Name: CHAN King Yeung
SID:1155119394
STAT4008 Assignment 4

---

Question 1

```
> time = c(18, 9, 28, 31, 39, 19, 45, 6, 8, 15, 23, 28, 7, 12, 9, 8, 2, 26, 10,
4, 3, 4, 18, 8, 3, 14, 3, 13, 13, 35)
> censor = c(1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 0)
> group = c(1, 3, 1, 3, 3, 3, 3, 3, 3, 3, 1, 1, 3, 2, 2, 2, 4, 2, 4, 2, 2, 2,
4, 4, 4, 4, 2, 4, 4, 2)
>
> coxModel = coxph(Surv(time, censor) ~ factor(group), ties = 'exact')
> summary(coxModel)
Call:
coxph(formula = Surv(time, censor) ~ factor(group), ties = "exact")

  n= 30, number of events= 23

                  coef exp(coef) se(coef)     z Pr(>|z|)
factor(group)2 1.3155    3.7266   0.8252 1.594   0.1109
factor(group)3 0.6355    1.8880   0.8355 0.761   0.4469
factor(group)4 1.5683    4.7985   0.8236 1.904   0.0569 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

               exp(coef) exp(-coef) lower .95 upper .95
factor(group)2     3.727     0.2683    0.7394     18.78
factor(group)3     1.888     0.5297    0.3671      9.71
factor(group)4     4.799     0.2084    0.9552     24.11

Concordance= 0.648  (se = 0.052 )
Likelihood ratio test= 5.85  on 3 df,   p=0.1
Wald test            = 5.19  on 3 df,   p=0.2
Score (logrank) test = 5.69  on 3 df,   p=0.1

>
> 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 4        2     4.92     1.736     2.433
group=2 9        7     5.11     0.697     0.954
group=3 9        6     8.16     0.571     0.961
group=4 8        8     4.80     2.126     2.965

 Chisq= 5.7  on 3 degrees of freedom, p= 0.1
```

With refer to the assumption that we have made in Assignment 3, we have 4 groups of individuals.

$H_0: \beta_i = 0 \; vs \; H_1: \beta_i \neq 0 \; \forall \; i \in \{group\ 2, group\ 3, group\ 4\}$

From the result given by above code, the p-values of wald test, likelihood ratio test and score test are greater than 0.05, thus, we do not reject $H_0$ at 5% level of significance. As we know a score test in cox regression model is equivalent to a k-sample log-rank test, test from `survdiff()`, it also suggests the same result.

---

Question 2

```
> data = read.csv('ass4.csv')
> set.seed(123457)
> data = data[sample(nrow(data), 1000), ]
> data$cens = 1 - data$cens
> data$hlstat3 = ifelse(data$hlstat == 3, 1, 0)
> data$hlstat4 = ifelse(data$hlstat == 4, 1, 0)
> data$hlstat5 = ifelse(data$hlstat == 5, 1, 0)
```

a)

```
> coxModel = coxph(Surv(lstay, cens) ~ age + trt + gender + marstat + hlstat3
+ hlstat4 + hlstat5, data = data)
> summary(coxModel)
Call:
coxph(formula = Surv(lstay, cens) ~ age + trt + gender + marstat +
    hlstat3 + hlstat4 + hlstat5, data = data)

  n= 1000, number of events= 808

             coef exp(coef)  se(coef)      z Pr(>|z|)
age     -0.005143  0.994870  0.004831 -1.065 0.287082
trt     -0.049818  0.951402  0.073156 -0.681 0.495881
gender   0.373212  1.452392  0.085414  4.369 1.25e-05 ***
marstat  0.150702  1.162650  0.098335  1.533 0.125389
hlstat3  0.098262  1.103252  0.098224  1.000 0.317126
hlstat4  0.342692  1.408735  0.099839  3.432 0.000598 ***
hlstat5  0.638442  1.893529  0.131099  4.870 1.12e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

        exp(coef) exp(-coef) lower .95 upper .95
age        0.9949     1.0052    0.9855     1.004
trt        0.9514     1.0511    0.8243     1.098
gender     1.4524     0.6885    1.2285     1.717
marstat    1.1626     0.8601    0.9588     1.410
hlstat3    1.1033     0.9064    0.9101     1.337
hlstat4    1.4087     0.7099    1.1584     1.713
hlstat5    1.8935     0.5281    1.4645     2.448

Concordance= 0.588  (se = 0.011 )
Likelihood ratio test= 62.1   on 7 df,   p=6e-11
Wald test            = 65.86  on 7 df,   p=1e-11
Score (logrank) test = 66.92  on 7 df,   p=6e-12
```

$H_0: \beta_i = 0 \; vs \; H_1: \beta_i \neq 0 \; \forall \; i \in parameters$

From the result given by above code, we have wald test value (65.86), likelihood ratio test value (62.1) and score test value (66.92). Their corresponding p-values are less than 0.05, thus, we reject $H_0$ at 5% level of significance.

b)

```
> sub = !coxModel$coefficients / sqrt(diag(coxModel$var)) >= qnorm(0.975)
> subPar = coxModel$coefficient[sub]
> subVar = coxModel$var[sub, sub]
> WT = t(subPar) %*% solve(subVar) %*% subPar
> pv_WT = 1 - pchisq(WT, length(subPar))
>
> coxModelSub = coxph(Surv(lstay, cens) ~ gender + hlstat4 + hlstat5, data =
data)
> LRT = 2 * (coxModel$loglik[2] - coxModelSub$loglik[2])
> pv_LRT = 1 - pchisq(LRT, length(coxModel$coefficients) -
length(coxModelSub$coefficients))
>
> initial = rep(0, length(coxModel$coefficients))
> initial[!sub] = coxModelSub$coefficients
> coxModelSco = coxph(Surv(lstay, cens) ~ age + trt + gender + marstat +
hlstat3 + hlstat4 + hlstat5, data = data, init = initial)
> ST = coxModelSco$score
> pv_ST = 1 - pchisq(coxModelSco$score, length(coxModel$coefficients) -
length(coxModelSub$coefficients))
>
> data.frame(testValue = c(WT, LRT, ST), pv = c(pv_WT, pv_LRT, pv_ST),
row.names = c('Wald', 'Likelihood', 'Score'))
           testValue        pv
Wald        5.430187 0.2459346
Likelihood  5.356401 0.2526432
Score       5.436134 0.2454006
```

The individual z-scores only describe the significance of that covariate on the model fitted. In other words, we cannot simply remove those covariates with p-value greater than 0.05. We have to construct a test to verify the significance of those covariates in terms of models. From code in (a), only 'gender', 'hlstat4' and 'hlstat5' are significant based on individual z-scores. Thus, we will test 'age', 'trt', 'marstat', and 'hlstat3' simultaneously, such that we have

$H_0: \beta_i = 0 \; vs \; H_1: \beta_i \neq 0 \; \forall \; i \in \{age, trt, marstat, hlstat3\}$

From the result given by above code, we have wald test value $(5.4302)$, likelihood ratio test value $(5.3564)$ and score test value $(5.4361)$. Their corresponding p-values $(0.2459,\ 0.2526$ and $0.2454$, respectively) are greater than 0.05, thus, we do not reject $H_0$ at 5% level of significance.

c)

```
> n = 2
> m = 3
> me = qnorm(0.975) * sqrt(coxModelSub$var[n, n] + coxModelSub$var[m, m] - 2
* coxModelSub$var[n, m])
> c(exp(coxModelSub$coefficients[n] - coxModelSub$coefficients[m] - me),
exp(coxModelSub$coefficients[n] - coxModelSub$coefficients[m] + me))
  hlstat4   hlstat4
0.5909952 0.9415761
```

The confidence interval is $[0.591, 0.9416]$ based on model with covariates 'gender', 'hlstat4' and 'hlstat5' only.