



Analiza przeżycia pacjentek poddanych leczeniu nowotworu piersi

Analiza Historii Zdarzeń z Wykorzystaniem
Narzędzi SAS

Dr Aneta Ptak-Chmielewska

Autorzy:

Bartłomiej Dupлага 63356

Jacek Dziwisz 78876



Struktura opracowania

I. Zdefiniowanie problemu badawczego.

II. Modelowanie nieparametryczne.

III. Modelowanie parametryczne.

IV. Modelowanie semiparametryczne.

V. Konkluzje.



I. Zdefiniowanie problemu badawczego

1. Dane :

- Zbiór danych *GBSG2* pochodzi z badania Sauerbrei, W., and P. Royston 1999.
- Zbiór składa się z grupy obserwacji 686 pacjentek, które zostały poddane leczeniu nowotworu piersi.
- Zmienne ujęte zbiorze:
 - *time* – zmienna określająca dzień od rozpoczęcia obserwacji w którym pacjentka umarła lub jej obserwacja została ucięta
 - *cens* – zmienna binarna określająca to czy obserwacja została ocenzurowana czy nie – {0,1}
 - *age* – zmienna ciągła, wiek pacjentki w latach
 - *tsize* – wielkość guza w [mm]
 - *menostat* – status pacjentki, przed menopauzą 0, po menopauzie 1
 - *pnodes* – liczba ognisk nowotworowych
 - *tgrade* – stopień zaawansowania guza: {1,2,3} – 3 najwyższy stopień
 - *progrec* – ilość receptorów progesteronowych w [fmol]



I. Zdefiniowanie problemu badawczego

- *estrec* – ilość receptorów estrogenowych w [fmol]
- *horTh* – zmienna binarna oznaczająca bycie poddanym terapii hormonalnej

2. Cel badawczy – ocena wpływu czynników na przeżywalność pacjentek poddanych terapii leczenia nowotworu piersi.

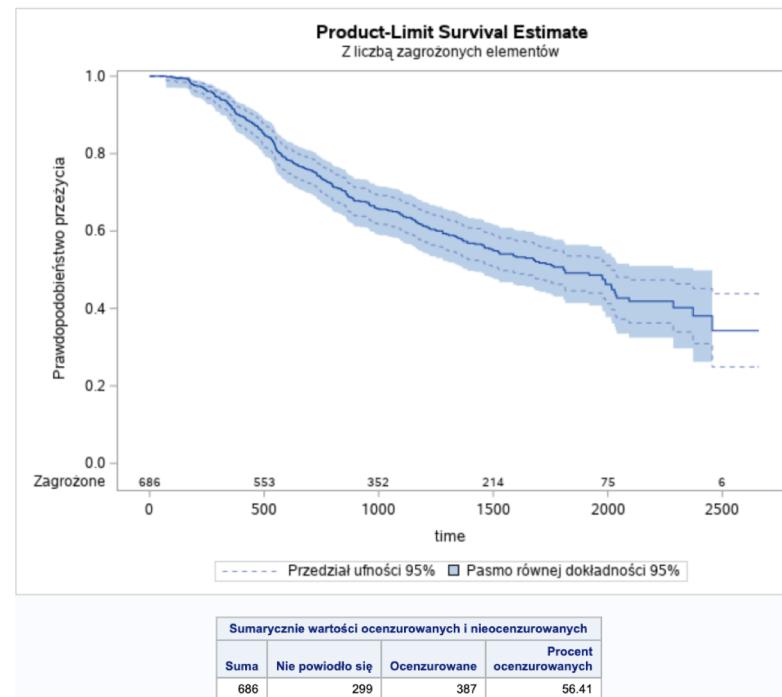
3. Hipotezy badawcze:

- 1) Im wyższy wiek pacjentki tym krótszy czas przeżycia podczas terapii nowotworowej.
- 2) Terapia hormonalna wydłuża czas przeżycia podczas leczenia nowotworu.

II. Modelowanie nieparametryczne.

1. Estymacja metodą Kaplan – Meiera.

- Mediana funkcji przeżycia wskazuje, że połowa pacjentów przeżyła do minimum do **1806** dnia badania.
- Prawdopodobieństwo przeżycia do końca obserwacji ($t=2700$) wyniosło na tym zbiorze **0.34**.

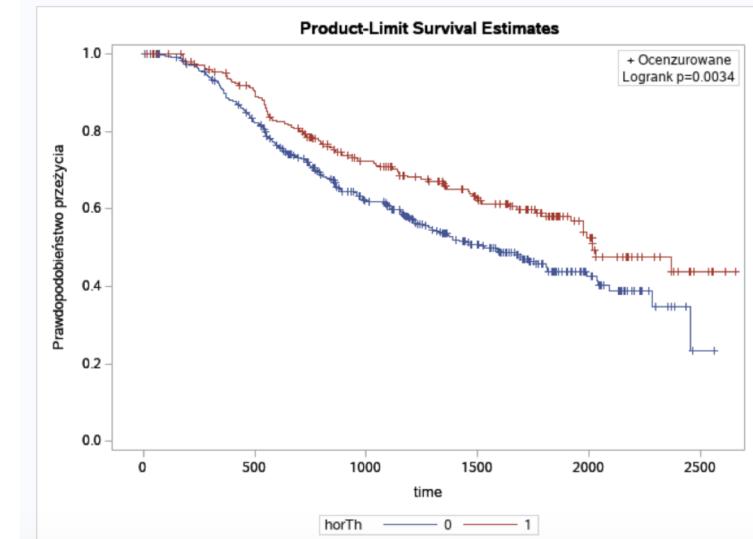


II. Modelowanie nieparametryczne.

2. Analiza czasu przeżycia pacjentek poddanych i niepoddanych terapii hormonalnej:

- Pacjentki poddane terapii hormonalnej ($horTh=1$) miały statystycznie dłuższy czas dożycia okresie obserwacji.
- Przy poziomie istotności 0.05 dla każdego z wykonanych testów (log-rang i Wilcoxon) należy odrzucić hipotezę zerową od braku różnic w grupach.

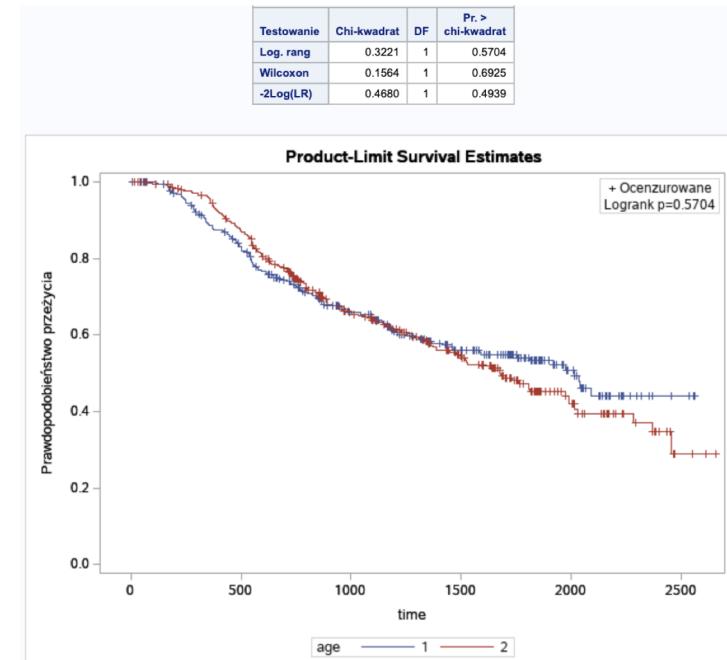
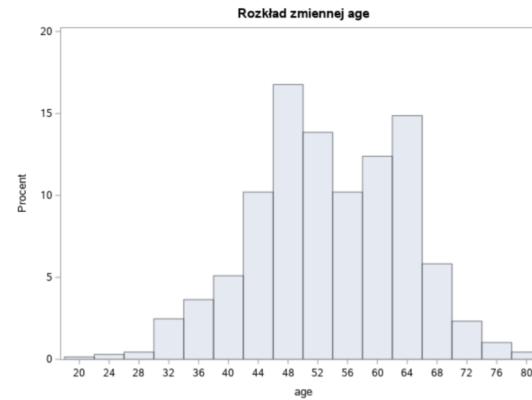
Test równości w warstwach			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kwadrat
Log. rang	8.5648	1	0.0034
Wilcoxon	8.3614	1	0.0038
-2Log(LR)	8.4808	1	0.0036



II. Modelowanie nieparametryczne.

2. Analiza czasu przeżycia pacjentek ze względu przynależności do grupy wiekowej:

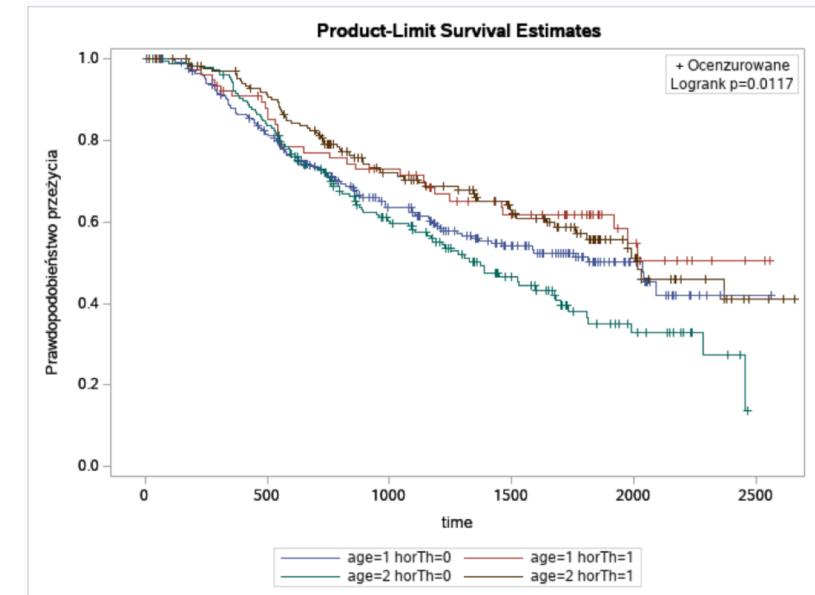
- Zmienna ciągła *age* została zdychotomizowana wzdłuż mediany wynoszącej 53. Empiryczny rozkład tej zmiennej przypominający rozkład dwumianowy uzasadnia taki wybór. Pacjentki poniżej 53 roku życia zostały przydzielonej do młodszej grupy wiekowej - 1, pacjentki powyżej 53 roku życia do starszej grupy wiekowej - 2.
- Estymacja nieparametryczna funkcji dożycia w grupach metodą K-M wykazała **brak** istotności różnic w przeżywalności wyszczególnionych grup dla przyjętego poziomu istotności 0.05.



II. Modelowanie nieparametryczne.

3. Analiza czasu przeżycia pacjentek w grupach wiekowych (**age**) z uwzględnieniem udziału w terapii hormonalnej (**horTh**):

- Wszystkie statystyki testowe nakazują odrzucenie hipotezy zerowej mówiącej o braku różnic w funkcji dożycia pomiędzy 4 rozróżnionymi grupami dla przyjętego poziomu istotności 0.05.
- Najbardziej narażoną grupą pacjentek są pacjentki starsze niepoddane terapii hormonalnej a najmniej pacjentki młodsze poddane terapii hormonalnej.



II. Modelowanie nieparametryczne.

3.

- Po przeliczeniu różnic w grupach z uwzględnieniem poprawki Tukeya dla testu Wilcoxona, brak podstaw do odrzucenia hipotezy zerowej o różnicach pomiędzy grupami 1 i 4, 2 i 3 oraz 3 i 4 – co potwierdza zasadność rozróżnienia pomiędzy grupą najbardziej i najmniej narażoną na ryzyko czyli grupą starszych kobiet niepoddanych terapii hormonalnej a grupą młodszych kobiet poddanych tej terapii (grupy 2 i 3).

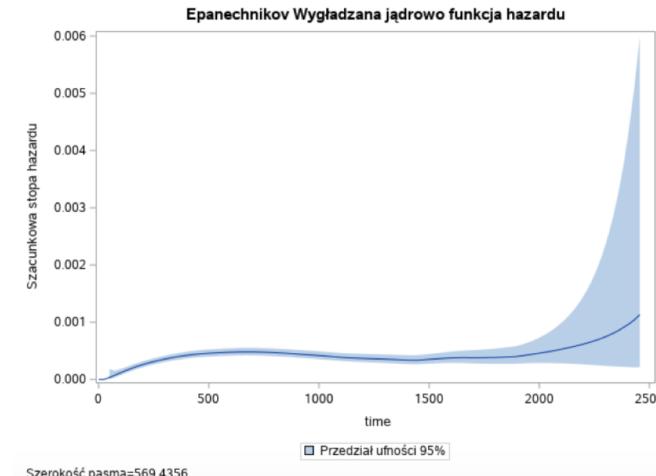
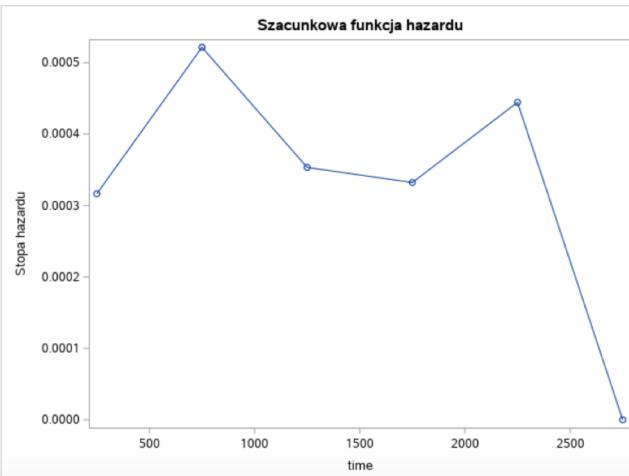
Poprawka na porównania wielokrotne dla testu Logrank				
Stratum_	Stratum_		Wartości p	
		Chi-kwadrat	Surowe	Tukey-Kramer
1	2	1.3228	0.2501	0.6583
1	3	1.5394	0.2147	0.6008
1	4	2.1501	0.1426	0.4580
2	3	7.9935	0.0047	0.0243
2	4	0.3996	0.5273	0.9217
3	4	8.5154	0.0035	0.0185

Poprawka na porównania wielokrotne dla testu Wilcoxon				
Stratum_	Stratum_		Wartości p	
		Chi-kwadrat	Surowe	Tukey-Kramer
1	2	1.9844	0.1589	0.4938
1	3	0.1371	0.7112	0.9827
1	4	4.2008	0.0404	0.1700
2	3	4.0170	0.0450	0.1863
2	4	1.3128	0.2519	0.6611
3	4	6.8404	0.0089	0.0442

Legenda warstw		
Warstwa	age	horTh
1	1	0
2	1	1
3	2	0
4	2	1

II. Modelowanie nieparametryczne.

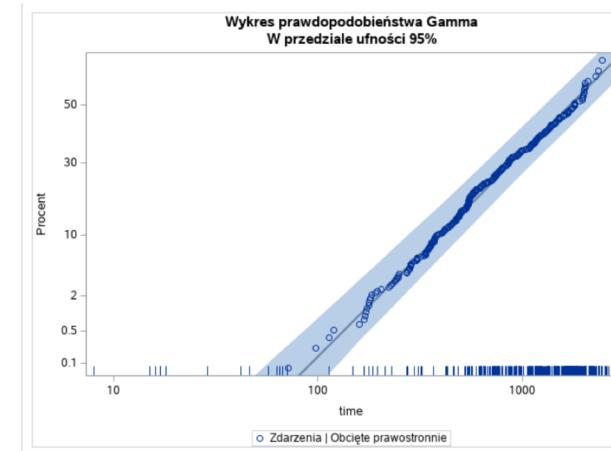
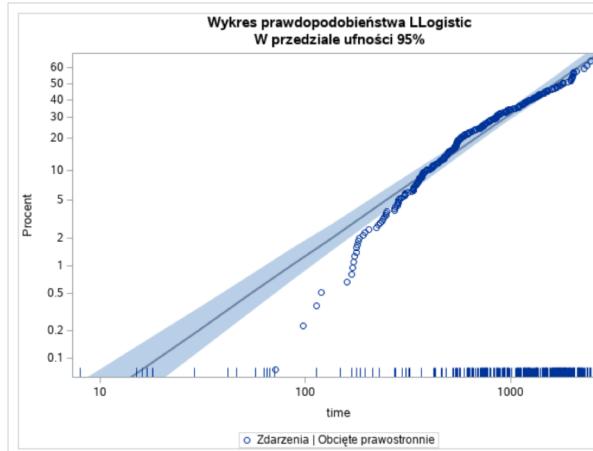
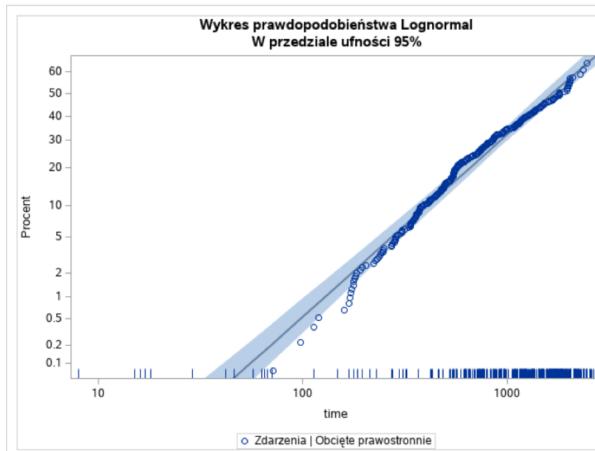
4. Funkcja hazardu wyestymowana z tablic trwania życia.
 - Niemonotoniczność empirycznej funkcji hazardu nakazuje odrzucenie parametrycznej analizy rozkładem wykładniczym lub Weibulla.
5. Wygładzona jądrowo funkcja hazardu.
 - Za szerokość pasma przyjęto 570 dni – co oznacza, że obserwacje o takiej rozpiętości przyjęto do obliczania hazardów w danych momentach czasowych w celu wygładzenia funkcji hazardu. Pasmo nie zostało ręcznie sprecyzowane lecz zostało zaproponowane przez optymalizator wygładzania.
 - Na wykres naniesiono 95% przedziału ufności.



III. Modelowanie parametryczne.

1. Ocena graficzna dopasowania modeli.

- Wstępnie najlepiej dopasowany zdaje się być model oparty o rozkład Gamma, następnie model log-normalny a na końcu log-logistyczny.





III. Modelowanie parametryczne.

2. W celu wyboru właściwego modelu należy policzyć statystykę $-2(LL_{\text{gamma}} - LL_{\text{log_norm}})$ a następnie przyrównać do wartości krytycznej obszaru wyznaczonego przez rozkład chi-kwadrat o jednym stopniu swobody dla poziomu istotności 0.05 czyli 3.84.

Log-normalny:

Statystyki dopasowania	
-2 log. wiarygodności	1352.822
AIC (jak najmniejsze)	1360.822
AICC (jak najmniejsze)	1360.880
BIC (jak najmniejsze)	1378.945

$$-2(LL_{\text{gamma}} - LL_{\text{log-norm}}) = \textcolor{blue}{16.62}$$

$$16.62 > 3.84$$

Wniosek: odrzucamy H_0 mówiącą o tym, że ograniczenie na parametrach jest prawdziwe.

Gamma:

Statystyki dopasowania	
-2 log. wiarygodności	1344.551
AIC (jak najmniejsze)	1354.551
AICC (jak najmniejsze)	1354.639
BIC (jak najmniejsze)	1377.205

Przyjmujemy zatem model **Gamma**, który jest istotnie lepiej dopasowany.

III. Modelowanie parametryczne.

3. Analiza estymacji modelu Gamma:

- Zmienna *age* okazała się nieistotna statystycznie w przeciwieństwie do zmiennej *horTh*.
- Dodatnie oszacowanie parametru zmiennej **horTh** wskazuje, że pacjentki poddane terapii hormonalnej mają 30% większy oczekiwany czas przeżycia od pacjentek niepoddanych terapii.
- Parametr *skali* większy od jeden potwierdza, że mamy do czynienia z malejącą funkcją hazardu.
- Parametr *ksztaltu* jest różny od 0 i od 1 – wniosek zatem jest taki, że nie mamy do czynienia ani z rozkładem log-normalnym ani Weibulla.

Analiza ocen parametrów maksymalnej wiarygodności						
Parametr	DF	Ocena	Błąd standardowy	Przedział ufności 95%	Chi-kwadrat	Pr. > chi-kw.
Intercept	1	6.9507	0.2226	6.5144 7.3870	975.03	<.0001
age	1	0.0243	0.1045	-0.1805 0.2290	0.05	0.8163
horTh	1	0.2986	0.1072	0.0886 0.5087	7.76	0.0053
Skala	1	1.2321	0.0572	1.1249 1.3495		
Kształt	1	-0.8047	0.2744	-1.3425 -0.2669		

IV. Modelowanie semiparametryczne.

1. Test globalny na istotność zmiennych.

- Wszystkie testy sugerują odrzucenie hipotezy zerowej o nieistotności globalnej zmiennych *age* oraz *horTh* a zatem przynajmniej jedna z nich jest istotna.
- Analiza ocen maksymalnej wiarygodności wskazuje, że obie zmienne są statystycznie istotne.
- Interpretacja: *bycie poddanym terapii hormonalnej zmniejsza hazard o 34%. Wzrost wieku pacjentki o 1 rok powoduje wzrost hazardu podczas leczenia nowotworu piersi o 29%*.

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Iloraz wiaryg.	10.5872	2	0.0050
Ocena	10.3492	2	0.0057
Wald	10.2580	2	0.0059

Analiza ocen maksymalnej wiarygodności						
Parametr	DF	Ocena parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi-kw.	Współczynnik hazardu
age	1	0.25234	0.12379	4.1551	0.0415	1.287
horTh	1	-0.42103	0.12787	10.8409	0.0010	0.656

IV. Modelowanie semiparametryczne.

2. Włączenie zmiennych zależnych od czasu:

- Globalny test Walda nakazuje odrzucić hipotezę o zbiorowej nieistotności oszacowań parametrów.
- Włączenie interakcji z czasem powoduje indywidualną nieistotność oszacowań zmiennych *age* i *horTh*.
- Efekt zmiennej *agetime* po 365 dniach (1 roku) od upłynięcia terapii:

$$-0.198 + 0.000579 * 365 = 0.013$$

Oznacza, że rosnący wiek pacjentki wystawia ją na większe ryzyko po 1 roku od rozpoczęcia terapii.

Testowanie globalnej hipotezy zerowej: BETA=0			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kw.
Iloraz wiaryg.	18.3297	4	0.0011
Ocena	17.6674	4	0.0014
Wald	17.3381	4	0.0017

Analiza ocen maksymalnej wiarygodności						
Parametr	DF	Ocena parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi-kw.	Współczynnik hazardu
age	1	-0.19789	0.23633	0.7011	0.4024	0.820
horTh	1	-0.43058	0.24432	3.1058	0.0780	0.650
agetime	1	0.0005787	0.0002648	4.7746	0.0289	1.001
horThtime	1	0.0000134	0.0002490	0.0029	0.9570	1.000



V. Konkluzje

Weryfikacja postawionych hipotez:

-- Im wyższy wiek pacjentki tym krótszy czas przeżycia podczas terapii nowotworowej

- Zarówno analiza nieparametryczna (test log-rang oraz test Wilcooxona) jak i parametryczna oparta na dwóch grupach wiekowych (starszej i młodszej) wykazały brak istotnych różnic w czasie dożycia w grupach wyszczególnionych ze względu na wiek. Analiza semiparametryczna wskazała istotność zmiennej wiek tylko w połączeniu z czasem (pacjentki leczone ponad **1 rok** są istotnie bardziej narażone na zdarzenie).

✓ Terapia hormonalna wydłuża czas przeżycia podczas leczenia nowotworu

- Analiza nieparametryczna i parametryczna potwierdziły, że terapia hormonalna jest skutecznym narzędziem wydłużającym czas życia pacjentek niezależnie od przynależności do grup wiekowych.
- Model Gamma wykazał, że pacjentki poddane tej terapii mają czas przeżycia o **ok. 30%** dłuższy od pacjentek, które nie są poddane terapii hormonalnej.



DZIĘKUJEMY ZA UWAGĘ!