Computer Assignment 2

Jan Alexandersson, Anton Stråhle & Max Sjödin

September 19, 2020

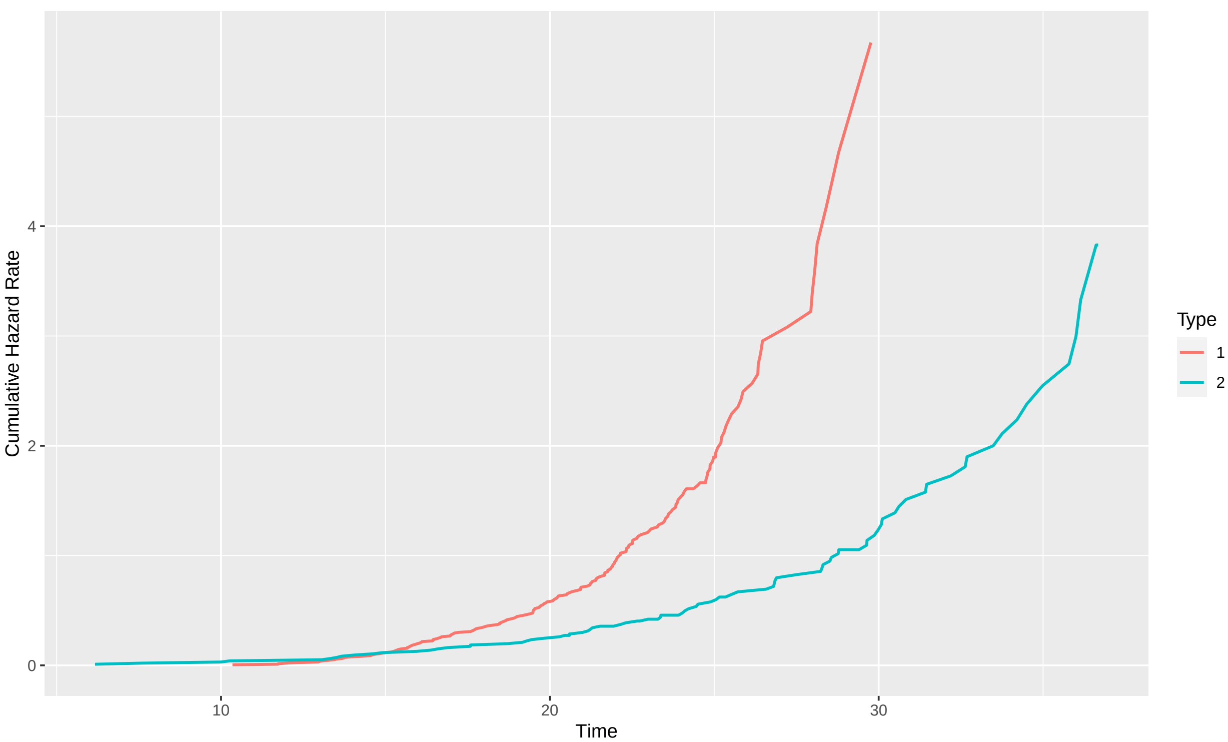
- Since you know that data is generated using certain Weibull distributions, you know whether or not the proportional hazards assumption is fulfilled - can you based on the Nelson-Aalen (or Kaplan-Meier) estimates say something about the assumption regarding proportional hazard?
- What does the logrank test say?
- Is there a difference between the two groups with respect to estimated regression coefficients?
- Comment on the comparison between the Kaplan-Meier curves and the Cox-curves.
- What is the effect of that the censoring is group dependent?

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.

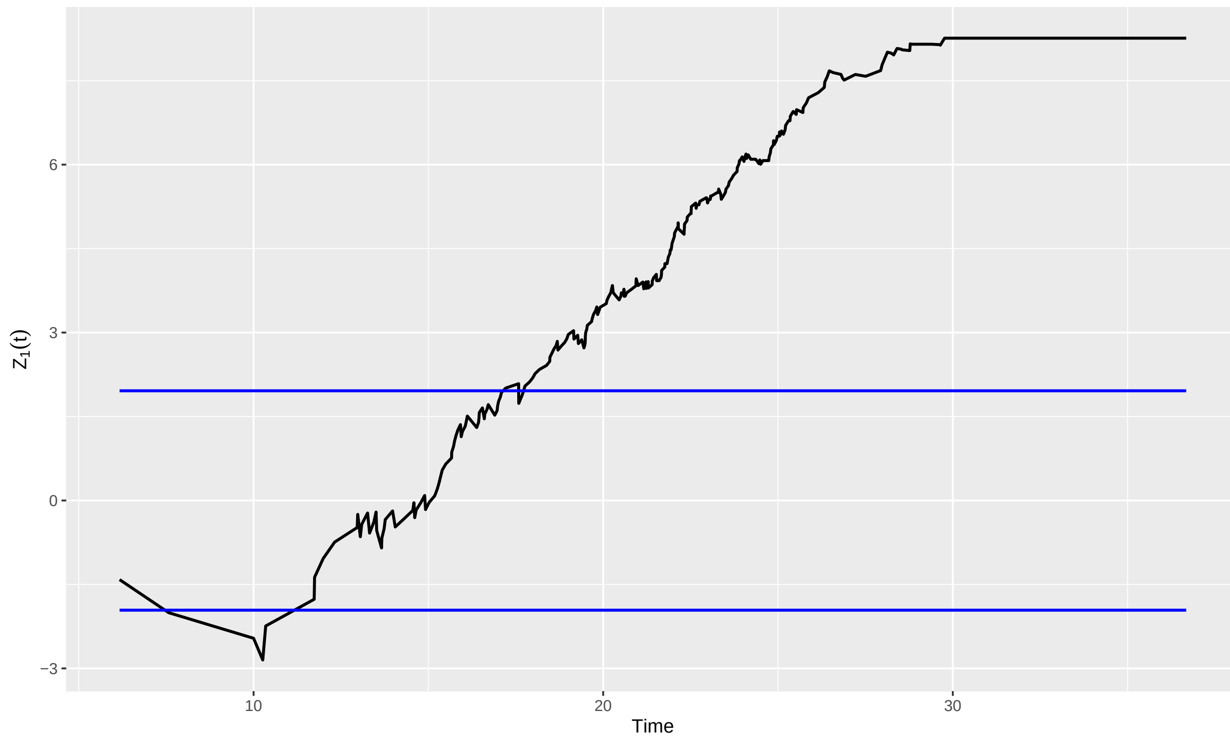
$$T_i^{(1)} \sim \text{WB}(5.5, 22.5)$$
$$T_i^{(2)} \sim \text{WB}(4.5, 28)$$

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



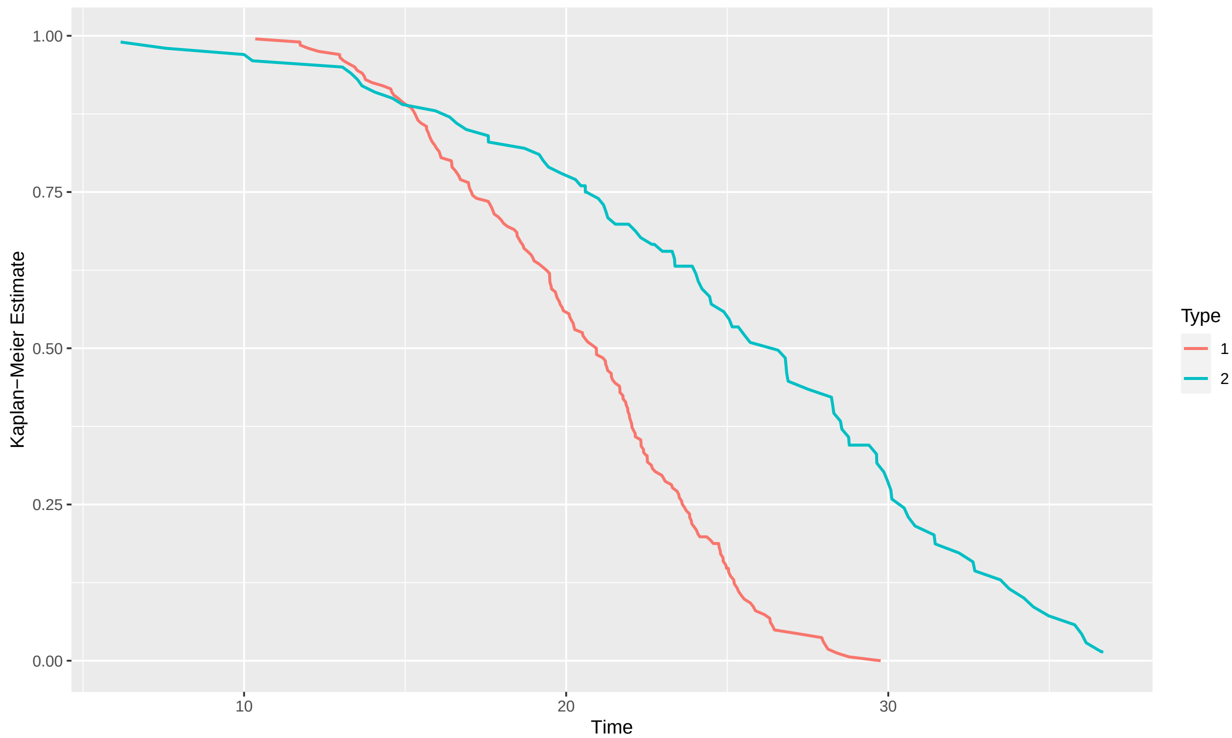
We see from the figure above that H_0 can be rejected on the 5%-level as $Z_1(t_0)$ exceeds $\Phi(0.975)$. Next up is to fit a Cox regression model to our data. We have that

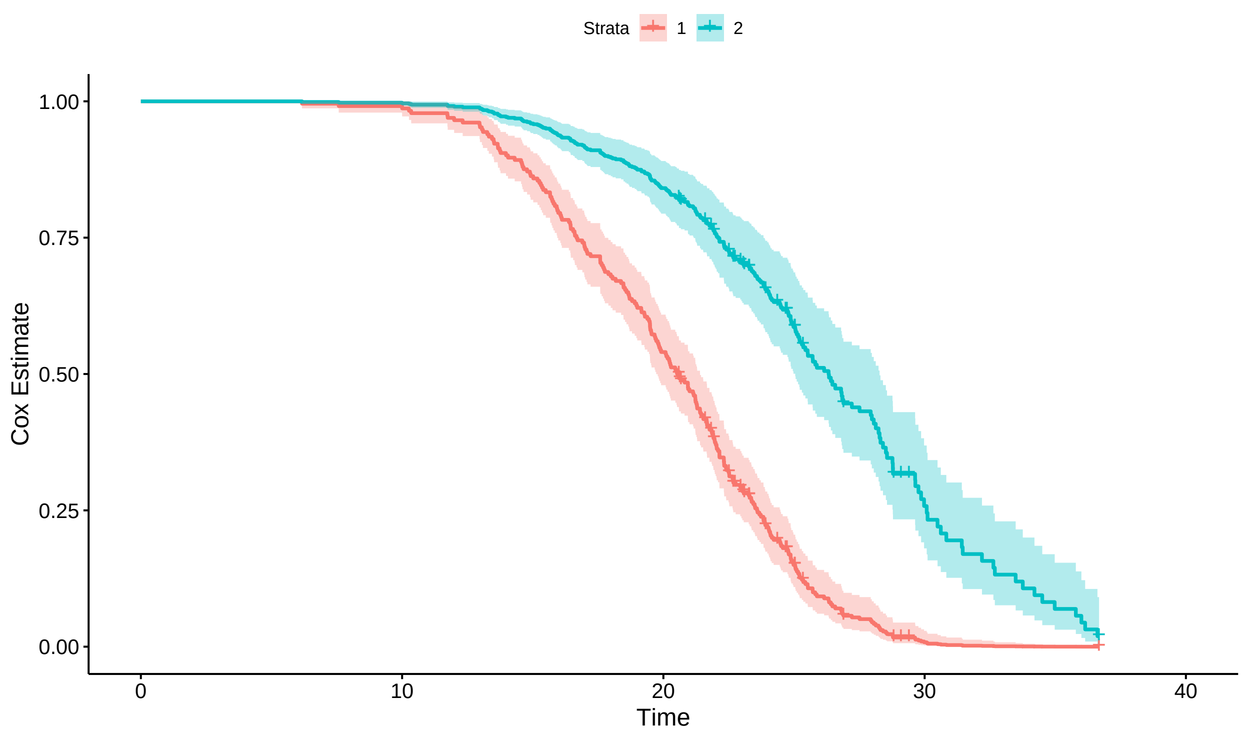
$$\alpha(t|x_i) = \alpha_0(t)e^{\beta x_i} \quad \text{where } x_i = \mathbf{1}_{\{\mathrm{i}\in\mathrm{type}\ 2\}}$$

After fitting we get the following,

```
## Call:
## coxph(formula = SurvObj ~ as.factor(Type), data = times)
##
##   n= 300, number of events= 274
##
##                      coef exp(coef) se(coef)      z Pr(>|z|)
## as.factor(Type)2 -1.2676    0.2815   0.1607 -7.888 3.07e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                  exp(coef) exp(-coef) lower .95 upper .95
## as.factor(Type)2    0.2815      3.552    0.2054    0.3857
##
## Concordance= 0.599  (se = 0.017 )
## Likelihood ratio test= 73.55  on 1 df,   p=<2e-16
## Wald test            = 62.22  on 1 df,   p=3e-15
## Score (logrank) test = 68.22  on 1 df,   p=<2e-16
```

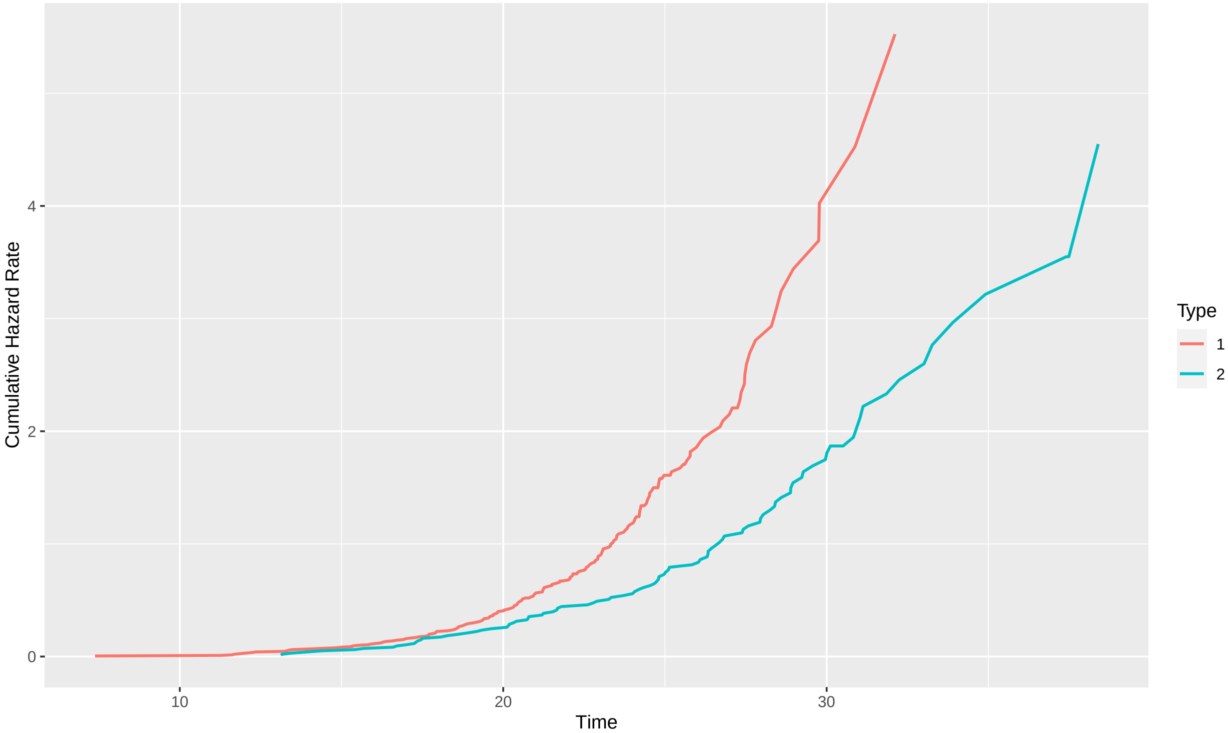
We see from the model summary above that the performed Wald test yields a highly significant difference between the two types. We can alos compare the survival curves obtained by the Kaplan-Meier estimator and those obtained via the Cox regression.

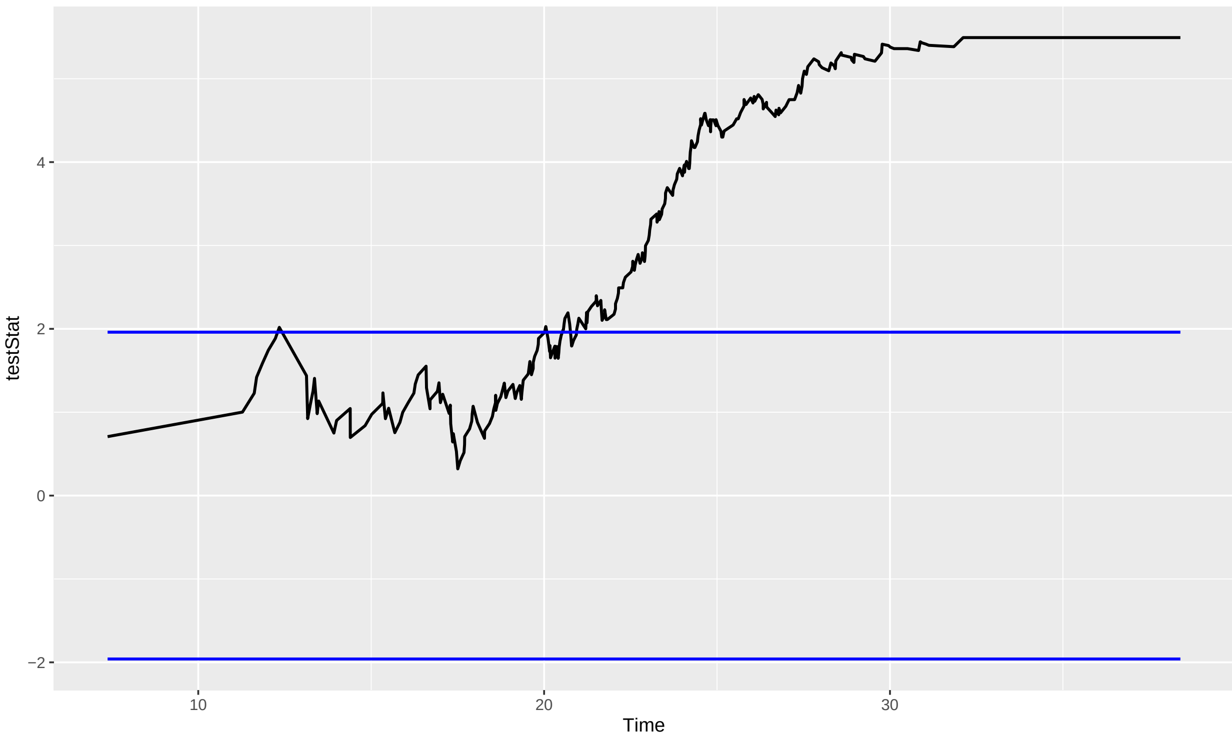


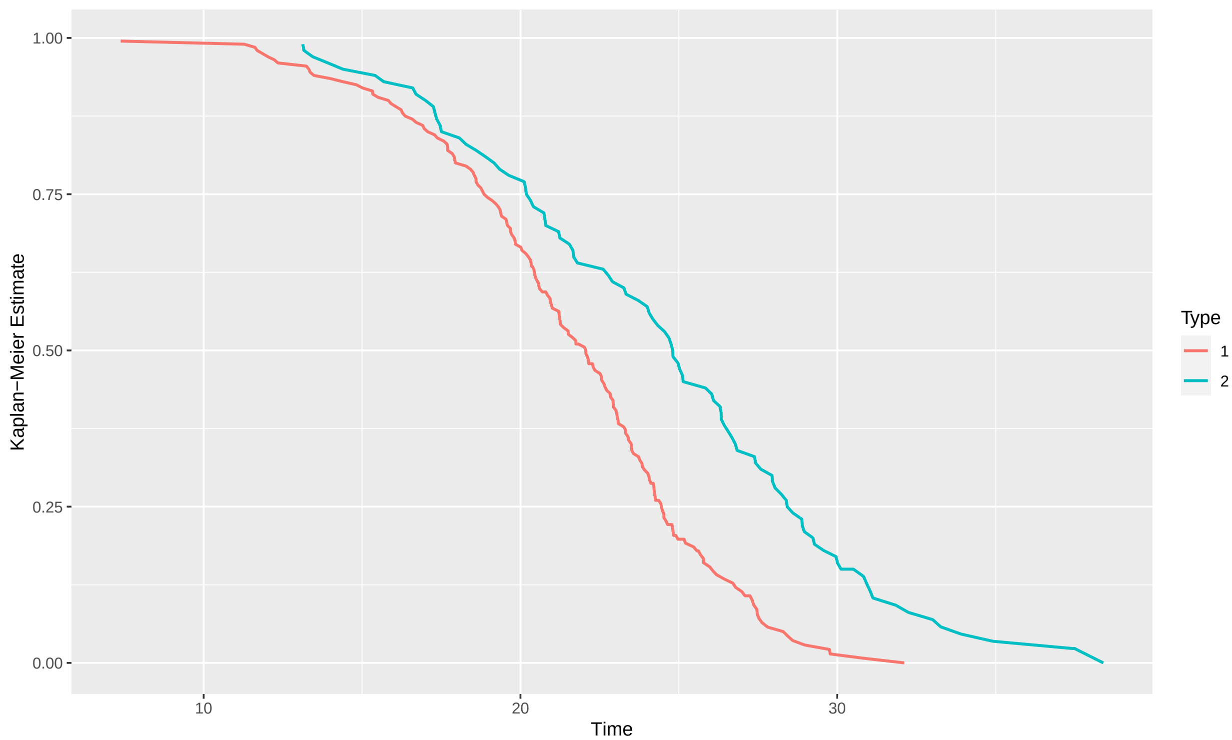


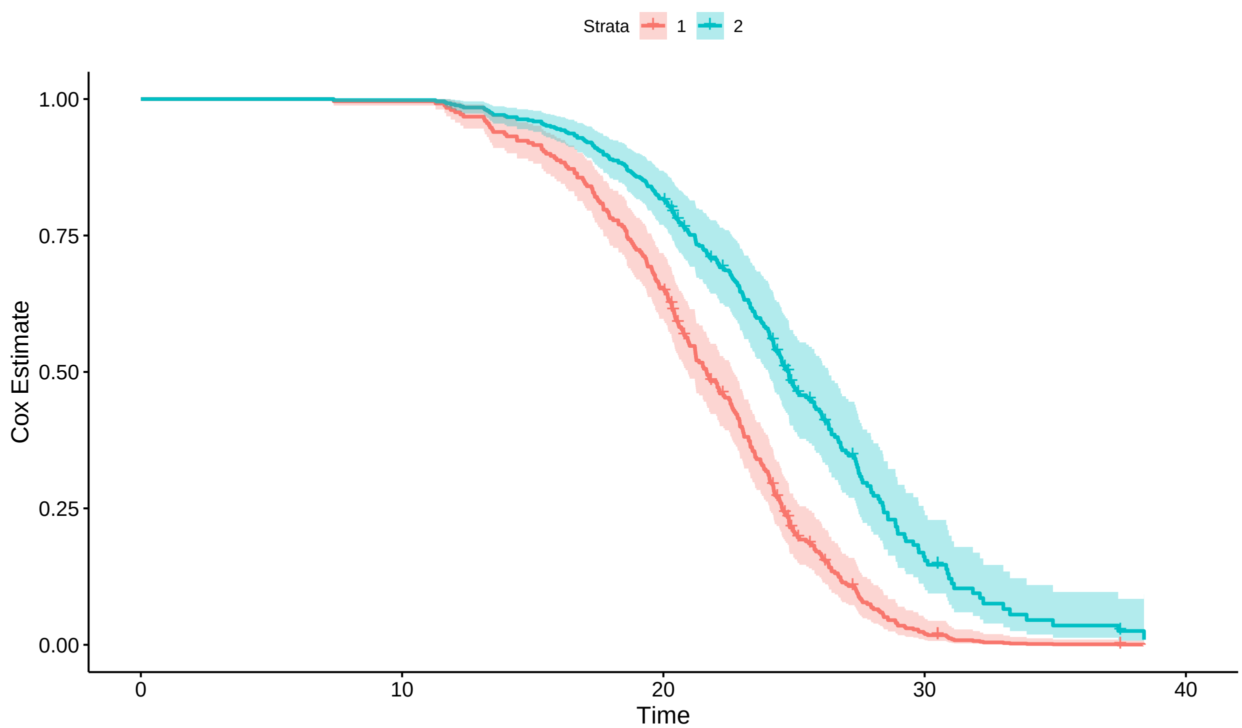
We see that the two figures are pretty much identical. We also note from the Kaplan-Meier curves (as well as from the those obtained via Cox regression) that the assumption of proportional hazards seems to be somewhat fulfilled as the effect of Type seems to be somehwta constant over time (ignoring the endpoints of the interval).

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 Unif$(20, 60)$ censoring times as before whilst changin the distribution of censoring times for type 2 individuals to 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).