R Notebook

1) Introduction

Description of the problem

It is generally known that infection is the chief reason for death and morbidity after burn injury, with it being responsible for 51% of the deaths, especially in the first 24 hours.

Indeed burn wounds lack epidermis and circulation, so they are the best culture media for bacterial growth. A few hours after the burn (normally 4-5 hours after) the wound surface becomes contaminated with many bacterial flora, which will start to grow and multiply, reaching the vascular and lymphatic vessels and starting to disperse. At this point, bacteraemia and sepsis start. However, due to the release of several cell mediators, burn patients have deficiencies in their immune system and cannot tolerate this invasion so, if not treated properly, septic shock, multiple organ failure and most probably death will follow.

Usually the most frequent bacteria found in burn wound culture are Staphylococcus (55%), Pseudomonas aeruginosa (14.29%), Enterococcus (12.24%), Escherichia coli (4%), Klebsiella and Proteus (both 2%).

Burn severity and the extent of burn injury are major risk factors for the development of infection in burn patients. As a matter of fact burns that involve a larger surface area of the body and burns that are deeper or more severe are more likely to become infected.

Goals

Our goal is to study how two treatments, a standard one (bathing care) and an alternative one (body cleansing), affect the time until a Staphylococcus infection.

The second one can be considered part of the *detoxification* practices, a series of alternative medicine systems whose purpose is the one to perform a physiological or medicinal removal of toxic substances from a living organism.

This study will focus on

long time effect of an infection which has not been adequately treated or has not been treated on time.

2) Descriptive analysis

First it's important to point out that 69% of the dataset consists of censored observations, which might lead to a lower accuracy and effectiveness of the result of our analyses.
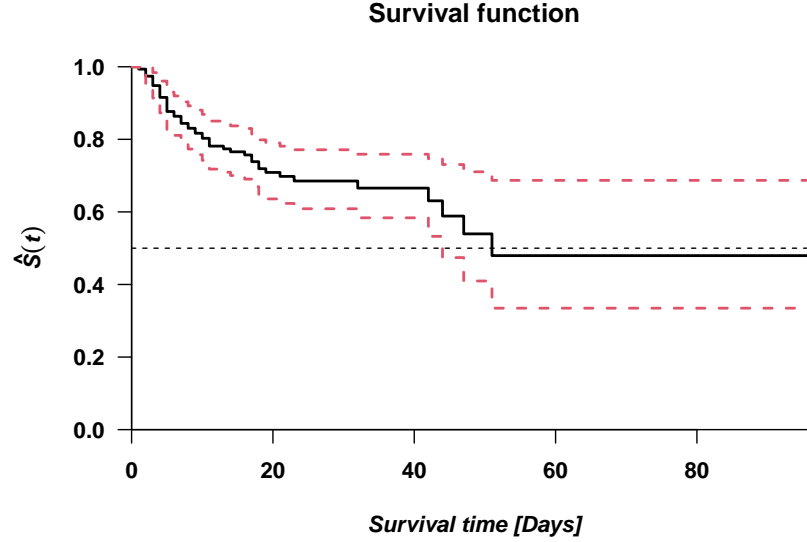
Let's start our preliminary analyses observing how our data are distributed:

DA METTER IN UN MFROW(2,2), MA GGPLOT FA LE BIZZE

And then we investigate about the correlation among themselves:

3) Nonparametric analysis

The overall survival function is

Survival function

and the median survival time is 51 days.

But if we observe the survival function for different levels of the categorical covariates we can notice very different patterns.

Since we want to study the effect of the two treatments on the patients, we'll study their effect both individually and in combination with the other covariates.

Treatment

We can immediately appreciate a difference of the survival function in the two treatments. They are stochastically ordered, with *body cleansing* being generally more effective than *bathing care* in fighting or delaying the insurgence of Staphylococcus infection. They are both quite smooth and stabilize respectively on the value 0.662 and 0.404 after 51 days. Therefore, to test the difference between these two survival functions we perform a Fleming-Harrington test with $\rho = 1$ and $\lambda = 1$. This test (with *pvalue* = 0.4) confirms at level 5% that there is a significant difference in the central part of the infection process.

Gender

Comparing the survival functions of men and women we notice that women show an overall higher survival probability, even though the two curves meet at around 40 days. We have also studied this trend through a bivariate analysis, dividing the observations both by gender and by treatment. This approach confirms a generally higher survival probability in women with respect to men and a remarkable difference in the two treatments, especially in the long term, to the point that men receiving *bathing care* reach $\hat{S}(t) = 0.16$ after 51 days. However, if we perform a Logrank test on the gender variable

$$\begin{cases} H_0 : S_{men}(t) = S_{women}(t) \\ H_1 : S_{men}(t) \neq S_{women}(t) \end{cases}$$

we get a χ^2 statistic of 2.4 on 1 degree of freedom, giving us a p-value of 0.1. Therefore, we don't have enough evidence to say that the survival times for male and female patients are different.

Body parts

We have then studied the influence of the location of the burn on the observed survival times. In almost all of these variables a graphical analysis suggests that the differences could be considered negligible, since the survival functions tend to overlap or follow very similar patterns. Only in the buttock variable we find a diverging pattern from day 0 to day 51, when the curves cross and remain quite close.

To corroborate this hypothesis we proceed to perform Logrank tests on these variables:

$$\begin{cases} H_0 : S_0(t) = S_1(t) \\ H_1 : S_0(t) \neq S_1(t) \end{cases}$$

It is probably relevant to remind the reader that the aim of this study is to analyse the treatments' effects in relation to the available covariates; however in this case no stratified test has been performed since the groups were quite balanced (maybe due to a particular design of the experiment), meaning that the two treatments were present in an approximately equal number of cases throughout the study of all the variables.

The tests are all not significant, most of them with very high p-values (in the range $[0.45, 0.73]$). It is interesting to point out that buttock is the one with the smallest pvalue (0.27), resulting the most significant among the location variables.

Survival functions according to burn type
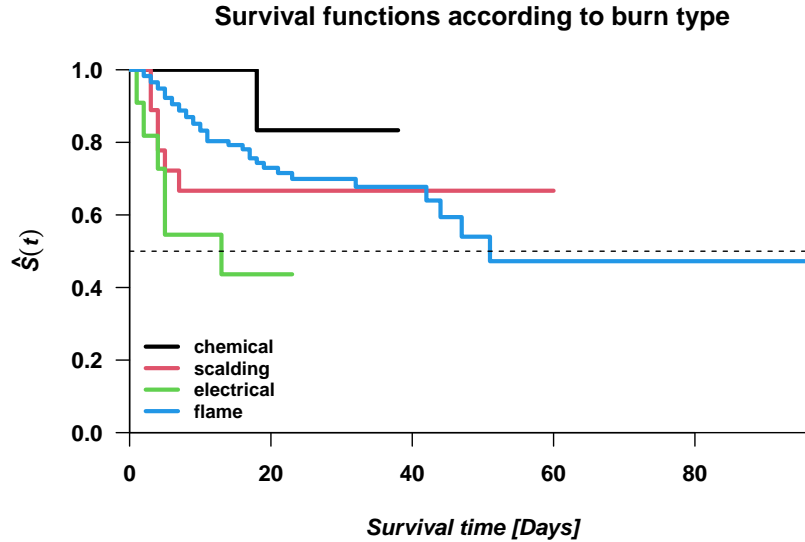
Burn type

Finally, we have compared the severity of the different kinds of burn, *Chemical*, *Scalding*, *Electrical* and *Flame*.

- The Scalding burn and the Electrical one seem to be the most critical ones in the first days after the accident with the Scalding reaching $\hat{S}(t) = 0.667$ after 7 days and the Electrical reaching $\hat{S}(t) = 0.436$ after only 13 days (very little compared to the median survival time of 51 days).
- The burn type with the highest survival probability is the Chemical, however this level has only one uncensored observation so we do not enough data to make a fair comparison between the groups.
- Lastly, the Flame burn shows a smoother trend, reaching $\hat{S}(t) = 0.473$ after 51 days. In general this variable presents quite diverse patterns. This is can be confirmed through a Logrank test:

$$\begin{cases} H_0 : S_{Chemical}(t) = S_{Scalding}(t) = S_{Electrical}(t) = S_{Flame}(t) \\ H_1 : \exists i, j \text{ s.t. } S_i(t) \neq S_j(t) \end{cases}$$

4) Fit of a parametric survival model

Fit of a Weibull, log-logistic, or lognormal model.

We'll fit the Weibull model:

$$Y = \ln T = \mu + \gamma_1 \cdot treatment + \gamma_2 \cdot gender + \gamma_3 \cdot head + \gamma_4 \cdot buttock + \gamma_5 \cdot torso + \gamma_6 \cdot upper\_leg$$

$$+\gamma_7 \cdot lower\_leg + \gamma_8 \cdot respiratory\_system + \gamma_9 \cdot burn\_type + \gamma_{10} \cdot burn\_percentage + \sigma \cdot W$$

Then we decide to apply a step-wise feature selection and reduce the model to a simpler one, using as classifying criterion the AIC

$$AIC = 2k - 2\ln(\text{maxLikelihood}), \text{ with } k = \text{n}° \text{ of variables}$$

which estimates the trade-off between the goodness of fit of the model and its simplicity.

The new model obtained is:

$$Y = \ln T = \mu + \gamma_1 \cdot treatment + \gamma_2 \cdot gender + \gamma_4 \cdot buttock + \gamma_9 \cdot burn\_type + \sigma \cdot W$$

We can notice how this result is somehow coherent with the tests we have previously performed on the survival times since both approaches have led to keeping and discarding the same sets of variables. / The same set of variable is kept after following the same procedure for a Log-logistic model. Therefore we consider ourselves satisfied with this feature selection phase and we can now proceed to inspect the residuals of the two models.
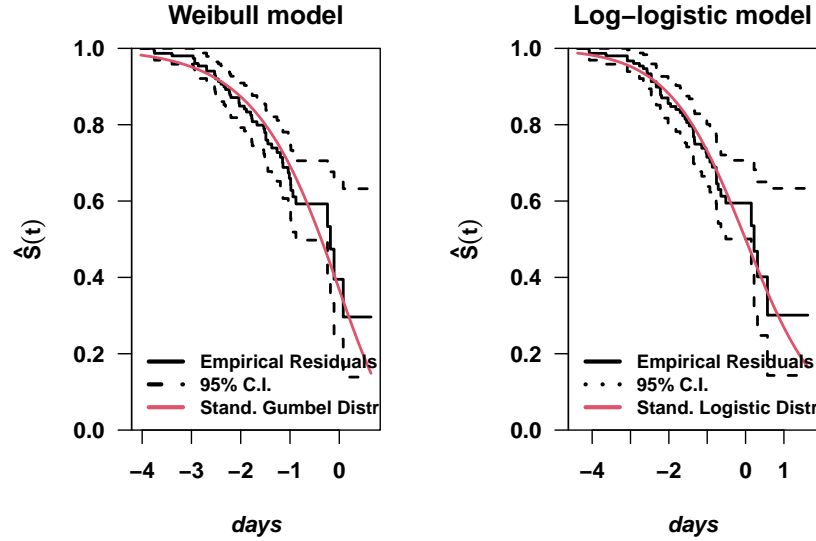
From a graphical analysis, in both the models the residuals seem to resemble quite much the theoretical ones, with almost all of the functions being inside the confidence intervals. Nonetheless we can notice how the Log-logistic model seems to have a slightly better behavior, being completely inside the intervals and closer to the empirical survival function.

This is also confirmed by the comparison of the logLikelihood of the models, with the Log-logistic having the highest value.

```
##    distribution Loglikelihood
## 1      Weibull      -241.627
## 2 Log-logistic      -239.684
```

Therefore we choose the Log-logistic model as the optimal one to describe our phenomenon and we proceed to interpret it.

Interpretation of the model fit.

It is known that a Log-logistic model can be written as a proportional odds model with respect to a baseline $Z = 0$ as:
$$\frac{S(t|Z)}{1 - S(t|Z)} = \exp(\frac{\gamma' Z}{\sigma}) \cdot \frac{S(t|Z = 0)}{1 - S(t|Z = 0)}$$

So that $\exp(\frac{\gamma_i Z_i}{\sigma})$ can be interpreted as the Odds Ratio between the ith covariate and the baseline: Man who received a burn of type *Chemical.*
While the general expression of the accelerating factor for a single variable Z_i is $AF = e^{\gamma_i Z_i}$

allows a representation as a proportional odds (PO) model.

The Odds Ratio represents the ratio of the odds of an event occurring in one group to the odds of the event occurring in another group.
In our case we can observe how, for each time t, the odds of surviving under the alternative treatment (body cleansing) are 2.37 times larger than the ones under the standard treatment (bathing care), confirming the considerations made above about the effectiveness of such treatments.
In the same way we can infer that, according to the model, the surviving odds for a woman are more than double the odds of a woman (precisely 2.25).
In addition, thanks to the low OR in the buttock variable ($OR = 0.44$) we can identify it as a **risk factor**: a person who burns himself on his buttock can see his survival odds reduced by almost half.
Finally, comparing the Odds Ratios of the burn type variable, we can once again confirm how grave an Electrical burn is with respect to the other types (with its odds being 5% of the chemical odds, 24% of the scalding odds and 17% of the flame odds). The Chemical reaffirms itself as the less dangerous, while Scalding and Flame can be deemed to be of almost comparable risk.

Analogous conclusions can be drawn by analyzing the Acceleration Factor since it indicates how the lifespan of an individual is accelerated or decelerated by a specific covariate. In particular if $AF_i < 1$, this means that the covariate is extending the survival time. Therefore we'll deduce that both being woman and using the alternative treatment will almost double the lifespan, while burning himself in the buttock and undergoing a Scalding/Electrical/Flame burn will reduce the lifespan respectively by a factor of 2, 4, 15 and 3.

5) Fit of a semi-parametric model

Fit of the proportional hazards models.

Interpretation of the model fit.

Interpretation of the model parameters in terms of relative hazards.

Analysis of the residuals to check the model's goodness-of-fit.

6) Conclusions