Survival Analysis HW 3

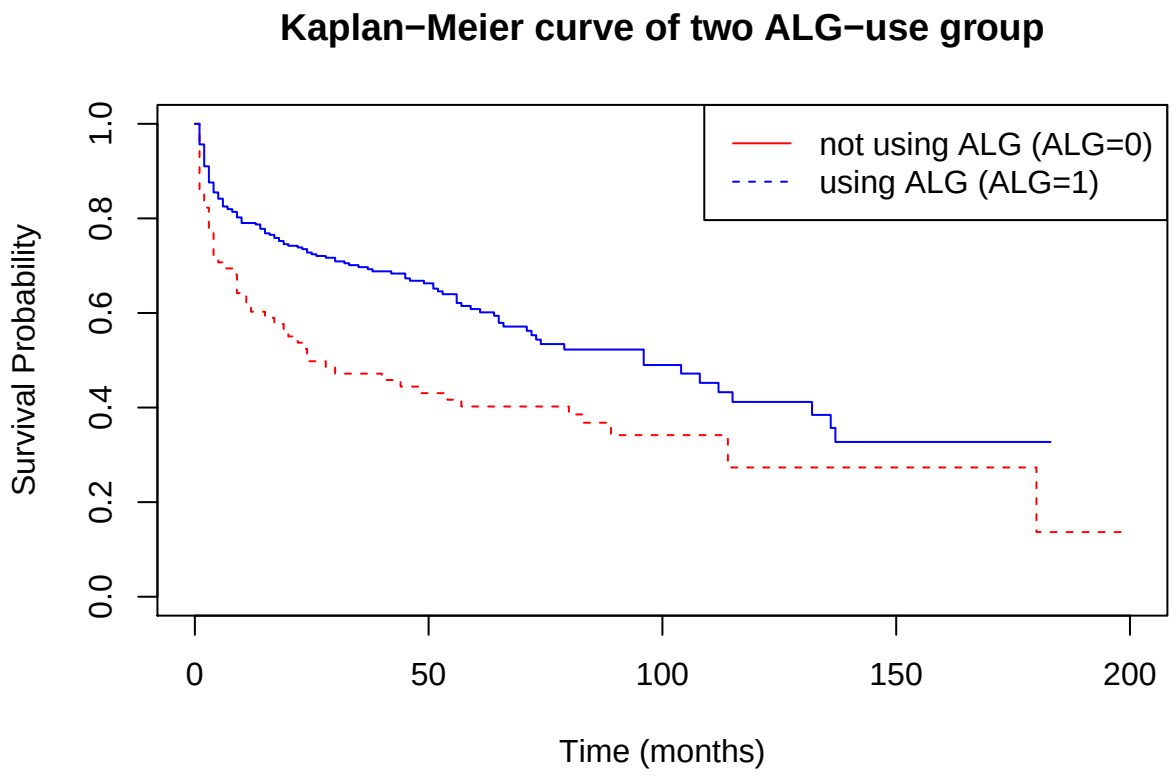
Joanna Chen

2/14/2020

Question 2

The data for this question come from 469 patients with kidney transplants (HW03_Kidney.xlsx). The primary endpoint is graft survival (FAIL=1 if failure occurred, =0 if censored) and time to graft failure was recorded in months (MONTH). #### a. Plot Kaplan-Meier survival curves, interpreting each plot in the context of the current study, and use log-rank tests to perform these comparisons: #### (1) Those using ALG, an immune suppression drug (ALG=1) vs. those not using ALG (ALG=0)

```
HW03_Kidney$ALG = factor(HW03_Kidney$ALG)
fit <- survfit(Surv(MONTH,FAIL)~ALG, data = HW03_Kidney,conf.type = "none")
plot(fit, xlab="Time (months)", ylab="Survival Probability",
   conf.int=FALSE, col=c("red", "blue"), lty=c(2,1))
legend("topright", c("not using ALG (ALG=0)", "using ALG (ALG=1) "), col=c("red", "blue"), lty=c(1,2))
title("Kaplan-Meier curve of two ALG-use group")
```



In the graph, we can see that the ALG using group has better survival experience comparing with not using ALG group becuase the blue curve is always above the red one.

```
survdiff(Surv(MONTH, FAIL) ~ ALG, data=HW03_Kidney) #log-rank test
```

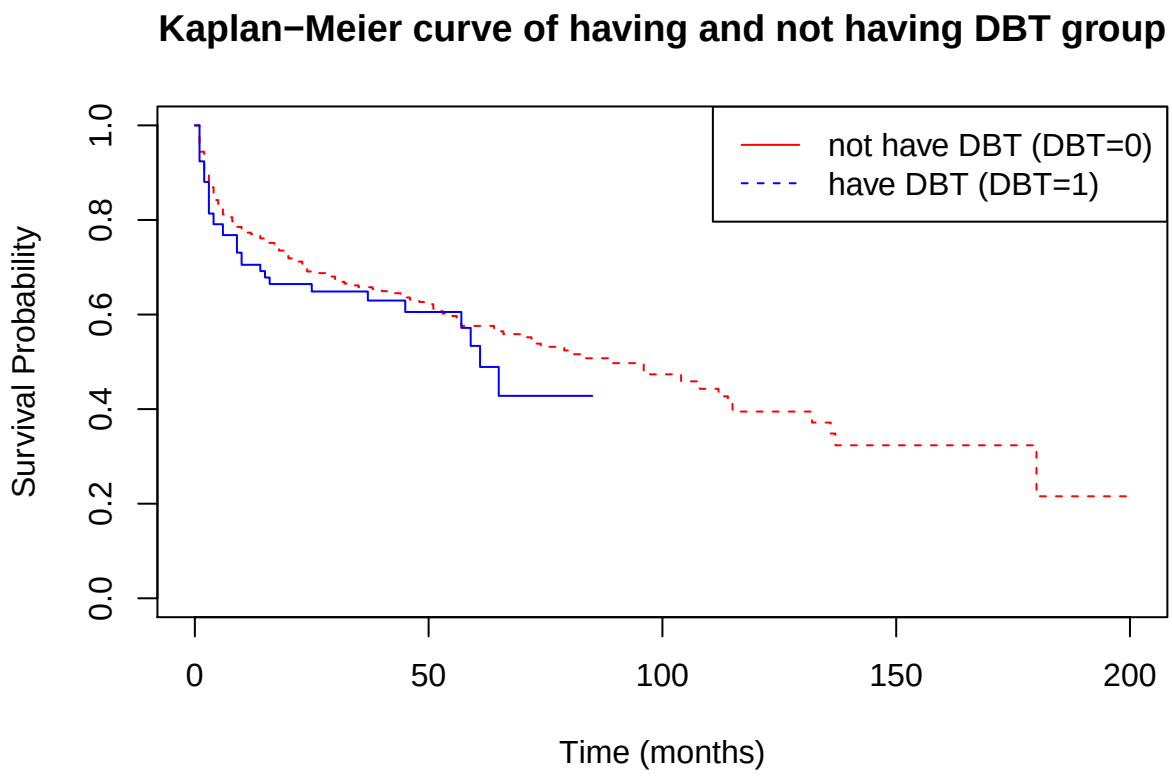
```
## Call:
## survdiff(formula = Surv(MONTH, FAIL) ~ ALG, data = HW03_Kidney)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## ALG=0  79       51     33.9      8.60      10.9
## ALG=1 390      141    158.1      1.85      10.9
##
##  Chisq= 10.9  on 1 degrees of freedom, p= 0.001
```

- Hypothesis: $H_0 : S_1(t) = S_2(t)$ for all $t \leq 200$ vs. $H_1 : S_1(t) \neq S_2(t)$ for some $t \leq 200$.
- Significant level: two-sided $\alpha = 0.05$
- Test statistic : $X^2_{\text{logrank}} = 10.9$
- Decision rule: At $\alpha = 0.05$, reject H_0 if $X^2_{\text{logrank}} \geq \chi^2_{1,0.95} = 3.84$
- Statistical conclusion about H_0:
  We compared the test statistic to the decision rule that $10.9 > 3.84$, therefore reject H_0 with $p = 0.001 < \alpha = 0.05$.
- The graft survival in two ALG use group are significantly different.

(2) Diabetics (DBT=1) vs. non-diabetics (DBT=0)

```
HW03_Kidney$DBT = factor(HW03_Kidney$DBT)
fit <- survfit(Surv(MONTH,FAIL)~DBT, data = HW03_Kidney,conf.type = "none")
plot(fit, xlab="Time (months)", ylab="Survival Probability",
   conf.int=FALSE, col=c("red", "blue"), lty=c(2,1))
legend("topright", c("not have DBT (DBT=0)", "have DBT (DBT=1) "), col=c("red", "blue"), lty=c(1,2))
title("Kaplan-Meier curve of having and not having DBT group")
```



In the graph, we can see that the group not having Diabetics has better survival experience comparing with the gorup having Diabetics for most of the time. But they still have some overlap around month 50.

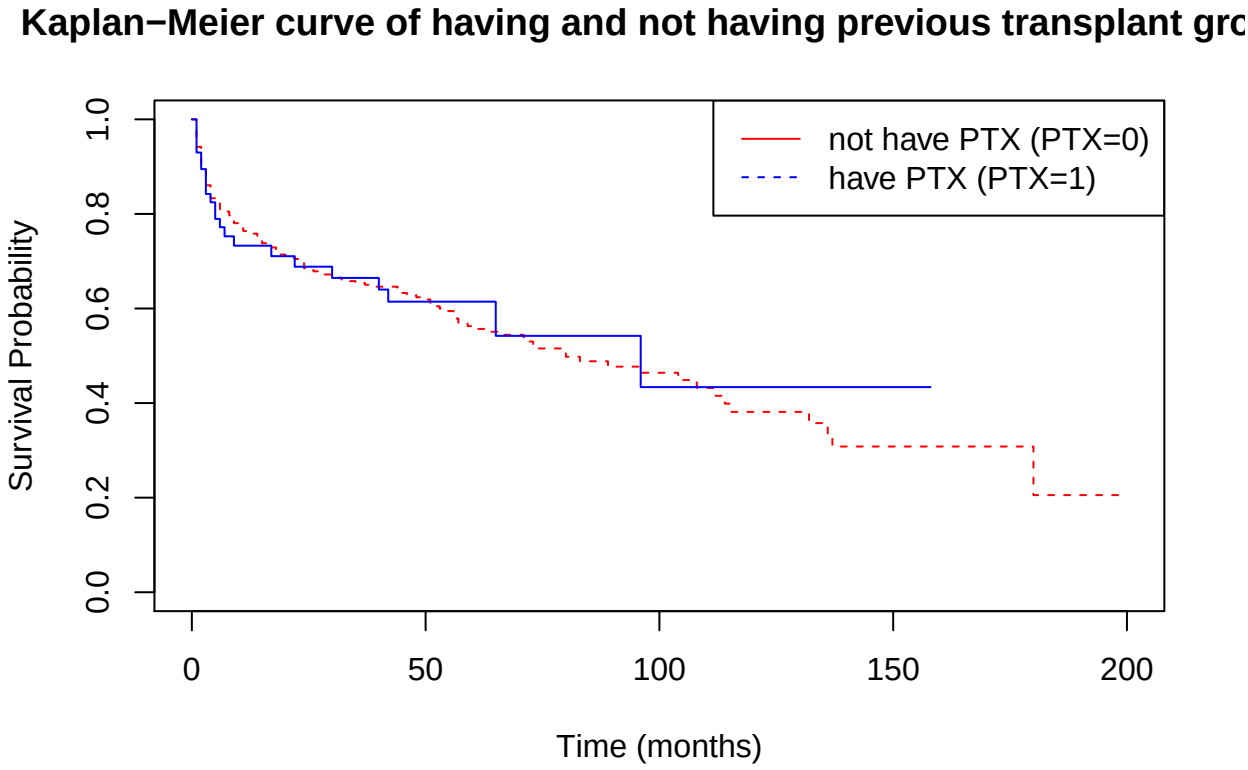
```r
survdiff(Surv(MONTH, FAIL) ~ DBT, data=HW03_Kidney) #log-rank test
```

```
## Call:
## survdiff(formula = Surv(MONTH, FAIL) ~ DBT, data = HW03_Kidney)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## DBT=0 377      156    160.9     0.147     0.952
## DBT=1  92       36     31.1     0.761     0.952
##
##  Chisq= 1  on 1 degrees of freedom, p= 0.3
```

- Hypothesis: $H_0 : S_1(t) = S_2(t)$ for all $t \leq 200$ vs. $H_1 : S_1(t) \neq S_2(t)$ for some $t \leq 200$.
- Significant level: two-sided $\alpha = 0.05$
- Test statistic : $X^2_{\text{logrank}} = 1$
- Decision rule: At $\alpha = 0.05$, reject H_0 if $X^2_{\text{logrank}} \geq \chi^2_{1,0.95} = 3.84$
- Statistical conclusion about H_0:
  We compared the test statistic to the decision rule that $1 < 3.84$, therefore fail to reject H_0 with $p = 0.3 > \alpha = 0.05$.
- The graft survival in two Diabetics group (with and without) are not significantly different.

(3) Previous transplants (PTX=1) vs. no previous transplants (PTX=0)

```r
HW03_Kidney$PTX = factor(HW03_Kidney$PTX)
fit <- survfit(Surv(MONTH,FAIL)~PTX, data = HW03_Kidney,conf.type = "none")
plot(fit, xlab="Time (months)", ylab="Survival Probability",
   conf.int=FALSE, col=c("red", "blue"), lty=c(2,1))
legend("topright", c("not have PTX (PTX=0)", "have PTX (PTX=1) "), col=c("red", "blue"), lty=c(1,2))
title("Kaplan-Meier curve of having and not having previous transplant group")
```



Interpret: The two curves overlap a lot which shows that two curves almost has no too much difference.

```
survdiff(Surv(MONTH, FAIL) ~ PTX, data=HW03_Kidney) #log-rank test
```

```
## Call:
## survdiff(formula = Surv(MONTH, FAIL) ~ PTX, data = HW03_Kidney)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## PTX=0 412      169    168.5   0.00135    0.0113
## PTX=1  57       23     23.5   0.00969    0.0113
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

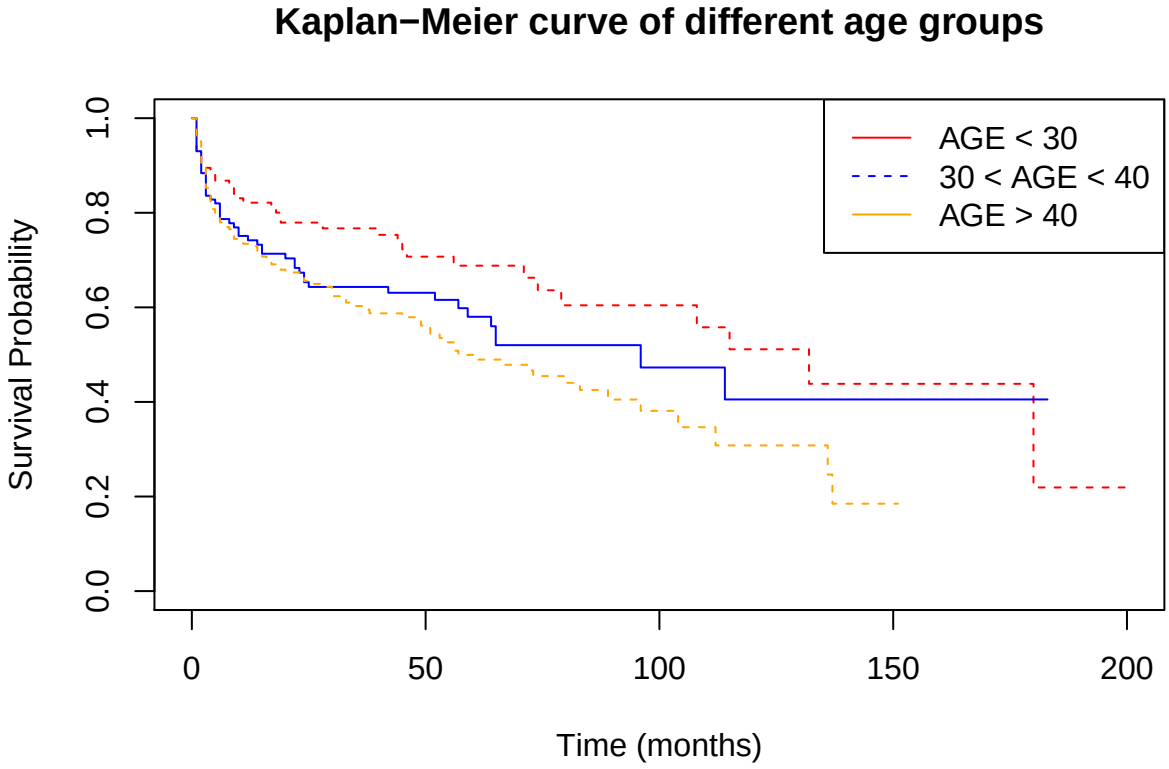
- Hypothesis: $H_0 : S_1(t) = S_2(t)$ for all $t \leq 200$ vs. $H_1 : S_1(t) \neq S_2(t)$ for some $t \leq 200$.
- Significant level: two-sided $\alpha = 0.05$
- Test statistic : $X^2_{\text{logrank}} = 0$
- Decision rule: At $\alpha = 0.05$, reject H_0 if $X^2_{\text{logrank}} \geq \chi^2_{1,0.95} = 3.84$
- Statistical conclusion about H_0:
  We compared the test statistic to the decision rule that $0 < 3.84$, therefore fail to reject H_0 with $p = 0.9 > \alpha = 0.05$.
- The graft survival in previous-transplants and no-previous-transplants groups are not significantly different.

(4) Divide the data set into three groups based on age at transplant (AGE):

AGE < 30, 30 <= AGE < 40, and AGE >= 40 and compare the three groups. If you observe a significant difference, identify which pairs of age categories are significantly different using a Bonferroni adjustment to the log-rank test.

```
# categorize age
setDT(HW03_Kidney)
HW03_Kidney[,agecat:=ifelse(AGE<30,1,ifelse(AGE>=30 & AGE <40, 2, ifelse(AGE >=40, 3, NA)))]
```

```
HW03_Kidney$agecat = factor(HW03_Kidney$agecat)
fit <- survfit(Surv(MONTH,FAIL)~agecat, data = HW03_Kidney,conf.type = "none")
plot(fit, xlab="Time (months)", ylab="Survival Probability",
   conf.int=FALSE, col=c("red", "blue","orange"), lty=c(2,1))
legend("topright", c("AGE < 30", "30 < AGE < 40","AGE > 40"), col=c("red", "blue","orange"), lty=c(1,2))
title("Kaplan-Meier curve of different age groups")
```



It looks like the younger age group has better survival experience than the older. Blue curve and organge curve have some overlap from 0 to ~30 months.

```
#log-rank test
survdiff(Surv(MONTH, FAIL) ~ agecat, data=HW03_Kidney)
```

```
## Call:
## survdiff(formula = Surv(MONTH, FAIL) ~ agecat, data = HW03_Kidney)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## agecat=1 115       37     52.9    4.7554    6.7846
## agecat=2 129       51     52.1    0.0212    0.0298
## agecat=3 225      104     87.1    3.2809    6.2089
##
##  Chisq= 8.4  on 2 degrees of freedom, p= 0.02
```

- Hypothesis: $H_0 : S_1(t) = S_2(t) = S_3(t)$ for all $t \leq 200$ vs. $H_1 : S_p(t) \neq S_q(t)$ for some pair p, q.

- Significant level: two-sided $\alpha = 0.05$
- Test statistic : $X^2_{\text{logrank}} = 8.4$
- Decision rule: At $\alpha = 0.05$, reject H_0 if $X^2_{\text{logrank}} \geq \chi^2_{2,0.95} = 5.99$
- Statistical conclusion about H_0:
  We compared the test statistic to the decision rule that $8.4 > 3.84$, therefore reject H_0 with $p = 0.02 < \alpha = 0.05$.
- The survival curves are not equal across all three groups.

We do observe a significant difference. Let us identify which pairs of age categories are significantly different using a Bonferroni adjustment to the log-rank test.

```
lrbon <- function(formula, data){
  logrank = survdiff(formula, data=data)
  ngroups = length(logrank$n)
  rawpvals = matrix(NA, ngroups, ngroups)
  adjpvals = matrix(NA, ngroups, ngroups)
  rownames(rawpvals) = dimnames(logrank$n)$groups
  colnames(rawpvals) = dimnames(logrank$n)$groups
  O = logrank$obs
  E = logrank$exp
  U = matrix((O-E), nrow=length(O))
  V = logrank$var
  for (i in 1:ngroups) {
    for (j in (1:ngroups)[-i]) {
      X2ij = (U[i] - U[j])^2/(V[i,i]+V[j,j]-2*V[i,j])
      rawpvals[i,j] <- 1-pchisq(X2ij,1)
      adjpvals[i,j] <- min(1,rawpvals[i,j]*ngroups*(ngroups-1)/2)
    }
  }
list("Raw p-values", rawpvals, "Bonferroni-adjusted p-values", adjpvals)
}
lrbon(formula=Surv(MONTH, FAIL) ~ agecat, data=HW03_Kidney)
```

```
## [[1]]
## [1] "Raw p-values"
##
## [[2]]
##             agecat=1  agecat=2    agecat=3
## agecat=1          NA 0.1430610 0.003936053
## agecat=2 0.143061025        NA 0.114075369
## agecat=3 0.003936053 0.1140754          NA
##
## [[3]]
## [1] "Bonferroni-adjusted p-values"
##
## [[4]]
##            [,1]      [,2]       [,3]
## [1,]         NA 0.4291831 0.01180816
## [2,] 0.42918307        NA 0.34222611
## [3,] 0.01180816 0.3422261         NA
```

The hypothesis and significance level are still the same as above.

- Log-rank test p = 0.02
- group number P = 3 with c = P(P − 1)/2 = 3 possible pairwise comparisons
- $\alpha^* = 0.05/c = 0.05/3 = 0.0167$ (On lecture 3 slides page 61, we can see that we use raw p value compare with α^*)
- There is a significant difference between age group 1 and age group 3, that is AGE < 30 and AGE >= 40 ($p = 0.004 < \alpha^* = 0.0167$).
- We cannot conclude there is a significant difference between age group 1 and 2 (AGE < 30 and 30 <= AGE < 40), and age group 2 and 3 (30 <= AGE < 40, and AGE >= 40) because their raw p-value is greater than α^*.

b. Perform a test for trend over the age categories that you created in part (a) using weights (1,2,3) for age groups < 30, 30 <= AGE < 40, and >= 40, respectively. We hypothesize that hazard of graft failure increases with age. How do your results from this question and part (a) compare?

```
lrtrend <- function(formula, weights, data){
  logrank = survdiff(formula, data=data)
  df = length(logrank$n) - 1
  O = logrank$obs
  E = logrank$exp
  U = matrix((O-E), nrow=length(O))
  V = logrank$var
  w = weights
  z = matrix(w, nrow=length(w))
  Xtrend = (t(z)%*%U)/sqrt(t(z)%*% V %*% z)

  cat("\nLog-rank Test: Chi^2(", df, " df) = ", logrank$chisq, ", p = ", 1-pchisq(logrank$chisq, df), s
  cat("\n Test Trend: Xtrend ~ N(0,1) = ", Xtrend, ", 2-sided p = ", 2*(1-pnorm(abs(Xtrend))),
      ", lower p = ", pnorm(Xtrend),
      ", upper p = ", 1-pnorm(Xtrend), sep="" )
}
lrtrend(formula=Surv(MONTH, FAIL) ~ agecat, weights=c(1, 2, 3), data=HW03_Kidney)
```

```
##
## Log-rank Test: Chi^2(2 df) = 8.354461, p = 0.01534094
##  Test Trend: Xtrend ~ N(0,1) = 2.883242, 2-sided p = 0.003936053, lower p = 0.998032, upper p = 0.00
```

- Since we hypothesize that hazard of graft failure increases with age, we are doing an upper-tail test - Hypothesis: $H_0 : S_1(t) = S_2(t) = S_3(t)$ for $t \leq 200$ vs. $H_1 : S_1(t) > S_2(t) > S_3(t)$ for $t \leq 200$
- Significant level: two-sided $\alpha = 0.05$
- Test statistic: $X_{trend} = 2.883242$
- At $\alpha = 0.05$, reject H0 in favor of 1-sided upper-tailed H1 if $z > z_{1.05} = z_{.95} = 1.645$
- Since $X_{trend} = 2.883242 > 1.645$, reject H0 with p = 0.001968026
- There is evidence to suggest age is positively associated with hazard of graft failure. The results extend the conclusion of part (a) that the survival curves are not equal across all three

groups. It gives us a trend of survival in survival functions,

c. Perform a stratified log-rank test for the effect of ALG controlling for diabetes.

```
# stratified log-rank test
survdiff(Surv(MONTH, FAIL) ~ ALG + strata(DBT), data = HW03_Kidney)
```

```
## Call:
## survdiff(formula = Surv(MONTH, FAIL) ~ ALG + strata(DBT), data = HW03_Kidney)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## ALG=0  79       51     33.3      9.37        12
## ALG=1 390      141    158.7      1.97        12
##
##  Chisq= 12  on 1 degrees of freedom, p= 5e-04
```

- Hypothesis: $H_0 : S_{1s}(t) = S_{2s}(t)$ for s = 1,2 $t \leq 200$ vs. $H_1 : H_0$ is not true
- Significant level: two-sided $\alpha = 0.05$
- Test statistic : $X_s^2 = 12$
- Decision rule: At $\alpha = 0.05$, reject H_0 if $X_s^2 \geq \chi^2_{1,0.95} = 3.84$
- Statistical conclusion about H_0: We compared the test statistic to the decision rule that $12 > 3.84$, therefore reject H_0 with $p = 5e - 04 = 5 * 10^{-4} < \alpha = 0.05$.
- Controlling for Diabetes, the graft survival in two ALG-use group are significantly different.

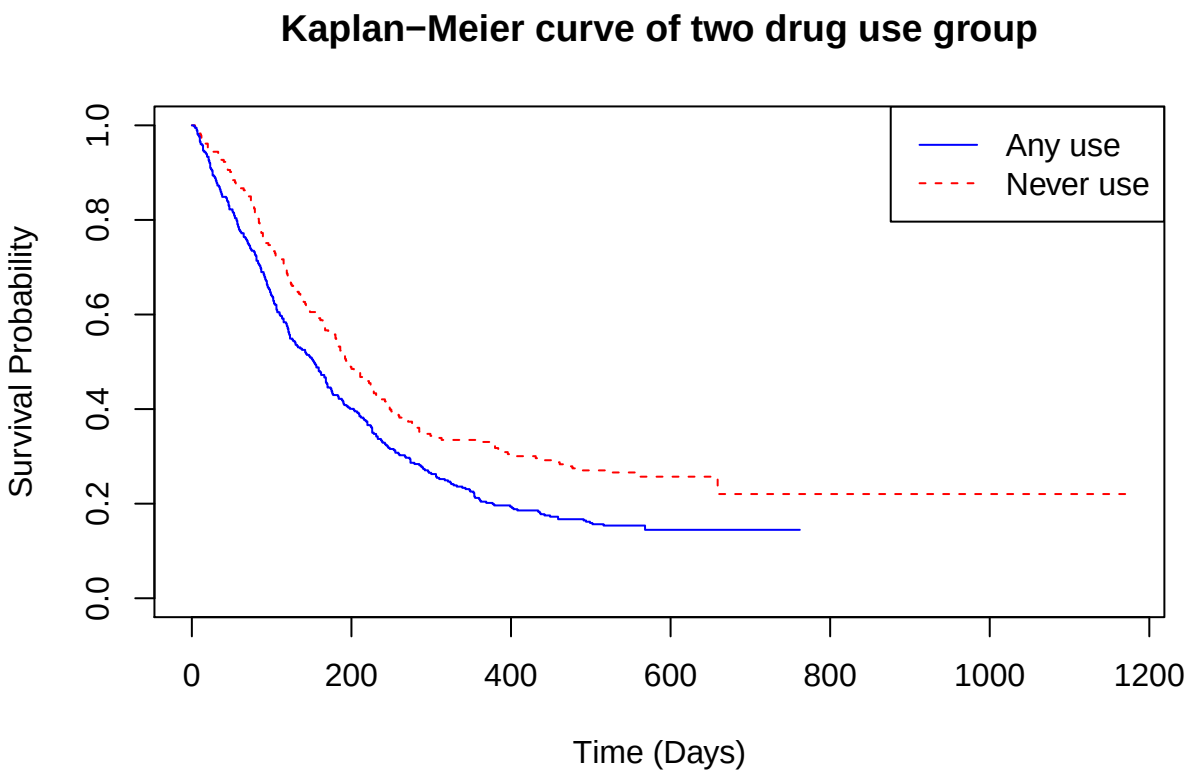
Question 3

Data for this question come from the University of Massachusetts AIDS Research UNIT (UMARU) IMPACT Study. This was a 5-year (1989-1994) collaborative research project comprised of two concurrent randomized trials of residential treatment for drug abuse. The purpose of the study was to compare treatment programs of different planned durations designed to reduce drug abuse and to prevent high-risk HIV behavior. The study sought to determine whether alternative residential treatment approaches are variable in effectiveness and whether efficacy depends on planned program duration. The primary outcome measure was time from randomization to treatment to self-reported return to drug use. The data are provided in an Excel file (HW03_Drug.xlsx) and the data key is provided at the end of this assignment. The primary risk factor for this question is a dichotomous variable (drug) that will require you to recode the history of IV drug use variable (ivhx) into none (never) versus any use.

```
HW03_Drug[,ivhx_new := ifelse(ivhx == 1, 0,1)] # ivhx_new = 0: Never use; ivhx_new = 1: Any use.
```

a. Plot Kaplan-Meier survival curves by your newly-created IV drug use variable. Interpret the plot in the context of the current study.

```
HW03_Drug$ivhx_new = factor(HW03_Drug$ivhx_new)
fit_km <- survfit(Surv(time,censor)~ivhx_new, data = HW03_Drug,conf.type = "none")
plot(fit_km, xlab="Time (Days)", ylab="Survival Probability",
   conf.int=FALSE, col=c("red", "blue"), lty=c(2,1))
legend("topright", c("Any use", "Never use"), col=c("blue", "red"), lty=c(1,2))
title("Kaplan-Meier curve of two drug use group")
```



In the graph, we can see that any-drug-use group is quicker to return to drug use than the without-any-drug-use group. #### b. Perform a hypothesis test (your choice) to determine if the overall survival experience is significantly different in the two IV drug use groups. The survival curves have shown the harzards are close to proportional. Therefore, I would like to perform a log-rank test with significance level 0.05.
- Hypothesis: $H_0 : S_1(t) = S_2(t)$ for all $t \leq 1172$ vs. $H_1 : S_1(t) \neq S_2(t)$ for some $t \leq 1172$.
- Significant level: two-sided $\alpha = 0.05$
- Test statistic : $X^2_{\text{logrank}} = 12$
- Decision rule: At $\alpha = 0.05$, reject H_0 if $X^2_{\text{logrank}} \geq \chi^2_{1,0.95} = 3.84$
- Statistical conclusion about H_0: We compared the test statistic to the decision rule that $12 > 3.84$, therefore reject H_0 with $p = 5e - 04 < \alpha = 0.05$.
- The overall survival experience in two drug-use groups are significantly different.

```
# log-rank test
survdiff(Surv(time, censor) ~ ivhx_new, data=HW03_Drug)
```

```
## Call:
## survdiff(formula = Surv(time, censor) ~ ivhx_new, data = HW03_Drug)
##
## n=610, 18 observations deleted due to missingness.
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## ivhx_new=0 233      173      211      6.82        12
## ivhx_new=1 377      320      282      5.10        12
##
##  Chisq= 12  on 1 degrees of freedom, p= 5e-04
```

c.

A potential confounder is depression severity of the subject. Create a categorical depression variable based on the subject's *Beck Depression Score* at admission. Note that higher scores indicate more severe depressive

symptoms. Use the cut-off score guidelines from the original Beck Depression Inventory (below) to create a Beck category variable (beckcat) for each patient: For a given subject, if their Beck Depression Inventory is missing, then their value of beckcat should also be missing. Perform a stratified log-rank test for the effect of drug controlling for depressive symptoms (*beckcat*). What is one limitation of using stratification to control for depressive symptoms?

```
HW03_Drug[,beckcat:=ifelse(beck>=0 & beck<=9,1,ifelse(beck>=10 & beck<=18,2,ifelse(beck>=19 & beck<=29,3
HW03_Drug$beckcat = factor(HW03_Drug$beckcat)
# stratified log-rank test
survdiff(Surv(time, censor) ~ ivhx_new + strata(beckcat), data = HW03_Drug)
```

```
## Call:
## survdiff(formula = Surv(time, censor) ~ ivhx_new + strata(beckcat),
##     data = HW03_Drug)
##
## n=595, 33 observations deleted due to missingness.
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## ivhx_new=0 228      168      202      5.65      10.1
## ivhx_new=1 367      310      276      4.13      10.1
##
##  Chisq= 10.1  on 1 degrees of freedom, p= 0.002
```

- Hypothesis: $H_0 : S_{1s}(t) = S_{2s}(t)$ for s = 1,2,3,4 $t \leq 1172$ vs. $H_1 : H0$ is not true
- Significant level: two-sided $\alpha = 0.05$
- Test statistic : $X_s^2 = 10.1$
- Decision rule: At $\alpha = 0.05$, reject H_0 if $X_s^2 \geq \chi^2_{1,0.95} = 3.84$
- Statistical conclusion about H_0:We compared the test statistic to the decision rule that $10.1 > 3.84$, therefore reject H_0 with $p = 0.002 < \alpha = 0.05$.
- Controlling for depression severity of the subject, the return to drug use survival probabilty in two drug-use groups are significantly different.

Limitations of using stratification to control for depressive symptoms:
The beck depression score was collected at admission. Using stratification, it is assumed to be the depression symptom of the subjects for a duration of time. Errors can be induced in outcome when a subject shows a different mood or symptom. There is a need to use a better approach to capture daily symptoms of depression for each participant. Moreover, although beck depression score seems a validated scale to screen for symptoms of depression, cannot firmly diagnose depression. Therefore, participants with high beck depression scores does not necessarily have depression and vice versa. [1]

Reference: [1] Sarda A, Munuswamy S, Sarda S, Subramanian V. Using Passive Smartphone Sensing for Improved Risk Stratification of Patients With Depression and Diabetes: Cross-Sectional Observational Study, JMIR Mhealth Uhealth 2019; 7(1):e11041. DOI: 10.2196/11041