

Analiza przeżycia
Raport 5

-

Modele proporcjonalnych szans

Romana Żmuda

8 lutego 2021

Spis treści

1	Zadanie do sprawozdania - Część 1	3
1.1	Zadanie 1 - Metoda proporcjonalnych szans - wszystkie zmienne	3
1.2	Zadanie 2 - Metoda proporcjonalnych szans bez zmiennych <i>age</i> i <i>meal.cal</i>	8
1.3	Zadanie 3, 4 - Weryfikacje hipotez do podanych wyżej modeli	9
1.4	Zadanie 5 - Wybór zmiennych do modelu	10
1.5	Zadanie 6 - funkcja przeżycia modelu z kryterium AIC	16
1.6	Zadanie 7 - hipoteza o proporcjonalności szans	17

1 Zadanie do sprawozdania - Część 1

1.1 Zadanie 1 - Metoda proporcjonalnych szans - wszystkie zmienne

W tej części znowu będziemy tworzyć semiparametryczne modele regresji, a dokładniej model proporcjonalnych szans. Model ten jest często przyjmowanym modelem regresji w przypadku, gdy założenie proporcjonalności hazardów nie jest spełnione. W statystyce szansą nazywamy iloraz prawdopodobieństwa sukcesu do prawdopodobieństwa porażki, należy jednak pamiętać, że u nas to prawdopodobieństwo ma interpretację w postaci: "Szansa na śmierć mając ten parametr wynosi ..."

Tak jak w poprzednim raporcie analizie poddamy zbiór danych *lung*, który dotyczy pacjentów z zaawansowanym rakiem płuc. . Zbiór zawiera informacje o 228 pacjentach, których zbiór charakterystyk obejmuje 8 następujących zmiennych:

- *inst* kod instytucji
- *time* czas przeżycia
- *status* cenzura (1. cenzura, 2. śmierć)
- *age* wiek
- *sex* płeć (1. mężczyzna, 2. kobieta)
- *ph.ecog* skala sprawności wg. lekarza (0-sprawność prawidłowa, 5-zgon)
- *ph.karno* skala sprawności wg. lekarza (sprawność prawidłowa - 100. zgon - 0)
- *pat.karno* skala sprawności wg. pacjenta
- *meal.cal* kalorie na posiłek
- *wt.loss* utrata masy ciała w ciągu ostatnich 6 miesięcy

Do modelu musimy założyć, że obserwowalne zmienne losowe mają rozkłady o ciągłych i różniczkowalnych dystrybuantach, a niektóre zmienne muszą być zmiennymi factor, należy też dokonać centralizacji zmiennych.

```
> dane<-data.frame(lung)
> dane$status <- as.factor(dane$status)
> dane$ph.ecog <-as.factor(dane$ph.ecog)
> dane$ph.karno <- as.factor(dane$ph.karno)
> dane$sex <- as.factor(dane$sex)
> dane$pat.karno <- as.factor(dane$pat.karno)
> dane<-subset(dane, (meal.cal != "NA") & (wt.loss != "NA"))
> dane$age <- dane$age - mean(dane$age)
> dane$meal.cal <- dane$meal.cal - mean(dane$meal.cal)
> dane$wt.loss <- dane$wt.loss - mean(dane$wt.loss)
> model<-prop.odds(Event(time, status == 2) ~ age + sex + ph.ecog + ph.karno
+                               + pat.karno + meal.cal + wt.loss,
+                               data = dane, n.sim = 500, profile = 1)
> model
```

Proportional Odds model

Did not converge, allow more iterations

Test for baseline

Test for nonparametric terms

Test for non-significant effects

Supremum-test of significance p-value $H_0: B(t)=0$

Baseline	6.27	0
----------	------	---

Test for time invariant effects

Kolmogorov-Smirnov test p-value $H_0: \text{constant effect}$

Baseline	1.72	0.086
----------	------	-------

Covariate effects

	Coef.	SE	Robust SE	$D^2 \log(L)^{-1}$	z	P-val	lower2.5%	upper97.5%
age	5.87e-03	0	0	0	Inf	0	5.87e-03	5.87e-03
sex2	-6.08e-01	0	0	0	-Inf	0	-6.08e-01	-6.08e-01
ph.ecog1	6.40e-01	0	0	0	Inf	0	6.40e-01	6.40e-01
ph.ecog2	1.32e+00	0	0	0	Inf	0	1.32e+00	1.32e+00
ph.ecog3	2.55e+00	0	0	0	Inf	0	2.55e+00	2.55e+00
ph.karno60	1.03e+00	0	0	0	Inf	0	1.03e+00	1.03e+00
ph.karno70	1.00e+00	0	0	0	Inf	0	1.00e+00	1.00e+00
ph.karno80	1.17e+00	0	0	0	Inf	0	1.17e+00	1.17e+00
ph.karno90	1.31e+00	0	0	0	Inf	0	1.31e+00	1.31e+00
ph.karno100	1.46e+00	0	0	0	Inf	0	1.46e+00	1.46e+00
pat.karno40	-3.52e-01	0	0	0	-Inf	0	-3.52e-01	-3.52e-01
pat.karno50	7.53e-01	0	0	0	Inf	0	7.53e-01	7.53e-01
pat.karno60	1.23e-01	0	0	0	Inf	0	1.23e-01	1.23e-01
pat.karno70	-1.74e-01	0	0	0	-Inf	0	-1.74e-01	-1.74e-01
pat.karno80	-2.81e-01	0	0	0	-Inf	0	-2.81e-01	-2.81e-01
pat.karno90	-6.89e-02	0	0	0	-Inf	0	-6.89e-02	-6.89e-02
pat.karno100	-5.68e-01	0	0	0	-Inf	0	-5.68e-01	-5.68e-01
meal.cal	-4.43e-05	0	0	0	-Inf	0	-4.43e-05	-4.43e-05
wt.loss	-1.39e-02	0	0	0	-Inf	0	-1.39e-02	-1.39e-02

Test of Goodness-of-fit

	sup hat U(t)	p-value H_0
age	86.400	0.308
sex2	3.870	0.432
ph.ecog1	4.780	0.288
ph.ecog2	4.880	0.192
ph.ecog3	0.531	0.330
ph.karno60	1.970	0.602
ph.karno70	3.350	0.340
ph.karno80	2.870	0.622
ph.karno90	2.150	0.852
ph.karno100	3.170	0.320

pat.karno40	0.777	0.580
pat.karno50	0.956	0.580
pat.karno60	4.250	0.136
pat.karno70	1.810	0.854
pat.karno80	2.370	0.706
pat.karno90	3.070	0.550
pat.karno100	2.520	0.524
meal.cal	3570.000	0.378
wt.loss	78.600	0.742

Widzimy, że model nie mógł się skonwertować, gdyż podano do modelu za mało iteracji. Sprawdźmy, co się stanie gdy zwiększymy ilość iteracji.

```
> model_1<-prop.odds(Event(time, status == 2) ~ age + sex + ph.ecog + ph.karno
+
+ pat.karno + meal.cal + wt.loss,
+
+ data = dane, n.sim = 5000, profile = 1)
> model_1
```

Proportional Odds model

Did not converge, allow more iterations

Test for baseline

Test for nonparametric terms

Test for non-significant effects

Supremum-test of significance p-value $H_0: B(t)=0$

Baseline	6.27	0
----------	------	---

Test for time invariant effects

Kolmogorov-Smirnov test p-value H_0 :constant effect

Baseline	1.72	0.107
----------	------	-------

Covariate effects

	Coef.	SE Robust	SE	D2log(L) ⁻¹	z	P-val	lower2.5%	upper97.5%
age	5.87e-03	0	0	0	Inf	0	5.87e-03	5.87e-03
sex2	-6.08e-01	0	0	0	-Inf	0	-6.08e-01	-6.08e-01
ph.ecog1	6.40e-01	0	0	0	Inf	0	6.40e-01	6.40e-01
ph.ecog2	1.32e+00	0	0	0	Inf	0	1.32e+00	1.32e+00
ph.ecog3	2.55e+00	0	0	0	Inf	0	2.55e+00	2.55e+00
ph.karno60	1.03e+00	0	0	0	Inf	0	1.03e+00	1.03e+00
ph.karno70	1.00e+00	0	0	0	Inf	0	1.00e+00	1.00e+00
ph.karno80	1.17e+00	0	0	0	Inf	0	1.17e+00	1.17e+00
ph.karno90	1.31e+00	0	0	0	Inf	0	1.31e+00	1.31e+00
ph.karno100	1.46e+00	0	0	0	Inf	0	1.46e+00	1.46e+00
pat.karno40	-3.52e-01	0	0	0	-Inf	0	-3.52e-01	-3.52e-01
pat.karno50	7.53e-01	0	0	0	Inf	0	7.53e-01	7.53e-01
pat.karno60	1.23e-01	0	0	0	Inf	0	1.23e-01	1.23e-01
pat.karno70	-1.74e-01	0	0	0	-Inf	0	-1.74e-01	-1.74e-01

pat.karno80	-2.81e-01	0	0	0 -Inf	0 -2.81e-01	-2.81e-01
pat.karno90	-6.89e-02	0	0	0 -Inf	0 -6.89e-02	-6.89e-02
pat.karno100	-5.68e-01	0	0	0 -Inf	0 -5.68e-01	-5.68e-01
meal.cal	-4.43e-05	0	0	0 -Inf	0 -4.43e-05	-4.43e-05
wt.loss	-1.39e-02	0	0	0 -Inf	0 -1.39e-02	-1.39e-02

Test of Goodness-of-fit

	sup hat U(t)	p-value H_0
age	86.400	0.272
sex2	3.870	0.426
ph.ecog1	4.780	0.299
ph.ecog2	4.880	0.176
ph.ecog3	0.531	0.326
ph.karno60	1.970	0.604
ph.karno70	3.350	0.360
ph.karno80	2.870	0.619
ph.karno90	2.150	0.846
ph.karno100	3.170	0.328
pat.karno40	0.777	0.555
pat.karno50	0.956	0.566
pat.karno60	4.250	0.159
pat.karno70	1.810	0.862
pat.karno80	2.370	0.718
pat.karno90	3.070	0.543
pat.karno100	2.520	0.503
meal.cal	3570.000	0.384
wt.loss	78.600	0.739

```
> model_2<-prop.odds(Event(time, status == 2) ~ age + sex + ph.ecog + ph.karno
+
+ pat.karno + meal.cal + wt.loss,
+
+ data = dane, n.sim = 10000, profile = 1)
> model_2
```

Proportional Odds model

Did not converge, allow more iterations

Test for baseline

Test for nonparametric terms

Test for non-significant effects

Supremum-test of significance p-value H_0: B(t)=0		
Baseline	6.27	0

Test for time invariant effects

Kolmogorov-Smirnov test p-value H_0:constant effect		
Baseline	1.72	0.106

Covariate effects

Coef.	SE Robust	SE D2log(L)^-1	z	P-val	lower2.5%	upper97.5%
-------	-----------	----------------	---	-------	-----------	------------

age	5.87e-03	0	0	0	Inf	0	5.87e-03	5.87e-03
sex2	-6.08e-01	0	0	0	-Inf	0	-6.08e-01	-6.08e-01
ph.ecog1	6.40e-01	0	0	0	Inf	0	6.40e-01	6.40e-01
ph.ecog2	1.32e+00	0	0	0	Inf	0	1.32e+00	1.32e+00
ph.ecog3	2.55e+00	0	0	0	Inf	0	2.55e+00	2.55e+00
ph.karno60	1.03e+00	0	0	0	Inf	0	1.03e+00	1.03e+00
ph.karno70	1.00e+00	0	0	0	Inf	0	1.00e+00	1.00e+00
ph.karno80	1.17e+00	0	0	0	Inf	0	1.17e+00	1.17e+00
ph.karno90	1.31e+00	0	0	0	Inf	0	1.31e+00	1.31e+00
ph.karno100	1.46e+00	0	0	0	Inf	0	1.46e+00	1.46e+00
pat.karno40	-3.52e-01	0	0	0	-Inf	0	-3.52e-01	-3.52e-01
pat.karno50	7.53e-01	0	0	0	Inf	0	7.53e-01	7.53e-01
pat.karno60	1.23e-01	0	0	0	Inf	0	1.23e-01	1.23e-01
pat.karno70	-1.74e-01	0	0	0	-Inf	0	-1.74e-01	-1.74e-01
pat.karno80	-2.81e-01	0	0	0	-Inf	0	-2.81e-01	-2.81e-01
pat.karno90	-6.89e-02	0	0	0	-Inf	0	-6.89e-02	-6.89e-02
pat.karno100	-5.68e-01	0	0	0	-Inf	0	-5.68e-01	-5.68e-01
meal.cal	-4.43e-05	0	0	0	-Inf	0	-4.43e-05	-4.43e-05
wt.loss	-1.39e-02	0	0	0	-Inf	0	-1.39e-02	-1.39e-02

Test of Goodness-of-fit

	sup hat U(t)	p-value H_0
age	86.400	0.281
sex2	3.870	0.435
ph.ecog1	4.780	0.294
ph.ecog2	4.880	0.186
ph.ecog3	0.531	0.324
ph.karno60	1.970	0.624
ph.karno70	3.350	0.352
ph.karno80	2.870	0.614
ph.karno90	2.150	0.853
ph.karno100	3.170	0.321
pat.karno40	0.777	0.558
pat.karno50	0.956	0.575
pat.karno60	4.250	0.158
pat.karno70	1.810	0.860
pat.karno80	2.370	0.724
pat.karno90	3.070	0.549
pat.karno100	2.520	0.505
meal.cal	3570.000	0.389
wt.loss	78.600	0.722

>

Widzimy, że nie ma możliwości dopasowania zmiennych do modelu, gdyż nawet w dużej ilości iteracji model nie konwertuje się.

1.2 Zadanie 2 - Metoda proporcjonalnych szans bez zmiennych *age* i *meal.cal*

Model budujemy z pomocą funkcji *prop.odds*, ale tym razem bez zmiennych *age* i *meal.cal*:

```
> model_bez<-prop.odds(Event(time, status == 2) ~ sex + ph.ecog
+                        + ph.karno + pat.karno + wt.loss,
+                        data = dane, n.sim = 500, profile = 1)
> model_bez
```

Proportional Odds model

Test for baseline

Test for nonparametric terms

Test for non-significant effects

Supremum-test of significance p-value $H_0: B(t)=0$

Baseline	0.397	0.884
----------	-------	-------

Test for time invariant effects

Kolmogorov-Smirnov test p-value H_0 :constant effect

Baseline	0.997	0.736
----------	-------	-------

Covariate effects

	Coef.	SE Robust	SE	D2log(L)^-1	z	P-val	lower2.5%
sex2	-1.0400	0.3460	0.3560	0.3250	-2.910	0.00367	-1.7200
ph.ecog1	0.6060	0.5090	0.5180	0.5300	1.170	0.24300	-0.3920
ph.ecog2	1.7300	0.8540	0.8910	0.8330	1.940	0.05190	0.0562
ph.ecog3	2.7100	1.4700	1.1500	1.6900	2.350	0.01890	-0.1710
ph.karno60	1.3100	1.6800	2.3200	1.3200	0.563	0.57400	-1.9800
ph.karno70	1.3900	1.6300	2.3200	1.3000	0.598	0.55000	-1.8000
ph.karno80	1.4700	1.6400	2.4100	1.3600	0.613	0.54000	-1.7400
ph.karno90	1.6300	1.6500	2.4200	1.3800	0.675	0.50000	-1.6000
ph.karno100	1.5400	1.7300	2.4700	1.4800	0.623	0.53300	-1.8500
pat.karno40	0.6430	2.5200	3.3300	2.2200	0.193	0.84700	-4.3000
pat.karno50	1.6100	1.4200	1.1900	1.6000	1.350	0.17600	-1.1700
pat.karno60	0.9570	1.0400	0.7590	1.3100	1.260	0.20700	-1.0800
pat.karno70	0.1190	1.1100	0.7840	1.3300	0.151	0.88000	-2.0600
pat.karno80	0.1140	1.1100	0.8010	1.3300	0.143	0.88600	-2.0600
pat.karno90	0.4550	1.1300	0.8330	1.3400	0.546	0.58500	-1.7600
pat.karno100	-0.3440	1.2300	0.9580	1.3800	-0.359	0.72000	-2.7500
wt.loss	-0.0114	0.0123	0.0121	0.0115	-0.937	0.34900	-0.0355
upper97.5%							
sex2	-0.3620						
ph.ecog1	1.6000						
ph.ecog2	3.4000						
ph.ecog3	5.5900						
ph.karno60	4.6000						
ph.karno70	4.5800						

ph.karno80	4.6800
ph.karno90	4.8600
ph.karno100	4.9300
pat.karno40	5.5800
pat.karno50	4.3900
pat.karno60	3.0000
pat.karno70	2.2900
pat.karno80	2.2900
pat.karno90	2.6700
pat.karno100	2.0700
wt.loss	0.0127
Test of Goodness-of-fit	
	sup hat U(t) p-value H_0
sex2	2.230 0.654
ph.ecog1	3.960 0.100
ph.ecog2	1.460 0.932
ph.ecog3	0.508 0.310
ph.karno60	1.470 0.640
ph.karno70	2.160 0.400
ph.karno80	2.480 0.404
ph.karno90	1.020 1.000
ph.karno100	1.710 0.488
pat.karno40	0.651 0.440
pat.karno50	0.583 0.552
pat.karno60	2.350 0.338
pat.karno70	2.300 0.342
pat.karno80	1.700 0.792
pat.karno90	2.000 0.674
pat.karno100	2.350 0.188
wt.loss	78.300 0.434

Tutaj model konwertuje się, jednak jest on nieracjonalny, gdyż niektóre zmienne charakterystyczne informujące o hipotetycznie gorszej sytuacji pacjenta zmniejszają jego szansę na śmierć, a więc na przeżycie. Bierze się to z faktu braku danych.

1.3 Zadanie 3, 4 - Weryfikacje hipotez do podanych wyżej modeli

Jako, że nie udało się stworzyć odpowiednich modeli, nie jesteśmy w stanie przeprowadzić weryfikacji hipotez podanych w zadaniach.

1.4 Zadanie 5 - Wybór zmiennych do modelu

Kryterium informacyjne AIC

Kryterium informacyjne AIC jest metodą forward, w skrócie od jednej zmiennej poprzez dodawanie kolejnych zmiennych, aż do momentu, gdy wartość AIC dla modelu jest mniejsza niż po dodaniu jakiegokolwiek kolejnej. Najpierw tworzymy krok pierwszy czyli weryfikacja, która zmienna musi być w modelu, mamy ich 7.

Krok 1

```
> model1 <- prop.odds(Event(time, status == 2) ~ age,
+                      data = dane, n.sim = 500, profile = 1)
> model2 <- prop.odds(Event(time, status == 2) ~ sex,
+                      data = dane, n.sim = 500, profile = 1)
> model3 <- prop.odds(Event(time, status == 2) ~ ph.ecog,
+                      data = dane, n.sim = 500, profile = 1)
> model4 <- prop.odds(Event(time, status == 2) ~ ph.karno,
+                      data = dane, n.sim = 500, profile = 1)
> model5 <- prop.odds(Event(time, status == 2) ~ pat.karno,
+                      data = dane, n.sim = 500, profile = 1)
> model6 <- prop.odds(Event(time, status == 2) ~ meal.cal,
+                      data = dane, n.sim = 500, profile = 1)
> model7 <- prop.odds(Event(time, status == 2) ~ wt.loss,
+                      data = dane, n.sim = 500, profile = 1)
> AIC_model1 <- -2*model1$loglike[1] + 2*1
> AIC_model2 <- -2*model2$loglike[1] + 2*2
> AIC_model3 <- -2*model3$loglike[1] + 2*4
> AIC_model4 <- -2*model4$loglike[1] + 2*6
> AIC_model5 <- -2*model5$loglike[1] + 2*8
> AIC_model6 <- -2*model6$loglike[1] + 2*1
> AIC_model7 <- -2*model7$loglike[1] + 2*1
```

Zmienna	Wartość AIC
age	1053.493
sex	1052.954
ph.ecog	1038.088
ph.karno	1055.103
pat.karno	1032.985
meal.cal	1057.064
wt.loss	1058.472

Najmniejsza wartość jest w przypadku *pat.karno* więc to ją bierzemy do dalszej analizy

Krok 2

```
> model51 <- prop.odds(Event(time, status == 2) ~ pat.karno + age,
+                       data = dane, n.sim = 500, profile = 1)
> model52 <- prop.odds(Event(time, status == 2) ~ pat.karno + sex,
+                       data = dane, n.sim = 500, profile = 1)
> model53 <- prop.odds(Event(time, status == 2) ~ pat.karno + ph.ecog,
+                       data = dane, n.sim = 500, profile = 1)
> model54 <- prop.odds(Event(time, status == 2) ~ pat.karno + ph.karno,
+                       data = dane, n.sim = 500, profile = 1)
> model56 <- prop.odds(Event(time, status == 2) ~ pat.karno + meal.cal,
+                       data = dane, n.sim = 500, profile = 1)
> model57 <- prop.odds(Event(time, status == 2) ~ pat.karno + wt.loss,
+                       data = dane, n.sim = 500, profile = 1)
> AIC_model51 <- -2*model51$loglike[1] + 2*9
> AIC_model52 <- -2*model52$loglike[1] + 2*10
> AIC_model53 <- -2*model53$loglike[1] + 2*12
> AIC_model54 <- -2*model54$loglike[1] + 2*14
> AIC_model56 <- -2*model56$loglike[1] + 2*9
> AIC_model57 <- -2*model57$loglike[1] + 2*9
```

Zmienna	Wartość AIC
age + pat.karno	1033.786
sex + pat.karno	1029.641
ph.ecog + pat.karno	1024.591
ph.karno + pat.karno	1040.814
meal.cal + pat.karno	1038.571
wt.loss + pat.karno	1034.978

Najmniejsza wartość jest dla zmiennych *ph.ecog* oraz *pat.karno*. Sprawdzamy, czy trzeba dodać jeszcze jakąś zmienną.

Krok 3

```
> model531 <- prop.odds(Event(time, status == 2) ~ pat.karno
+                       + ph.ecog + age, data = dane,
+                       n.sim = 500, profile = 1)
> model532 <- prop.odds(Event(time, status == 2) ~ pat.karno
+                       + ph.ecog + sex, data = dane,
+                       n.sim = 500, profile = 1)
> model534 <- prop.odds(Event(time, status == 2) ~ pat.karno
+                       + ph.ecog + ph.karno, data = dane,
+                       n.sim = 500, profile = 1)
> model536 <- prop.odds(Event(time, status == 2) ~ pat.karno
+                       + ph.ecog + meal.cal, data = dane,
+                       n.sim = 500, profile = 1)
> model537 <- prop.odds(Event(time, status == 2) ~ pat.karno
```

```

+           + ph.ecog + wt.loss, data = dane,
+           n.sim = 500, profile = 1)
> AIC_model531 <- -2*model531$loglike[1] + 2*13
> AIC_model532 <- -2*model532$loglike[1] + 2*14
> AIC_model534 <- -2*model534$loglike[1] + 2*19
> AIC_model536 <- -2*model536$loglike[1] + 2*13
> AIC_model537 <- -2*model537$loglike[1] + 2*13

```

Zmienna	Wartość AIC
age + pat.karno + ph.ecog	1026.188
sex + pat.karno + ph.ecog	1018.58
ph.karno + pat.karno + ph.ecog	1037.034
meal.cal + pat.karno + ph.ecog	1031.709
wt.loss + pat.karno + ph.ecog	1026.444

Widzimy, że do modelu dodajemy jeszcze zmienną sex.

Krok 4

```

> model5321 <- prop.odds(Event(time, status == 2) ~ pat.karno
+           + ph.ecog + sex + age, data = dane,
+           n.sim = 500, profile = 1)
> model5324 <- prop.odds(Event(time, status == 2) ~ pat.karno
+           + ph.ecog + sex + ph.karno, data = dane,
+           n.sim = 500, profile = 1)
> model5326 <- prop.odds(Event(time, status == 2) ~ pat.karno
+           + ph.ecog + sex + meal.cal, data = dane,
+           n.sim = 500, profile = 1)
> model5327 <- prop.odds(Event(time, status == 2) ~ pat.karno
+           + ph.ecog + sex + wt.loss, data = dane,
+           n.sim = 500, profile = 1)
> AIC_model5321 <- -2*model5321$loglike[1] + 2*15
> AIC_model5324 <- -2*model5324$loglike[1] + 2*21
> AIC_model5326 <- -2*model5326$loglike[1] + 2*15
> AIC_model5327 <- -2*model5327$loglike[1] + 2*15

```

Zmienna	Wartość AIC
age + pat.karno + ph.ecog + sex	1020.492
ph.karno + pat.karno + ph.ecog + sex	1031.4
meal.cal + pat.karno + ph.ecog + sex	1028.326
wt.loss + pat.karno + ph.ecog + sex	1019.838

Najmniejszą wartością jest dodając zmienną *wt.loss*, jednak przypomnijmy sobie ile wynosiła wartość AIC, bez tej zmiennej:

Widzimy, że bez tej zmiennej model jest lepszy ostateczny model zbudujemy na:

- sex

	Zmienna	Wartość AIC
	pat.karno + ph.ecog + sex	1018.58
	pat.karno + ph.ecog + sex + wt.loss	1019.838

- *pat.karno*
- *ph.ecog*

Kryterium BIC

Kryterium BIC jest metodą podobną do kryterium AIC, w skrócie od jednej zmiennej poprzez dodawanie kolejnych zmiennych, aż do momentu, gdy wartość BIC dla modelu jest mniejsza niż po dodaniu jakiegokolwiek kolejnej. Najpierw tworzymy krok pierwszy czyli weryfikacja, która zmienna musi być w modelu, mamy ich 7.

Krok 1

```
> model1 <- prop.odds(Event(time, status==2)~age,
+                      data=dane, n.sim=500, profile=1)
> model2 <- prop.odds(Event(time, status==2)~sex,
+                      data=dane, n.sim=500, profile=1)
> model3 <- prop.odds(Event(time, status==2)~ph.ecog,
+                      data=dane, n.sim=500, profile=1)
> model4 <- prop.odds(Event(time, status==2)~ph.karno,
+                      data=dane, n.sim=500, profile=1)
> model5 <- prop.odds(Event(time, status==2)~pat.karno,
+                      data=dane, n.sim=500, profile=1)
> model6 <- prop.odds(Event(time, status==2)~wt.loss,
+                      data=dane, n.sim=500, profile=1)
> model7 <- prop.odds(Event(time, status==2)~meal.cal,
+                      data=dane, n.sim=500, profile=1)
> n <- nrow(dane)
> BIC_model1 <- -2*model1$loglike[1] + log(n)*1
> BIC_model2 <- -2*model2$loglike[1] + log(n)*2
> BIC_model3 <- -2*model3$loglike[1] + log(n)*4
> BIC_model4 <- -2*model4$loglike[1] + log(n)*6
> BIC_model5 <- -2*model5$loglike[1] + log(n)*8
> BIC_model6 <- -2*model6$loglike[1] + log(n)*1
> BIC_model7 <- -2*model7$loglike[1] + log(n)*1
>
```

Najmniejsza wartość jest w przypadku *ph.ecog* więc to ją bierzemy do dalszej analizy

Zmienna	Wartość BIC
age	1050.685
sex	1059.237
ph.ecog	1045.543
ph.karno	1073.924
pat.karno	1058.133
wt.loss	1061.614
meal.cal	1060.208

Krok 2

```
> model31 <- prop.odds(Event(time, status==2)~ph.ecog + age,
+                        data=dane, n.sim=500, profile=1)
> model32 <- prop.odds(Event(time, status==2)~ph.ecog + sex,
+                        data=dane, n.sim=500, profile=1)
> model34 <- prop.odds(Event(time, status==2)~ph.ecog + ph.karno,
+                        data=dane, n.sim=500, profile=1)
> model35 <- prop.odds(Event(time, status==2)~ph.ecog + pat.karno,
+                        data=dane, n.sim=500, profile=1)
> model36 <- prop.odds(Event(time, status==2)~ph.ecog + wt.loss,
+                        data=dane, n.sim=500, profile=1)
> model37 <- prop.odds(Event(time, status==2)~ph.ecog + meal.cal,
+                        data=dane, n.sim=500, profile=1)
> BIC_model31 <- -2*model31$loglike[1] + log(n)*5
> BIC_model32 <- -2*model32$loglike[1] + log(n)*6
> BIC_model34 <- -2*model34$loglike[1] + log(n)*10
> BIC_model35 <- -2*model35$loglike[1] + log(n)*12
> BIC_model36 <- -2*model36$loglike[1] + log(n)*5
> BIC_model37 <- -2*model37$loglike[1] + log(n)*5
```

Zmienna	Wartość BIC
age + ph.ecog	1041.513
sex + ph.ecog	1040.501
ph.karno + ph.ecog	1052.007
pat.karno + ph.ecog	1074.514
wt.loss + ph.ecog	1046.28
meal.cal + ph.ecog	1042.686

Najmniejsza wartość jest dla zmiennych *ph.ecog* oraz *sex*. Sprawdzamy, czy trzeba dodać jeszcze jakąś zmienną.

Krok 3

```
> model3.2.1 <- prop.odds(Event(time, status==2)~ph.ecog + sex + age
+                        , data=dane, n.sim=500, profile=1)
> model3.2.4 <- prop.odds(Event(time, status==2)~ph.ecog + sex + ph.karno
+                        , data=dane, n.sim=500, profile=1)
```

```

> model3.2.5 <- prop.odds(Event(time, status==2)~ph.ecog + sex + pat.karno
+
+ , data=dane, n.sim=500, profile=1)
> model3.2.6 <- prop.odds(Event(time, status==2)~ph.ecog + sex + wt.loss,
+
+ data=dane, n.sim=500, profile=1)
> model3.2.7 <- prop.odds(Event(time, status==2)~ph.ecog + sex + meal.cal,
+
+ data=dane, n.sim=500, profile=1)
> BIC_model3.2.1 <- -2*model3.2.1$loglike[1] + log(n)*7
> BIC_model3.2.4 <- -2*model3.2.4$loglike[1] + log(n)*12
> BIC_model3.2.5 <- -2*model3.2.5$loglike[1] + log(n)*14
> BIC_model3.2.6 <- -2*model3.2.6$loglike[1] + log(n)*7
> BIC_model3.2.7 <- -2*model3.2.7$loglike[1] + log(n)*7

```

Zmienna	Wartość BIC
age + sex + ph.ecog	1055.332
ph.karno + sex + ph.ecog	1079.888
pat.karno + sex + ph.ecog	1062.563
wt.loss + sex+ ph.ecog	1054.659
meal.cal+ sex + ph.ecog	1061.12

Najmniejszą wartością jest dodając zmienną *wt.loss*, jednak przypomnijmy sobie ile wynosiła wartość AIC, bez tej zmiennej:

Zmienna	Wartość BIC
ph.ecog + sex	1040.501
ph.ecog + sex + wt.loss	1054.659

Widzimy, że bez tej zmiennej model jest lepszy ostateczny model zbudujemy na:

- *sex*
- *ph.ecog*

1.5 Zadanie 6 - funkcja przeżycia modelu z kryterium AIC

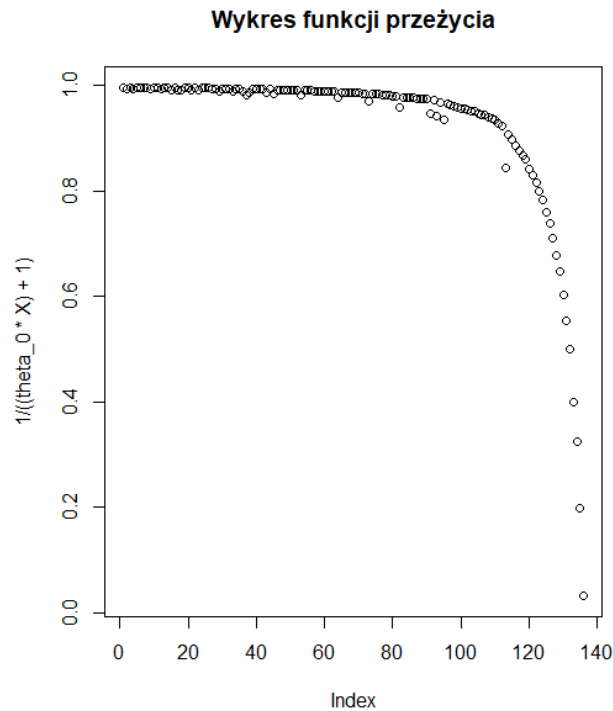
Na rysunku poniżej 1.5 umieściłam wykres funkcji przeżycia przy wartościach :

- $sex = 2$
- $ph.ecog = 2$
- $pat.karno = 50$

```
> lung<-data.frame(lung)
> lung %<>% as.data.table()
> lung<-lung[, `:=`(sex = as.factor(sex),
+               ph.ecog = as.factor(ph.ecog),
+               ph.karno = as.factor(ph.karno),
+               pat.karno = as.factor(pat.karno)))]
> lung<-lung[, `:=`(age = age - mean(age, na.rm = TRUE),
+               meal.cal = meal.cal - mean(meal.cal, na.rm = TRUE),
+               wt.loss = wt.loss - mean(wt.loss, na.rm = TRUE)))]
> lung %>% str
```

```
Classes 'data.table' and 'data.frame':      228 obs. of  10 variables:
 $ inst      : num  3 3 3 5 1 12 7 11 1 7 ...
 $ time      : num  306 455 1010 210 883 ...
 $ status    : num  2 2 1 2 2 1 2 2 2 2 ...
 $ age       : num  11.55 5.55 -6.45 -5.45 -2.45 ...
 $ sex       : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 2 2 1 1 ...
 $ ph.ecog   : Factor w/ 4 levels "0","1","2","3": 2 1 1 2 1 2 3 3 2 3 ...
 $ ph.karno  : Factor w/ 6 levels "50","60","70",...: 5 5 5 5 6 1 3 2 3 3 ...
 $ pat.karno : Factor w/ 8 levels "30","40","50",...: 8 7 7 4 7 6 4 6 6 5 ...
 $ meal.cal  : num  246 296 NA 221 NA ...
 $ wt.loss   : num  NA 5.17 5.17 1.17 -9.83 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

```
> model <- nltm(Surv(time,status==2) ~ age + ph.ecog + pat.karno,
+               data = lung, nlt.model = "P0")
> theta_0 <- (1-model$surv)/model$surv
> X <- exp(sum(model$coefficients*c(2, 2, 50)))
```

1.6 Zadanie 7 - hipoteza o proporcjonalności szans

W ostatnim zadaniu mamy zweryfikować hipotezę o proporcjonalności szans.

```
> model532 <- prop.odds(Event(time, status == 2) ~ sex
+                          + ph.ecog + pat.karno, data = dane,
+                          n.sim = 500, profile = 1)
> wald.test(model532)
```

	lin.comb	se	lower	upper	pval
[1,]	-0.97776596	0.3195623	-1.6040965	-0.3514355	0.002215557
[2,]	0.50166711	0.3743404	-0.2320265	1.2353608	0.180201041
[3,]	1.45134289	0.5960004	0.2832036	2.6194822	0.014886133
[4,]	2.36912429	1.1940913	0.0287483	4.7095003	0.047251443
[5,]	-0.40339676	1.8938872	-4.1153474	3.3085539	0.831327463
[6,]	1.54975246	1.3963573	-1.1870575	4.2865624	0.267062017
[7,]	0.88251736	1.0264108	-1.1292108	2.8942455	0.389894264
[8,]	-0.02261225	1.0753674	-2.1302936	2.0850691	0.983223744
[9,]	0.02048933	1.0692799	-2.0752607	2.1162394	0.984712029
[10,]	0.39364603	1.0965262	-1.7555059	2.5427979	0.719599785
[11,]	-0.36312091	1.1729613	-2.6620829	1.9358410	0.756883256

Wald test

data:

chisq = 31.307, df = 11, p-value = 0.0009843

Na podstawie uzyskanych danych, korzystając z testu Walda, na poziomie istotności 0.05 (patrzmy na wartości *pval*) mamy podstawy do odrzucenia hipotezy o proporcjonalności szans w

przyjętym przez nas modelu. Ponieważ wartość niektórych zmiennych jest mniejsza od przyjętego poziomu istotności, czyli poniżej 0.05. Widzimy możliwość wystąpienia problemu momentu, a więc niektóre zmienne mogą wymagać naprawy, ponieważ istnieją współczynniki zależne od czasu.