

BIOSTATYSTYKA – PRACA DOMOWA 3

MARTA SOMMER – BSMAD

Zadanie 1.

Rozważamy dane dotyczące raka pęcherza. Model warunkowy (PWP) przedstawiony na wykładzie (`mod.wyk`) zakłada ten sam efekt leczenia dla wszystkich typów zdarzeń. Skonstruuję model (`mod.moj`), który dopuszcza możliwość, że efekt ten jest różny dla różnych typów zdarzeń. Oto współczynniki w moim modelu:

```
## Call:
## coxph(formula = Surv(time1, time2, status) ~ rx1 + rx2 + rx3 +
##       rx4 + size + number + cluster(patid) + strata(enum), data = b3)
##
##
##              coef exp(coef) se(coef) robust se      z      p
## rx1      -0.41835     0.658   0.3042    0.2936 -1.425 0.150
## rx2      -0.45184     0.636   0.3981    0.4371 -1.034 0.300
## rx3      -0.06741     0.935   0.5985    0.4107 -0.164 0.870
## rx4       0.20436     1.227   0.6881    0.5028  0.406 0.680
## size     -0.00745     0.993   0.0732    0.0633 -0.118 0.910
## number   0.11421     1.121   0.0540    0.0513  2.227 0.026
##
## Likelihood ratio test=7.47  on 6 df, p=0.28  n= 178, number of events= 112
```

Zwróćmy najpierw uwagę na statystykę ilorazu wiarygodności ($p = 0.2799$). Oznacza ona, że model jest źle dopasowany i współczynniki nie są istotne. Podobnie zresztą wskazują testy Walda ($p = 0.1983$), score ($p = 0.2511$) i Robust ($p = 0.1489$). Zatem na efekt leczenia nie ma wpływu rodzaj zdarzenia.

Używając skalowanych reszt Schoenfelda, sprawdzimy jeszcze, czy spełnione jest założenie proporcjonalnych hazardów:

```
##              rho  chisq    p
## rx1      -0.0593 0.37922 0.538
## rx2       0.1281 2.05291 0.152
## rx3     -0.0531 0.15751 0.691
## rx4       0.0430 0.10867 0.742
## size     -0.0056 0.00269 0.959
## number  -0.0376 0.14146 0.707
## GLOBAL          NA 2.84062 0.829
```

P-value wszędzie jest duże, zatem nie mamy podstaw do odrzucenia hipotezy o spełnianiu założenia proporcjonalnych hazardów. Model skonstruowany powyżej ma zatem sens i w ogóle mógł być zbudowany.

Przeprowadźmy jeszcze test hipotezy zerowej, że efekt leczenia jest taki sam dla wszystkich typów zdarzeń. Użyję statystyki testu ilorazu wiarygodności do hipotez liniowych i przyrównam ją do rozkładu χ^2 z trzema stopniami swobody, gdyż takiego rzędu jest nasza macierz restrykcji. P-value mojego testu jest duże (0.8118099), zatem nie mam podstaw do odrzucenia hipotezy zerowej.

Zadanie 2.

Będziemy rozważać dane dotyczące czasu do wznowy po uzyskaniu remisji dla par chorych na białaczkę. Pacjenci byli dobierani w pary tak, by pochodzili z tego samego ośrodka medycznego i mieli ten sam typ uzyskanej remisji, dlatego zastosujemy do nich model podatności. Ponadto chcemy uwzględnić efekt remisji, więc użyjemy warstwowania zmiennej `partial`.

```
## Call:
## coxph(formula = Surv(time, status) ~ treat + frailty(patid) +
##       strata(partial), data = 1)
##
##
## n= 42, number of events= 31
##
```

```
##           coef se(coef) se2   Chisq DF p
## treat      -1.71 0.437    0.437 15.3  1 9.3e-05
## frailty(patid)           0.0  0 9.5e-01
##
##      exp(coef) exp(-coef) lower .95 upper .95
## treat    0.181      5.52    0.077    0.427
##
## Iterations: 6 outer, 35 Newton-Raphson
##      Variance of random effect= 5e-07   I-likelihood = -68.6
## Degrees of freedom for terms= 1 0
## Concordance= 0.808 (se = 0.076 )
## Likelihood ratio test= 17.8 on 1 df,  p=2.53e-05
```

Test największej wiarygodności ($p = 2.526e - 05$) wskazuje na to, że mamy podstawy do odrzucenia hipotezy o nieistotności modelu. Zatem model ma sens. Zmienna **treat** jest istotna ($9.3e - 05$). Zatem efekt leczenia w parach jest istotny. Możemy również odczytać, że leczenie przy wzroście o jednostkę (czyli jeśli mamy aktywny lek) zmniejsza ryzyko wznowy 0.18 razy, czyli wydłuża czas do wystąpienia zdarzenia 5,5 raza.

Sprawdźmy jeszcze na koniec, czy spełnione są założenia PH:

```
##           rho chisq      p
## treat -0.0411 0.0492 0.825
```

P-value jest duże, zatem założenie proporcjonalnych hazardów jest spełnione.

Zadanie 3.

Analizujemy dane dotyczące czasów do pojawienia się guzów gruczołów mlekowych u 48 samic szczurów. Dopasuję model proporcjonalnych średnich (AG), biorąc pod uwagę korelację zmiennej **ratid**.

```
## Call:
## coxph(formula = Surv(time1, time2, status) ~ treat + cluster(ratid),
##       data = m)
##
##      n= 252, number of events= 210
##
##           coef exp(coef) se(coef) robust se      z Pr(>|z|)
## treat -0.685      0.504    0.152    0.212 -3.23   0.0012 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## treat    0.504      1.98    0.333    0.764
##
## Concordance= 0.587 (se = 0.02 )
## Rsquare= 0.083 (max possible= 0.998 )
## Likelihood ratio test= 21.9 on 1 df,  p=2.85e-06
## Wald test            = 10.4 on 1 df,  p=0.00123
## Score (logrank) test = 21 on 1 df,  p=4.5e-06,  Robust = 8.03 p=0.00461
##
## (Note: the likelihood ratio and score tests assume independence of
## observations within a cluster, the Wald and robust score tests do not).
```

Z testu ilorazu wiarygodności ($p = 2.847e - 06$) wynika, że model ma sens. Zmienna **treat** jest istotna. Zmniejsza ona ryzyko o około połowę, czyli czas do wznowy zwiększa się 1.98 raza.

Sprawdźmy jeszcze (testem skalowanych reszt Schoenfelda), czy są spełnione założenia PH:

```
##          rho chisq      p
## treat -0.038 0.604 0.437
```

P-value jest duże, zatem nie mamy podstaw do odrzucenia hipotezy o spełnionych założeniach proporcjonalnych hazardów i mogliśmy użyć powyższego modelu.

Zadanie 4.

Będziemy analizować dane dla 865 pacjentów chorych na ziarnicę złośliwą (chorobę Hodgkina), którzy byli leczeni radioterapią lub radioterapią łączoną z chemioterapią. Usuwam z danych zmienną `stnum`, która odpowiada za numer pacjenta. Obserwowano czas do pierwszego zdarzenia. Zdarzeniem mógł być nawrót choroby, zgon lub pojawienie się innego pierwotnego nowotworu, czyli mamy tu do czynienia z konkurującym ryzykiem. Interesuje nas efekt leczenia na czas do pojawienia się innego pierwotnego nowotworu. W tym celu najpierw dopasujemy model PH dla funkcji sub-hazardu:

```
## Call:
## coxph(formula = Surv(firsttime, event_cause == 1) ~ cmt, data = m)
##
##      n= 865, number of events= 93
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## cmt 0.59      1.80      0.21 2.81      0.005 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## cmt      1.8      0.554      1.19      2.72
##
## Concordance= 0.593 (se = 0.027 )
## Rsquare= 0.009 (max possible= 0.699 )
## Likelihood ratio test= 7.61 on 1 df,  p=0.00581
## Wald test              = 7.88 on 1 df,  p=0.005
## Score (logrank) test = 8.1 on 1 df,  p=0.00441
```

```
## Call:
## coxph(formula = Surv(firsttime, event_cause == 2) ~ cmt, data = m)
##
##      n= 865, number of events= 386
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## cmt -0.499      0.607      0.125 -4.01      6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## cmt      0.607      1.65      0.475      0.775
##
## Concordance= 0.547 (se = 0.012 )
## Rsquare= 0.02 (max possible= 0.997 )
## Likelihood ratio test= 17.6 on 1 df,  p=2.72e-05
## Wald test              = 16.1 on 1 df,  p=6.03e-05
## Score (logrank) test = 16.4 on 1 df,  p=5.05e-05
```

Zarówno dla modelu z chorobą Hodgkina, jak i dla modelu z innym nowotworem leczenie jest istotne. W pierwszym przypadku przy leczeniu radioterapią i chemioterapią na raz ryzyko zdarzenia maleje 0.6 raza. W drugim przypadku rośnie 1,8 raza! W obu modelach test największej wiarygodności ma małe p-value, więc jest istotny statystycznie, czyli modele mają sens. Sprawdźmy jeszcze założenia modelu hazardów:

```
##      rho chisq      p
## cmt -0.175  2.87 0.0904
##      rho chisq      p
## cmt 0.0127 0.062 0.803
```

Zarówno w pierwszym, jak i drugim modelu nie mamy podstaw sądzić, że założenia nie są spełnione.

Powyższy model nie brał pod uwagę pozostałych zmiennych z modelu. Dopasujmy więc teraz model PH dla funkcji hazardu sub-dystrybuanty, biorąc pod uwagę wszystkie zmienne:

```
(mod.inny <- crr(ftime = firsttime, fstatus = event_cause, cov1 = mac, failcode = 1))

## convergence: TRUE
## coefficients:
##      cmt      male  nodes mediast  age30 clinstg
## 0.7164 -0.3682 -0.5415  0.2912 -0.2128 -0.1678
## standard errors:
## [1] 0.2346 0.2075 0.3949 0.1693 0.2248 0.2550
## two-sided p-values:
##      cmt      male  nodes mediast  age30 clinstg
## 0.0023  0.0760  0.1700  0.0850  0.3400  0.5100
```

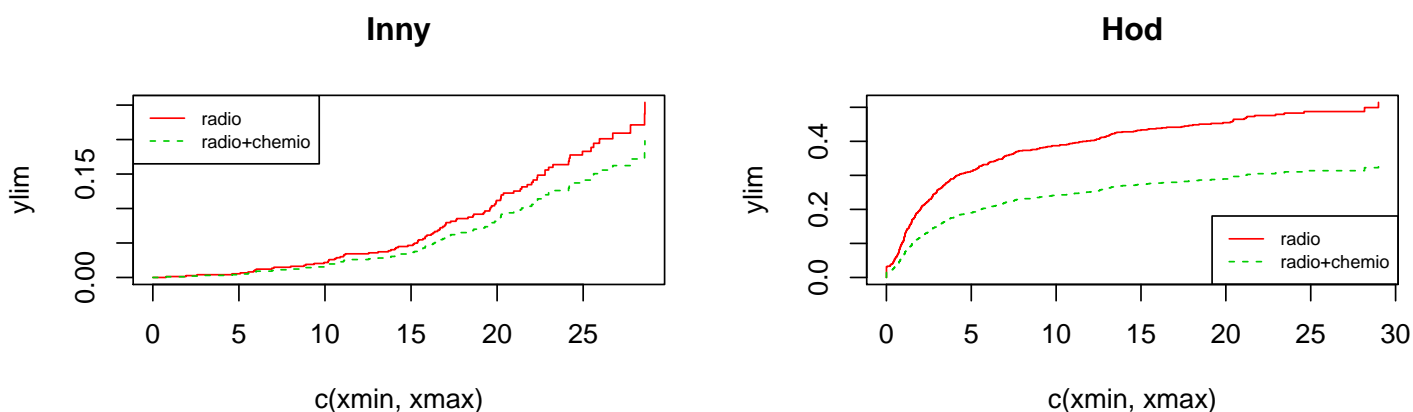
Dla innych nowotworów z testu największej wiarygodności widać, że model jest dobrze dopasowany. Istotny jest tylko efekt leczenia, inne zmienne już nie. Przy czym leczenie chemioterapią i radioterapią jednocześnie zwiększa ryzyko zachorowania dwukrotnie w porównaniu do leczenia samą radioterapią.

```
(mod.hod <- crr(ftime = firsttime, fstatus = event_cause, cov1 = mac, failcode = 2))

## convergence: TRUE
## coefficients:
##      cmt      male  nodes mediast  age30 clinstg
## -0.5335  0.1418  0.3152 -0.2382 -0.4689  0.2095
## standard errors:
## [1] 0.13650 0.10520 0.19770 0.09801 0.10410 0.11550
## two-sided p-values:
##      cmt      male  nodes mediast  age30 clinstg
## 9.3e-05 1.8e-01 1.1e-01 1.5e-02 6.6e-06 7.0e-02
```

Dla choroby Hodgkina z testu największej wiarygodności widać, że model jest dobrze dopasowany. Istotny jest efekt leczenia, ale też zmienna `mediast` i `clinstg`. Przy czym efekt leczenia chemioterapią zmniejsza ryzyko wznowy o połowę. Wyższy stopień zaawansowania nowotworu zwiększa ryzyko 1,2 raza, zaś wzrost stopnia zajęcia śródpiersia zmniejsza (co jest mało intuicyjne) ryzyko wznowy 0,79 raza.

Graficzne oszacowania funkcji skumulowanych częstości dla modelu:



Syntaks R-a

```
# zad.1

library("survival")
b3 <- read.csv("C:\\Users\\Marta\\Desktop\\Marta\\studia\\rok4\\Biostatystyka\\pd3\\bladder3.csv")
mod.moj <- coxph(Surv(time1,time2,status) ~ rx1 + rx2 + rx3 + rx4 + size + number + cluster(patid)+ strata(enum),
                data=b3)
mod.wyk <- coxph(Surv(time1,time2,status) ~ rx + size + number + cluster(patid) + strata(enum),
                data=b3)

mod.moj
zal_mod1 <- cox.zph(mod.moj, transform="identity")
zal_mod1
s <- 2*(mod.moj$loglik[2]-mod.wyk$loglik[2])
p <- 1-pchisq(s,3)

# zad.2

l <- read.csv("C:\\Users\\Marta\\Desktop\\Marta\\studia\\rok4\\Biostatystyka\\pd3\\leukemia.csv")
mod <- coxph(Surv(time,status) ~ treat + frailty(patid)+strata(partial), data=l)
summary(mod)
zz <- cox.zph(mod, transform="identity")
zz

# zad.3

m <- read.csv("C:\\Users\\Marta\\Desktop\\Marta\\studia\\rok4\\Biostatystyka\\pd3\\mammary.csv")
mod <- coxph(Surv(time1,time2,status) ~ treat + cluster(ratid), data=m)
summary(mod)
zz <- cox.zph(mod, transform="identity")
zz

# zad.4

library("cmprsk")
m <- read.csv("C:\\Users\\Marta\\Desktop\\Marta\\studia\\rok4\\Biostatystyka\\pd3\\hodgkin1.csv")
attach(m)
summary(coxph.inny <- coxph(Surv(firsttime, event_cause == 1) ~ cmt, data=m))
summary(coxph.hod <- coxph(Surv(firsttime, event_cause == 2) ~ cmt, data=m))
cox.zph(coxph.inny, transform="identity")
cox.zph(coxph.hod, transform="identity")
mac <- model.matrix(~ cmt + male + nodes + mediast + age30 + clinstg)[,-1]
(mod.inny <- crr(ftime=firsttime, fstatus=event_cause, cov1=mac,
                failcode=1))
(mod.hod <- crr(ftime=firsttime, fstatus=event_cause, cov1=mac,
                failcode=2))

par(mfrow=c(1,2))
innypred <- predict(mod.inny, cov1=rbind(0,1))
plot(innypred,lty=1:2,color=2:3, main="Inny")
legend("topleft",lty=1:2,col=2:3,c("radio","radio+chemio"),cex=0.7)
hodpred <- predict(mod.hod, cov1=rbind(0,1))
plot(hodpred,lty=1:2,color=2:3, main="Hod")
legend("bottomright",lty=1:2,col=2:3,c("radio","radio+chemio"),cex=0.7)
```