t0) ~ vol0 + surface0 + lumen0 + fill0 +
    drainage0, dist = "weibull")
                        Value Std. Error       z        p
(Intercept)           6.94879   0.474654 14.6397 1.57e-48
vol0                 -0.00237   0.000439 -5.4039 6.52e-08
surface0polyurethan  -0.74302   0.421297 -1.7636 7.78e-02
surface0textured      0.02586   0.375829  0.0688 9.45e-01
lumen0single          0.51836   0.234185  2.2135 2.69e-02
fill0other           -0.43669   0.663071 -0.6586 5.10e-01
fill0mixed            0.23406   0.340100  0.6882 4.91e-01
fill0hydrogel        -0.53076   0.272887 -1.9450 5.18e-02
fill0saline solution -1.25356   0.191684 -6.5397 6.16e-11
drainage0no          -0.31523   0.106438 -2.9616 3.06e-03
Log(scale)            0.12802   0.026571  4.8180 1.45e-06

Scale= 1.14

Weibull distribution
Loglik(model)= -6816.9   Loglik(intercept only)= -6873
Chisq= 112.15 on 9 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 885
```

With these coefficients, our resulting estimate for the hazard function is given by

$$\hat{\lambda}(t|x) = \frac{0.00197}{t}\exp\left(\frac{\log t}{1.14}\right)1.00298^{vol}1.91894^{I_{poly}}0.97757^{I_{text}}0.63464^{I_{single}}$$
$$1.46677^{I_{other}}0.81439^{I_{mixed}}1.59294^{I_{hydro}}3.00301^{I_{saline}}1.31853^{I_{no}}.$$

Considering the above mentioned output, we can see that the loglikelihood of the resulting model `finalparam` is given by -6816.9 , whereas in comparison the loglikelihood of a model for the response `t0` with just an intercept is given by -6873. Furthermore, we are given the value of `Chisq`, which is by definition the test statistic for the likelihood ratio test for the null hypothesis $H_0 : \beta_j = 0 \; \forall j$. One can easily calculate this by taking double the difference of the above mentioned loglikelihood values. Due to the fact that `Chisq` under H_0 is supposed to approximately follow a χ^2_9 distribution, we see that the resulting value of 112.15 is way too high and we thus reject H_0 with an associated p-value of basically 0. Furthermore, we are given the number of Newton-Raphson iterations, which indicates the number of iterations of the Newton-Raphson algorithm it took for this model to reach convergence. Since it only took 6 iterations, we believe that there are no issues in that regard. We also obtain an estimate for σ in this output, which is given by the scale 1.14.

One needs to remember that with all of these different predictors, we have a different baseline than before; our reference in this model is an implant with `smooth surface`,

`double lumen`, `siliconegel filling` and `drainage yes`. Thus the increase of the hazard of 91.89% for an implant with `polyurethan` surface instead of `smooth` means that all other baseline settings remain the same in that instance. We see that an increase of 100 cm^3 of volume immediately results in an increase of 34.66% of the hazard. Picking an implant with a `textured` surface is yet again to be seen as an improvement rewarded by a decrease of 2.24% in the hazard which is far less than initially expected given the results of `volsurfaceparam`; this however can be explained, since this still corresponds to a change in `surface` only, leaving all the other predictors untouched. Again, `single` is the preferable option for `lumen`, since choosing `single` decreases the hazard by 36.54%. When it comes to `filling`, we see increases in the hazard all around: 46.68% for `other`, 59.59% for `hydrogel` and a record of a little over 200% for `saline solution`. Contrary to our expectations however, `mixed` shows a decrease in the hazard of 18.56%. Unsurprisingly, `drainage no` shows an increase in the hazard of about 31.85%.

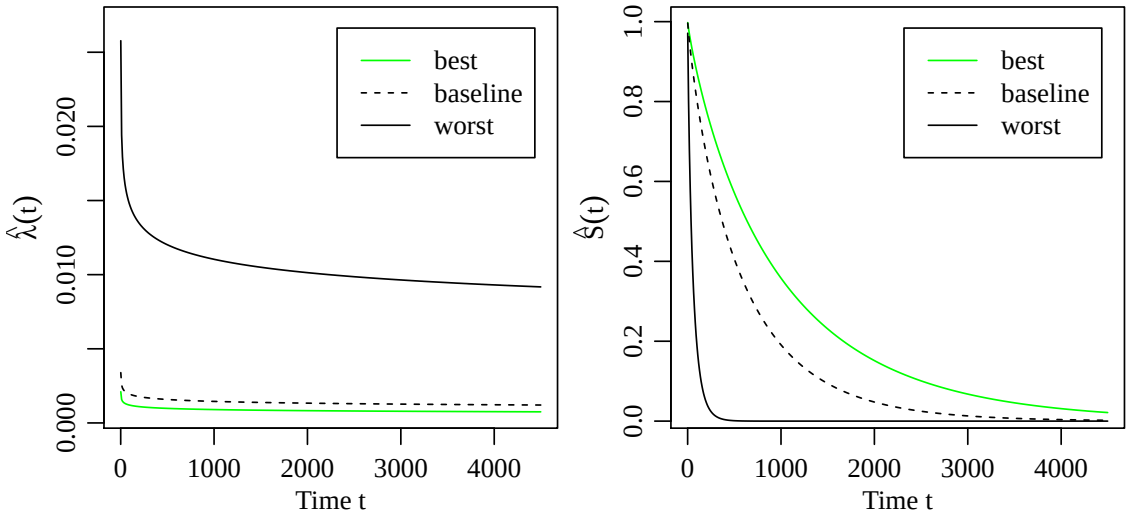

Figure 5.11: The estimated hazard function (right) and survival function (left).

As seen in Figure 5.11, the differences in the hazard and survival functions are astounding. Here, *best* refers to implants with `textured` surface, `siliconegel` filling, `single` lumen and drainage `yes`. The *baseline* setting refers to the reference setting in `finalparam`, namely an implant with `smooth` surface, `double` lumen, `siliconegel` filling and drainage `yes`. In that case, *worst* refers to an implant with `polyurethan` surface, `double` lumen, `saline solution` filling and drainage `no`.

Contrary to the corresponding Cox model in Section 3.2.5, this model does not seem to violate any assumptions. We can't really compare the individual contribution to the hazard of each predictor between the two models, given that the Cox model doesn't fulfill the proportional hazards assumptions. Still, we want to point out that in the parametric approach, we can confidently state the individual effects of each predictor in the full model,

which we couldn't do in the Cox model. Thus, it seems as though in this instance the more restrictive approach of assuming a distribution for T has paid off.

Chapter 6

Summary

Inspired by the the survival times of breast implants taken from the Austrian Silicone Registry, we began this thesis by asking the question: How long do breast implants last in the human body and can we find out which predictors influence the length of the survival in which way? We had several predictors to our disposal, most notably the surface of implants, the volume of an implant, the lumen, the type of filling of an implant and an indicator stating whether during the implant surgery any drainage had taken place.

When picking a model for the given data, we were faced with many different options due to the large amount of different approaches when it comes to modeling survival time observations. After eventually deciding on the Cox proportional hazards model due to its large popularity and high flexibility it turned out that the non-parametric approach of the Cox model was perfectly adequate to fit our given data, if we only include very few predictors in the model at once. While the proportional hazard model performed admirably with two predictors, a model including all of the relevant predictors failed to meet necessary conditions.

This lead us to a parametric approach. By using a more restrictive concept that is based on explicitly fixing a distribution for the survival times, the parametric approach gave a different point of view on the data. For these models, we realize that we were not only perfectly capable of fitting models with two predictors, but also fitting models with all relevant predictors at once– something the Cox model had failed to do.

In conclusion, we find that the survival times of the implant data we analyzed depend heavily on the above mentioned predictors. An increase in the volume of an implant has a negative effect on the survival times; we found that an increase of 100 cm^3 in volume can increase the hazard for a subject by $24.52\% - 34.66\%$, depending on the model. Furthermore, the surface of an implant is highly relevant to its survival time. A `textured` surface is to be preferred, as such implants tend to reduce the hazard for a subject by $27.03\% - 30.87\%$ compared to its counterparts. The preferable filling for an implant is without a doubt to be found in `siliconegel`. The alternative options tend to always increase the hazard, some options even tripling the risk for the subject. Furthermore we can tell that having drainage during the initial surgery is to be preferred to having no drainage; the latter increases the hazard between 25.36% to almost 31% for the patient.

Appendix A

The Partial Likelihood Approach

In many real life applications, one doesn't stand a chance of fully classifying the likelihood without making potentially overly restrictive assumptions. As Cox (1975) points out, such issues can arise especially if there are unknown nuisance parameters involved, as it is the case in proportional hazard models.

Simply put, let us consider a sequence of observations y that are from the population of a random variable Y, which can be expressed by a sequence of $(X_1, Z_1, \ldots, X_n, Z_n)$. Furthermore, let the density of Y depend on a parameter $\theta = (\beta, \phi)$, where ϕ represents our nuisance parameter and β is the parameter of interest– in case of the proportional hazard model, one may see ϕ as the baseline hazard $\lambda_0(t)$ which is both unknown and not necessarily the parameter of interest in this context. Let us also define by $x_{(i)} = (x_1, \ldots, x_i)$ and $z_{(i)} = (z_1, \ldots, z_i)$. Then one can conclude

$$\begin{aligned} L(\theta|y) &= \prod_{i=1}^{n} f_Y(y_i|\theta) = \prod_{i=1}^{n} f_{(X,Z)}(x_i, z_i|\theta, x_{(i-1)}, z_{(i-1)}) \\ &= \prod_{i=1}^{n} f_{X|Z}(x_i|\theta, x_{(i-1)}, z_{(i)}) f_Z(z_i|\theta, x_{(i-1)}, z_{(i-1)}) \\ &= P(\theta)Q(\theta). \end{aligned}$$

The best case scenario would be $P(\theta) = P(\beta)$ and $Q(\theta) = Q(\phi)$, namely P doesn't depend on the nuisance parameter and Q doesn't depend on β, which is our parameter of interest. In this case, one can easily omit the function Q and maximize P in order to find the MLE for β, which tends to simplify the process significantly. Even though that is not necessarily the case when considering proportional hazard models, it is still common practice to consider P only when searching for a proper estimate for β. This of course implies some information loss that needs to be taken into account. In his paper, Cox also argues that under broad assumptions the properties of maximum likelihood estimates can be transferred onto maximum partial likelihood estimates– namely, the use of likelihood ratio tests and their resulting asymptotic chi-squared distributions can be applied as usual.

Bibliography

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society B*, *34*(2), 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276.

Crowder, M. (2012). *Multivariate Survival Analysis and Competing Risks*. Boca Raton, Florida: Chapman and Hall/CRC.

Kalbfleisch, J. D., & Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York; Chichester: John Wiley & Sons, Inc.

Kaplan, E., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457-481.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Liu, X. (2012). *Survival Analysis: Models and Applications*. Chichester: John Wiley & Sons, Inc.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, *69*(1), 239-241.

Therneau, T. M. (2015). A Package for Survival Analysis in S [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=survival` (version 2.38)

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer.