

# **Temat projektu: Modele nieparametryczne w analizie historii zdarzeń**

Autor: Sebastian Boruch

## Spis treści

1. Wprowadzenie .....	3
2. Opis danych .....	3
2.1. Źródło .....	3
2.2. Cechy zbioru .....	3
3. Estymacja modeli nieparametrycznych.....	7
3.1. Metoda tradycyjna.....	7
3.2. Metoda tradycyjna ze stratyfikacją.....	10
3.3. Metoda Kaplana-Meiera .....	15
4. Porównanie modeli i wnioski .....	17
5. Bibliografia.....	17
6. Kody SAS.....	17

# 1. Wprowadzenie

Leczenie narkomanów to jeden z kluczowych problemów medycyny w ostatnich dekadach. Rozwój technologii, osamotnienie czy nadmiar obowiązków sprzyja rozwojowi czynników stresogennych, co z kolei prowadzi do szukania substancji uśmierzających dyskomfort psychiczny jak np. narkotyki. Wzrost dochodów per capita i wzrastająca dostępność środków odurzających (np. poprzez zakup w internecie) powoduje, że ryzyko zachorowania rośnie. W poniższym opracowaniu skupię się na czynnikach pozwalających na wyleczenie uzależnionego. W pierwszej części przedstawię dane, na których przeprowadzę analizę- interpretacja i testowanie wszelkich dostępnych testów istotności oraz testu homogeniczności funkcji dożycia w warstwach, opis zbioru danych, analizę zmiennych i definicję modelowanego zjawiska. W drugiej części stworzę tablice przetrwania metodą tradycyjną i Kaplana-Meiera za pomocą procedury LIFETEST systemu SAS.

## 2. Opis danych

### 2.1. Źródło

Australijskie badanie przeprowadzone przez Caplehorn i Bell (1991) z departamentu zdrowia publicznego Uniwersytetu w Sydney porównywało czas przeżycia w dwóch klinikach leczenia metadonem dla osób uzależnionych od heroiny. Czas przeżycia pacjenta określono jako czas w dniach, aż pacjent powrócił do nałogu. Obie kliniki różniły się w zależności od ogólnej polityki leczenia. Celem było zidentyfikowanie czynników, które wpływają na czas przeżycia: klinika, maksymalna dzienna dawka metadonu i (nie)obecność w więzieniu.

Proces leczenia pacjentów metadonem badano w kohorcie 238 osób uzależnionych od heroiny, którzy weszli do programu terapii uzależnień od lutego 1986 do sierpnia 1987 r. Wszyscy pacjenci zostali ocenieni w tym samym instytucie i odesłani do jednej z dwóch placówek leczniczych w celu rozpoczęcia terapii.

Czym jest metadon?

Metadon wykorzystywany jest w leczeniu osób uzależnionych od opiatów, m.in. heroiny i morfiny. Substancja znosi objawy abstynencyjne, ale jednocześnie sama posiada duży potencjał uzależniający, dlatego jej stosowanie powinno odbywać się pod ścisłą kontrolą lekarską. W razie przedawkowania metadonu pacjent naraża się na kolejne uzależnienie, którego skutki mogą być dużo poważniejsze niż w przypadku zażywania heroiny. Dlatego indywidualny dobór dawki jest tu kluczowy dla wyleczenia uzależnionego.

### 2.2. Cechy zbioru

Początki badania sięgają lutego 1986, lecz nie jest podana dokładna data zakończenia zbierania danych. Biorąc pod uwagę fakt, że najdłuższy okres przeżycia sięga 1100 dni, a wyniki badań opublikowane zostały w roku 1991 autor ocenia, że okres zbierania danych do analizy zakończył się w roku 1990. W zbiorze nie występują braki danych.

Nazwa zmiennej	Opis	Typ	Wartości
<b>Klinika</b>	Placówka, w której pacjent był leczony	Jakościowa	1- klinika nr 1 2- klinika nr 2
<b>Status</b>	Przyczyna zakończenia naliczania dni pobytu pacjenta w klinice	Binarna	0- pacjent opuścił klinikę (cenzura prawostronna) 1- powrót do nałogu
<b>Dni przeżycia</b>	Liczba dni pobytu w klinice	Ilościowa ciągła	od 2 do 1076
<b>Więzienie</b>	Czy pacjent przebywał kiedykolwiek w więzieniu?	Binarna	0- nie 1- tak
<b>Dawka</b>	Maksymalna dzienna dawka metadonu (mg/dzień)	ilościowa	20-110

*Tabela 1. charakterystyka zmiennych*

Zdarzeniem (porażką) jest tu powrót do nałogu przez pacjenta (status=1). Cenzorowanie następuje gdy zmienna status=0. Porażka wystąpiła w 63% przypadków:

Status	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
0	88	36.97	88	36.97
1	150	63.03	238	100.00

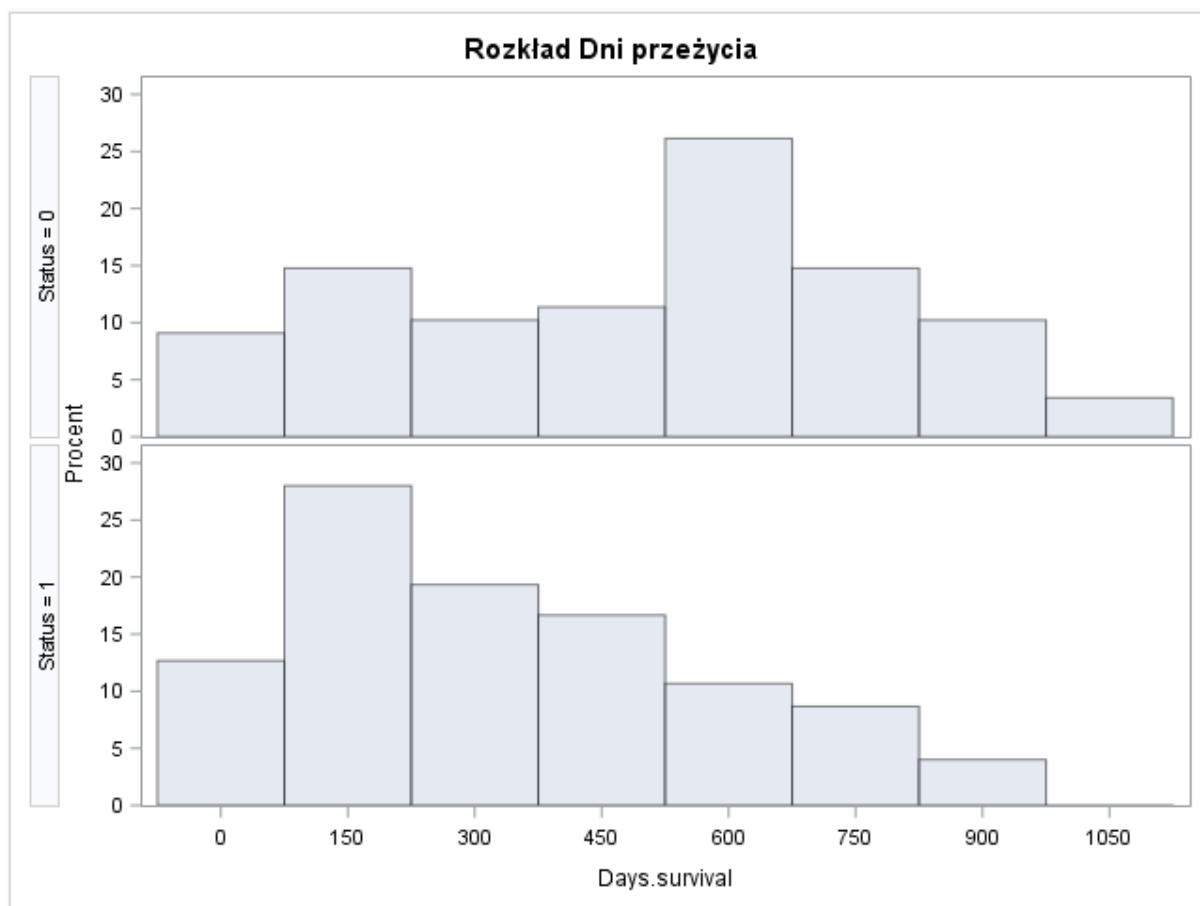
*Tabela 2. Rozkład obserwacji, dla zmiennej status*

Zmienna analizowana: Dni przeżycia Days.survival									
Status	N obs.	Srednia	Odch. std.	Minimum	Maksimum	N	Dolny kwartyl	Mediana	Górny kwartyl
0	88	499.8295455	283.6670876	2.0000000	1076.00	88	300.0000000	536.5000000	698.5000000
1	150	345.5133333	241.3917626	7.0000000	899.0000000	150	157.0000000	283.0000000	522.0000000

*Tabela 3. Miary statystyczne zmiennej czasu z podziałem na porażkę lub jej brak*

Jeżeli zaszło zdarzenie, to przeciętny czas przeżycia pacjenta wynosił 345 dni. Mediana to tylko 283, więc rozkład jest prawostronnie asymetryczny.

Jeżeli do zdarzenia nie doszło, to średni czas przeżycia pacjenta w klinice wyniósł 500 dni i jest to mniej niż mediana (537) więc rozkład jest lewostronnie asymetryczny.



Wykres 1. Rozkład dni przeżycia na podstawie zmiennej STATUS

Z wykresu można wnioskować, że osoby które wracają do nałogu, czynią to przeważnie po około pół roku pobytu w klinice. Osoby, które do nałogu nie powracają, muszą przeważnie przebywać w klinice około półtora roku.

Poza zmienną STATUS występują również 3 inne zmienne objaśniające model. Jedną z nich jest zmienna KLINIKA.

Clinic				
Klinika	Liczebność	Procent	Liczebność skumulowana	Procent skumulowany
1	122	81.33	122	81.33
2	28	18.67	150	100.00

Tabela 4. Rozkład dla zmiennej KLINIKA, gdzie zaszło zdarzenie

	Status		N
	0	1	
	N	N	
Klinika			
1	41	122	163
2	47	28	75

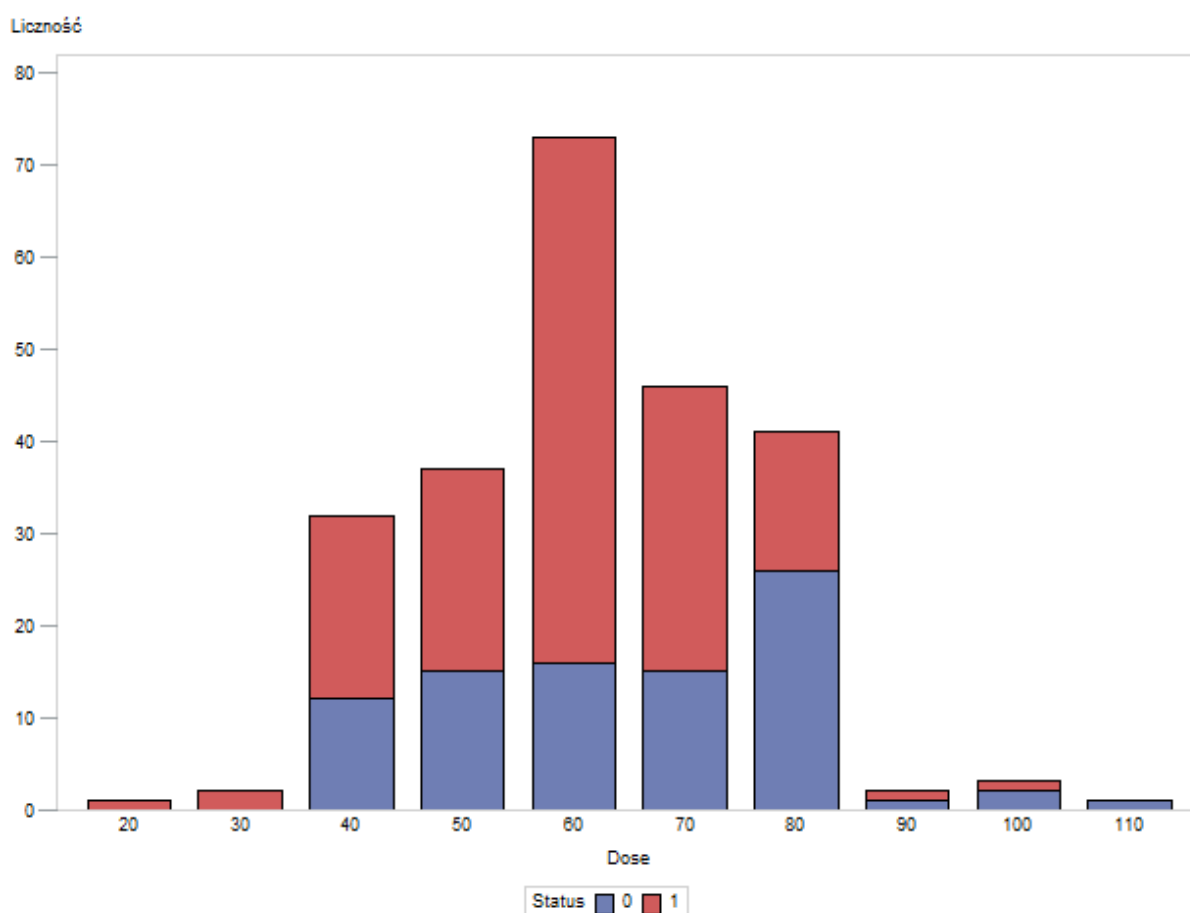
Tabela 5. Zagregowana tabela liczebności zmiennych KLINIKA i STATUS

Powyższe tabele sugerują, że zmienna KLINIKA jest istotna statystycznie. Ryzyko zdarzenia znacząco wzrasta przy pobycie pacjenta w klinice 1. Odsetek pacjentów, którzy wrócili do nałogu w klinice 2. jest znacząco niższy.

		Status	
		0	1
Więzienie			
0	N	46	81
1	N	42	69

Tabela 6. Zagregowana tabela liczebności zmiennych WIEZIENIE i STATUS

Kolejną zmienną biorącą udział w badaniu była zmienna WIEZIENIE. Dostarcza ona informacji czy dany pacjent kiedykolwiek przebywał w zakładzie karnym. Po analizie powyższej tabeli można zasugerować wniosek, iż pobyt w więzieniu nie wpływa istotnie na ryzyko powrotu do nałogu. Jeżeli jednak należałoby wyciągnąć jakieś wnioski, to można stwierdzić, że pacjenci którzy przebywali wcześniej w więzieniu, mają mniejszą skłonność do powrotu do nałogu.



Wykres 2. Rozkład zmiennej DAWKA z uwzględnieniem zmiennej STATUS

Ostatnią zmienną uwzględnioną w badaniu była zmienna DAWKA, która informowała o przyjmowanej przez pacjenta dawce metadonu. Po analizie powyższego wykresu można stwierdzić, że odsetek porażki przewyższa znacząco odsetek przeżycia do dawki 70 mg/dzień. Powyżej tej dawki przeważa ilość pacjentów, którzy nie powrócili do nałogu. Rozkład zmiennej jest zbliżony do rozkładu normalnego.

Współczynniki korelacji Pearsona, N = 238 Prawd. >  r  przy H0: rho=0			
	Klinika	Więzenie	Dawka
Klinika	1.00000	0.01851	0.14744
		0.7763	0.0229
Więzenie	0.01851	1.00000	0.02377
	0.7763		0.7153
Dawka	0.14744	0.02377	1.00000
	0.0229	0.7153	

Tabela 7. Tabela korelacji zmiennych

Powyższa tabela korelacji ukazuje, że nie występuje silna korelacja pomiędzy żadną ze zmiennych. Co za tym idzie, nie trzeba usuwać żadnej ze zmiennych ani dodawać nowych. Nie trzeba również dodawać efektów interakcji.

### 3. Estymacja modeli nieparametrycznych

W niniejszym rozdziale przedstawiono estymację modeli nieparametrycznych i konstrukcję tablic trwania życia przy użyciu kolejno metody tradycyjnej z użyciem stratyfikacji oraz metody Kaplana-Meiera. Kroków tych dokonano za pomocą procedury PROC LIFETEST.

#### 3.1. Metoda tradycyjna

Tradycyjne tablice trwania życia są metodą, która dostarcza wyników dla funkcji dożycia, funkcji gęstości i wskaźnika hazardu. Jednym z wymogów tej metody jest uzyskanie takiej struktury danych, aby zmienna czasowa miała przedziały o podobnej długości. Zmienną czasową jest tu zmienna DNI PRZEŻYCIA a zmienną cenzurującą zmienna STATUS. Jeżeli zmienna STATUS przyjmie wartość 0- obserwacja jest obcięta, jeżeli wartość 1- nieudana (zachodzi porażka).

Przedział		Oceny przeżycia z tablic wymieralności								
[Dolny,	Górny]	Liczba nieudanych	Liczba obciętych	Minimalna liczebność próby	Warunkowe prawdopodobieństwo niepowodzenia	Błąd standardowy prawdopodobieństwa warunkowego	Przeżycie	Niepowodzenie	Błąd standardowy przeżycia	Mediana dalszego trwania
0	70	19	7	234.5	0.0810	0.0178	1.0000	0	0	497.4
70	140	13	6	209.0	0.0622	0.0167	0.9190	0.0810	0.0178	462.9
140	210	24	6	190.0	0.1263	0.0241	0.8618	0.1382	0.0227	417.9
210	280	18	2	162.0	0.1111	0.0247	0.7530	0.2470	0.0287	426.9
280	350	11	7	139.5	0.0789	0.0228	0.6693	0.3307	0.0316	417.7
350	420	11	5	122.5	0.0898	0.0258	0.6165	0.3835	0.0329	
420	490	10	5	106.5	0.0939	0.0283	0.5612	0.4388	0.0339	
490	560	14	10	89.0	0.1573	0.0386	0.5085	0.4915	0.0346	
560	630	7	13	63.5	0.1102	0.0393	0.4285	0.5715	0.0351	
630	700	6	5	47.5	0.1263	0.0482	0.3812	0.6188	0.0355	
700		17	22	28.0	0.6071	0.0923	0.3331	0.6669	0.0360	

Tabela 8. Podstawowe wyniki estymacji modelu nieparametrycznego metodą tradycyjną.

Zmienna czasu została podzielona na 70-dniowe interwały z górną granicą 700 dni. Każdy z przedziałów ma satysfakcjonującą dużą liczebność i wystarczająco mały odsetek danych odciętych bądź nieudanych, co pozwala przeprowadzić analizę.

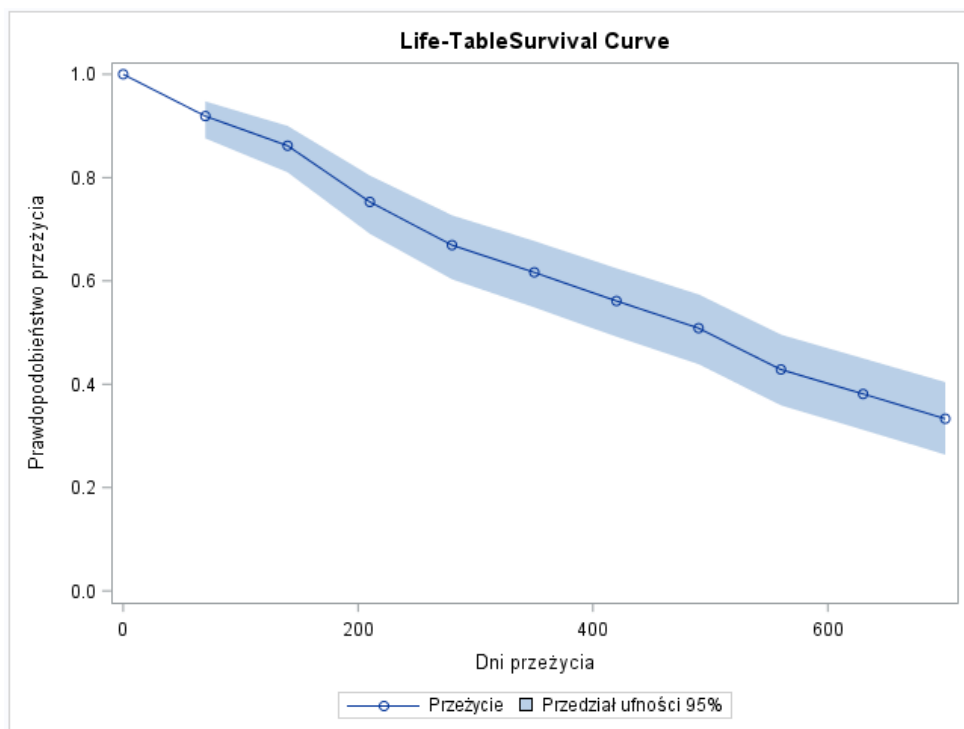
Błąd standardowy mediany	Obliczane w punkcie środkowym przedziału			
	PDF	Błąd standardowy PDF	Hazard	Błąd standardowy hazardu
28.5757	0.00116	0.000255	0.001206	0.000277
27.8163	0.000817	0.000220	0.000917	0.000254
27.3594	0.00156	0.000300	0.001926	0.000392
42.9946	0.00120	0.000269	0.001681	0.000395
41.1843	0.000754	0.000221	0.001173	0.000353
	0.000791	0.000231	0.001343	0.000405
	0.000753	0.000231	0.001407	0.000445
	0.00114	0.000291	0.002439	0.000649
	0.000675	0.000247	0.001667	0.000629
	0.000688	0.000270	0.001926	0.000785

*Tabela 9. Podstawowe wyniki estymacji obliczane w punkcie środkowym przedziału*

Autor podjął się analizy przedziału, gdzie występuje najwięcej porażek (140-210):

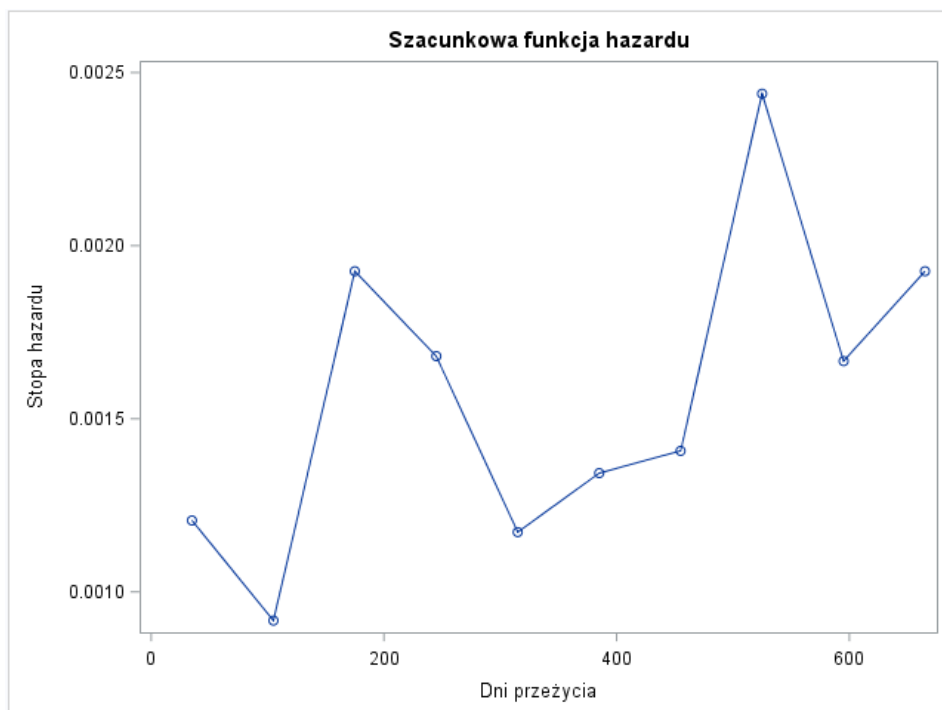
- liczba nieudanych: **24** - porażki, czyli liczba pacjentów, która powróciła do nałogu w tym przedziale czasu
- liczba obciążonych: **6** - liczba obserwacji ocenianych- pacjent opuścił klinikę
- minimalna liczebność próby- wielkość efektywna próby w tym przedziale to 190:
  - metoda obliczenia:  $209 - 13 - \frac{1}{2} * 6 - \frac{1}{2} * 6 = 190$
  - Całkowita liczebność próby (populacja) – suma z kolumny liczba nieudanych do przedziału poprzedzającego włącznie – suma z kolumny liczba obciążonych do przedziału poprzedzającego włącznie -  $\frac{1}{2} * \text{liczba obciążonych dla badanego przedziału}$ .
- Warunkowe prawdopodobieństwo niepowodzenia: **0,1263** – warunkowe prawdopodobieństwo zajścia zdarzenia, a zatem powrotu do nałogu dla badanego przedziału wynosi 12,63%, zatem prawdopodobieństwo dożycia 210. dnia (koniec przedziału) wynosi  $100\% - 12,63\% = 87,37\%$
- Przeżycie: **0,8618** prawdopodobieństwo, że osoba nie wróci do nałogu do początku badanego przedziału (prawdopodobieństwo dożycia przez pacjenta dnia 140.) wynosi 86,18%
- Niepowodzenie: **0,1382** – prawdopodobieństwo, że osoba wróci do nałogu przed początkiem badanego przedziału (prawdopodobieństwo śmierci pacjenta przed dniem 140) wynosi 13,82%
- Mediana dalszego trwania życia: **417,9** i błąd standardowy mediany: **27,36** – w badanym przedziale połowa pacjentów przeżyje czas nie dłuższy niż 417,9, a pozostała połowa czas dłuższy niż 417,9 dnia, z błędem na poziomie 27,36 dnia.
- Hazard: **0,0019** - ryzyko, że przypadek, który przeżył do początku danego przedziału ulegnie w tym przedziale zdarzeniu





Wykres 3. Krzywa przeżycia w zależności od zmiennej czasowej

Krzywa spada monotonicznie w każdym przedziale. Na wykresie widać również przedział ufności 95%, który nie powiększa się znacznie wraz ze wzrostem zmiennej czasu. Nie występuje tu heteroskedastyczność, a co za tym idzie krzywa jest istotna statystycznie.



Wykres 4. Funkcja hazardu dla metody tradycyjnej



Wykres 5. Funkcja gęstości prawdopodobieństwa dla modelu tradycyjnego

Funkcja hazardu nie wykazuje się monotonicznością- ani nie rośnie ani nie maleje przez cały okres trwania badania. Występują tu 2 maksima- jedno przy ok. 180 dniach i drugie przy ok. 500 dniach. Biorąc pod uwagę rozkład zmiennej dni przeżycia z poprzedniego rozdziału, warto dodać, że ok. dnia 500. występuje znaczący wzrost danych ocenzorowanych, co ma znaczący wpływ na występujący w tym okresie odsetek niepowodzeń. Gdyby jednak przeprowadzić linię klasycznej regresji liniowej przez tę funkcję, należałoby stwierdzić, że funkcja hazardu rośnie w czasie, co oznacza że ryzyko porażki wzrasta w czasie.

### 3.2. Metoda tradycyjna ze stratyfikacją

Kolejnym krokiem w analizie metodą tradycyjną jest wykorzystanie zmiennej warstwującej. Autor zdecydował się użyć jako zmiennej stratyfikującej zmiennej DAWKA, jako że patrząc po jej rozkładzie(patrz- wykres 2)- może ona mieć największy wpływ na zmienną STATUS. Poza tym, już po samej analizie danych zagregowanych można ocenić, że pobyt w więzieniu miał mały wpływ na powrót do nałogu, a pacjenci z kliniki 1 o wiele częściej doświadczali porażki niż pacjenci z kliniki 2.

Zmienna została podzielona na 3 warstwy:

- wartości poniżej 60 mg/dzień
- wartość dominująca- 60 mg/dzień
- wartości powyżej 60 mg/dzień

UWAGA: proszę nie sugerować się wartością 60.5 przy Warstwie 2- jest to umowny środek przedziału oszacowany przez SAS Enterprise Miner, ale uwzględnia tylko dawkę 60 mg/dzień.

Warstwa 1: Dawka < 60

Przedział		Oceny przeżycia z tablic wymieralności							
[Dolny,	Górny]	Liczba nieudanych	Liczba obciążonych	Minimalna liczebność próby	Warunkowe prawdopodobieństwo niepowodzenia	Błąd standardowy prawdopodobieństwa warunkowego	Przeżycie	Niepowodzenie	Błąd standardowy przeżycia
0	70	11	4	91.0	0.1209	0.0342	1.0000	0	0
70	140	8	6	75.0	0.1067	0.0356	0.8791	0.1209	0.0342
140	210	11	1	63.5	0.1732	0.0475	0.7853	0.2147	0.0437
210	280	8	2	51.0	0.1569	0.0509	0.6493	0.3507	0.0520
280	350	4	2	41.0	0.0976	0.0463	0.5475	0.4525	0.0549
350	420	7	1	35.5	0.1972	0.0668	0.4940	0.5060	0.0556
420	490	6	1	27.5	0.2182	0.0788	0.3966	0.6034	0.0555
490	560	3	5	18.5	0.1622	0.0857	0.3101	0.6899	0.0535
560	630	1	2	12.0	0.0833	0.0798	0.2598	0.7402	0.0521
630	700	4	0	10.0	0.4000	0.1549	0.2382	0.7618	0.0521
700	.	2	4	4.0	0.5000	0.2500	0.1429	0.8571	0.0483

Warstwa 2: Dawka = 60.5

Przedział		Oceny przeżycia z tablic wymieralności							
[Dolny,	Górny]	Liczba nieudanych	Liczba obciążonych	Minimalna liczebność próby	Warunkowe prawdopodobieństwo niepowodzenia	Błąd standardowy prawdopodobieństwa warunkowego	Przeżycie	Niepowodzenie	Błąd standardowy przeżycia
0	70	8	3	50.5	0.1584	0.0514	1.0000	0	0
70	140	2	0	41.0	0.0488	0.0336	0.8416	0.1584	0.0514
140	210	5	2	38.0	0.1316	0.0548	0.8005	0.1995	0.0565
210	280	4	0	32.0	0.1250	0.0585	0.6952	0.3048	0.0658
280	350	2	4	26.0	0.0769	0.0523	0.6083	0.3917	0.0705
350	420	3	1	21.5	0.1395	0.0747	0.5615	0.4385	0.0724
420	490	1	0	18.0	0.0556	0.0540	0.4832	0.5168	0.0751
490	560	3	0	17.0	0.1765	0.0925	0.4563	0.5437	0.0756
560	630	0	3	12.5	0	0	0.3758	0.6242	0.0752
630	700	1	0	11.0	0.0909	0.0867	0.3758	0.6242	0.0752
700	.	8	2	9.0	0.8889	0.1048	0.3416	0.6584	0.0757

Warstwa 3: Dawka >= 61

Przedział		Oceny przeżycia z tablic wymieralności							
[Dolny,	Górny]	Liczba nieudanych	Liczba obciążonych	Minimalna liczebność próby	Warunkowe prawdopodobieństwo niepowodzenia	Błąd standardowy prawdopodobieństwa warunkowego	Przeżycie	Niepowodzenie	Błąd standardowy przeżycia
0	70	0	0	93.0	0	0	1.0000	0	0
70	140	3	0	93.0	0.0323	0.0183	1.0000	0	0
140	210	8	3	88.5	0.0904	0.0305	0.9677	0.0323	0.0183
210	280	6	0	79.0	0.0759	0.0298	0.8803	0.1197	0.0339
280	350	5	1	72.5	0.0690	0.0298	0.8134	0.1866	0.0408
350	420	1	3	65.5	0.0153	0.0152	0.7573	0.2427	0.0451
420	490	3	4	61.0	0.0492	0.0277	0.7457	0.2543	0.0459
490	560	8	5	53.5	0.1495	0.0488	0.7091	0.2909	0.0482
560	630	6	8	39.0	0.1538	0.0578	0.6030	0.3970	0.0536
630	700	1	5	26.5	0.0377	0.0370	0.5103	0.4897	0.0572
700	.	7	16	15.0	0.4667	0.1288	0.4910	0.5090	0.0582

Tabela 10. Oceny przeżycia dla warstw

Sumarycznie wartości obciążonych i nieobciążonych					
Warstwa	Dawka	Suma	Niezaliczone	Obciążone	Procent obciążonych
1	<60	93	65	28	30.11
2	60.5	52	37	15	28.85
3	>=61	93	48	45	48.39
Suma		238	150	88	36.97

Tabela 11. Sumaryczne wartości obciążonych i nieobciążonych

Statystyki rangowe		
Dawka	Log. rang	Wilcoxon
<60	23.950	3518.0
60.5	7.096	941.0
>=61	-31.046	-4459.0

Macierz kowariancji dla statystyk log. rang			
Dawka	<60	60.5	>=61
<60	28.7464	-8.1112	-20.6352
60.5	-8.1112	23.8054	-15.6942
>=61	-20.6352	-15.6942	36.3293

Macierz kowariancji dla statystyk Wilcoxona			
Dawka	<60	60.5	>=61
<60	720963	-217083	-503880
60.5	-217083	523233	-306150
>=61	-503880	-306150	810031

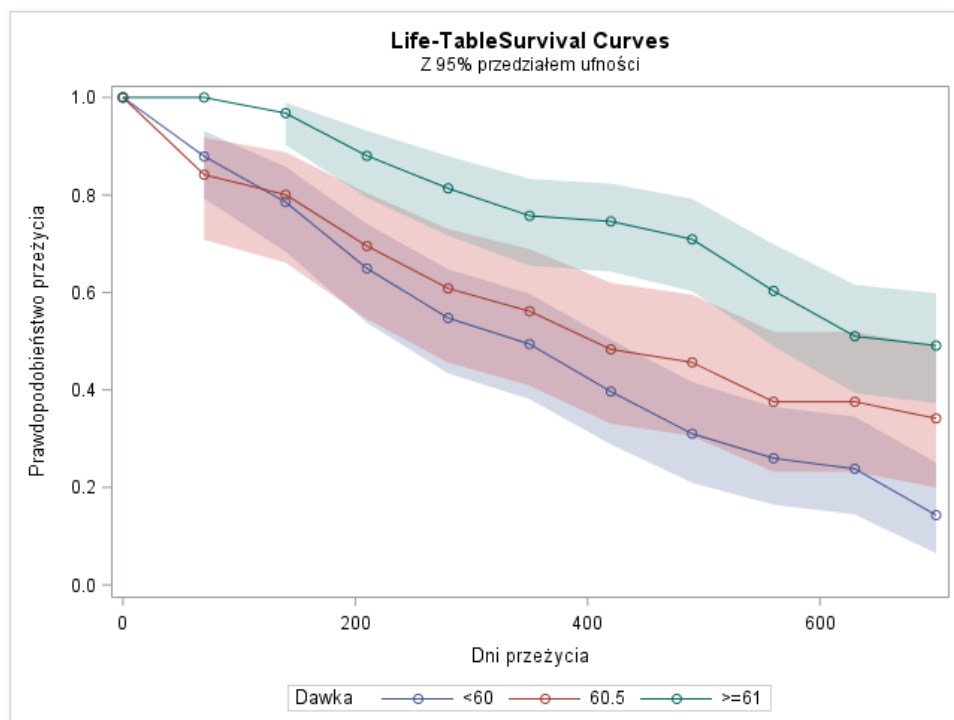
Test równości w warstwach			
Test	Chi-kwadrat	DF	Pr. > chi-kwadrat
Log. rang	28.8734	2	<.0001
Wilcoxon	25.9049	2	<.0001
-2Log(LR)	23.0536	2	<.0001

Tabela 12. Testy homogeniczności krzywych przeżycia w warstwach

Ponieważ funkcja przeżycia daje pełne rozeznanie doświadczenia przeżycia każdej grupy, naturalnym podejściem do odpowiedzi na to pytanie o homogeniczność jest przetestowanie hipotezy zerowej, stanowiącej że funkcje, które przeżyły są takie same w obu grupach; to jest,  $S_1(t) = S_2(t)$  dla wszystkich  $t$ , gdzie indeksy dolne rozróżniają dwie grupy. PROC LIFETEST oblicza dwie alternatywne statystyki do testowania tej hipotezy zerowej: test log-rank (znany również jako test Mantela-Haenszela) i test Wilcoxona. Trzeci test, statystyka ilorazu wiarygodności, jest obliczany przy dodatkowym założeniu, że czasy zdarzeń mają rozkład wykładniczy. Postawione hipotezy są w tym przypadku następujące:

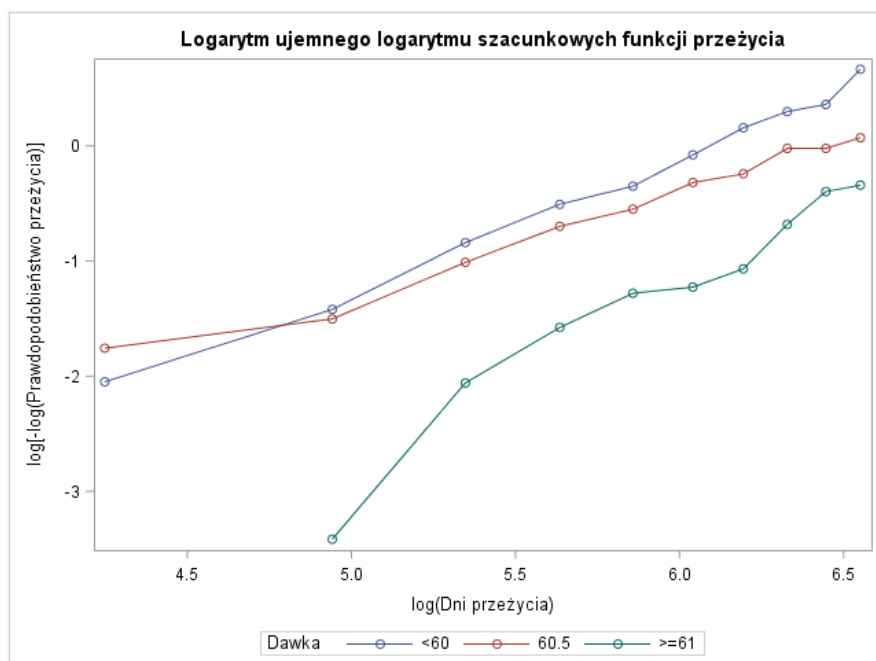
- $H_0$ : warstwy są homogeniczne
- $H_1$ : warstwy są heterogeniczne

Odrzucamy  $H_0$ ; funkcje dożycia w warstwach są heterogeniczne; istotnie statystycznie różnicuje się czas dożycia, czyli czas powrotu do nałogu.



Wykres 6. Funkcja dożycia z podziałem na warstwy

Powyższy wykres potwierdza wniosek z testów homogeniczności, mianowicie- warstwy są heterogeniczne. Na powyższym wykresie widać, że funkcja dożycia maleje najwolniej dla największych dawek ( $> 60$  mg/dzień), a najszybciej dla dawek najmniejszych. Na tej podstawie można wysunąć wniosek, że im większa dawka metadonu, tym mniejsze prawdopodobieństwo powrotu przez pacjenta do nałogu.

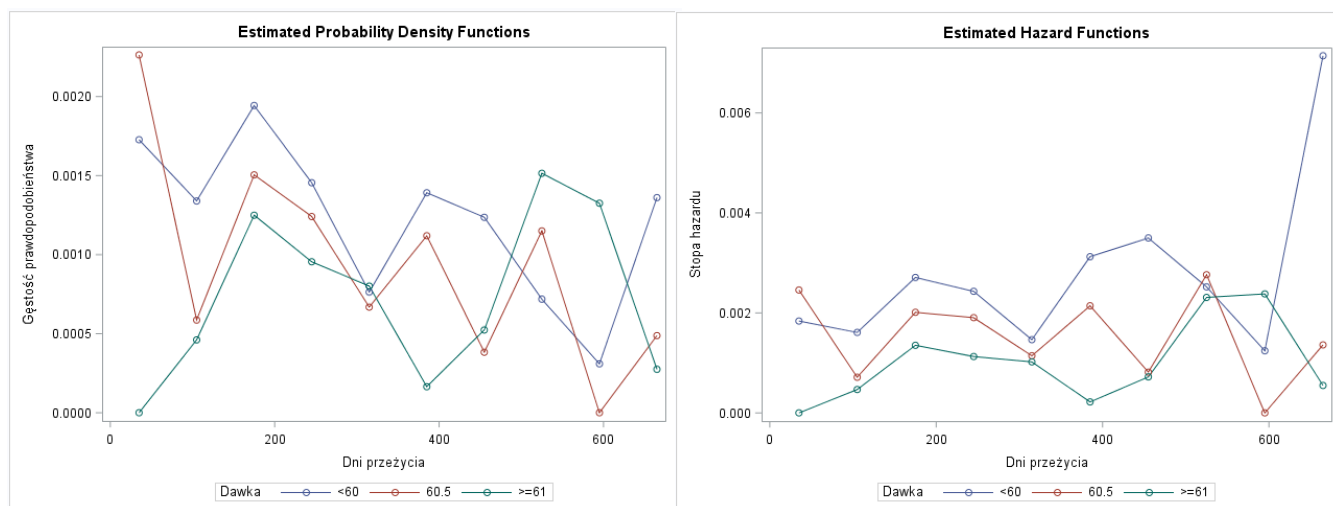


Wykres 7. Log ujemnego logu funkcji przeżycia

Rysunek pokazuje logarytm ujemnego logarytmu funkcji przeżycia. Aby sprawdzić w przybliżeniu, jak kształtuje się rozkład zmiennej czasowej, można posłużyć się poniższymi założeniami:

- Jeśli rozkład wykładniczy byłby odpowiedni, krzywa LLS powinna być linią prostą o nachyleniu równym 1;
- Jeśli zaś rozkład Weibulla byłby odpowiedni, krzywa LLS powinna być linią prostą o nachyleniu mniejszym bądź większym od 1.

W tym badaniu wykres LLS wygląda w przybliżeniu liniowo, jednak o nachyleniu mniejszym niż 1, co sugeruje graficznie, że rozważany rozkład czasu przeżycia to rozkład Weibulla – jednak na tym etapie analizy (analiza nieparametryczna) nie jesteśmy w stanie tego stwierdzić. Ze względu na wahania wykresu w końcowej fazie, być może konieczne będzie zastosowanie innego rozkładu, np. log-logistycznego lub log-normalnego.



*Wykres 8. Funkcje prawdopodobieństwa i hazardu dla metody tradycyjnej ze stratyfikacją*

Obie funkcje cechują się brakiem monotoniczności. Można jednak zauważyć pewne podobieństwa: o okresie od 100 do 300 dni, obie funkcje dla wszystkich warstw zachowują się w podobny sposób. Może to oznaczać, że w tym okresie czynniki inne niż dawka mają wpływ na proces leczenia pacjenta. Po ok. 330 dniach funkcja prawdopodobieństwa „rozchodzi się” osiągając minimum globalne dla warstwy  $>60$  mg/dzień. Następnie jednak, dla tejże warstwy funkcja szybko rośnie, osiągając swe maksimum około dnia 520, podczas gdy dla najmniejszej dawki funkcja znacząco maleje. Sugerować to może, że większe dawki mają sens tylko do około roku terapii, następnie najlepsze wydaje się zmniejszenie dawki. Wartości funkcji powyżej 600 dni są trudno interpretowalne przez małą liczebność próby w tym przedziale czasu.

### 3.3. Metoda Kaplana-Meiera

Kolejną metodą nieparametryczną analizy historii zdarzeń jest metoda Kaplana-Meiera. W odróżnieniu od metody tradycyjnej metoda ta nie wymaga arbitralnego przyjęcia punktów w czasie, w którym dokonywana jest estymacja funkcji przeżycia. Funkcja przeżycia liczona jest w każdym punkcie, w którym wystąpiło co najmniej jedno zdarzenie. Funkcja przeżycia estymowana jest w sposób następujący:

$$\hat{S}(t) = \sum_{t_j < t} \frac{n_j - d_j}{n_j},$$

gdzie  $n_j$  – liczba pacjentów, którzy nie wrócili do nałogu do momentu  $t_j$ ,  $d_j$  – liczba pacjentów, którzy powrócili do nałogu w momencie  $t_j$ .

Analiza rozpoczęta została od przeprowadzenia estymacji dla wszystkich 238 pacjentów w próbie. Z uwagi na dużą liczbę wydarzeń, tabela z wartościami funkcji przeżycia zawiera dużą liczbę informacji. Z tego powodu poniżej przedstawiono wybrane wartości tejże funkcji.

Oceny przeżycia Kaplana-Meiera					
Dni przeżycia	Przeżycie	Niepowodzenie	Błąd standardowy przeżycia	Liczba nieudanych	Liczba pozostałych
0.00	1.0000	0	0	0	238
7.00	0.9958	0.00424	0.00423	1	235
109.00	0.8882	0.1118	0.0207	26	201
204.00	0.7660	0.2340	0.0282	53	167
302.00	0.6400	0.3600	0.0324	80	136
428.00	0.5559	0.4441	0.0339	97	108
504.00	0.4977	0.5023	0.0346	108	91
612.00	0.3889	0.6111	0.0355	126	52
708.00	0.3251	0.6749	0.0362	134	38
1076.00 *	0.1362	0.8638	.	150	0

Tabela 13. Wartości funkcji przeżycia dla metody Kaplana-Meiera

#### Statystyki agregujące zmiennej czasowej Dni przeżycia

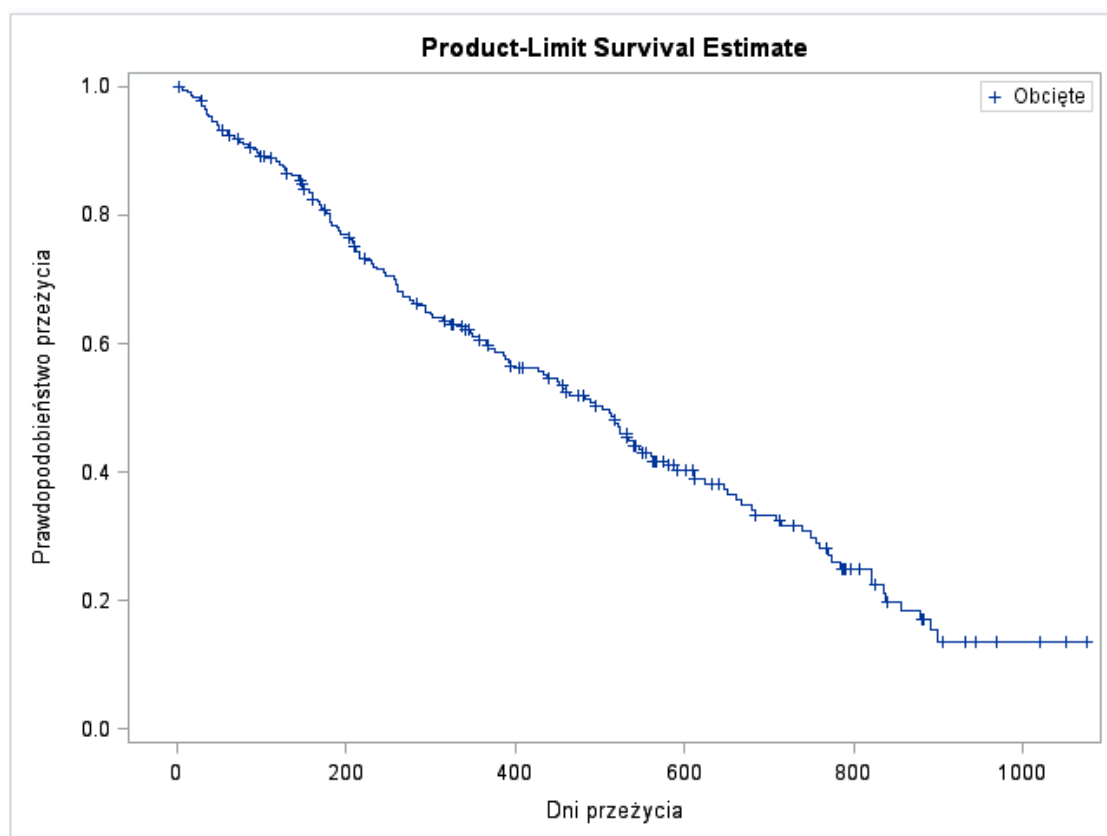
Oceny kwartylowe				
Procent	Ocena punktowa	Przedział ufności 95%		
		Transform.	[Dolna	Górna]
75	821.00	LOGLOG	714.00	878.00
50	504.00	LOGLOG	394.00	550.00
25	212.00	LOGLOG	175.00	259.00

Średnia	Błąd standardowy
494.30	20.89

Tabela 14. Statystyki agregujące dla metody Kaplana-Meiera

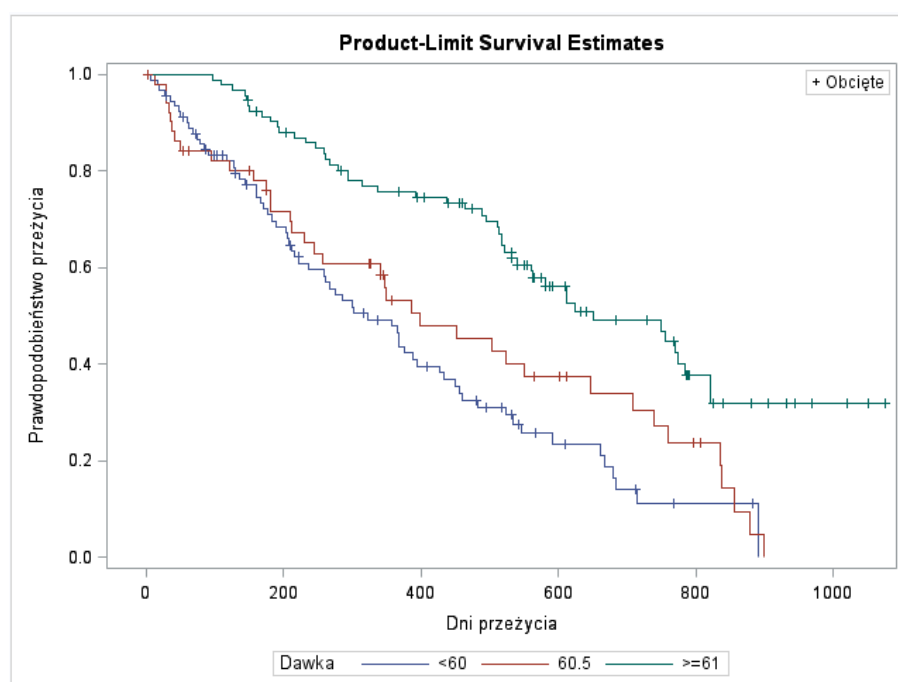
Średni czas dożycia wynosi 494 dni, średnia wynosi 504 dni. Średnia jest większa niż mediana, więc mamy do czynienia z rozkładem prawostronnie asymetrycznym.



Wykres 9. Funkcja dożycia w metodzie Kaplana-Meiera

Na podstawie wyestymowanej funkcji dożycia można stwierdzić, że prawdopodobieństwo dożycia maleje w tempie 10% na 100 dni.

Kolejnym etapem było użycie zmiennej stratyfikującej DAWKA- analogicznie do metody tradycyjnej. Wyniki testów homogeniczności okazały się być bliźniacze jak dla metody tradycyjnej, więc autor uznał ponowne publikowanie tych samych wyników za niepotrzebne.



Wykres 10. Funkcja dożycia metodą Kaplana-Meiera ze zmienną stratyfikującą



Interpretacja wyników przy użyciu zmiennej warstwującej DAWKA jest analogiczna jak przy metodzie tradycyjnej- funkcja dożycia maleje w mniejszym tempie dla wyższych dawek metadonu, a maleje szybciej dla dawek mniejszych. Nie występuje homogeniczność funkcji przeżycia w warstwach.

## 4. Porównanie modeli i wnioski

Próbka, na której budowane były analizy miała 238 obserwacji. Tak duża próba wystarczyła, aby dokonać estymacji ze stosunkowo dużą precyzją. W obu modelach błąd standardowy funkcji przeżycia dla górnych wartości zmiennej czasowej nie przekraczał wartości 8%. Świadczy to o dużej precyzji dokonanych szacunków.

Funkcja przeżycia w obu modelach miała zbliżony przebieg.

Głównym wnioskiem odkrytym dzięki stratyfikacji jest wpływ dawki metadonu na ryzyko powrotu do nałogu narkotykowego. Analiza pokazała, że większe dawki metadonu niż 60 mg/dzień korzystnie wpływają na terapię pacjenta objętego leczeniem uzależnienia od heroiny. Jednakże w dłuższym okresie czasu, najkorzystniejszą strategią jest stopniowe zmniejszanie dawki leku.

## 5. Bibliografia

1. SAS Institute; "SAS/STAT 13.2 User's Guide"; 2014
2. E.Frątczak, U.Sienkiewicz, H.Babiker; „Analiza historii zdarzeń – elementy teorii, wybrane przykłady zastosowań”; Szkoła Główna Handlowa w Warszawie; Wydanie 1 2009
3. <https://www.ncbi.nlm.nih.gov/pubmed/1988793>

## 6. Kody SAS

```
/* Importowanie zbioru ADDICTS */
DATA WORK.ADDICTS;
  LENGTH
    Klinika          3
    Status            3
    'Dni przeżycia'n  8
    'Więzienie'n     3
    Dawka             3 ;
  DROP
    ID ;
  LABEL
    Klinika          = " "
    Status            = " "
    'Dni przeżycia'n = " "
```

```

        'Więzienie'n      = " "
        Dawka              = " " ;
FORMAT
        Klinika            BEST12.
        Status             BEST12.
        'Dni przeżycia'n  BEST12.
        'Więzienie'n      BEST12.
        Dawka              BEST12. ;
INFORMAT
        Klinika            BEST12.
        Status             BEST12.
        'Dni przeżycia'n  BEST12.
        'Więzienie'n      BEST12.
        Dawka              BEST12. ;
INFILE 'C:\Users\Sebastian\AppData\Local\Temp\SEG3012\ADDICTS-
3d9e43b25aa64f6c82fe4d186a99dcc2.txt'
        LRECL=17
        ENCODING="WLATIN2"
        TERMSTR=CRLF
        DLM='7F'x
        MISSOVER
        DSD ;
INPUT
        ID                 : $1.
        Klinika            : BEST32.
        Status             : BEST32.
        'Dni przeżycia'n  : BEST32.
        'Więzienie'n      : BEST32.
        Dawka              : BEST32. ;
RUN;

/* Tabela jednoczynnikowa dla STATUS */
PROC SQL;
        CREATE VIEW WORK.SORT AS
                SELECT T.Status
                FROM WORK.ADDICTS as T
;
QUIT;

PROC FREQ DATA=WORK.SORT
        ORDER=INTERNAL
;
        TABLES Status / SCORES=TABLE;
RUN;

/* Statystyki agregujące dla STATUS i Dni przeżycia*/
PROC SQL;
        CREATE VIEW WORK.SORTTempTableSorted AS
                SELECT T."Dni przeżycia"n, T.Status
                FROM WORK.ADDICTS as T
;
QUIT;

```

```

PROC MEANS DATA=WORK.SORTTempTableSorted
    FW=12
    PRINTALLTYPES
    CHARTYPE
    QMETHOD=OS
    NWAY
    VARDEF=DF
        MEAN
        STD
        MIN
        MAX
        N
        Q1
        MEDIAN
        Q3 ;
    VAR "Dni przeżycia"n;
    CLASS Status / ORDER=UNFORMATTED ASCENDING;

RUN;
ODS GRAPHICS ON;
TITLE;
/*-----
 * Generuje histogramy za pomocą PROC UNIVARIATE..
 */

TITLE;
TITLE1 "Statystyki agregujące";
TITLE2 "Histogramy";
PROC UNIVARIATE DATA=WORK.SORTTempTableSorted NOPrint ;
    VAR "Dni przeżycia"n;

    CLASS Status ;
        HISTOGRAM ;

RUN; QUIT;
ODS GRAPHICS OFF;

/* Tabela jednoczynnikowa dla KLINIKA */

PROC SQL;

    CREATE VIEW WORK.SORTk AS
        SELECT T.Klinika
        FROM WORK.ADDICTS as T
;
QUIT;

PROC FREQ DATA=WORK.SORTk
    ORDER=INTERNAL
;
    TABLES Klinika / SCORES=TABLE;
RUN;

/* Tabela zagregowane dla KLINIKA I STATUS */
PROC TABULATE
DATA=WORK.ADDICTS

```

```

        ;

        CLASS Klinika / ORDER=UNFORMATTED MISSING;
        CLASS Status / ORDER=UNFORMATTED MISSING;
        TABLE /* Wymiar wiersza */
Klinika,
/* Wymiar kolumny */
Status*
    N
N          ;
        ;

RUN;

/* Zagregowana tabela dla zmiennych więzienie i status */

PROC TABULATE
DATA=WORK.ADDICTS

        ;

        CLASS Status / ORDER=UNFORMATTED MISSING;
        CLASS "Więzienie"n / ORDER=UNFORMATTED MISSING;
        TABLE /* Wymiar wiersza */
'Więzienie'n*
    N,
/* Wymiar kolumny */
Status          ;
        ;

RUN;

/* Rozkład zmiennej DAWKA z uwzględnieniem zmiennej STATUS */

PROC SQL;
    CREATE VIEW WORK.SORTTempTableSorted AS
        SELECT T.Dawka, T.Status
        FROM WORK.ADDICTS as T
    ;
QUIT;
Legend1
    FRAME
    ;
Axis1
    STYLE=1
    WIDTH=1
    MINOR=NONE
    ;
Axis2
    STYLE=1
    WIDTH=1
    ;
TITLE;
TITLE1 "Wykres słupkowy";
PROC GCHART DATA=WORK.SORTTempTableSorted

```

```

;
    VBAR
        Dawka
    /
    SUBGROUP=Status
    CLIPREF
FRAME TYPE=FREQ
    LEGEND=LEGEND1
    COUTLINE=BLACK
    RAXIS=AXIS1
    MAXIS=AXIS2
;

RUN; QUIT;

TITLE;

/* Współczynniki korelacji Pearsona */
PROC SQL;
    CREATE VIEW WORK.SORTTempTableCORR AS
        SELECT T.Klinika, T."Więzenie"n, T.Dawka
        FROM WORK.ADDICTS as T
;
QUIT;
TITLE1 "Analiza korelacji";
PROC CORR DATA=WORK.SORTTempTableCORR
    PLOTS=NONE
    PEARSON
    VARDEF=DF
;
    VAR Klinika "Więzenie"n Dawka;
RUN;

TITLE;

/* Tabele przeżycia metodą tradycyjną */
ODS GRAPHICS ON;

%_eg_conditional_dropds(WORK.TMP0TempTableInput,
                        WORK.TMP2TempTablePlot0);

PROC SQL;
    CREATE VIEW WORK.TMP0TempTableInput AS
        SELECT T."Dni przeżycia"n, T.Status, T.Dawka,
        T."Więzenie"n
        FROM WORK.ADDICTS as T
;
QUIT;

/*-----
    Sekcja analizy
*/

TITLE;
TITLE1 "Analiza tabel przeżycia";

```

```

/*-----
    Uruchom PROC LIFETEST, aby przeprowadzić analizę.
*/
PROC LIFETEST DATA=WORK.TMP0TempTableInput
    ALPHA=0.05
    METHOD=LIFE
    INTERVALS= 70 140 210 280 350 420 490 560 630 700
    PLOTS(ONLY)=SURVIVAL( CL )
    PLOTS(ONLY)=LOGSURV
    PLOTS(ONLY)=LOGLOGS
    PLOTS(ONLY)=HAZARD
    PLOTS(ONLY)=DENSITY
;
    TEST "Więzienie"n;
    TIME "Dni przeżycia"n * Status (0);

RUN; TITLE;
RUN; QUIT;
%_eg_conditional_dropds(WORK.TMP0TempTableInput,
                        WORK.TMP2TempTablePlot0);
TITLE;
ODS GRAPHICS OFF;

/* Tabele przeżycia metodą tradycyjną ze stratyfikacją */
ODS GRAPHICS ON;
%_eg_conditional_dropds(WORK.TMP0TempTableInput,
                        WORK.TMP2TempTablePlot0);

PROC SQL;
    CREATE VIEW WORK.TMP0TempTableInput AS
        SELECT T."Dni przeżycia"n, T.Status, T.Dawka,
T."Więzienie"n
        FROM WORK.ADDICTS as T
;
QUIT;

TITLE;
TITLE1 "Analiza tabel przeżycia";

/*-----
    Uruchom PROC LIFETEST, aby przeprowadzić analizę.
*/
PROC LIFETEST DATA=WORK.TMP0TempTableInput
    ALPHA=0.05
    METHOD=LIFE
    INTERVALS= 70 140 210 280 350 420 490 560 630 700
    PLOTS(ONLY)=SURVIVAL( CL )
    PLOTS(ONLY)=LOGSURV
    PLOTS(ONLY)=LOGLOGS
    PLOTS(ONLY)=HAZARD
    PLOTS(ONLY)=DENSITY
;
    STRATA Dawka      ( 60 61 );
    TEST "Więzienie"n;

```

```

        TIME "Dni przeżycia"n * Status (0);

RUN; TITLE;
RUN; QUIT;
%_eg_conditional_dropds(WORK.TMP0TempTableInput,
                        WORK.TMP2TempTablePlot0);
TITLE;
ODS GRAPHICS OFF;

/* Metoda Kaplana-Meiera */

ODS GRAPHICS ON;

%_eg_conditional_dropds(WORK.TMP0TempTableInput,
                        WORK.TMP2TempTablePlot0);

PROC SQL;
    CREATE VIEW WORK.TMP0TempTableInput AS
        SELECT T."Dni przeżycia"n, T.Status, T.Dawka
        FROM WORK.ADDICTS as T;
QUIT;

TITLE;
TITLE1 "Analiza tabel przeżycia";
PROC LIFETEST DATA=WORK.TMP0TempTableInput
    ALPHA=0.05
    INTERVALS= 100 200 300 400 500 600 700 PLOTS (ONLY)=SURVIVAL
    PLOTS (ONLY)=LOGLOGS;
    TIME "Dni przeżycia"n * Status (0);
RUN; TITLE;
QUIT;
%_eg_conditional_dropds(WORK.TMP0TempTableInput,
                        WORK.TMP2TempTablePlot0);
ODS GRAPHICS OFF;

/* Metoda Kaplana-Meiera ze stratyfikacją */

ODS GRAPHICS ON;

%_eg_conditional_dropds(WORK.TMP0TempTableInput,
                        WORK.TMP2TempTablePlot0);

PROC SQL;
    CREATE VIEW WORK.TMP0TempTableInput AS
        SELECT T."Dni przeżycia"n, T.Status, T.Dawka
        FROM WORK.ADDICTS as T;
QUIT;

TITLE;
TITLE1 "Analiza tabel przeżycia";
PROC LIFETEST DATA=WORK.TMP0TempTableInput
    ALPHA=0.05

```

```

INTERVALS= 100 200 300 400 500 600 700 PLOTS (ONLY)=SURVIVAL
PLOTS (ONLY)=LOGLOGS;
STRATA Dawka      ( 60 61);
TIME "Dni przeżycia"n * Status (0);

RUN;TITLE;
QUIT;
%_eg_conditional_dropds(WORK.TMP0TempTableInput,
                        WORK.TMP2TempTablePlot0);
ODS GRAPHICS OFF;

```