

Analiza czasu przeżycia pacjentów z rakiem nerki

Piotr Dzierża

5 października 2020

1 Wstęp

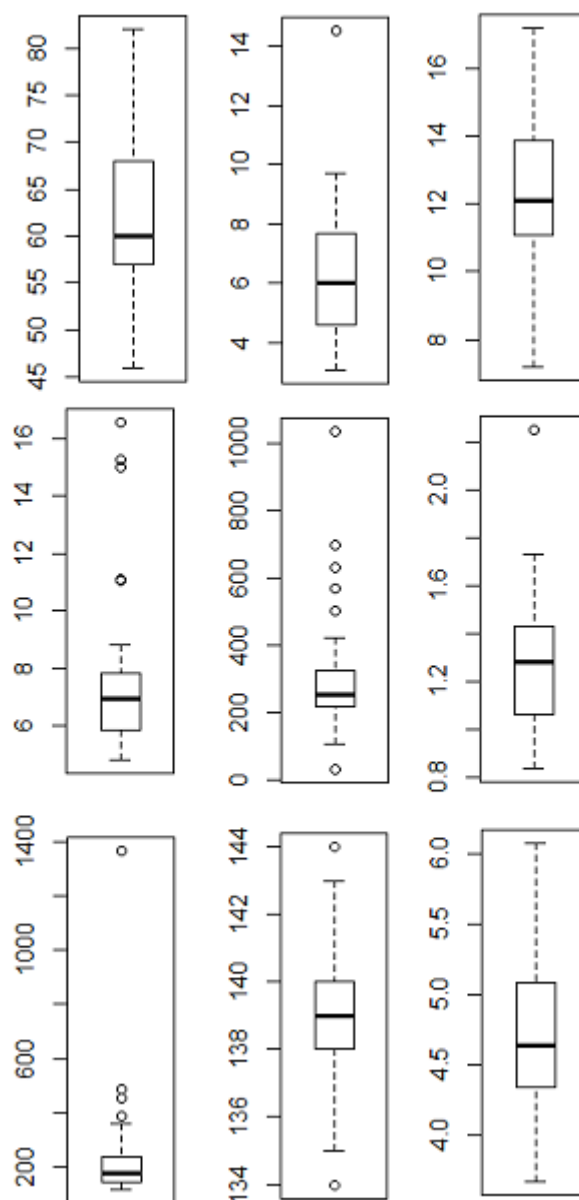
Przeprowadzona zostanie analiza rzeczywistych danych dotyczących czasu przeżycia od wdrożenia eksperymentalnego leczenia raka nerki, które było zastosowane na grupie 33 pacjentów o bardzo złych rokowaniach.

Na początku zrobimy przegląd statystyk podstawowych.

Tabela 1.1: Statystyki podstawowe

Zmienna	Min.	D. Kwartył	Med.	Śr.	G. Kwartył	Maks.	Od.std
Wiek	46.00	57.00	60.00	62.52	68.00	82.00	8.95
Fibrynogen	3.10	4.60	6.00	6.41	7.70	14.50	2.54
Hb	7.20	11.10	12.10	12.18	13.90	17.20	2.21
WBC	4.840	5.870	6.960	7.721	7.830	16.550	2.95
PLT	31.0	215.0	250.0	307.9	324.0	1034.0	193.13
Kreatynina	0.840	1.060	1.280	1.282	1.430	2.250	0.29
LDH	120.0	147.0	176.0	243.1	238.0	1364.0	220.81
Sód	134.0	138.0	139.0	138.9	140.0	144.0	2.54
Potas	3.680	4.340	4.640	4.743	5.090	6.070	0.55

Podczas obliczania wyników wymienionych w tabeli 1.1 konieczne było usunięcie 6 obserwacji dla fibrynogenu, które nie posiadały wartości. Dodatkowo wiemy, że u 5 osób zdiagnozowano cukrzycę, tyle samo osób cierpi na chorobę autoimmunologiczną. Rozkład czynników ryzyka to 7, 12, 9, 5 dla zero, jednego, dwóch i trzech czynników ryzyka odpowiednio.



Rysunek 1.1: Boxploty zmiennych od kolejno: wiek, fibrynogen, Hb, WBC, PLT, kreatynina, LDH, sód, potas

Z rysunku 1.1 możemy odczytać, że fibrynogen, WBC, PLT, kreatynina, LDH i sód posiadają obserwacje odstające.

Sprawdzimy teraz korelacje między zmiennymi. Biorąc pod uwagę wszystkie obserwacje, najmocniej skorelowane są PLT i WBC ze współczynnikiem korelacji Pearsona równym 0.69, WBC i LDH na poziomie 0.67, PLT i potas o wartości 0.56. Patrząc na korelacje z fibrynogenem przy odrzuceniu niepełnych obserwacji najwyższą wartość osiągnęła para z PLT na poziomie równym 0.43. Pozostałe współczynniki są poniżej 0.4. W celu zmniejszenia liczby zmiennych objaśniających, z dalszej analizy usunięte będą: PLT, Fibrynogen, LDH. Pozostawiony zostanie wskaźnik WBC, którego właściwości są najbardziej miarodajne w przypadku tych danych. Odrzucenie fibrynogenu pozwala na pozostawienie wszystkich obserwacji, tylko w tym przypadku pojawiały się dane niekompletne.

2 Model Coxa

W dalszej części pracy spróbujemy dopasować model proporcjonalnych hazardów Coxa. Spójrzmy najpierw na wydruk funkcji *coxph* dla wszystkich zmiennych.

```
n= 33, number of events= 16
```

	coef	exp(coef)	se(coef)	z	Pr(> z)	
wiek	0.09243	1.09684	0.04651	1.987	0.04687	*
Cukrzyca	0.68763	1.98899	1.75448	0.392	0.69511	
Autoim	0.67274	1.95960	1.28722	0.523	0.60123	
Liczba_czynników_ryzykab	2.70008	14.88097	1.11743	2.416	0.01568	*
Liczba_czynników_ryzykac	2.07156	7.93721	1.50867	1.373	0.16972	
Liczba_czynników_ryzykad	5.03817	154.18790	1.78921	2.816	0.00486	**
Hb	0.25410	1.28931	0.30419	0.835	0.40352	
WBC	-0.04299	0.95793	0.24693	-0.174	0.86180	
Kreatynina	-4.12613	0.01615	2.32205	-1.777	0.07558	.
Sód	0.18112	1.19855	0.20568	0.881	0.37856	
Potas	-0.19691	0.82127	1.09282	-0.180	0.85701	

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
wiek	1.09684	0.911709	1.0012783	1.202
Cukrzyca	1.98899	0.502768	0.0638586	61.951
Autoim	1.95960	0.510308	0.1572103	24.426
Liczba_czynników_ryzykab	14.88097	0.067200	1.6652278	132.981
Liczba_czynników_ryzykac	7.93721	0.125989	0.4125614	152.703
Liczba_czynników_ryzykad	154.18790	0.006486	4.6245393	5140.816
Hb	1.28931	0.775611	0.7102788	2.340
WBC	0.95793	1.043923	0.5903924	1.554
Kreatynina	0.01615	61.937519	0.0001704	1.530
Sód	1.19855	0.834339	0.8008993	1.794
Potas	0.82127	1.217633	0.0964437	6.993

Concordance= 0.819 (se = 0.101)
 Rsquare= 0.425 (max possible= 0.872)
 Likelihood ratio test= 18.25 on 11 df, p=0.07594
 Wald test = 12.9 on 11 df, p=0.2999
 Score (logrank) test = 19.42 on 11 df, p=0.05401

Rysunek 2.1: Wydruk funkcji *coxph* dla modelu ze wszystkimi zmiennymi

Patrząc na wyniki testów Likelihood ratio test, Walda i Score przedstawionych na rysunku 2.1, na poziomie istotności 0.05, nie ma podstaw do odrzucenia hipotezy, że wszystkie współczynniki w zadanym modelu równają się zero.

Używając funkcji *step(model)* wyznaczmy najlepszy model według kryterium Akaike'a. Najmniejszą wartość AIC osiągnął model ze zmiennymi Wiek, Autoim, Liczba czynników ryzyka, Kreatynina.

```

              coef exp(coef) se(coef)      z Pr(>|z|)
wiek          0.07617   1.07914  0.03773   2.019  0.04351 *
Autoim        1.29831   3.66310  0.72175   1.799  0.07205 .
Liczba_czynników_ryzykab  2.11800   8.31451  0.91118   2.324  0.02010 *
Liczba_czynników_ryzykac  1.64959   5.20483  1.02856   1.604  0.10876
Liczba_czynników_ryzykad  4.23672  69.18039  1.67865   2.524  0.01161 *
Kreatynina    -3.56854   0.02820  1.24827  -2.859  0.00425 **
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
wiek          1.0791   0.92666   1.002221   1.1620
Autoim        3.6631   0.27299   0.890206  15.0733
Liczba_czynników_ryzykab  8.3145   0.12027   1.393917  49.5948
Liczba_czynników_ryzykac  5.2048   0.19213   0.693260  39.0766
Liczba_czynników_ryzykad 69.1804   0.01445   2.576996 1857.1726
Kreatynina     0.0282  35.46483   0.002442   0.3256

Concordance= 0.808 (se = 0.101 )
Rsquare= 0.397 (max possible= 0.872 )
Likelihood ratio test= 16.72 on 6 df,  p=0.01037
wald test               = 11.89 on 6 df,  p=0.06438
Score (logrank) test = 16.01 on 6 df,  p=0.01368

```

Rysunek 2.2: Wydruk funkcji *coxph* dla modelu z Wiek, Autoim, Liczba czynników ryzyka, Kreatynina

Z rysunku 2.2 można odczytać, że test Walda na poziomie istotności 0.05 orzeka brak podstaw do odrzucenia hipotezy o zerowości wszystkich współczynników. Poza tym współczynniki przy Autoim oraz jednej z liczb czynników ryzyka są zerami według testu na tym samym poziomie co poprzednie. Spróbujemy odrzucić którąś z tych zmiennych i ponownie wykonamy testy.

```

              coef exp(coef) se(coef)      z Pr(>|z|)
wiek          0.07747   1.08055  0.03579   2.165  0.03041 *
Liczba_czynników_ryzykab  1.73238   5.65410  0.82702   2.095  0.03620 *
Liczba_czynników_ryzykac  1.25759   3.51693  0.92789   1.355  0.17531
Liczba_czynników_ryzykad  3.48925  32.76134  1.52558   2.287  0.02219 *
Kreatynina    -3.49089   0.03047  1.17004  -2.984  0.00285 **
---
              exp(coef) exp(-coef) lower .95 upper .95
wiek          1.08055   0.92545   1.007353   1.1591
Liczba_czynników_ryzykab  5.65410   0.17686   1.117902  28.5972
Liczba_czynników_ryzykac  3.51693   0.28434   0.570616  21.6762
Liczba_czynników_ryzykad 32.76134   0.03052   1.647356 651.5322
Kreatynina     0.03047  32.81512   0.003076   0.3019

Concordance= 0.83 (se = 0.101 )
Rsquare= 0.34 (max possible= 0.872 )
Likelihood ratio test= 13.72 on 5 df,  p=0.01747
wald test               = 11.32 on 5 df,  p=0.04531
Score (logrank) test = 13.64 on 5 df,  p=0.01804

```

Rysunek 2.3: Wydruk funkcji *coxph* dla modelu bez Autoim

Z rysunku 2.3 można odczytać, że na poziomie $\alpha = 0.05$ powinno się odrzucić hipotezę o zerowości wszystkich współczynników. Jednakże, dla czynników ryzyka osiągających wartość c , wciąż powinno się przyjmować, że współczynnik przy tej zmiennej równa się 0.

```

              coef exp(coef) se(coef)      z Pr(>|z|)
wiek          0.07136   1.07397  0.03832   1.862   0.0626 .
Kreatynina  -1.84817   0.15753  0.91905  -2.011   0.0443 *
Autoim       0.78172   2.18523  0.65224   1.199   0.2307
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
wiek          1.0740     0.9311    0.9963    1.1577
Kreatynina    0.1575     6.3482    0.0260    0.9542
Autoim        2.1852     0.4576    0.6086    7.8466

Concordance= 0.725 (se = 0.101 )
Rsquare= 0.232 (max possible= 0.872 )
Likelihood ratio test= 8.72 on 3 df,  p=0.03332
wald test               = 8.09 on 3 df,  p=0.04425
Score (logrank) test = 9.34 on 3 df,  p=0.02511

```

Rysunek 2.4: Wydruk funkcji *coxph* dla modelu bez liczby czynników ryzyka

Z rysunku 2.4 można odczytać, że ten model również ma wadę poprzedników. Dla mode-

```

              coef exp(coef) se(coef)      z Pr(>|z|)
wiek          0.07663   1.07965  0.03713   2.064   0.0390 *
Kreatynina  -2.04674   0.12915  0.88799  -2.305   0.0212 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
wiek          1.0796     0.9262    1.00386    1.1612
Kreatynina    0.1292     7.7426    0.02266    0.7362

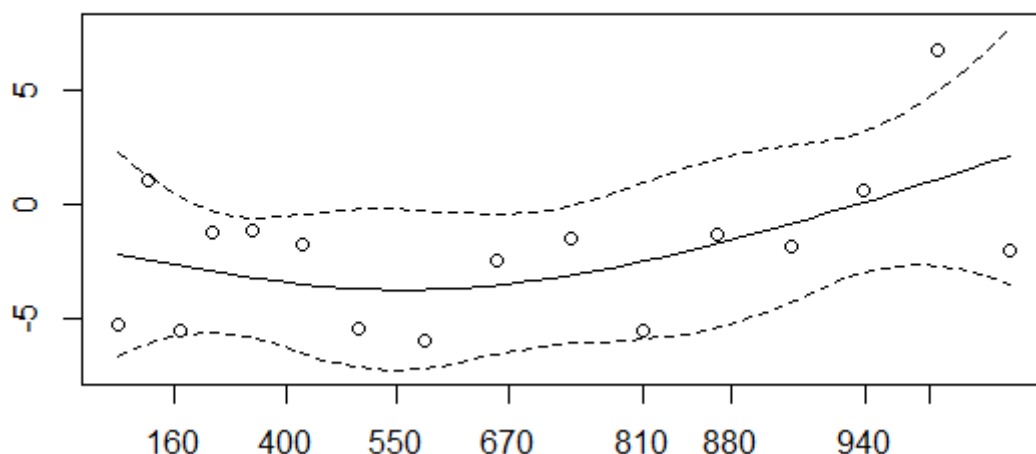
Concordance= 0.725 (se = 0.101 )
Rsquare= 0.201 (max possible= 0.872 )
Likelihood ratio test= 7.39 on 2 df,  p=0.02488
wald test               = 6.9 on 2 df,  p=0.03182
Score (logrank) test = 7.4 on 2 df,  p=0.02475

```

Rysunek 2.5: Wydruk funkcji *coxph* dla modelu bez liczby czynników ryzyka i Autoim

lu przedstawionego na rysunku 2.5 wszystkie testy na poziomie 0.05 wykazały niezerowanie współczynników. Model ten jest jednak słabszy pod względem kryterium Akaike'a od tego opisanego rysunkiem 2.3 oraz oczywiście tego z rysunku 2.2. Jednakże nie ma on w sobie zerowych współczynników bez względu na wartości czynników ryzyka i to on będzie poddawany dalszej analizie.

Sprawdzimy teraz założenie proporcjonalności hazardów tego modelu. Użyjemy do tego funkcji *cox.zph*.

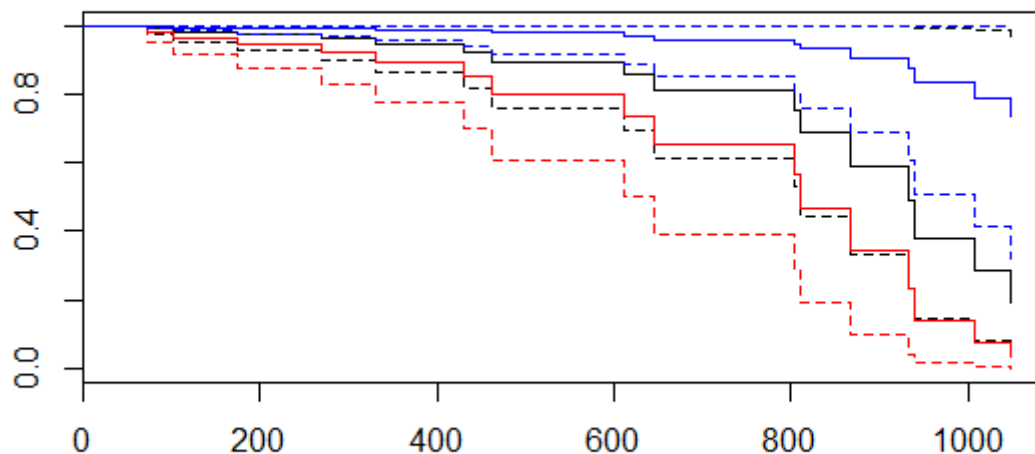


Rysunek 2.6: Wykres funkcji *cox.zph* z resztami Schoenfelda i transformacją "K-M"

Tylko jedna reszta na rysunku 2.6 nieznacznie wychodzi poza zakres przerywanych linii, co oznacza, że można przyjąć, że założenie proporcjonalności hazardów jest spełnione.

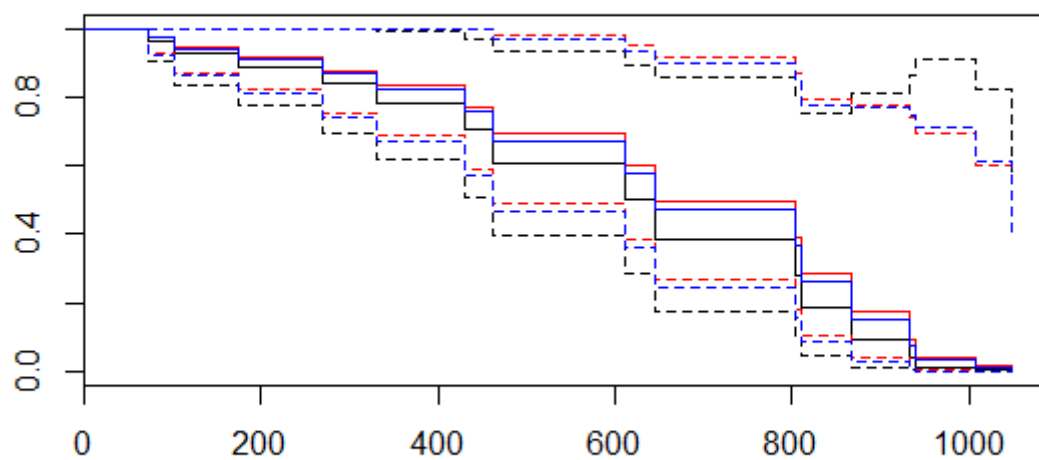
Przyglądając się współczynnikom z rysunku 2.5 można powiedzieć, że czas przeżycia rośnie wraz z wiekiem oraz maleje przy wzroście kreatyniny. Przy wzroście wieku o jeden, hazard należy przemnożyć przez 1.07965, a przy wzroście o jednostkę kreatyniny hazard należy przemnożyć przez 0.12915.

Spójrzmy teraz na 3 przykładowe wykresy przeżycia.



Rysunek 2.7: Przykładowe wykresy przeżycia

Wykres niebieski odnosi się do osoby dwudziestoletniej o kreatynie na poziomie 1.05, czarny do osoby w wieku 50 lat i kreatynie 1.35 oraz czerwony dla 70 lat i 1.75 kreatyniny. Widzimy, że funkcje zachowują się według przewidywań, osoby o mniejszej kreatynie mają większą funkcję przeżycia. Spójrzmy na jeszcze jedno porównanie.



Rysunek 2.8: Przykładowe wykresy przeżycia dla tych samych poziomów kreatyniny

Tutaj czarny kolor odpowiada osobie w wieku 62 lat, niebieski osobie 59-letniej, a czerwony 58-letniej. Wszystkie te osoby figurują w danych, mają ten sam poziom kreatyniny równy 1.06. Im osoba młodsza, tym większa funkcja przeżycia.