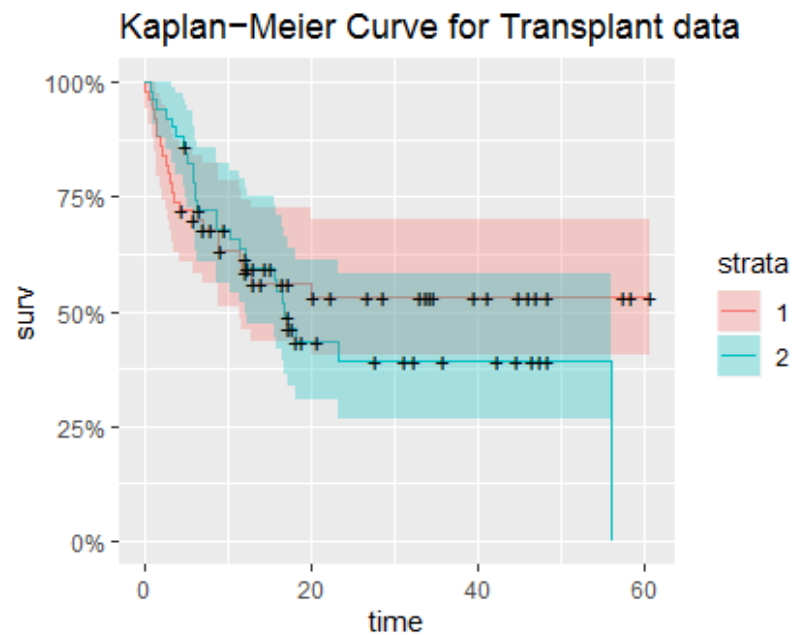


Group: Zining Fan, Mutian Wang, Siyuan Wang

UNI: zf2234, mw3386, sw3418

Graded Homework 2- Exercise 1

1. Yes, random censoring is the right choice. Because it's unlikely that transplantations happen at the same time. That is, we record the data independently at different times.
2. There seems a difference between the two groups. The first type of transplant, allogeneic, seems to be more efficient. However, the significance is unknown.



3. The signs of the fitted parameters agree with my intuition. The sign of the transplant type is negative, so the surviving time of type 1 transplant is longer.

Call:

```
survreg(formula = Surv(df$V1, df$V3) ~ df$V2, data = df, dist = "exponential")
```

	Value	Std. Error	z	p
(Intercept)	4.066	0.466	8.72	<2e-16
df\$V2	-0.325	0.285	-1.14	0.25

Scale fixed at 1

Exponential distribution

Loglik(model)= -228 Loglik(intercept only)= -228.6

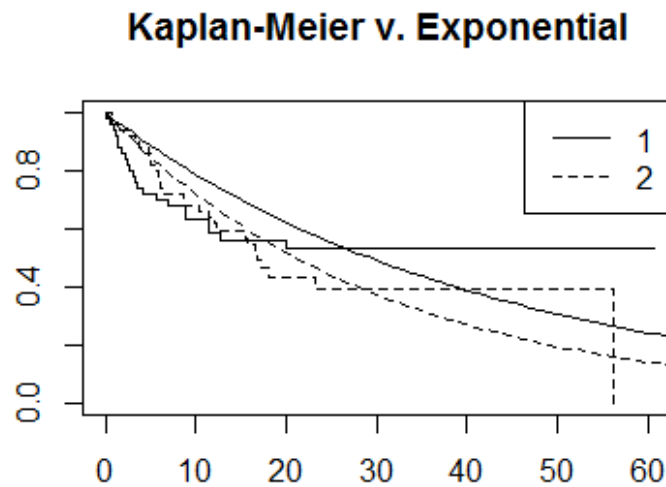
Chisq= 1.31 on 1 degrees of freedom, p= 0.25

Number of Newton-Raphson Iterations: 5

n= 101

4. The p-value is 0.25 for the two groups. Therefore, we cannot observe a significant difference under the 5% significance level. This conclusion depends on parametric model assumptions, since we are fitting an exponential model.
5. Kaplan-Meier curve (non parametric model) is a step function, while exponential curve (parametric model) is a smooth curve.

When survival time < 25 , these two models show similar results, but exponential model slightly overestimates the probability of survival; when time > 25 , Kaplan-Meier curve flattens, while exponential curve continues decreasing. That is, exponential model does not fit very well when time > 25 .



6. Call:

```
survreg(formula = Surv(df$V1, df$V3) ~ df$V2, data = df, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	4.342	0.698	6.22	5e-10
df\$V2	-0.374	0.420	-0.89	0.3742
Log(scale)	0.388	0.123	3.15	0.0016

Scale= 1.47

Weibull distribution

Loglik(model)= -222 Loglik(intercept only)= -222.4

Chisq= 0.8 on 1 degrees of freedom, p= 0.37

Number of Newton-Raphson Iterations: 5

n= 101

```
pchisq(2*(model_w$loglik[2]-model_exp$loglik[2]), df = 1)
[1] 0.9994362
```

Considering the coefficients and p-value of this model, it seems that it does not fit the data better than the exponential model. Moreover, the p-value of likelihood ratio test is 0.999, providing no evidence against the exponential model. This can also be visually confirmed by comparing the Weibull curve and exponential curve - the below figure shows a very similar result as the previous figure.

