stat-633-a3

2023-03-22

question 1

part a)

```
library(KMsurv)
library(survival)
data(burn)
data = burn
head(data,1)
```

```
##   Obs Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z9 Z10 Z11 T1 D1 T2 D2 T3 D3
## 1   1  0  0  0 15  0  0  1  1  0   0   2 12  0 12  0 12  0
```

```
coxph.fit = coxph(Surv(T3, D3)~factor(Z1), data = data, method = 'breslow')
summary(coxph.fit)
```

```
## Call:
## coxph(formula = Surv(T3, D3) ~ factor(Z1), data = data, method = "breslow")
##
##   n= 154, number of events= 48
##
##                coef exp(coef) se(coef)      z Pr(>|z|)
## factor(Z1)1 -0.5599    0.5713   0.2933 -1.909   0.0563 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef) exp(-coef) lower .95 upper .95
## factor(Z1)1    0.5713       1.75    0.3215     1.015
##
## Concordance= 0.566  (se = 0.039 )
## Likelihood ratio test= 3.71  on 1 df,   p=0.05
## Wald test            = 3.64  on 1 df,   p=0.06
## Score (logrank) test = 3.74  on 1 df,   p=0.05
```

```
s = summary(coxph.fit)
# wald test
betahat=s$coefficients[1,1]
se_beta=s$coefficients[1,3]
z_stat=betahat/se_beta
p.val=2*(1-pnorm(abs(z_stat)))
cat("Wald Z-Test Statistics = ", z_stat, "\n")
```

```
## Wald Z-Test Statistics =  -1.908666
```

```
cat("p-value of the Wald test = ", p.val, "\n")
```

```
## p-value of the Wald test =  0.05630523
```

Since the $p - value > 0.05 = \alpha$, we failed to reject the null hypothesis, meaning that there is a difference between the distributions of the times to staphylococcus infection in the two disinfectant groups.

part b)

```
library(KMsurv)
library(survival)
data = burn
head(data,1)
```

```
##   Obs Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z9 Z10 Z11 T1 D1 T2 D2 T3 D3
## 1   1  0  0  0 15  0  0  1  1  0   0   2 12  0 12  0 12  0
```

```
coxph.fit = coxph(Surv(T3, D3)~factor(Z1)+Z4, data = data, method = 'breslow')
summary(coxph.fit)
```

```
## Call:
## coxph(formula = Surv(T3, D3) ~ factor(Z1) + Z4, data = data,
##     method = "breslow")
##
##   n= 154, number of events= 48
##
##                   coef exp(coef)  se(coef)      z Pr(>|z|)
## factor(Z1)1 -0.523278  0.592575  0.295806 -1.769   0.0769 .
## Z4           0.007186  1.007212  0.007142  1.006   0.3143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef) exp(-coef) lower .95 upper .95
## factor(Z1)1    0.5926     1.6876    0.3319     1.058
## Z4             1.0072     0.9928    0.9932     1.021
##
## Concordance= 0.585  (se = 0.046 )
## Likelihood ratio test= 4.66  on 2 df,   p=0.1
## Wald test            = 4.68  on 2 df,   p=0.1
## Score (logrank) test = 4.79  on 2 df,   p=0.09
```

```
s = summary(coxph.fit)
# wald test
betahat=s$coefficients[1,1]
se_beta=s$coefficients[1,3]
z_stat=betahat/se_beta
p.val=2*(1-pnorm(abs(z_stat)))
cat("Wald Z-Test Statistics = ", z_stat, "\n")
```

```
## Wald Z-Test Statistics =  -1.768991
```

```
cat("p-value of the Wald test = ", p.val, "\n")
```

```
## p-value of the Wald test =  0.07689545
```

Since the $0.07 = p - value > 0.05 = \alpha$, we failed to reject the null hypothesis, meaning that there is a difference between the distributions of the times to staphylococcus infection in the two disinfectant groups, adjusting for the total area burned.

We can also find that adding the adjustment of total area burned, the estimate of *beta* and its standard error are also changed as well as the wald test statistics. This makes sense as now we have one additional variable needs to be considered in the model, and total area burned also influcence the fitted value as well.

part c)

```
library(KMsurv)
library(survival)
data = burn
head(data,1)
```

```
##   Obs Z1 Z2 Z3 Z4 Z5 Z6 Z7 Z8 Z9 Z10 Z11 T1 D1 T2 D2 T3 D3
## 1   1  0  0  0 15  0  0  1  1  0   0   2 12  0 12  0 12  0
```

```
# Adjust for gender
coxph.fit2 = coxph(Surv(T3, D3)~factor(Z1)+factor(Z2), data = data, method = 'breslow')
summary(coxph.fit2)
```

```
## Call:
## coxph(formula = Surv(T3, D3) ~ factor(Z1) + factor(Z2), data = data,
##     method = "breslow")
##
##   n= 154, number of events= 48
##
##                 coef exp(coef) se(coef)      z Pr(>|z|)
## factor(Z1)1 -0.6049    0.5462   0.2943 -2.055   0.0398 *
## factor(Z2)1 -0.6483    0.5229   0.3902 -1.661   0.0966 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef) exp(-coef) lower .95 upper .95
## factor(Z1)1    0.5462      1.831    0.3068    0.9723
## factor(Z2)1    0.5229      1.912    0.2434    1.1236
##
## Concordance= 0.617  (se = 0.038 )
## Likelihood ratio test= 6.86  on 2 df,   p=0.03
## Wald test            = 6.5  on 2 df,   p=0.04
## Score (logrank) test = 6.66  on 2 df,   p=0.04
```

```
s = summary(coxph.fit2)
# wald test
betahat=s$coefficients[1,1]
se_beta=s$coefficients[1,3]
z_stat=betahat/se_beta
p.val=2*(1-pnorm(abs(z_stat)))
cat("Wald Z-Test Statistics = ", z_stat, "\n")
```

```
## Wald Z-Test Statistics =  -2.05534
```

```
# If the p-value is less than 0.05, we reject the null hypothesis
cat("p-value of the Wald test = ", p.val, "\n")
```

```
## p-value of the Wald test =  0.03984613
```

```r
# Adjust for race
coxph.fit3 = coxph(Surv(T3, D3)~factor(Z1)+factor(Z3), data = data, method = 'breslow')
summary(coxph.fit3)
```

```
## Call:
## coxph(formula = Surv(T3, D3) ~ factor(Z1) + factor(Z3), data = data,
##     method = "breslow")
##
##   n= 154, number of events= 48
##
##                 coef exp(coef) se(coef)      z Pr(>|z|)
## factor(Z1)1 -0.5885    0.5552   0.2936 -2.004   0.0450 *
## factor(Z3)1  2.1592    8.6639   1.0112  2.135   0.0327 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef) exp(-coef) lower .95 upper .95
## factor(Z1)1    0.5552     1.8013    0.3122    0.9871
## factor(Z3)1    8.6639     0.1154    1.1939   62.8737
##
## Concordance= 0.633  (se = 0.036 )
## Likelihood ratio test= 13.56  on 2 df,   p=0.001
## Wald test            = 8.45  on 2 df,   p=0.01
## Score (logrank) test = 10.48  on 2 df,   p=0.005
```

```r
s = summary(coxph.fit3)
# wald test
betahat=s$coefficients[1,1]
se_beta=s$coefficients[1,3]
z_stat=betahat/se_beta
p.val=2*(1-pnorm(abs(z_stat)))
cat("Wald Z-Test Statistics = ", z_stat, "\n")
```

```
## Wald Z-Test Statistics =  -2.00427
```

```r
# If the p-value is less than 0.05, we reject the null hypothesis
cat("p-value of the Wald test = ", p.val, "\n")
```

```
## p-value of the Wald test =  0.04504115
```

```r
# Adjust for total surface area burned
coxph.fit11 = coxph(Surv(T3, D3)~factor(Z1)+Z4, data = data, method = 'breslow')
summary(coxph.fit11)
```

```
## Call:
## coxph(formula = Surv(T3, D3) ~ factor(Z1) + Z4, data = data,
##     method = "breslow")
##
##   n= 154, number of events= 48
##
##                  coef exp(coef)  se(coef)      z Pr(>|z|)
```

```
## factor(Z1)1 -0.523278  0.592575  0.295806 -1.769   0.0769 .
## Z4           0.007186  1.007212  0.007142  1.006   0.3143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef) exp(-coef) lower .95 upper .95
## factor(Z1)1    0.5926     1.6876    0.3319     1.058
## Z4             1.0072     0.9928    0.9932     1.021
##
## Concordance= 0.585  (se = 0.046 )
## Likelihood ratio test= 4.66  on 2 df,   p=0.1
## Wald test            = 4.68  on 2 df,   p=0.1
## Score (logrank) test = 4.79  on 2 df,   p=0.09
```

```
s = summary(coxph.fit11)
# wald test
betahat=s$coefficients[1,1]
se_beta=s$coefficients[1,3]
z_stat=betahat/se_beta
p.val=2*(1-pnorm(abs(z_stat)))
cat("Wald Z-Test Statistics = ", z_stat, "\n")
```

```
## Wald Z-Test Statistics =  -1.768991
```

```
# If the p-value is less than 0.05, we reject the null hypothesis
cat("p-value of the Wald test = ", p.val, "\n")
```

```
## p-value of the Wald test =  0.07689545
```

```
# Adjust for type of burn
coxph.fit11 = coxph(Surv(T3, D3)~factor(Z1)+factor(Z11), data = data, method = 'breslow')
summary(coxph.fit11)
```

```
## Call:
## coxph(formula = Surv(T3, D3) ~ factor(Z1) + factor(Z11), data = data,
##     method = "breslow")
##
##   n= 154, number of events= 48
##
##                 coef exp(coef) se(coef)      z Pr(>|z|)
## factor(Z1)1  -0.5956    0.5513   0.2968 -2.007   0.0448 *
## factor(Z11)2  1.1209    3.0675   1.0828  1.035   0.3006
## factor(Z11)3  2.2434    9.4249   1.0838  2.070   0.0385 *
## factor(Z11)4  0.9847    2.6769   1.0160  0.969   0.3325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(Z1)1     0.5513     1.8140    0.3081    0.9862
## factor(Z11)2    3.0675     0.3260    0.3674   25.6126
## factor(Z11)3    9.4249     0.1061    1.1265   78.8540
## factor(Z11)4    2.6769     0.3736    0.3654   19.6098
```

```
##
## Concordance= 0.641  (se = 0.042 )
## Likelihood ratio test= 11.34  on 4 df,   p=0.02
## Wald test            = 12.28  on 4 df,   p=0.02
## Score (logrank) test = 13.53  on 4 df,   p=0.009
```

```
s = summary(coxph.fit11)
# wald test
betahat=s$coefficients[1,1]
se_beta=s$coefficients[1,3]
z_stat=betahat/se_beta
p.val=2*(1-pnorm(abs(z_stat)))
cat("Wald Z-Test Statistics = ", z_stat, "\n")
```

```
## Wald Z-Test Statistics =  -2.006637
```

```
# If the p-value is less than 0.05, we reject the null hypothesis
cat("p-value of the Wald test = ", p.val, "\n")
```

```
## p-value of the Wald test =  0.04478836
```

part d)

```
# forward selection
upper.model = coxph(Surv(T3, D3)~factor(Z1)+factor(Z2)+factor(Z3)+Z4+factor(Z11),
                    data = data, method = 'breslow')
null.model = coxph(Surv(T3, D3)~factor(Z1), data = data, method = 'breslow')

library(MASS)
forward = stepAIC(null.model, direction = 'forward',
                  scope = list(lower = null.model, upper = upper.model),
                  test = 'Chisq', alpha = 0.05)
```

```
## Start:  AIC=437.55
## Surv(T3, D3) ~ factor(Z1)
##
##               Df    AIC    LRT  Pr(Chi)
## + factor(Z3)   1 429.70 9.8488 0.001699 **
## + factor(Z11)  3 435.92 7.6308 0.054291 .
## + factor(Z2)   1 436.40 3.1545 0.075717 .
## <none>           437.55
## + Z4           1 438.60 0.9524 0.329098
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=429.7
## Surv(T3, D3) ~ factor(Z1) + factor(Z3)
##
##               Df    AIC    LRT Pr(Chi)
## + factor(Z11)  3 427.76 7.9395 0.04728 *
## + factor(Z2)   1 428.71 2.9957 0.08349 .
## <none>           429.70
## + Z4           1 431.29 0.4093 0.52232
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=427.76
## Surv(T3, D3) ~ factor(Z1) + factor(Z3) + factor(Z11)
##
##              Df    AIC     LRT Pr(Chi)
## + factor(Z2)  1 427.62 2.14734  0.1428
## <none>          427.76
## + Z4          1 429.16 0.60561  0.4364
##
## Step:  AIC=427.62
## Surv(T3, D3) ~ factor(Z1) + factor(Z3) + factor(Z11) + factor(Z2)
##
##        Df    AIC     LRT Pr(Chi)
## <none>    427.62
## + Z4    1 429.24 0.37527  0.5401
```

```
summary(forward)
```

```
## Call:
```

```
## coxph(formula = Surv(T3, D3) ~ factor(Z1) + factor(Z3) + factor(Z11) +
##     factor(Z2), data = data, method = "breslow")
##
##   n= 154, number of events= 48
##
##                  coef exp(coef) se(coef)      z Pr(>|z|)
## factor(Z1)1  -0.6461    0.5241   0.2989 -2.162   0.0306 *
## factor(Z3)1   2.2707    9.6863   1.0260  2.213   0.0269 *
## factor(Z11)2  1.5729    4.8204   1.0872  1.447   0.1480
## factor(Z11)3  2.0474    7.7481   1.0894  1.879   0.0602 .
## factor(Z11)4  1.0109    2.7482   1.0173  0.994   0.3204
## factor(Z2)1  -0.5516    0.5760   0.3966 -1.391   0.1642
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(Z1)1     0.5241     1.9080    0.2918    0.9415
## factor(Z3)1     9.6863     0.1032    1.2966   72.3593
## factor(Z11)2    4.8204     0.2075    0.5723   40.5986
## factor(Z11)3    7.7481     0.1291    0.9160   65.5401
## factor(Z11)4    2.7482     0.3639    0.3742   20.1841
## factor(Z2)1     0.5760     1.7360    0.2648    1.2531
##
## Concordance= 0.719  (se = 0.037 )
## Likelihood ratio test= 23.65  on 6 df,   p=6e-04
## Wald test            = 18.67  on 6 df,   p=0.005
## Score (logrank) test = 21.95  on 6 df,   p=0.001
```

Initial: we have $Z1$ variable.

First variable to be added: $Z3$ as its p-value is the smallest with the value of 0.0017.

Second variable to be added: $Z11$ as its p-value is the smallest with the value of 0.047.

When we try to add the third variable, none of the candidate variables has the p-value less than 0.05, therefore, we terminate the process of adding variables.

So, we have the proposed model with **Disinfectant**, **Type of Burn** and **Race**

```
my.model = coxph(Surv(T3, D3)~factor(Z1)+factor(Z3)+factor(Z11), data = data, method = 'breslow')
summary(my.model)
```

```
## Call:
## coxph(formula = Surv(T3, D3) ~ factor(Z1) + factor(Z3) + factor(Z11),
##     data = data, method = "breslow")
##
##   n= 154, number of events= 48
##
##                  coef exp(coef) se(coef)      z Pr(>|z|)
## factor(Z1)1  -0.6008    0.5483   0.2978 -2.017   0.0437 *
## factor(Z3)1   2.2696    9.6752   1.0256  2.213   0.0269 *
## factor(Z11)2  1.5565    4.7422   1.0865  1.433   0.1520
## factor(Z11)3  2.1510    8.5935   1.0858  1.981   0.0476 *
## factor(Z11)4  0.9994    2.7167   1.0163  0.983   0.3254
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(Z1)1     0.5483     1.8237    0.3059    0.9831
## factor(Z3)1     9.6752     0.1034    1.2961   72.2238
## factor(Z11)2    4.7422     0.2109    0.5638   39.8856
## factor(Z11)3    8.5935     0.1164    1.0230   72.1842
## factor(Z11)4    2.7167     0.3681    0.3707   19.9115
##
## Concordance= 0.705  (se = 0.037 )
## Likelihood ratio test= 21.5  on 5 df,   p=7e-04
## Wald test            = 17.05  on 5 df,   p=0.004
## Score (logrank) test = 20.16  on 5 df,   p=0.001
```

The local Wald test of the primary hypothesis of no difference between the times to staphylococcus infection for the two disinfectant groups has a p-value of 0.044, which suggests that the times to staphylococcus infection are different for the two disinfectant groups after adjustment for the type of the burn and race of the patient.

question 2

See hand written solution