Dylan Wong Xuan Kai **H00280633**

F79SU Survival Models Assignment 1 The Kaplan-Meier estimate and the Cox model

Motive of this report is to find the **effectiveness** of **Drug A** and **Drug B** in **prolonging life** of patience with critical illness, especially as to **which drug** has a **better effect** (A or B).

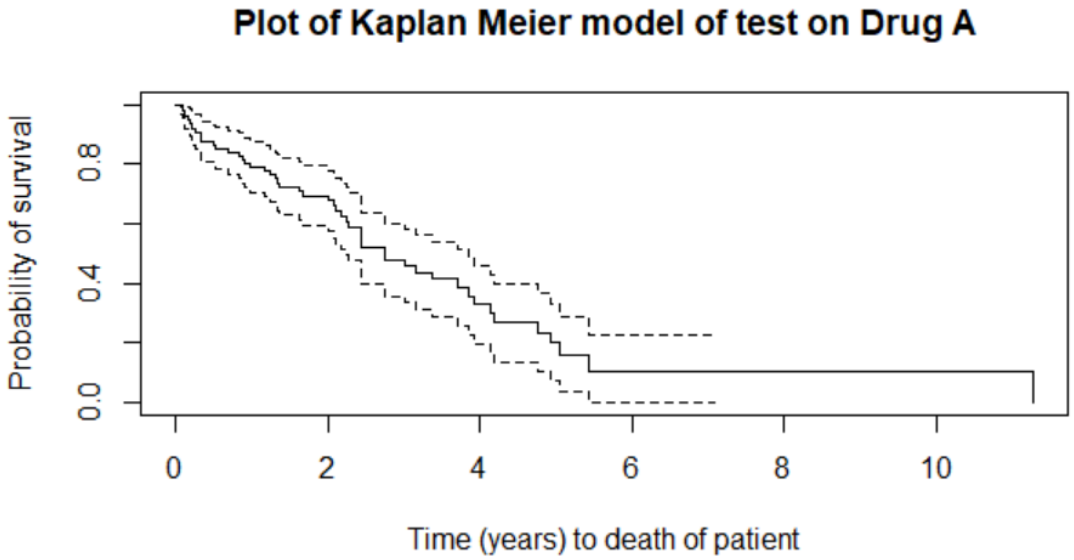
->Below are the **data on survival time** and **observation censor** of **total 200** patience with critical illness, **100** being tested on **Drug A** and the remaining **100** on **Drug B (sorted time)**:

```
> # Data of patience taking Drug A(Group 1) and Drug B(Group 2)
> # (ascending time of survival time)
> # spacing<- extra space added betwwen data from Drug A and B, better display
> spacing<-rep('  ', 50)
> info<-data.frame(TimeA[1:50], CensorA[1:50],  spacing, TimeA[51:100], CensorA[51:100], spacir
nsorB[1:50], spacing, TimeB[51:100], CensorB[51:100])
> colnames(info)<-c('Time A', 'Censor A', '','Time A', 'Censor A', '', 'Time B', 'Censor B', '
 B')
> info
       Time A Censor A       Time A Censor A          Time B Censor B        Time B Censor B
1  0.07240635        1     1.554987        0     0.005086028        1      1.334870        0
2  0.08342388        1     1.564967        0     0.071782765        1      1.335615        1
3  0.11234312        1     1.628609        1     0.095871025        1      1.340119        1
4  0.11612431        0     1.631933        0     0.099541413        1      1.349385        0
5  0.12011516        1     1.681064        1     0.133952988        1      1.428375        0
6  0.16371761        1     1.796158        0     0.136134064        1      1.468703        1
7  0.17478802        0     1.889893        0     0.154483167        1      1.491773        0
8  0.20070034        1     1.899897        0     0.168293644        1      1.498272        0
9  0.20078649        0     1.972549        0     0.226139019        1      1.574506        1
10 0.20631517        1     1.999519        1     0.228717548        0      1.652626        0
11 0.20811855        0     2.074251        1     0.284477680        1      1.688401        0
12 0.21563318        1     2.080623        0     0.285259314        1      1.705896        1
13 0.22498245        0     2.101855        1     0.306839133        0      1.842118        1
14 0.26069079        0     2.168116        1     0.321993122        1      1.861493        0
15 0.26247702        1     2.251990        1     0.348444383        0      1.868172        0
16 0.31477739        0     2.280236        1     0.364303771        1      1.934057        0
17 0.32314490        1     2.407178        0     0.368673102        0      2.020904        1
18 0.32860748        1     2.430806        1     0.372953590        1      2.057039        1
19 0.33271034        1     2.435299        1     0.402040827        0      2.059549        1
20 0.35333079        0     2.441007        1     0.418985265        0      2.127876        0
21 0.38454526        0     2.445578        1     0.440629892        0      2.188078        0

22 0.45342925        0     2.451854        0     0.451009583        0      2.218204        0
23 0.46093332        0     2.539017        0     0.453033052        0      2.219723        0
24 0.49705391        1     2.654979        0     0.499678295        0      2.547020        1
25 0.50298932        0     2.748101        1     0.508831345        1      2.620590        0
26 0.51579809        1     2.758011        1     0.557970719        0      2.628334        1
27 0.53462564        0     2.860839        0     0.568029767        0      2.637690        1
28 0.57832921        0     3.014392        1     0.639291006        1      2.654829        1
29 0.58242260        0     3.155662        1     0.647427394        1      2.884552        1
30 0.70129219        1     3.307138        0     0.651355031        0      2.977014        0
31 0.70172913        0     3.371916        1     0.719825377        0      3.042795        1
32 0.79554272        0     3.500634        0     0.764582506        1      3.080421        1
33 0.81831940        0     3.604220        0     0.798133594        1      3.129787        0
34 0.83285199        1     3.662254        0     0.837246266        0      3.225847        1
35 0.87308211        1     3.681378        0     0.876095579        1      3.242139        1
36 0.89595591        0     3.714990        1     0.914673154        1      3.561781        1
37 0.91840291        1     3.775484        0     0.924460002        1      3.567059        1
38 0.98396634        1     3.850910        1     0.936121206        1      3.875249        0
39 1.18214217        0     3.916233        1     0.964558557        1      4.340806        1
40 1.18324340        1     4.131737        1     0.985014762        0      4.437808        0
41 1.24953095        0     4.194458        1     1.003823339        1      4.513830        1
42 1.25463020        1     4.381892        0     1.027531668        1      4.968941        1
43 1.31457173        1     4.750416        1     1.183309088        1      5.387658        1
44 1.34252366        1     4.927642        1     1.237164424        1      6.162521        0
45 1.36701222        1     4.953324        0     1.254004150        0      6.363470        0
46 1.37713322        0     5.055069        1     1.255913211        1      6.396934        1
47 1.39099050        0     5.120375        0     1.293944721        1      7.889979        1
48 1.42273550        0     5.425368        1     1.294025170        1      8.601420        0
49 1.46445131        0     7.076652        0     1.302307444        1      8.728188        1
50 1.49791799        0    11.281452        1     1.310323548        1      9.053923        0
```

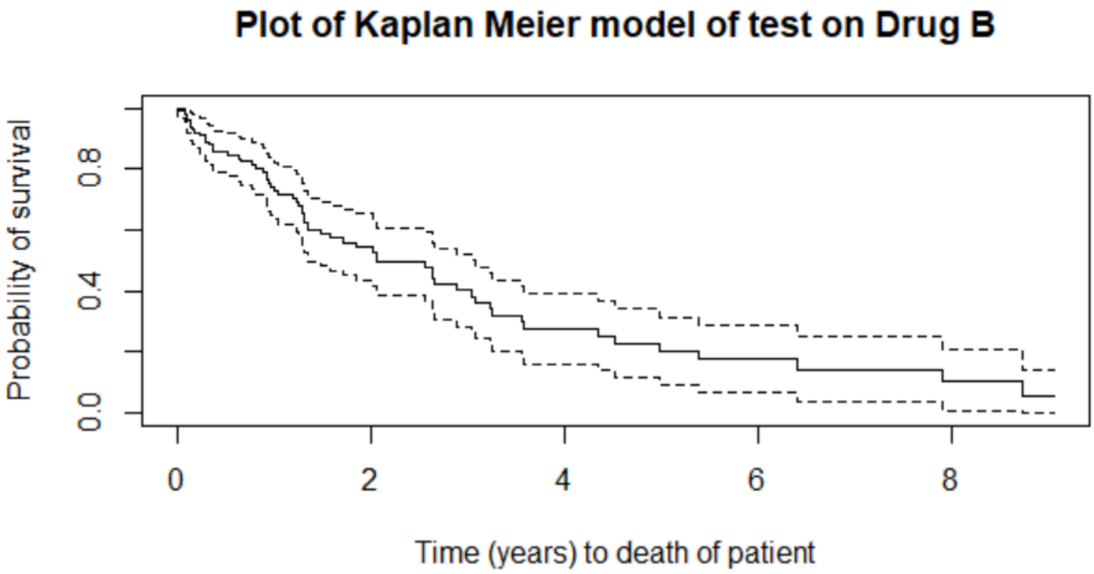
->Below is the Graphical output of **Kaplan Meier model** for **Drug A with plain conf.level**:

```
plot(K_M_model_A, xlab='Time (years) to death of patient', ylab='Probability of survival', main='Plot of Kaplan Meier model of test on Drug A')
```



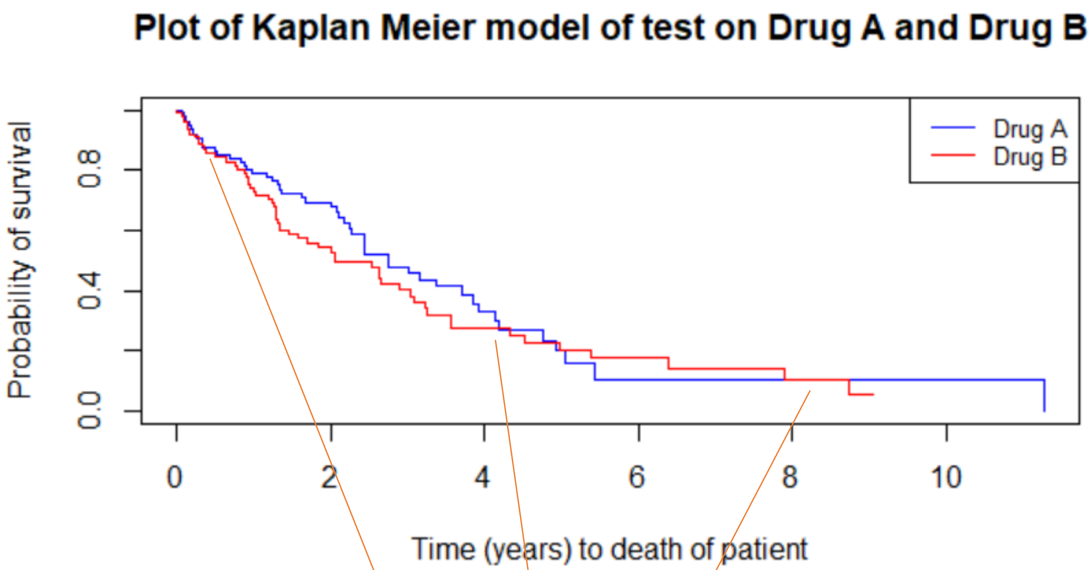
->Below is the Graphical output of **Kaplan Meier model** for **Drug B with plain conf.level**:

```
plot(K_M_model_B, xlab='Time (years) to death of patient', ylab='Probability of survival', main='Plot of Kaplan Meier model of test on Drug B')
```



->Below is the Graphical output of **Kaplan Meier model** for **Drug B**:

```
# Plot both Kaplan Meier model of Drug A and Drug B together
K_M_model_A_unbound<-survfit(Surv(TimeA, CensorA)~1, conf.type='none')
K_M_model_B_unbound<-survfit(Surv(TimeB, CensorB)~1, conf.type='none')
plot(K_M_model_A_unbound, main='Plot of Kaplan Meier model of test on Drug A and Drug B',
     xlab='Time (years) to death of patient', ylab='Probability of survival',
     col='blue')
lines(K_M_model_B_unbound, col='red')
legend('topright', legend=c('Drug A', 'Drug B'), col=c('blue', 'red'), lty=c(1,1), cex=0.8)
```



From Kaplan Meier model **plot** (both Drug A and B) **above**, observation suggests that **probability of survival** with **Drug A** is **higher** in **early stages** as the **survival curve** of **Drug A** is **above Drug B between** around **1 year to 4 years since start period**, then in **later periods both survival curves** start to **converge** as the effectiveness of both drugs start to **wear out**, also **suggesting** that **Drug A** has **a higher tendency** of **prolonging life of patients**. It is **also observed** that an **outlier exists** for test with **Drug A**, with one of the patients surviving up to **11.281452 years**. Perhaps testing **without the last patient survival time (patient that survived the longest)** for both drug test **might yield more accurate** findings **but** that would **also remove the last patient** survival time of test with **Drug B**, thus **not recommended** as it would generally **affect the validity of the test** but still worth noting.

->To test **difference** between **survival curves** of **both drug type** (A and B):

```
> survdiff(Surv(Time, Censor)~Group)
Call:
survdiff(formula = Surv(Time, Censor) ~ Group)

          N Observed Expected (O-E)^2/E (O-E)^2/V
Group=1 100       51     56.2     0.476     0.986
Group=2 100       60     54.8     0.487     0.986

 Chisq= 1  on 1 degrees of freedom, p= 0.3
```

As seen from above test, something worth noting is the assumed **outlier** of test with **Drug A** (11. 281452 years) has a **large influence** on the **expected survival time**, thus **Group 1** (Drug A) having **higher expected survival times** compared to those from **Group 2** (Drug B). In addition to that**,** test on **Drug A** has **more censored** survival times compared to **Drug B**, with **most censored** survival times of test on **Drug A** being **in early periods** (between 0 to 4 years, 47 censored) and **not in later periods** (5 to 11 years, 2 censored), **also** with **last censored** survival time of **Drug A** being **7.076652 years**, having **big time gap between the outlier** (11. 281452 years). But since expected values (survival times) are greatly influenced by outliers, it is best to determine drug life prolong effectiveness with other outlier resistant measures, like median.

Note: Observed=uncensored

->**Kaplan Meier model return data** for test on **Drug A and Drug B**:

```
> K_M_model_A
Call: survfit(formula = Surv(TimeA, CensorA) ~ 1, conf.type = "plain")

      n  events  median 0.95LCL 0.95UCL
 100.00   51.00    2.75    2.25    3.85
>
> K_M_model_B
Call: survfit(formula = Surv(TimeB, CensorB) ~ 1, conf.type = "plain")

      n  events  median 0.95LCL 0.95UCL
 100.00   60.00    2.06    1.34    3.04
```

As seen from above, note that the **confidence level gap** of test with **Drug A** (3.85-2.25=1.60) is only **slightly wider** than with **Drug B** (3.04-1.34=1.56), only **by 0.04**. The **median** is also **higher** for test on **Drug A** (2.75) **compared to** **Drug B** (2.06), remembering that **median** is a **resistant measure of centre** (not affected by outliers). In short, this **suggests** that **despite** having **outliers** and **higher expected survival times**, test on **Drug A** is **more effective** in prolonging life of patients **compared to Drug B** based on median values.

->Probability of patient taking **Drug A survive** for **more than 2.3 years**:

As shown below, estimated probability of surviving above 2.3 years would be **0.591**

```
> # Prob of patient taking Drug A surviving for more thn 2.3 years
> Ans_1=summary(survfit(Surv(TimeA, CensorA)~1), times = 2.3)
> Ans_1
Call: survfit(formula = Surv(TimeA, CensorA) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  2.3     34      32    0.591  0.0584        0.487        0.717
```

->Probability of patient taking **Drug B survive** for **less than/equal to 3.1 years**:

As shown below, estimated probability of surviving above 2.3 years would be

1-Pr(Surviving more than 3.1 years)= 1-0.362= **0.638**

```
> # Prob of patient taking Drug B surviving for less than/equal to 3.1 years
> Ans_2=summary(survfit(Surv(TimeB, CensorB)~1), times =3.1 )
> Ans_2
Call: survfit(formula = Surv(TimeB, CensorB) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  3.1     18      49    0.362  0.0602        0.261        0.501
```

-> **Cox Model** return data for **both drug type** (A and B):

```
> # Cox Model on both drug type (A and B)
> # Group=c(rep(1,100), rep(2,100)), 100 for Drug A(1) then 100 for Drug B(2)
> # Time=Survival time of all patience (all Drug A then B, ascending time)
> # Censor=Observation censor of all patience (all Drug A then B, ascending time)
> # corresponding to patient survival times
> Cox_Model_A<-coxph(Surv(Time, Censor)~Group)
> summary(Cox_Model_A)
Call:
coxph(formula = Surv(Time, Censor) ~ Group)

  n= 200, number of events= 111 

         coef exp(coef) se(coef)     z Pr(>|z|)
Group2 0.1911    1.2106   0.1928 0.991    0.321

       exp(coef) exp(-coef) lower .95 upper .95
Group2     1.211      0.826    0.8297     1.766

Concordance= 0.535  (se = 0.028 )
Likelihood ratio test= 0.99  on 1 df,   p=0.3
Wald test            = 0.98  on 1 df,   p=0.3
Score (logrank) test = 0.99  on 1 df,   p=0.3
```

beta_hat is required to describe a hazard with different characteristic (Drug B) in proportion to baseline hazard (Drug A) which equals to beta_hat multiplicative effect on baseline hazard, with exponential term accounting for the patient differences (formula shown below).

As seen from above return data of Cox Model, maximum partial likelihood estimate (**beta_hat value** from coef) is **0.1911**, meaning that the hazard rate (instantaneous force of mortality, rate of dying) of test with Drug B will be **1.2106** (exponential(Bhat) from exp(coef)) times higher than the hazard rate with Drug A by the below formula:

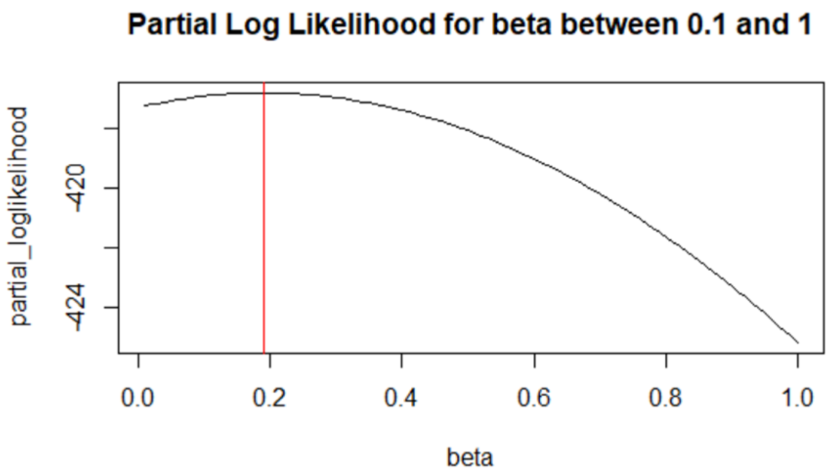
$$\lambda_B(t) = \lambda_A(t)\boldsymbol{e}^{\boldsymbol{beta\_hat}} => \lambda_B(t) = \lambda_A(t)\boldsymbol{e}^{\mathbf{0.1911}}; => \lambda_B(t) = \lambda_A(t) * \mathbf{1.2106}$$

Summary:

Relationship between **hazard of Drug A and Drug B** above **matches conclusions** from previous **plots and summary data**(confidence level, expected lifetime, median) of **Kaplan Meier model suggesting** that **patients** who take **Drug A** will **live** a **longer** life **compared to Drug B**. In order **to ensure** the fact that **beta_hat value** (0.1911) is indeed the beta value that **maximizes** the **maximum partial likelihood estimate** (thus being **covariate** value that best **describes** the **relationship** between **hazard rate** with test on **Drug A and Drug B**), **plots** of **partial log likelihood and score function against beta** has been **displayed** below to **observe validity** of such **assumption**.

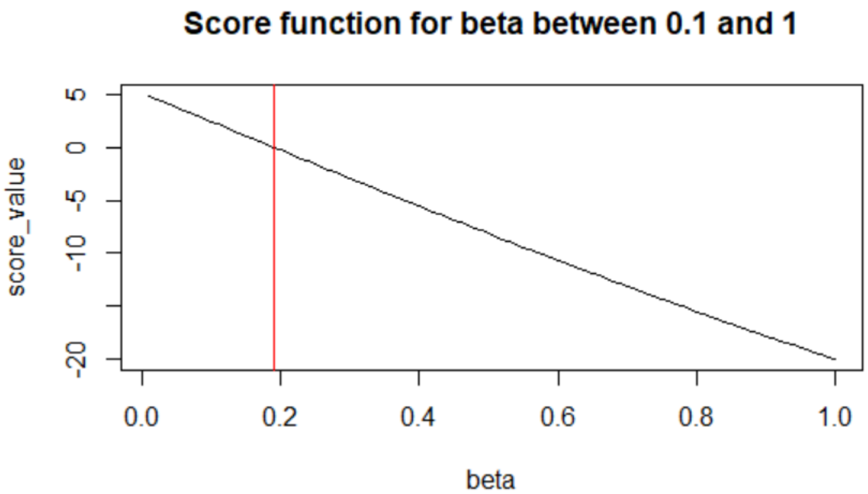
->Plot of partial log likelihood against beta values (using function 'partial_log_lik_plot'):

Red line is to indicate partial log likelihood of beta_hat (0.1911), seemed to give maximum value for partial log likelihood. This can be further confirmed with the next plot with score function against beta values.



->Plot of score function against beta values (using function 'score_func_plot'):

Red line is to indicate score value of beta_hat (0.1911), being approximately 0 from plot. This can be further confirmed with next return data of function 'score_func'



->Check exact score function value with beta_hat value using function 'score_func'

score function value of beta_hat (0.1911) is approximately 0, thus confirming that beta_hat maximizes partial likelihood function, also being covariate value that best describes the relationship between hazards for Drug A and B.

```
> score_func(info_df, beta_hat)
       Group2
-9.880985e-15
```

Conclusion:

It is clear that **evidence**(by summary data of Kaplan Meier model and Cox Regression model, hazard function relationship equation, then having check validity of beta_hat value with plots of partial log likelihood and score function against beta values) **suggests** that taking **Drug A** would be **best** in **prolonging life** of **patients** with the critical illness.

Note: All explanations on R code workings are written in the R code comments