homework05

Jiapeng Wang

2023-10-31

Some Code for Question1:

```
z <- read.csv("lungcancer.csv")
z$city <- as.factor(z$city)
z$age <- as.factor(z$age)
mean(z$cases/z$pop)
```

```
## [1] 0.01190005
```

```
glm(cases/pop ~ city + age, weight=pop, family='binomial', data=z)
```

```
##
## Call:  glm(formula = cases/pop ~ city + age, family = "binomial", data = z,
##     weights = pop)
##
## Coefficients:
## (Intercept)  cityHorsens  cityKolding    cityVejle     age55-59     age60-64
##     -5.6262      -0.3345      -0.3764      -0.2760       1.1070       1.5291
##    age65-69     age70-74       age75+
##      1.7819       1.8727       1.4289
##
## Degrees of Freedom: 23 Total (i.e. Null);  15 Residual
## Null Deviance:       131
## Residual Deviance: 23.64     AIC: 137.7
```

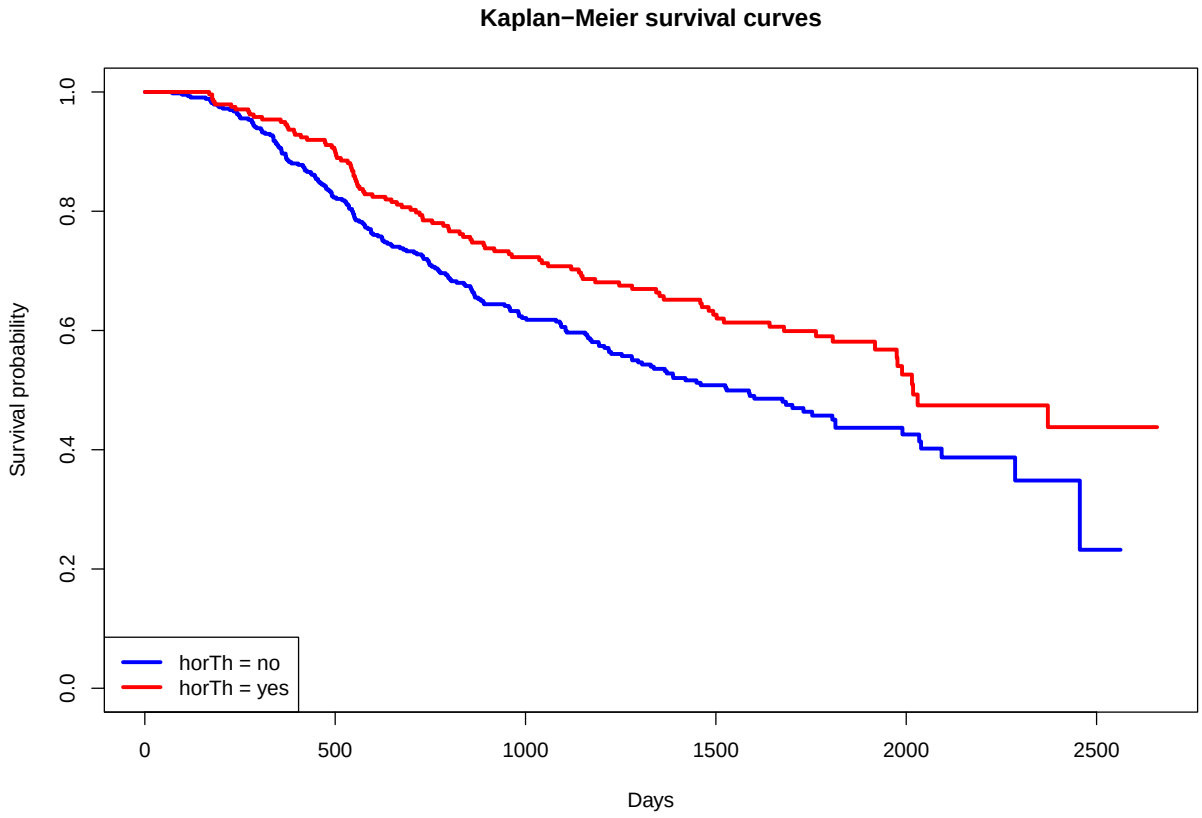
```
glm(cases ~ city + age, offset=log(pop), family='poisson', data=z)
```

```
##
## Call:  glm(formula = cases ~ city + age, family = "poisson", data = z,
##     offset = log(pop))
##
## Coefficients:
## (Intercept)  cityHorsens  cityKolding    cityVejle     age55-59     age60-64
##     -5.6321      -0.3301      -0.3715      -0.2723       1.1010       1.5186
##    age65-69     age70-74       age75+
##      1.7677       1.8569       1.4197
##
## Degrees of Freedom: 23 Total (i.e. Null);  15 Residual
## Null Deviance:       129.9
## Residual Deviance: 23.45     AIC: 137.8
```

Question2:

1.Reproduce the plot on page 117:

```
z <- read.csv("cancerdata.csv")
leg.txt <- c("horTh = no","horTh = yes")
leg.col <- c("blue","red")
leg.lty <- rep(1,2)
y <- z$time
stat <- z$death
treat <- z$horTh
fit.bytreat <- survfit(Surv(y,stat) ~ treat, conf.type="none")
plot(fit.bytreat,conf.int=FALSE,
     col=leg.col,mark.time=FALSE,
     lwd=3,
     main='Kaplan-Meier survival curves',
     xlab='Days',
     ylab='Survival probability')
legend("bottomleft",legend=leg.txt,lty=leg.lty,col=leg.col,lwd=3)
```



2.fit a Cox PH model to the breastcancer data with two-factor interactions with treatment added:

```
summary(coxph(Surv(time, death)~ treat , data=z))
```

```
## Call:
## coxph(formula = Surv(time, death) ~ treat, data = z)
```

```
##
##   n= 686, number of events= 299
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## treatyes -0.3640    0.6949   0.1250 -2.911   0.0036 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## treatyes    0.6949      1.439    0.5438    0.8879
##
## Concordance= 0.543  (se = 0.014 )
## Likelihood ratio test= 8.82  on 1 df,   p=0.003
## Wald test            = 8.47  on 1 df,   p=0.004
## Score (logrank) test = 8.57  on 1 df,   p=0.003
```

```
summary(coxph(Surv(time, death)~ . , data=z))
```

```
## Call:
## coxph(formula = Surv(time, death) ~ ., data = z)
##
##   n= 686, number of events= 299
##
##                   coef  exp(coef)   se(coef)      z Pr(>|z|)
## horThyes    -0.3372029  0.7137640  0.1289618 -2.615 0.008929 **
## age         -0.0093924  0.9906516  0.0092733 -1.013 0.311136
## menostatPre -0.2672772  0.7654609  0.1833366 -1.458 0.144882
## tsize        0.0077164  1.0077463  0.0039497  1.954 0.050739 .
## tgrade       0.2802894  1.3235128  0.1060553  2.643 0.008221 **
## pnodes       0.0498939  1.0511596  0.0074094  6.734 1.65e-11 ***
## progrec     -0.0022378  0.9977647  0.0005758 -3.887 0.000102 ***
## estrec       0.0001674  1.0001674  0.0004477  0.374 0.708431
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef) exp(-coef) lower .95 upper .95
## horThyes       0.7138     1.4010    0.5543    0.9190
## age            0.9907     1.0094    0.9728    1.0088
## menostatPre    0.7655     1.3064    0.5344    1.0964
## tsize          1.0077     0.9923    1.0000    1.0156
## tgrade         1.3235     0.7556    1.0751    1.6293
## pnodes         1.0512     0.9513    1.0360    1.0665
## progrec        0.9978     1.0022    0.9966    0.9989
## estrec         1.0002     0.9998    0.9993    1.0010
##
## Concordance= 0.688  (se = 0.015 )
## Likelihood ratio test= 101.9  on 8 df,   p=<2e-16
## Wald test            = 115.3  on 8 df,   p=<2e-16
## Score (logrank) test = 120.1  on 8 df,   p=<2e-16
```

```
formula = as.formula(Surv(time, death) ~ horTh + horTh:age + horTh:tsize +
                       horTh:tgrade + horTh:pnodes + horTh:progrec +
                       horTh:estrec + horTh:menostat)
```

```
model = coxph(formula, data=z)
summary(model)
```

```
## Call:
## coxph(formula = formula, data = z)
##
##   n= 686, number of events= 299
##
##                            coef  exp(coef)   se(coef)      z Pr(>|z|)
## horThyes             -1.3741340  0.2530587  1.3865227 -0.991  0.32165
## horThno:age          -0.0125766  0.9875021  0.0110689 -1.136  0.25587
## horThyes:age         -0.0026820  0.9973216  0.0179199 -0.150  0.88103
## horThno:tsize         0.0073572  1.0073843  0.0050913  1.445  0.14844
## horThyes:tsize        0.0090903  1.0091317  0.0067944  1.338  0.18093
## horThno:tgrade        0.2334551  1.2629561  0.1267457  1.842  0.06549 .
## horThyes:tgrade       0.4214016  1.5240963  0.2014613  2.092  0.03646 *
## horThno:pnodes        0.0424975  1.0434135  0.0090625  4.689 2.74e-06 ***
## horThyes:pnodes       0.0680320  1.0703995  0.0165265  4.117 3.85e-05 ***
## horThno:progrec      -0.0014300  0.9985710  0.0006370 -2.245  0.02477 *
## horThyes:progrec     -0.0042567  0.9957523  0.0013203 -3.224  0.00126 **
## horThno:estrec       -0.0001919  0.9998081  0.0006600 -0.291  0.77128
## horThyes:estrec       0.0001113  1.0001113  0.0006125  0.182  0.85577
## horThno:menostatPre  -0.3176198  0.7278795  0.2126709 -1.493  0.13531
## horThyes:menostatPre -0.2341843  0.7912160  0.3794376 -0.617  0.53711
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                      exp(coef) exp(-coef) lower .95 upper .95
## horThyes                0.2531     3.9517   0.01671    3.8321
## horThno:age             0.9875     1.0127   0.96631    1.0092
## horThyes:age            0.9973     1.0027   0.96290    1.0330
## horThno:tsize           1.0074     0.9927   0.99738    1.0175
## horThyes:tsize          1.0091     0.9910   0.99578    1.0227
## horThno:tgrade          1.2630     0.7918   0.98515    1.6191
## horThyes:tgrade         1.5241     0.6561   1.02690    2.2620
## horThno:pnodes          1.0434     0.9584   1.02504    1.0621
## horThyes:pnodes         1.0704     0.9342   1.03628    1.1056
## horThno:progrec         0.9986     1.0014   0.99733    0.9998
## horThyes:progrec        0.9958     1.0043   0.99318    0.9983
## horThno:estrec          0.9998     1.0002   0.99852    1.0011
## horThyes:estrec         1.0001     0.9999   0.99891    1.0013
## horThno:menostatPre     0.7279     1.3739   0.47977    1.1043
## horThyes:menostatPre    0.7912     1.2639   0.37611    1.6645
##
## Concordance= 0.693  (se = 0.015 )
## Likelihood ratio test= 112.2  on 15 df,   p=<2e-16
## Wald test            = 116  on 15 df,   p=<2e-16
## Score (logrank) test = 122.9  on 15 df,   p=<2e-16
```

Conclusion:

First, from the reproduction of plot, we get that horTh is effective to the survival probability. We can also prove this by the p-value = 0.0036 in summary(coxph(Surv(time, death)~ treat , data=z)). Thus, the coefficient of 'treat' has statistical difference from 0.

Also, from the Cox PH model(same as page 119 in slides), we can denote that tgrade, pnodes and progrec have significant effect as well.

Then from the model with two-factor interactions, we consider the influence caused by horTh and other factors together. From the model summary, it can be observed that the variables pnodes and progrec have a significant impact on the survival probability under this situation. Besides, one particular thing is that pnodes has a significant effect on patients regardless of whether they are treated with horTh. However, the influence of progrec varies depending on whether the patient is treated with horTh or not. When the patient is treated with horTh, he/she is much likely to be influenced by progrec.