

Assignment 3

Zhiwen Tan

3/14/2022

Question 1

a) To get the 95% confidence interval for median survival time, that is find $t_{0.5}$. We first need to fit a Weibull model for the data at stress level $750N/mm^2$, the result below is the model fit result. Then we have $\hat{\sigma}$, $\hat{\mu}$ and estimate \hat{Var} . Then let $Y = \log T$, and we can calculate $y_{0.5}$. After this step, we can calculate the $Var(y_{0.5})$. Now, we can use 95% CI for $y_{0.5}$. Finally, we can use $t_{0.5} = \exp(y_{0.5})$ to get 95% confidence interval for median($t_{0.5}$) is [5142.356, 14100.130].

```
##
## Call:
## survreg(formula = Surv(x, delta) ~ 1)
##           Value Std. Error      z      p
## (Intercept)  8.798        0.234 37.55 <2e-16
## Log(scale)  -0.376        0.267 -1.41   0.16
##
## Scale= 0.686
##
## Weibull distribution
## Loglik(model)= -86.8   Loglik(intercept only)= -86.8
## Number of Newton-Raphson Iterations: 6
## n= 10

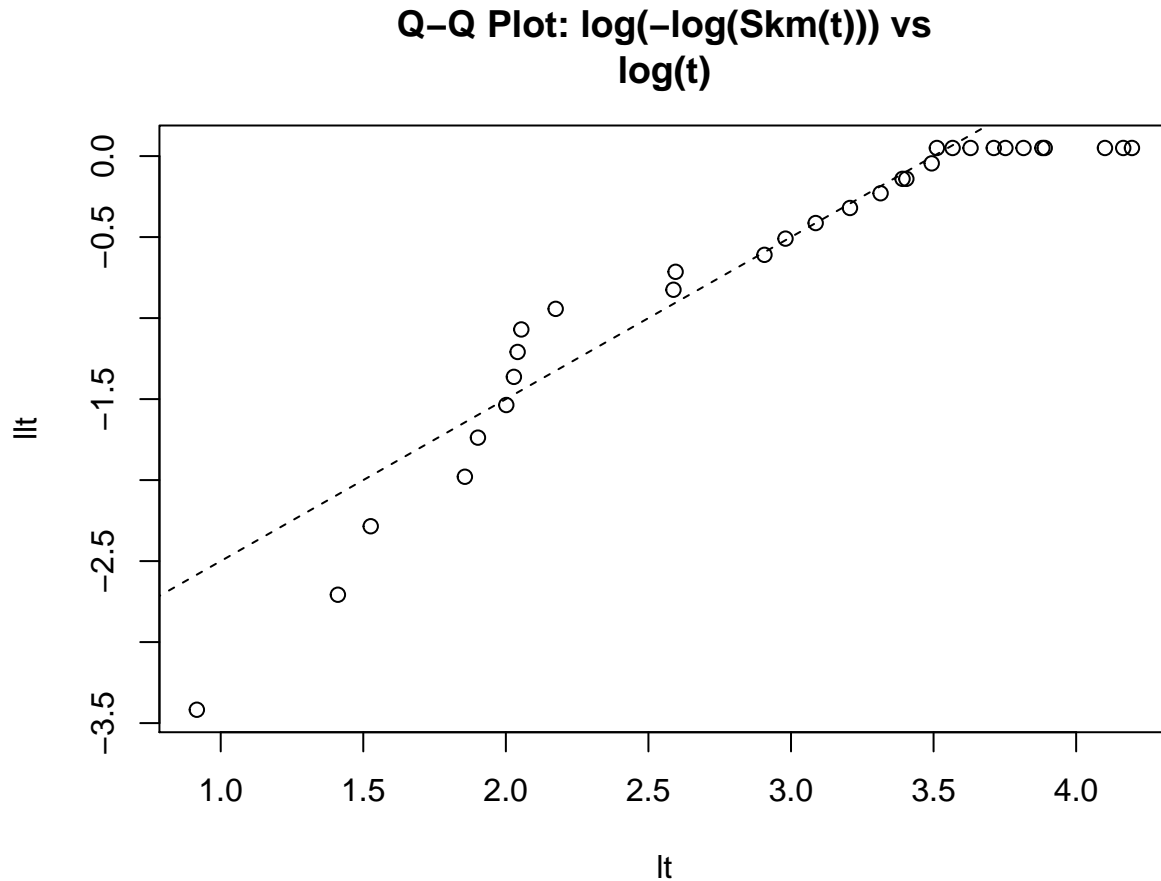
## (Intercept) (Intercept)
##    5142.356   14100.130
```

b) To get the confidence interval for the probability that the type of spring survives over 6000 thousand cycles at the stress level $750N = mm^2$. that is find the $S(6000)$. The first step is to find the CI for $\log[-\log(S(t))]$, this is equal to $(\log(t) - \mu)e^{-\phi}$. Then we can find $var((\log(t) - \mu)e^{-\phi})$ using δ -method. Now, we have the 95% CI for $\log[-\log(S(6000))]$, Then we can calculate CIl and CIu by using $\exp(-\exp(\hat{\psi}_u))$ and $\exp(-\exp(\hat{\psi}_l))$. Finally, we can get the result 95% CI is [0.108, 0.550]

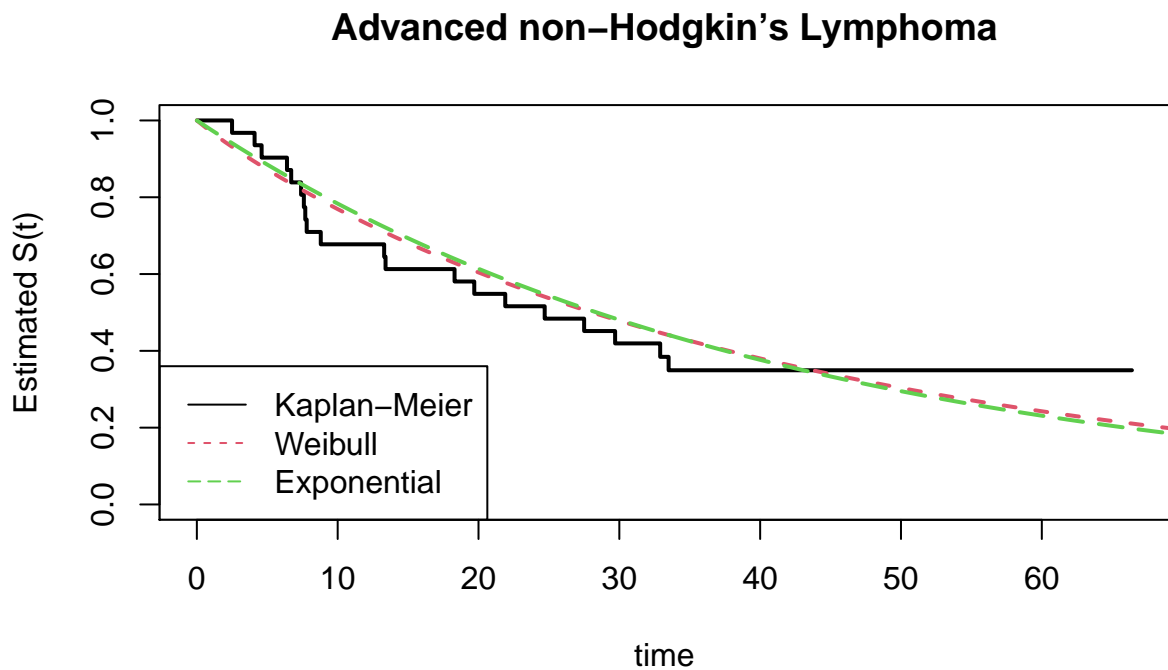
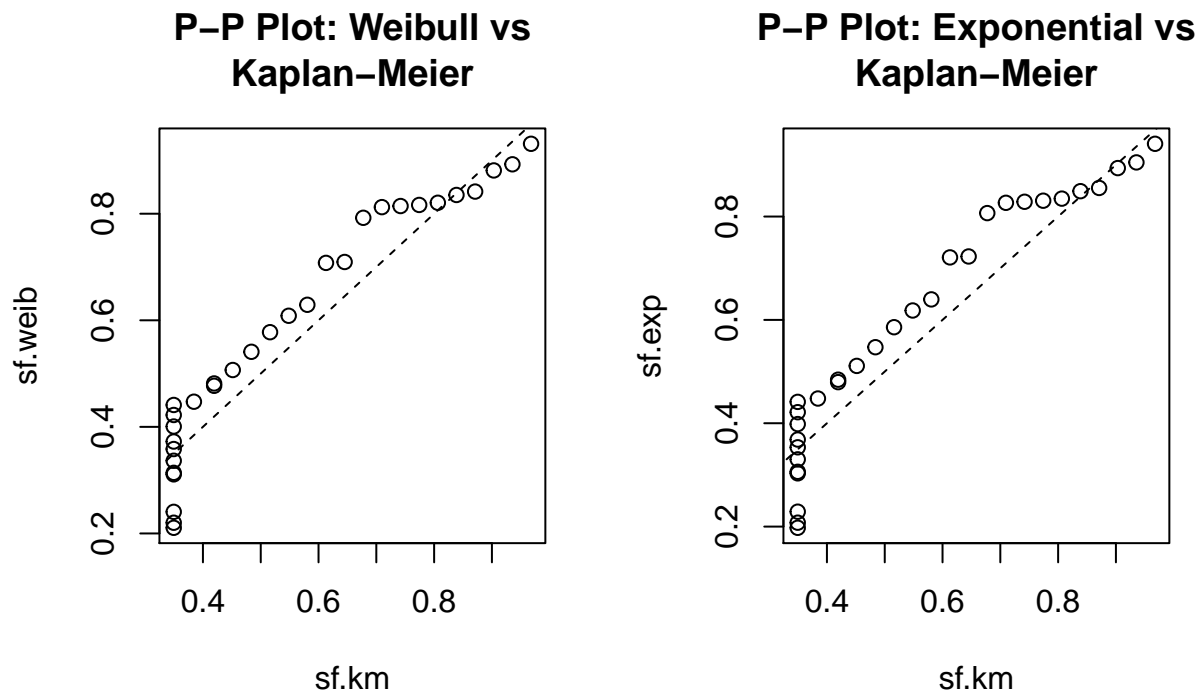
```
## [1] 0.1078917 0.5496670
```

Question 2

a) If weibull distribution fits well, then the Q-Q plot for $\log(-\log(Skm(t)))$ vs $\log(t)$ should be show an approximatly linear relationship. However, form the plot below, we can see the Q-Q plot is not very linear, this means weibull distribution may not fit the data very well.



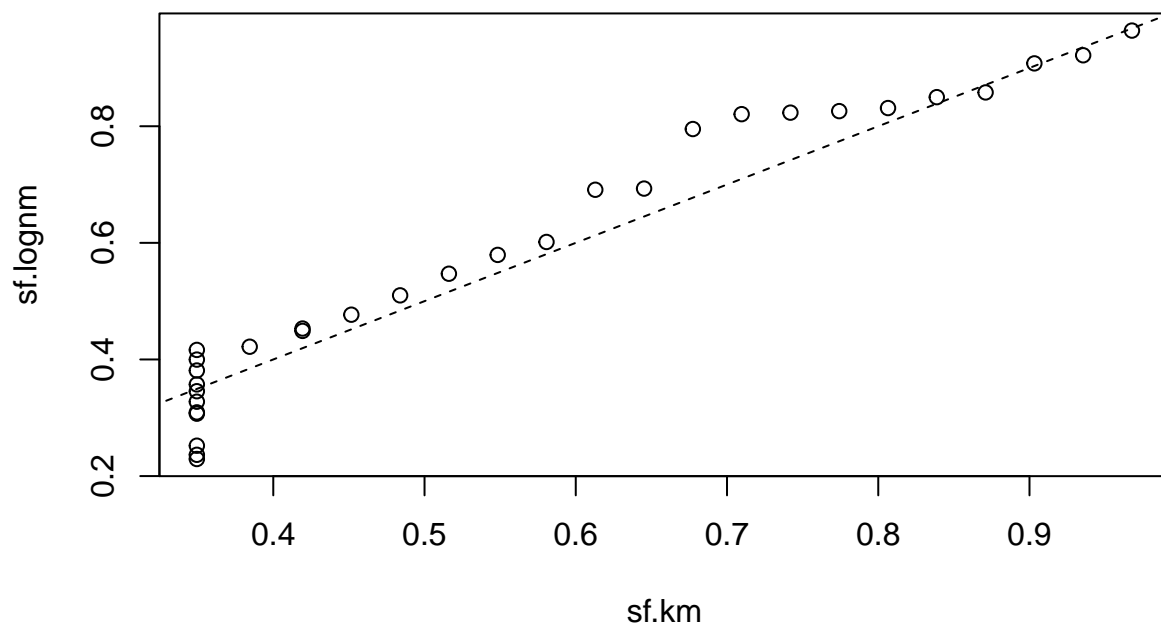
b) From the P-P plot, the difference between weibull and exponential estimation are very small and all patterns are similar. From the third plot, we can also see the survival curve for weibull and exponential distribution are very close to each other. However, we can see none of the P-P plot shows a linear pattern and we can also see neither weibull nor exponential distribution can give very accurate estimation between $t=5$ to $t=40$. Therefore, exponential distribution is not good enough for the data.



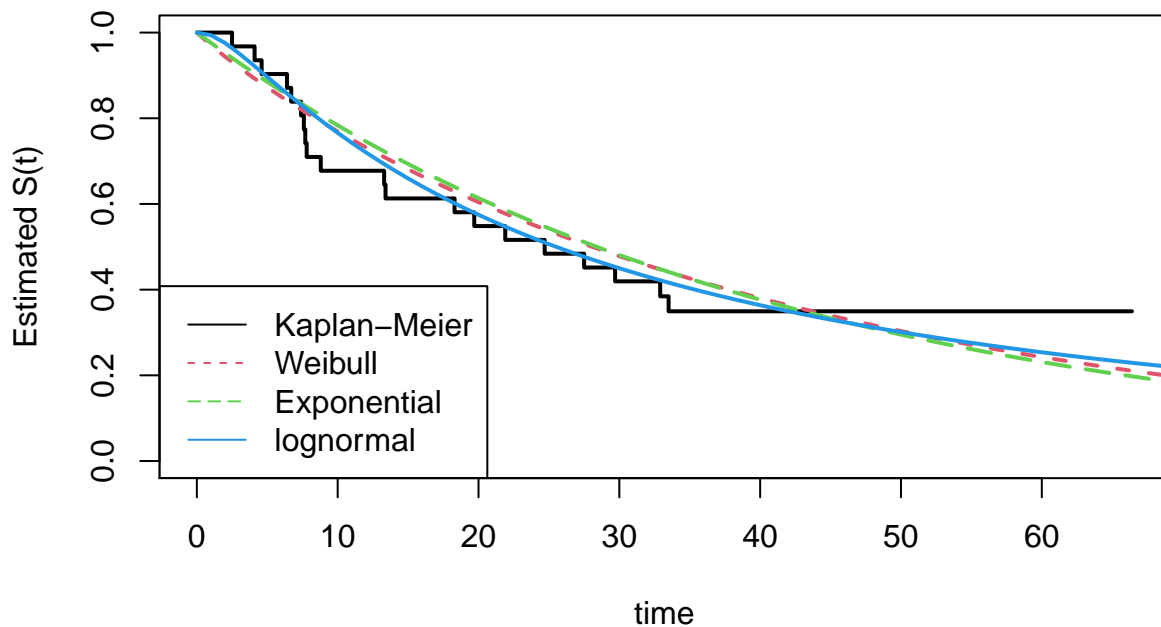
- c) From the P-P plot, we can see the points are more closer to the line. therefore, even though the data itself is not perfect and the P-P plot does not show a clear linear pattern, the log normal distribution fit is still better than weibull and exponential distribution. From the second plot, we can clear see the log normal distribution survival curve is more closer to the KM curve compare to exponential and weibull distribution. To conclude, Weibull and exponential model fits the data reasonably well for small and

large survival times, but they do not fit so well for intermediate survival times. log normal model fits the data better in the entire range.

P-P Plot: LogNormal vs Kaplan–Meier



Advanced non–Hodgkin’s Lymphoma



Question 3

a) To test equality of failure time distributions at stress levels 850 and 950 N/mm^2 , we need to test $S_1(t) = S_2(t)$. The first step is to test $H_0 : \sigma_1 = \sigma_2$, because R code will only provide ϕ , so we need to check if $H_0 : \phi_1 = \phi_2$. To test this we need to fit a full model and two separate models for two levels. Then we can get the `loglike(model)` value, add this value together for 850 and 950 levels and minus the full model, we get 0.34 with a p-value 0.31. This means we can not reject the H_0 , that is $\sigma_1 = \sigma_2$. The next step is to check $H_0 : \mu_1 = \mu_2$. Because we already have the model, so we can just use `loglike(model) - loglike(intercept)` to get the observed value which equals to 33.68. The p-value is 6.5e-09, this can be obtained from the full model. Because $6.5e-09 < 0.05$, so we can reject the null hypothesis. This means $\mu_1 \neq \mu_2$. The test suggests that for describing the survival times of the 2 groups, we should consider a Weibull model with $\sigma_1 = \sigma_2$, but $\mu_1 \neq \mu_2$.

```
##
## Call:
## survreg(formula = Surv(time, status) ~ as.factor(stress), data = data1)
##               Value Std. Error      z      p
## (Intercept)      5.9111      0.0492 120.3 <2e-16
## as.factor(stress)950 -0.7033      0.0683 -10.3 <2e-16
## Log(scale)       -1.8812      0.1746 -10.8 <2e-16
##
## Scale= 0.152
##
## Weibull distribution
## Loglik(model)= -103.1   Loglik(intercept only)= -120
## Chisq= 33.68 on 1 degrees of freedom, p= 6.5e-09
## Number of Newton-Raphson Iterations: 6
## n= 20

##
## Call:
## survreg(formula = Surv(time, status) ~ 1, data = data1[data1$stress ==
##      850, ])
##               Value Std. Error      z      p
## (Intercept)   5.9176      0.0454 130.35 < 2e-16
## Log(scale)   -1.9951      0.2510  -7.95 1.9e-15
##
## Scale= 0.136
##
## Weibull distribution
## Loglik(model)= -54.2   Loglik(intercept only)= -54.2
## Number of Newton-Raphson Iterations: 6
## n= 10

##
## Call:
## survreg(formula = Surv(time, status) ~ 1, data = data1[data1$stress ==
##      950, ])
##               Value Std. Error      z      p
## (Intercept)   5.2007      0.0559  92.98 < 2e-16
## Log(scale)   -1.7879      0.2442  -7.32 2.5e-13
##
## Scale= 0.167
```

```
##  
## Weibull distribution  
## Loglik(model)= -48.8   Loglik(intercept only)= -48.8  
## Number of Newton-Raphson Iterations: 6  
## n= 10  
  
## Observed value is 0.3486495  
  
## p-value = 0.3173105  
  
## Observed value is 33.67751  
  
## p-value = 6.5e-09
```