ST525 HW 5

Nora Quick

Question 1

Part (a)

```
#15
t15 <- matrix(c(0,1,5,4), ncol = 2)

rownames(t15) <- c('Group A','Group B')
colnames(t15) <- c('d','a')

t15 <- as.table(t15)
t15
```

```
##         d a
## Group A 0 5
## Group B 1 4
```

```
#18
t18 <- matrix(c(0,1,4,3), ncol = 2)

rownames(t18) <- c('Group A','Group B')
colnames(t18) <- c('d','a')

t18 <- as.table(t18)
t18
```

```
##         d a
## Group A 0 4
## Group B 1 3
```

```
#19
t19 <- matrix(c(0,2,3,1), ncol = 2)

rownames(t19) <- c('Group A','Group B')
colnames(t19) <- c('d','a')

t19 <- as.table(t19)
t19
```

```
##         d a
## Group A 0 3
## Group B 2 1
```

```r
#20
t20 <- matrix(c(0,1,3,0), ncol = 2)

rownames(t20) <- c('Group A','Group B')
colnames(t20) <- c('d','a')

t20 <- as.table(t20)
t20
```

```
##         d a
## Group A 0 3
## Group B 1 0
```

```r
#23
t23 <- matrix(c(1,0,1,0), ncol = 2)

rownames(t23) <- c('Group A','Group B')
colnames(t23) <- c('d','a')

t23 <- as.table(t23)
t23
```

```
##         d a
## Group A 1 1
## Group B 0 0
```

Part (b)

```r
dA <- c(0,0,0,0,1)
eA <- c(1/2, 1/2, 1/2, 3/4, 1/2)
vA <- c(1/4, 1/4, 1/4, 3/8, 1/4) #??? wasn't explained in ANY material for the class???
dA
```

```
## [1] 0 0 0 0 1
```

```r
eA
```

```
## [1] 0.50 0.50 0.50 0.75 0.50
```

```r
vA
```

```
## [1] 0.250 0.250 0.250 0.375 0.250
```

```r
dA <- sum(dA)
eA <- sum(eA)
vA <- sum(vA)
dA
```

```
## [1] 1
```

```
eA
```

```
## [1] 2.75
```

```
vA
```

```
## [1] 1.375
```

part (c)

```
logA <- ((1 - 2.75)^2) / 1.375
logA
```

```
## [1] 2.227273
```

part (d)

```
p_val <- 1 - pchisq(2.227273, df=1)
p_val
```

```
## [1] 0.135593
```

Based on the resulting p-value I would conclude to fail to reject the null hypothesis that the two groups hit remission at the same time/rate.

Question 2

Part (a)

```
#10
t10 <- matrix(c(0,1,7,5), ncol = 2)

rownames(t10) <- c('Group A','Group B')
colnames(t10) <- c('d','a')

t10 <- as.table(t10)
t10
```

```
##         d a
## Group A 0 7
## Group B 1 5
```

```
#14
t14 <- matrix(c(0,1,7,4), ncol = 2)

rownames(t14) <- c('Group A','Group B')
colnames(t14) <- c('d','a')

t14 <- as.table(t14)
t14
```

```
##         d a
## Group A 0 7
## Group B 1 4
```

```
#15
t15 <- matrix(c(0,1,7,3), ncol = 2)

rownames(t15) <- c('Group A','Group B')
colnames(t15) <- c('d','a')

t15 <- as.table(t15)
t15
```

```
##         d a
## Group A 0 7
## Group B 1 3
```

```
#16
t16 <- matrix(c(1,0,6,3), ncol = 2)

rownames(t16) <- c('Group A','Group B')
colnames(t16) <- c('d','a')

t16 <- as.table(t16)
t16
```

```
##         d a
## Group A 1 6
## Group B 0 3
```

```
#18
t18 <- matrix(c(0,1,5,2), ncol = 2)

rownames(t18) <- c('Group A','Group B')
colnames(t18) <- c('d','a')

t18 <- as.table(t18)
t18
```

```
##         d a
## Group A 0 5
## Group B 1 2
```

```
#20
t20 <- matrix(c(1,0,2,2), ncol = 2)

rownames(t20) <- c('Group A','Group B')
colnames(t20) <- c('d','a')

t20 <- as.table(t20)
t20
```

```
##         d a
## Group A 1 2
## Group B 0 2
```

```
#21
t21 <- matrix(c(0,1,2,0), ncol = 2)

rownames(t21) <- c('Group A','Group B')
colnames(t21) <- c('d','a')

t21 <- as.table(t21)
t21
```

```
##         d a
## Group A 0 2
## Group B 1 0
```

```
#28
t28 <- matrix(c(1,0,1,0), ncol = 2)

rownames(t28) <- c('Group A','Group B')
colnames(t28) <- c('d','a')

t28 <- as.table(t28)
t28
```

```
##         d a
## Group A 1 1
## Group B 0 0
```

Part (b)

```
dA <- c(0,0,0,1,0,1,0,1)
eA <- c(7/13, 7/12, 7/11, 6/10, 5/8, 3/5, 1)
vA <- c(7/26, 7/24, 7/22, 6/20, 5/16, 3/10, 0) #???
dA
```

```
## [1] 0 0 0 1 0 1 0 1
```

```
eA
```

```
## [1] 0.5384615 0.5833333 0.6363636 0.6000000 0.6250000 0.6000000 1.0000000
```

```
vA
```

```
## [1] 0.2692308 0.2916667 0.3181818 0.3000000 0.3125000 0.3000000 0.0000000
```

```
dA <- sum(dA)
eA <- sum(eA)
vA <- sum(vA)
dA
```

```
## [1] 3
```

```
eA
```

```
## [1] 4.583159
```

```
vA
```

```
## [1] 1.791579
```

part (c)

```
logA <- ((3 - 4.58)^2) / 1.792
logA
```

```
## [1] 1.39308
```

part (d)

```
p_val <- 1 - pchisq(1.39308, df=1)
p_val
```

```
## [1] 0.2378856
```

Based on the resulting p-value I conclude to fail to reject the null hypothesis that the two drugs have the same results.

Question 3

part (a)

```
tongue <- read.csv('tongue.csv')
head(tongue)
```

```
##   type time delta
## 1    1    1     1
## 2    1    3     1
## 3    1    3     1
## 4    1    4     1
## 5    1   10     1
## 6    1   13     1
```

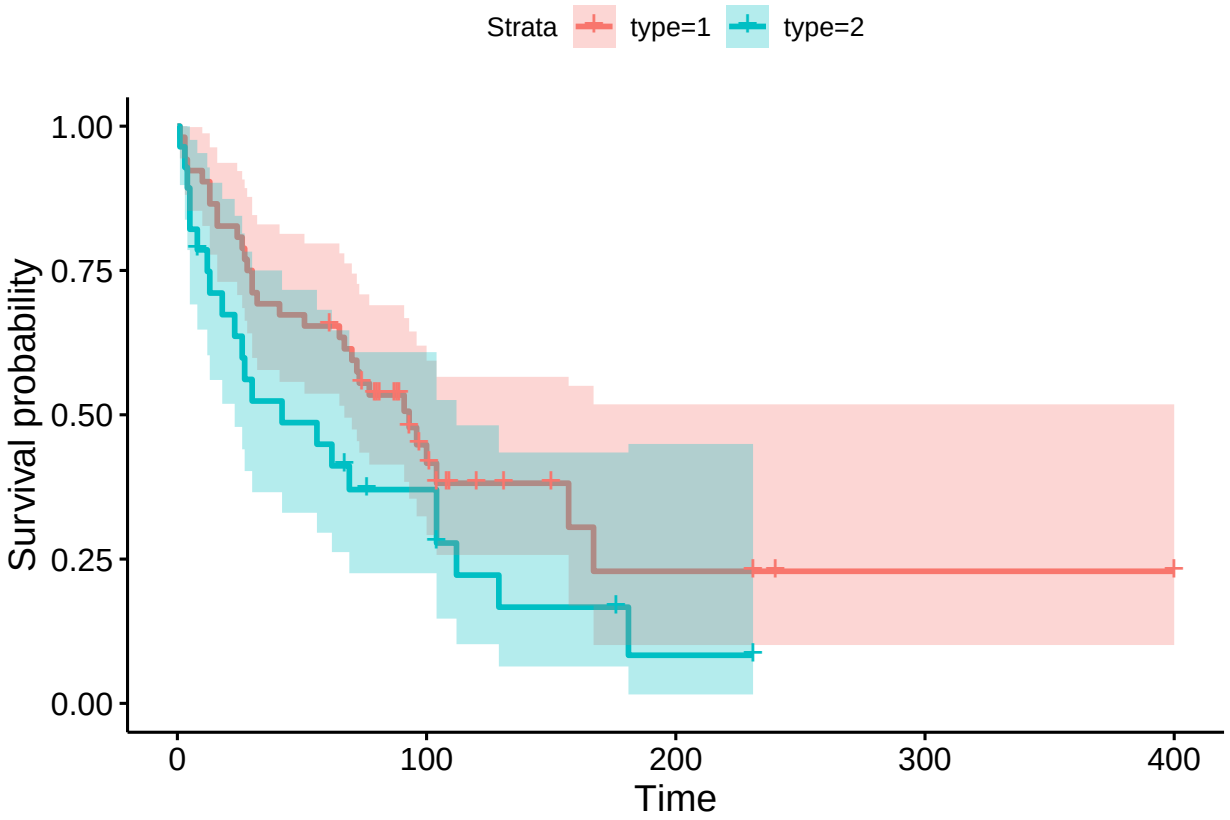
part (b)

```
library(survival)
library(survminer)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: ggpubr
```

```
censored <- Surv(tongue$time, tongue$delta)
fit1 <- survfit(Surv(time, delta) ~ type, data = tongue)
ggsurvplot(fit = fit1, data = tongue, conf.int = TRUE)
```



part (c) The null hypothesis is that there is no difference between the two types of tumor. The alternative hypothesis is that there is a difference btween the two types of tumors.

```
logrank<-survdiff(Surv(time, delta) ~ type, data = tongue)
logrank
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ type, data = tongue)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## type=1 52       31     36.6     0.843      2.79
## type=2 28       22     16.4     1.873      2.79
##
##  Chisq= 2.8  on 1 degrees of freedom, p= 0.09
```

```
peto<-survdiff(Surv(time, delta) ~ type, rho=1, data = tongue)
peto
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ type, data = tongue, rho = 1)
```

```
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## type=1 52     20.2     24.4     0.731       3.3
## type=2 28     15.1     10.9     1.643       3.3
##
##  Chisq= 3.3  on 1 degrees of freedom, p= 0.07
```

part (d) Based on the output from part (c) I would conclude that, yes, it is consistent with the plot given in part (b).

Question 4

part (a)

```
smoke <- read.csv("pharmacoSmoking-old.csv")
head(smoke)
```

```
##    id ttr relapse grp age gender race employment yearsSmoking levelSmoking
## 1  21  41       0   2  36      1    4          1           26             1
## 2 113  14       1   2  41      1    4          2           27             1
## 3  39   5       1   1  25      0    4          2           12             1
## 4  80  16       1   1  54      1    4          1           39             1
## 5  87   0       1   1  45      1    4          2           30             1
## 6  29 157       0   1  43      1    2          1           30             1
##    admitdate      fdate priorAttempts longestNoSmoke
## 1 11/20/2005 12/31/2005             0              0
## 2  6/16/2005  6/30/2005             3             90
## 3   5/9/2005  5/14/2005             3             21
## 4 10/26/2005 11/11/2005             0              0
## 5  9/27/2005  9/27/2005             0              0
## 6   7/6/2005 12/10/2005             2           1825
```

```
smoke$admitdate <- smoke$admitdate %>%
  sapply(function(x) x[1]) %>% as.Date(format = c("%m/%d/%y"))

smoke$fdate <- smoke$fdate %>%
  sapply(function(x) x[1]) %>% as.Date(format = c("%m/%d/%y"))

smoke$time <- difftime(smoke$fdate , smoke$admitdate, units = 'days') %>% as.numeric()

head(smoke)
```

```
##    id ttr relapse grp age gender race employment yearsSmoking levelSmoking
## 1  21  41       0   2  36      1    4          1           26             1
## 2 113  14       1   2  41      1    4          2           27             1
## 3  39   5       1   1  25      0    4          2           12             1
## 4  80  16       1   1  54      1    4          1           39             1
## 5  87   0       1   1  45      1    4          2           30             1
## 6  29 157       0   1  43      1    2          1           30             1
##    admitdate      fdate priorAttempts longestNoSmoke time
## 1 2020-11-20 2020-12-31             0              0   41
```

```
## 2 2020-06-16 2020-06-30             3              90   14
## 3 2020-05-09 2020-05-14             3              21    5
## 4 2020-10-26 2020-11-11             0               0   16
## 5 2020-09-27 2020-09-27             0               0    0
## 6 2020-07-06 2020-12-10             2            1825  157
```

part (b)

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

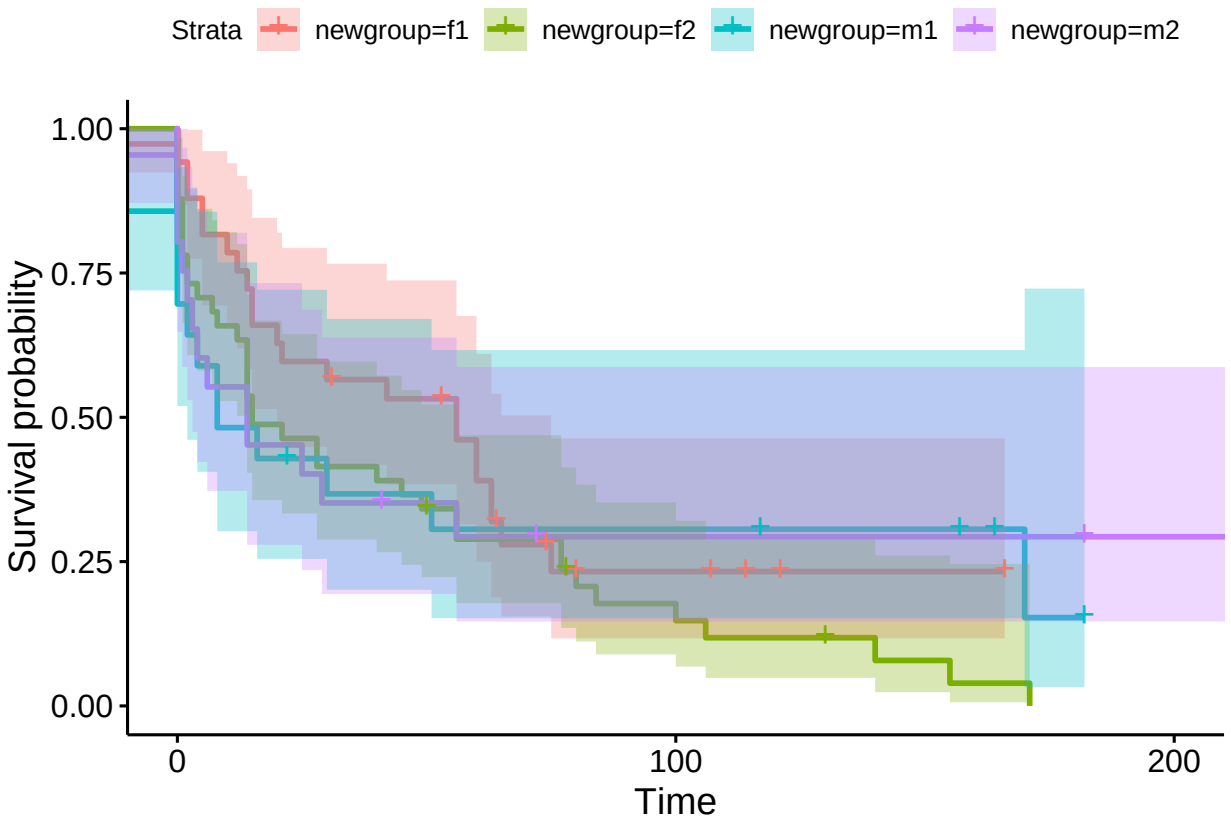
```
smoke$f1 <- (smoke$gender == 0 & smoke$grp == 1)
smoke$f2 <- (smoke$gender == 0 & smoke$grp == 2)
smoke$m1 <- (smoke$gender == 1 & smoke$grp == 1)
smoke$m2 <- (smoke$gender == 1 & smoke$grp == 2)

smoke <- smoke %>% mutate(newgroup = case_when(f1 == TRUE ~ 'f1',
                          f2 == TRUE ~ 'f2',
                          m1 == TRUE ~ 'm1',
                          m2 == TRUE ~ 'm2',
                          TRUE ~ 'na'))

head(smoke)
```

```
##    id ttr relapse grp age gender race employment yearsSmoking levelSmoking
## 1  21  41       0   2  36      1    4          1           26            1
## 2 113  14       1   2  41      1    4          2           27            1
## 3  39   5       1   1  25      0    4          2           12            1
## 4  80  16       1   1  54      1    4          1           39            1
## 5  87   0       1   1  45      1    4          2           30            1
## 6  29 157       0   1  43      1    2          1           30            1
##    admitdate      fdate priorAttempts longestNoSmoke time    f1    f2    m1
## 1 2020-11-20 2020-12-31             0              0   41 FALSE FALSE FALSE
## 2 2020-06-16 2020-06-30             3             90   14 FALSE FALSE FALSE
## 3 2020-05-09 2020-05-14             3             21    5  TRUE FALSE FALSE
## 4 2020-10-26 2020-11-11             0              0   16 FALSE FALSE  TRUE
## 5 2020-09-27 2020-09-27             0              0    0 FALSE FALSE  TRUE
## 6 2020-07-06 2020-12-10             2           1825  157 FALSE FALSE  TRUE
##      m2 newgroup
## 1  TRUE       m2
## 2  TRUE       m2
## 3 FALSE       f1
```

```
## 4 FALSE       m1
## 5 FALSE       m1
## 6 FALSE       m1
```

```
#smoke$newgroup[smoke$newgroup == 'f1'] <- 1
#smoke$newgroup[smoke$newgroup == 'f2'] <- 2
#smoke$newgroup[smoke$newgroup == 'm1'] <- 3
#smoke$newgroup[smoke$newgroup == 'm2'] <- 4
#smoke
```

```
#censored <- Surv(smoke$time, smoke$relapse)
# NOT seperating groups?
fit1 <- survfit(Surv(time, relapse) ~ newgroup, data = smoke)
ggsurvplot(fit = fit1, data = smoke, conf.int = TRUE)
```



part (c) Null hypothesis: There is no difference in time between admission and relapse based on gender and treatement group. Alternative hypothesis: There is a difference in time between admission and relapse based on gender and treatement group.

```
logrank<-survdiff(Surv(time, relapse) ~ newgroup, data = smoke)
logrank
```

```
## Call:
## survdiff(formula = Surv(time, relapse) ~ newgroup, data = smoke)
##
```

```
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## newgroup=f1 39       23     28.0  0.907807  1.421756
## newgroup=f2 42       38     31.5  1.361350  2.218132
## newgroup=m1 22       14     13.9  0.000732  0.000915
## newgroup=m2 22       14     15.6  0.163900  0.215540
##
##  Chisq= 2.6  on 3 degrees of freedom, p= 0.5
```

```
peto<-survdiff(Surv(time, relapse) ~ newgroup, rho=1, data = smoke)
peto
```

```
## Call:
## survdiff(formula = Surv(time, relapse) ~ newgroup, data = smoke,
##     rho = 1)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## newgroup=f1 39     12.7    17.77     1.472     3.224
## newgroup=f2 42     21.8    19.93     0.172     0.397
## newgroup=m1 22     10.3     8.22     0.523     0.914
## newgroup=m2 22     10.3     9.14     0.154     0.274
##
##  Chisq= 3.5  on 3 degrees of freedom, p= 0.3
```

My conclusion based on these results is that there is moderate evidence that there is no significant difference in the groups based on gender and treatment group.

part (d)

```
pairwise_survdiff(Surv(time, relapse) ~ newgroup, data = smoke, p.adjust.method = "bonferroni" )
```

```
##
##  Pairwise comparisons using Log-Rank test
##
## data:  smoke and newgroup
##
##    f1 f2 m1
## f2 1  -  -
## m1 1  1  -
## m2 1  1  1
##
## P value adjustment method: bonferroni
```

This output shows that there is no significant difference between the genders and treatment group. In other words, we fail to reject the null hypothesis.