# STAT4008
# Survival Modeling

## Assignment 3 Solution

## 1 Q1

Test $H_0 : h(t) = 0.2$.

We have $\chi^2 = \frac{(O-E)^2}{E} = \frac{\left[\sum_{i=1}^{12} d_i - (\sum_{i=1}^{12} T_i)(0.2)\right]^2}{(0.2)(\sum_{i=1}^{12} T_i)} = \frac{[8-(230)(0.2)]^2}{(0.2)(230)}$ = 31.4 > 3.84 = $\chi^2_{1,0.95}$ and so null is rej.

Note: denominator equals E = $\sum_{i=1}^{12} H_0(T_j) = \sum_{i=1}^{12} \int_0^{T_j} h_0(t)\, dt = \sum_{i=1}^{12} \int_0^{T_j} 0.2\, dt = \sum_{i=1}^{12} T_i\ (0.2)$

(If you follow the example in tutorial 5, you will obtain $\chi^2$ = 0.05128 and null is not rejected, but please don't follow that approach in subsequent tests and exams.)

## 2 Q2

Test $H_0 : h_1(t) = h_2(t)$.
Consider the two tables below:

| i | ti | ni | di | n1i | d1i | n2i | d2i | w1i | w2i | d1i-w1i | d2i-w2i |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.00 | 21.00 | 3.00 | 9.00 | 2.00 | 12.00 | 1.00 | 1.29 | 1.71 | 0.71 | -0.71 |
| 2 | 3.00 | 18.00 | 3.00 | 7.00 | 1.00 | 11.00 | 2.00 | 1.17 | 1.83 | -0.17 | 0.17 |
| 3 | 4.00 | 13.00 | 3.00 | 5.00 | 1.00 | 8.00 | 2.00 | 1.15 | 1.85 | -0.15 | 0.15 |
| 4 | 5.00 | 8.00 | 3.00 | 3.00 | 2.00 | 5.00 | 1.00 | 1.12 | 1.88 | 0.88 | -0.88 |
| 5 | 6.00 | 3.00 | 1.00 | 1.00 | 0.00 | 2.00 | 1.00 | 0.33 | 0.67 | -0.33 | 0.33 |
| 6 | 7.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 |

| i | [V11]i | [V12]i | [V22]i |
|---|---|---|---|
| 1 | 0.66 | -0.66 | 0.66 |
| 2 | 0.63 | -0.63 | 0.63 |
| 3 | 0.59 | -0.59 | 0.59 |
| 4 | 0.50 | -0.50 | 0.50 |
| 5 | 0.22 | -0.22 | 0.22 |
| 6 | 0.00 | 0.00 | 0.00 |

For Log rank test:
$U_L = \sum_{i=1}^{6}(d_{1i} - w_{1i}) = 0.9354$ and $V_L = \sum_{i=1}^{6}([V11]_i) = 2.60648$.
So, $\chi^2_L = U_L^2/V_L = 0.335 < 3.84 = \chi^2_{1,0.95}$

(Alternative Method) For Wilcoxon test:
$U_W = \sum_{i=1}^{6} n_i(d_{1i} - w_{1i}) = 16$ and $V_W = \sum_{i=1}^{6} n_i^2([V11]_i) = 629.5664$.
So, $\chi^2_W = U_L^2/V_L = 0.4066 < 3.84 = \chi^2_{1,0.95}$

In both methods, $H_0$ is NOT rejected at % siginificance.

Alternatively, you can use the following R-code:

Survdiff(formula=Surv(time,status)~group)      #chisq=0.3 and p=0.6

# 3 Q3

When two prognostic factors are joinly considered, there are 4 groups of data. If non-parametric method is used, R output:

```
Call:
survdiff(formula = Surv(data, c) ~ x1 + x2)

              N Observed Expected (O-E)^2/E (O-E)^2/V
x1=0, x2=0  4        2     4.92     1.736     2.433
x1=0, x2=1  9        6     8.16     0.571     0.961
x1=1, x2=0  9        7     5.11     0.697     0.954
x1=1, x2=1  8        8     4.80     2.126     2.965

 Chisq= 5.7  on 3 degrees of freedom, p= 0.128
```

$\Longrightarrow$ $p - value > 0.05$, $H_0$ is NOT rejected at % siginificance.

Alternatively, since the wordings are not clear in the questions, the followings are also accepted.

```
> survdiff(Surv(set3$time,set3$status)~set3$x1)
Call:
survdiff(formula = Surv(set3$time, set3$status) ~ set3$x1)

          N Observed Expected (O-E)^2/E (O-E)^2/V
set3$x1=0 13 8 13.08 1.97 5.03
set3$x1=1 17 15 9.92 2.61 5.03

 Chisq= 5 on 1 degrees of freedom, p= 0.02

> survdiff(Surv(set3$time,set3$status)~set3$x2)
Call:
survdiff(formula = Surv(set3$time, set3$status) ~ set3$x2)

          N Observed Expected (O-E)^2/E (O-E)^2/V
set3$x2=0 13 9 10 0.1071 0.203
set3$x2=1 17 14 13 0.0829 0.203

 Chisq= 0.2 on 1 degrees of freedom, p= 0.7
```

# 4 Q4

Q4a)Testing $H_0 : h_1(t) = h_2(t) = h_3(t)$ for group 1, 2, 3.
If use non-parametric method, R output:

```
Call:
survdiff(formula = Surv(d$Time, rep(1, nrow(d))) ~ Group, data = d)

         N Observed Expected (O-E)^2/E (O-E)^2/V
Group=1 51       51    40.40     2.783     6.448
Group=2 26       26    34.92     2.278     4.697
Group=3  7        7     8.68     0.327     0.497

 Chisq= 6.5  on 2 degrees of freedom, p= 0.0396
```

$\implies p-value < 0.05$, $H_0$ is rejected at % siginificance.

(Alternative method) If cox model is used, R-output:

```
Call:
coxph(formula = Surv(d$Time, rep(1, nrow(d))) ~ factor(Group), data = d)

                 coef exp(coef) se(coef)     z     p
factor(Group)2 -0.541     0.582    0.243 -2.23 0.026
factor(Group)3 -0.558     0.572    0.440 -1.27 0.200

Likelihood ratio test=5.88  on 2 df, p=0.0529  n= 84, number of events= 84
```

$\implies p-value > 0.05$, $H_0$ is marginally not rejected at % siginificance.

Q4b) Testing $H_0 : h_1(t) = h_2(t) = h_3(t)$ for treatment 1, 2, 3
If use non-parametric method, R output:

```
Call:
survdiff(formula = Surv(d$Time, rep(1, nrow(d))) ~ Treatment, data = d)

             N Observed Expected (O-E)^2/E (O-E)^2/V
Treatment=1 15       15     12.6   0.47562     0.712
Treatment=2 33       33     33.6   0.00914     0.018
Treatment=3 36       36     37.9   0.09428     0.205

 Chisq= 0.7  on 2 degrees of freedom, p= 0.692
```

$\implies p-value > 0.05$, $H_0$ is NOT rejected at % siginificance.

(Alternative method) If cox model is used, R-output:

```
Call:
coxph(formula = Surv(d$Time, rep(1, nrow(d))) ~ as.factor(Treatment), data = d)

                        coef exp(coef) se(coef)      z    p
as.factor(Treatment)2 -0.187     0.829    0.322 -0.582 0.56
as.factor(Treatment)3 -0.233     0.792    0.319 -0.731 0.47

Likelihood ratio test=0.52  on 2 df, p=0.771  n= 84, number of events= 84
```

$\implies p-value > 0.05$, $H_0$ is NOT rejected at % siginificance.

# 5 Q5

**data = read.csv("ass3q5.csv", header = T)**

**set.seed(123457)**

**draw = sample(nrow(data),100)**

**subsample = data[draw, ]**

**survival_data = Surv(subsample$Time, event = subsample$Status)**

**survdiff(formula = survival_data ~ subsample$Smoking.Status)**

**Output, you should be able to obtain this,**

```
Call:
survdiff(formula = Surv(sub$Time, sub$Status) ~ sub$Smoking.Status)

                      N Observed Expected (O-E)^2/E (O-E)^2/V
sub$Smoking.Status=0 32       26     34.4     2.043      4.20
sub$Smoking.Status=1 32       26     21.5     0.920      1.34
sub$Smoking.Status=2 36       33     29.1     0.531      0.88

 Chisq= 4.3  on 2 degrees of freedom, p= 0.1
```

**Or this**

```
> survdiff(Surv(sub$Time,sub$Status)~sub$Smoking.Status)
Call:
survdiff(formula = Surv(sub$Time, sub$Status) ~ sub$Smoking.Status)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
sub$Smoking.Status=0 37 29 33.5 0.5990 1.077
sub$Smoking.Status=1 28 24 21.0 0.4236 0.604
sub$Smoking.Status=2 35 31 29.5 0.0757 0.127

 Chisq= 1.2 on 2 degrees of freedom, p= 0.6
```