Computer Assignment 2

Jan Alexandersson, Anton Stråhle & Max Sjödin

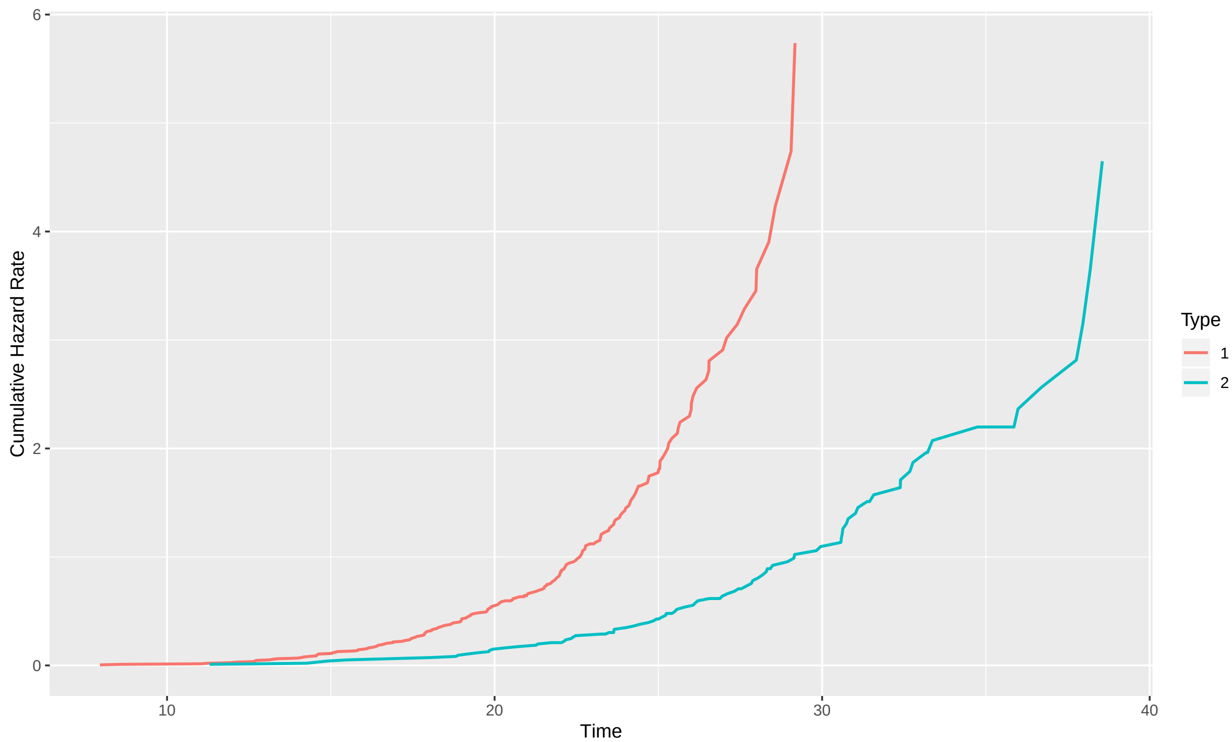
September 19, 2020

In this computer asssignment we aim to compare the survival rates of two groups using several different methods. The population is split (not evenly) into two types, henceforth called types 1 and types 2 containing 200 and 100 individuals respectivley. Their survival times are generated randomly from Weibull distributions as follows.

$$T_i^{(1)} \sim \mathrm{WB}(5.5, 22.5)$$
$$T_i^{(2)} \sim \mathrm{WB}(4.5, 28)$$

We also implement independent censoring using $\mathrm{Unif}(20, 60)$-censoring times.

We start of by plotting the Nelson-Aalen estimates of the two groups



From the figure above we see that the cumulative hazard rate for individuals of type 1 greatly exceeds that of those of type 2. We continue on by carrying out a logrank test for $\mathrm{H}_0 : \alpha_1(t) = \alpha_2(t)$ for all $t \in [0, t_0]$ where t_0 is the upper limit of the study time interval.

H_0 can be rejected on the 5%-level as $\frac{Z_1(t_0)}{\sqrt{V_{11}(t_0)}} = 9.8879152$ exceeds $\Phi(0.975)$. Next up is to fit a Cox regression model to our data. We have that

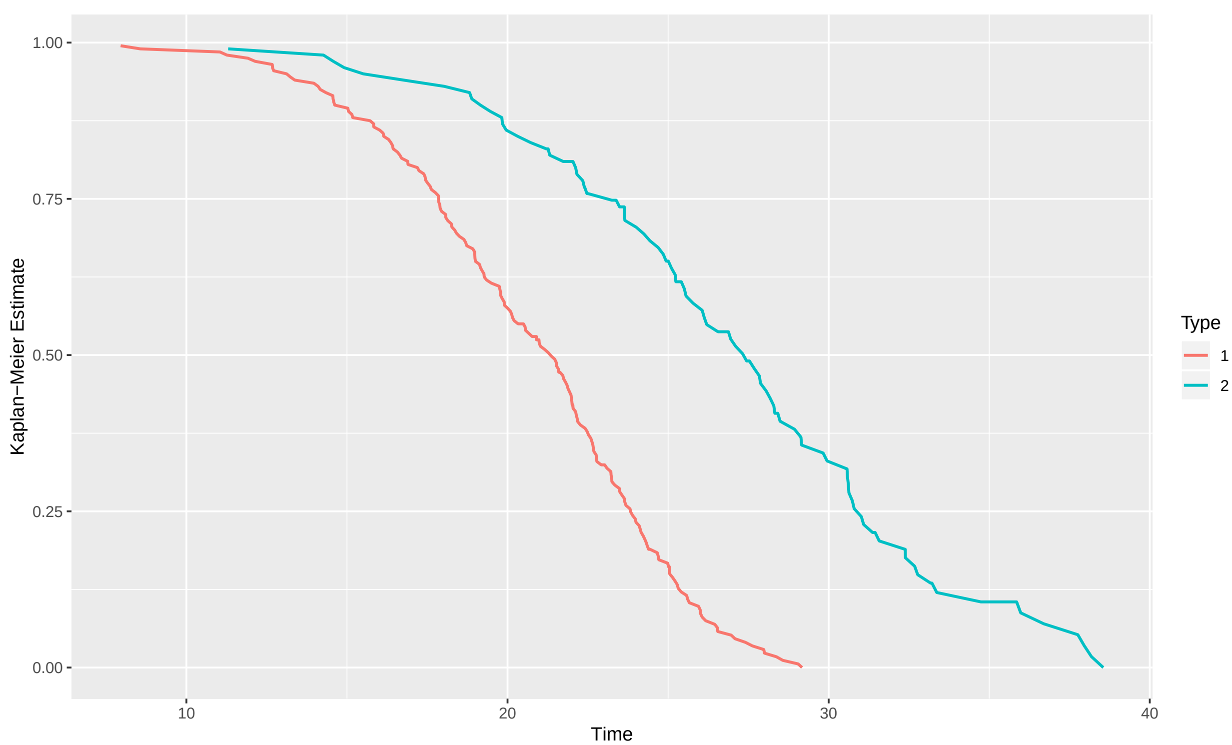
$$\alpha(t|x_i) = \alpha_0(t)e^{\beta x_i} \quad \text{where } x_i = \mathbf{1}_{\{\text{i}\in\text{type 2}\}},$$

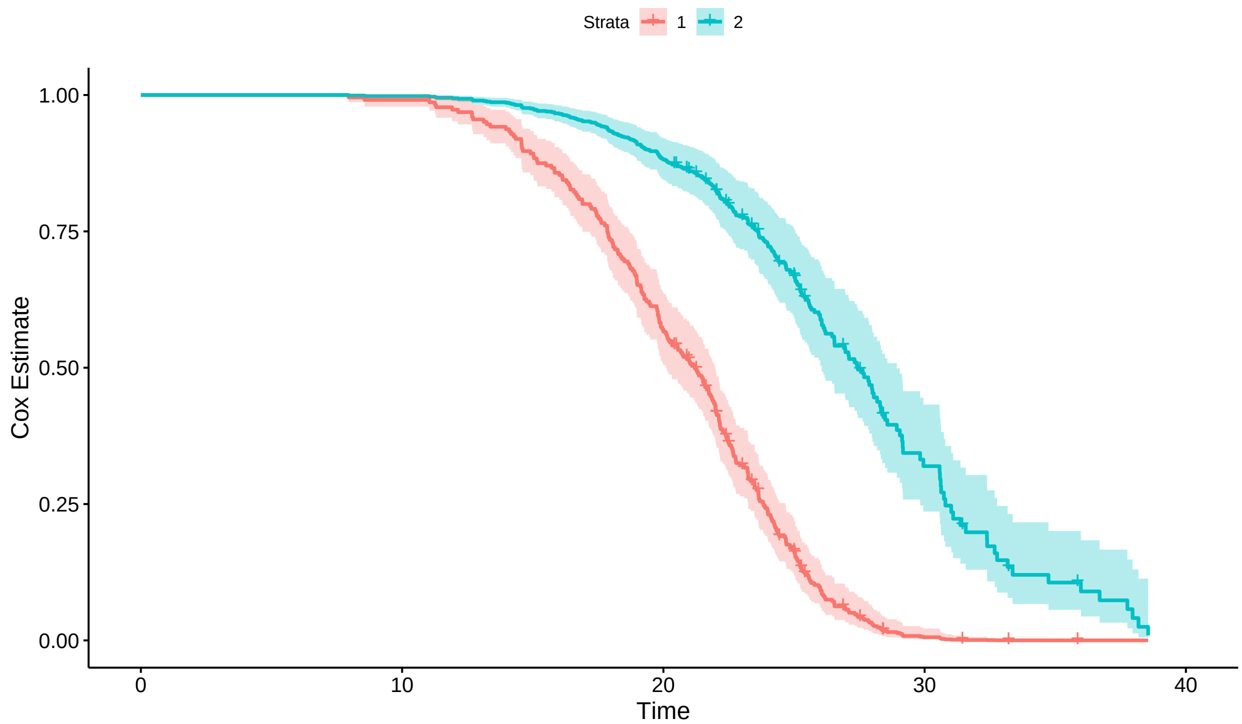
where $\alpha(t|x_i)$ is the hazard rate for individual i and $\alpha_0(t)$ is the baseline hazard rate. Since, $x_i = 0$ for type 1 we see that $\alpha_0(t)$ corresponds to the hazard rate of type 1 and e^{β} is the relative risk for type 2.

After fitting we get the following,

```
## Call:
## coxph(formula = SurvObj ~ as.factor(Type), data = times)
##
##   n= 300, number of events= 275
##
##                    coef exp(coef) se(coef)      z Pr(>|z|)
## as.factor(Type)2 -1.505     0.222    0.162 -9.293   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                  exp(coef) exp(-coef) lower .95 upper .95
## as.factor(Type)2     0.222      4.505    0.1616    0.3049
##
## Concordance= 0.636  (se = 0.015 )
## Likelihood ratio test= 103.6  on 1 df,   p=<2e-16
## Wald test            = 86.36  on 1 df,   p=<2e-16
## Score (logrank) test = 97.77  on 1 df,   p=<2e-16
```

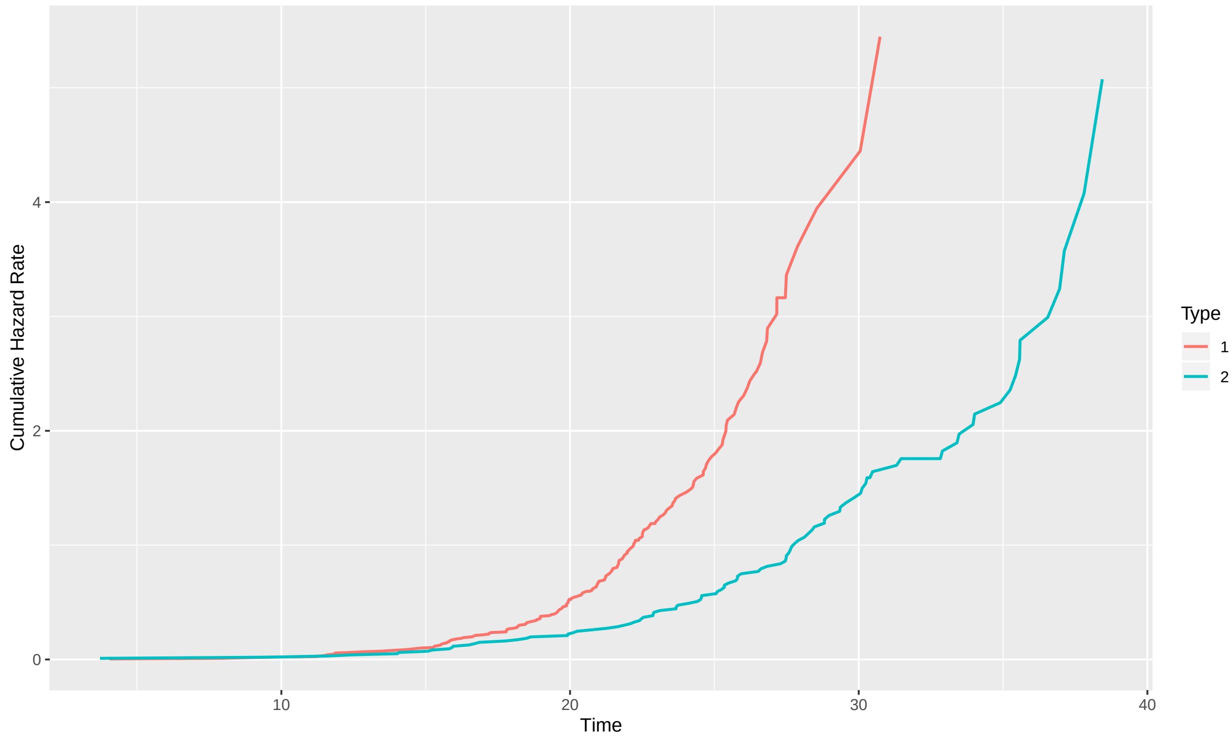
From the summary we see that we have $\hat{\beta}$ = -1.5052002 and the estimated relative risk $e^{\hat{\beta}}$ = 0.2219729. We see from the model summary above that the performed Wald test yields a highly significant difference between the two types. We can also compare the survival curves obtained by the Kaplan-Meier estimator and those obtained via the Cox regression.

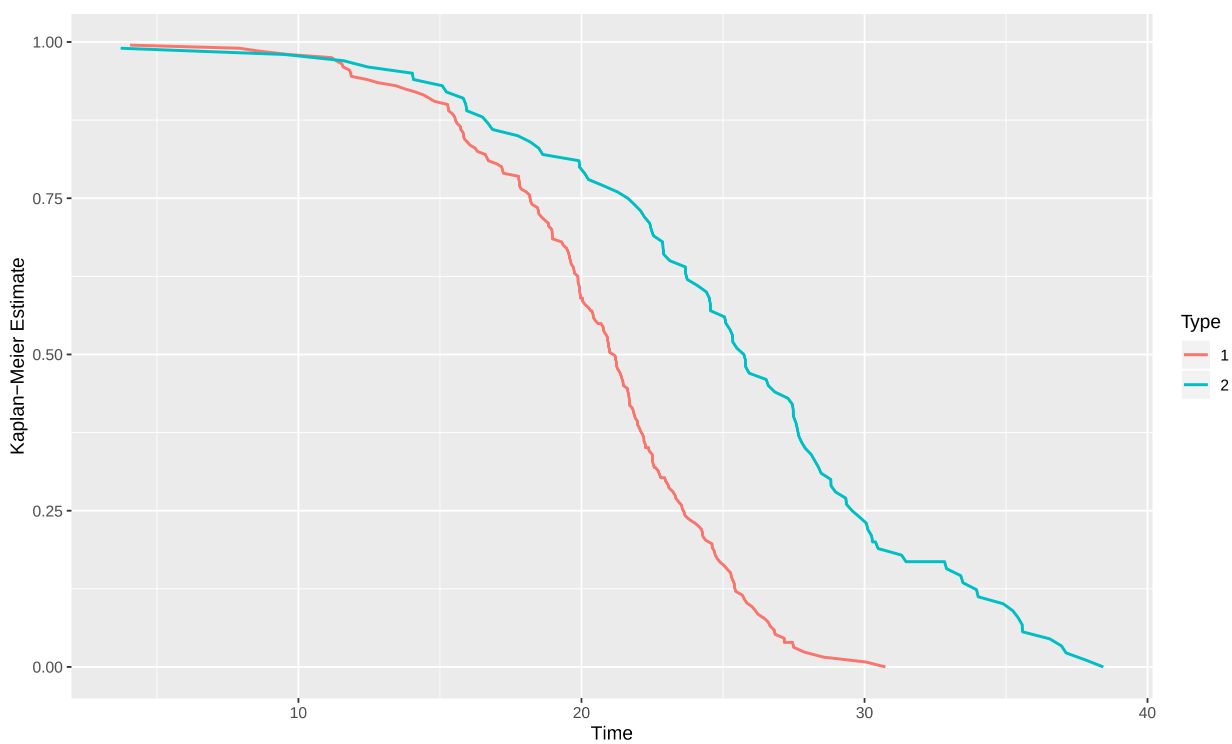


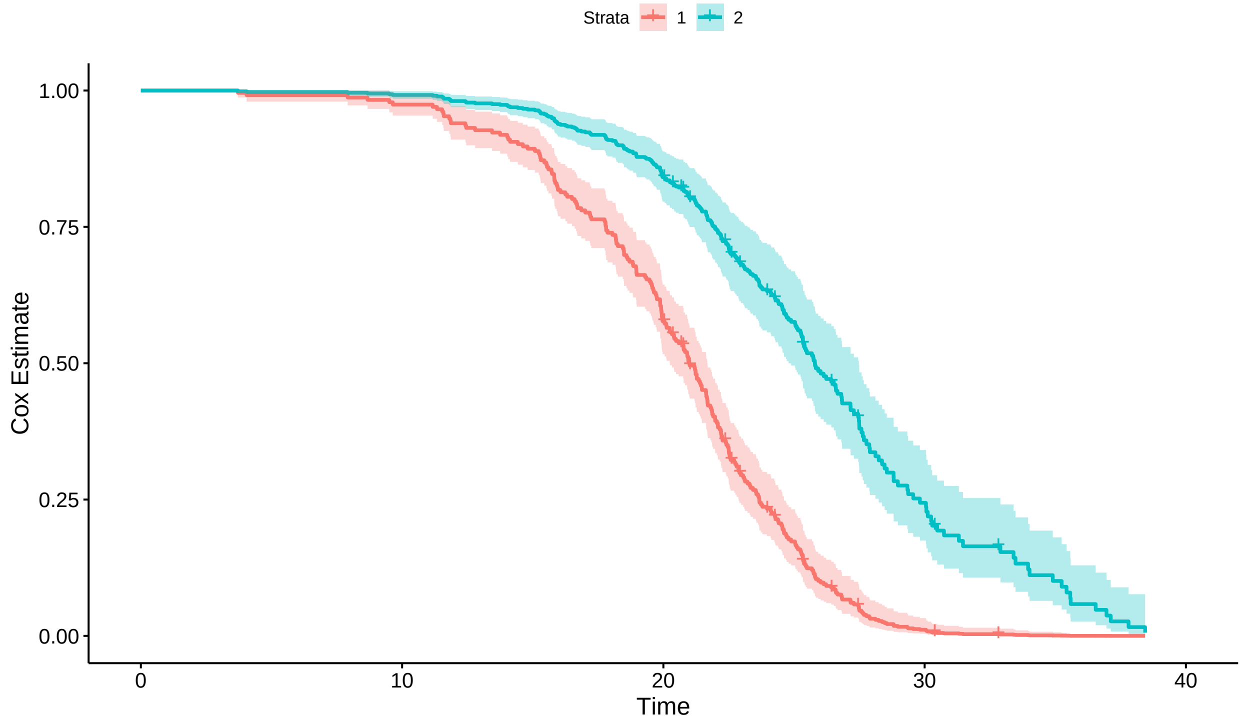


We see that the two figures are pretty much identical. We also note from the Kaplan-Meier curves that the assumption of proportional hazards seems to be somewhat fulfilled as the effect of Type seems to be somewhat constant over time (ignoring the endpoints of the interval) and since the two lines dont seem to cross.

Lastly we want to examine the differences that occur when we use dependent instead of independent censoring. As such we censor individuals of type 1 with $\text{Unif}(20, 60)$ censoring times as before whilst changing the distribution of censoring times for type 2 individuals to $\text{Unif}(30, 60)$







Visually there are not many differences when comparing the results. The main takeaway is that we have fewer "early" censorings for type 2 indviduals in the case of dependent censoring which leads to even lower cumulative hazard rates (as can be seen when comparing the figures including the Nelson-Aalen estimates) and higher survival rates (as can be seen when comparing the survival curves obtained through Kaplan-Meier and Cox regerssion). Also, when redoing the previous tests, both the logrank and when refitting the Cox model we get very similar results. We do however note that the assumption of proprotional hazard can be questioned as the two lines actually intersect. We also note that the log-rank test statistic is 8.0294715 which means that we can reject H_0 again.