

Temat projektu: Modele semiparametryczne

Spis treści

1. Wstęp	3
2. Analiza zbioru.....	3
2.1. Model Coxa hazardów proporcjonalnych.....	3
2.2. Weryfikacja założenia proporcjonalnych hazardów modelu Coxa	5
2.3. Weryfikacja założenia o liniowości	6
2.4. Usunięcie obserwacji odstających.....	6
3. Remodelowanie	7
3.1. Model hazardów proporcjonalnych po usunięciu obserwacji odstających	8
3.2. Weryfikacja założenia proporcjonalnych hazardów po remodelowaniu	9
3.3. Weryfikacja założenia o liniowości po remodelowaniu	9
4. Bibliografia.....	9
5. Kody SAS.....	10

1. Wstęp

Zbiór danych został opisany w poprzednich referatach autora (modele parametryczne i nieparametryczne).

W tym referacie autor przeprowadzi najpierw analizę hazardów proporcjonalnych za pomocą modelu Coxa, zinterpretuje wyniki oraz zweryfikuje założenia o liniowości i proporcjonalności hazardów. Następnie, za pomocą reszt ocen, usunie obserwacje odstające i wykona ponowną analizę zbioru danych bez obserwacji odstających.

Analiza zostanie wykonana za pomocą procedury PHREG w programie SAS. Procedura PHREG służy do estymacji modeli semiparametrycznych – modeli regresji Coxa, nazywanych najczęściej modelami proporcjonalnego hazardu. Metoda zaproponowana przez Coxa nie wymaga wyboru konkretnego rozkładu prawdopodobieństwa danych przeżycia (modele parametryczne tego wymagają). Możliwe jest również włączenie do analizy zmiennych zależnych od czasu. Procedura PHREG informuje o jakości dopasowania modelu do danych dzięki charakterystykom takim jak: - $2\ln(L)$, kryterium AIC, oraz kryterium SBC. Wskazane jest kierowanie się kryterium SBC, ponieważ kryterium AIC przy dużych zbiorach danych często wybiera modele ze zbyt dużą ilością zmiennych objaśniających. Testowana jest również hipoteza o statystycznej nieistotności parametrów. W tym celu przeprowadza się test stosunku wiarygodności, test Walda oraz test punktowy. Wszystkie trzy statystyki mają rozkład Chi-kwadrat z p stopniami swobody, gdzie p jest wymiarem wektora szacowanych parametrów.

Ze względu na występowanie danych połączonych, autor zdecydował się na użycie metody Efrona konstrukcji funkcji częściowej wiarygodności.

2. Analiza zbioru

2.1. Model Coxa hazardów proporcjonalnych

Procedura PHREG			
Informacje o modelu			
Zbiór	WORK.ADDICTS		
Zmienna zależna	Dni przeżycia		
Zmienna obciążenia	Status		
Wartości obciążenia	0		
Rangi wiązane	EFRON		
Wczytano obserwacji		238	
Użyto obserwacji		238	
Podsumowanie liczby zdarzeń i wartości obciążonych			
Suma	Zdarzenie	Obciążone	Procent obciążonych
238	150	88	36.97

Statystyki dopasowania							
Kryterium	Bez zmiennych towarzyszących			Ze zmiennymi towarzyszącymi			
-2 LOG L	1411.079			1346.518			
AIC	1411.079			1352.518			
SBC	1411.079			1361.550			

Testowanie globalnej hipotezy zerowej: BETA=0			
Test	Chi-kwadrat	DF	Pr. > chi-kw.
Iloraz wiaryg.	64.5603	3	<.0001
Ocena	56.3155	3	<.0001
Wald	54.1196	3	<.0001

Analiza ocen maksymalnej wiarygodności							
Parametr	DF	Ocena parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi-kw.	Współczynnik hazardu	Przedział ufności współczynnika hazardu 95%
Klinika	1	-1.00979	0.21488	22.0830	<.0001	0.364	0.239 0.555
Dawka	1	-0.03537	0.00638	30.7430	<.0001	0.965	0.953 0.977
Więzenie	1	0.32654	0.16722	3.8131	0.0509	1.386	0.999 1.924

Tabela 1. Statystyki modelu Coxa proporcjonalnych hazardów

Statystyki dopasowania oceniają jakość dopasowania modelu do danych, bazując na metodzie największej wiarygodności. Model dąży do minimalizacji tych wartości. Statystyki te mogą być oceniane podczas porównywania różnych modeli.

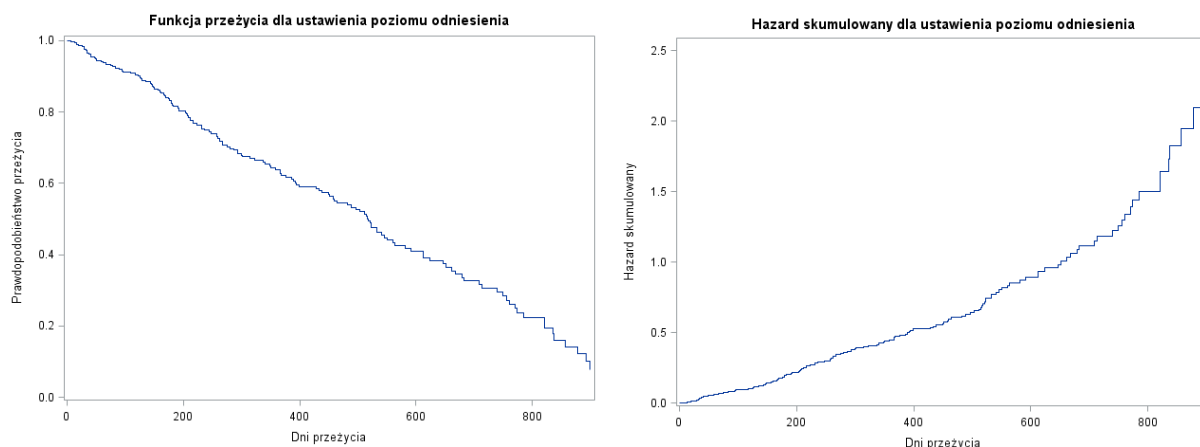
Tabela zatytułowana „Testowanie globalnej hipotezy zerowej BETA = 0” przedstawia wartości statystyki Chi-kwadrat, liczbę stopni swobody i wartości p-value dla testów: ilorazu wiarygodności, oceny oraz Walda. Hipoteza zerowa w każdym teście mówi o tym, że wszystkie zmienne są nieistotne statystycznie. Hipoteza alternatywna, że minimum jedna zmienna jest istotna statystycznie. Na przyjętym poziomie istotności $\alpha=0,05$ należy odrzucić hipotezę zerową na rzecz hipotezy alternatywnej:

- H0: każda zmienna jest nieistotna statystycznie
- H1: co najmniej jedna zmienna jest istotna statystycznie

Wartość współczynnika hazardu przy zmiennej KLINIKA wynosi 0,36 co oznacza, że u pacjentów z kliniki 2 ryzyko zajścia zdarzenia jest o 64% mniejsze w stosunku do pacjentów z kliniki 1. Zmienna jest istotna statystycznie – na poziomie istotności $\alpha=0,05$ i wartości p-value <0,0001 należy odrzucić hipotezę zerową mówiącą o braku istotności statystycznej zmiennej na rzecz hipotezy alternatywnej mówiącej, że zmienna jest statystycznie istotna.

Wartość współczynnika hazardu przy zmiennej DAWKA wynosi 0,965, zatem ryzyko wystąpienia zdarzenia wraz ze zwiększeniem dawki metadonu o 1 mg/dzień jest o 3,5% mniejsze. Zmienna jest istotna statystycznie na poziomie istotności $\alpha=0,05$.

Wartość współczynnika hazardu przy zmiennej WIĘZIENIE wynosi 1,386, co oznacza, że u pacjentów, którzy w przeszłości przebywali w więzieniu ryzyko zajścia zdarzenia rośnie o 39%, w porównaniu z tymi, którzy w więzieniu nigdy nie byli. Zmienna balansuje na granicy istotności- autor ustawił poziom istotności dla pozostania w modelu (selekcja krokowa) na 5%, zatem zmienne, których p-value jest większe niż 0,05, powinny zostać usunięte z modelu. Wartość p-value dla zmiennej WIĘZIENIE wynosi 0,0509 a mimo to SAS nie usunął tej zmiennej. Zatem jeśli system nie usunął tej zmiennej, autor uznaje ją za istotną statystycznie.



Wykres 1. Funkcja przeżycia i hazard skumulowany modelu hazardów proporcjonalnych

2.2. Weryfikacja założenia proporcjonalnych hazardów modelu Coxa

Analiza ocen maksymalnej wiarygodności						
Parametr	DF	Ocena parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi-kw.	Współczynnik hazardu
Klinika	1	0.02689	0.34895	0.0059	0.9386	1.027
Klinika_t	1	-0.00305	0.0009621	10.0417	0.0015	0.997
Dawka	1	-0.04403	0.01129	15.2124	<.0001	0.957
Dawka_t	1	0.0000270	0.0000282	0.9208	0.3373	1.000
Więzenie	1	0.44056	0.29714	2.1984	0.1382	1.554
Więzenie_t	1	-0.0001443	0.0007498	0.0371	0.8473	1.000

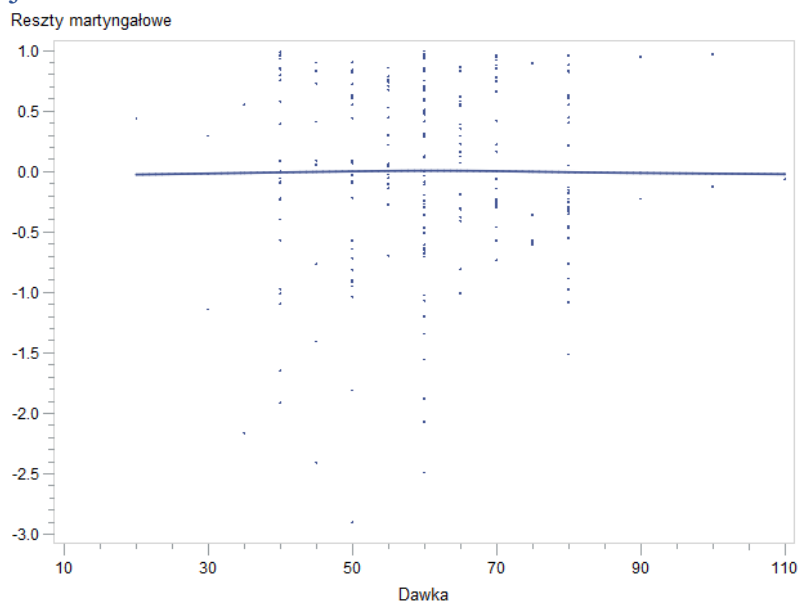
Tabela 2. Analiza ocen parametrów modelu ze zmiennymi zależnymi od czasu

W celu zweryfikowania założenia o proporcjonalności hazardów wykorzystano metodę polegającą na stworzeniu zmiennych będących interakcją zmiennych z modelu ze zmienną czasową oraz ocena istotności nowo zdefiniowanych zmiennych.

Zmienne Dawka_t i Więzenie_t są nieistotne statystycznie na każdym przyjętym poziomie istotności (wartości p-value odpowiednio: 0,3373; 0,8473) co oznacza, że zmienne DAWKA oraz WIĘZIENIE nie podlegają zmianom w czasie czyli założenie o proporcjonalności hazardów dla tych zmiennych jest spełnione.

Zmienna Klinika_t jest istotna statystycznie, zatem KLINIKA podlega zmianom w czasie a więc założenie o proporcjonalności hazardów dla tej zmiennej nie jest spełnione. Zatem, globalne założenie proporcjonalności hazardów nie jest spełnione.

2.3. Weryfikacja założenia o liniowości

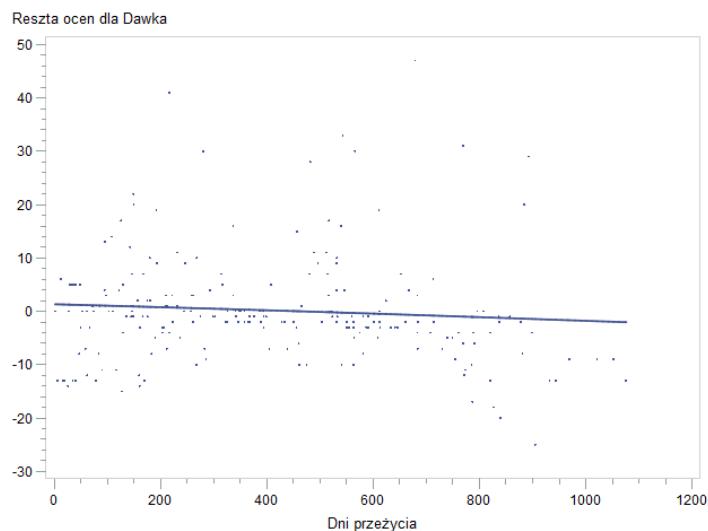


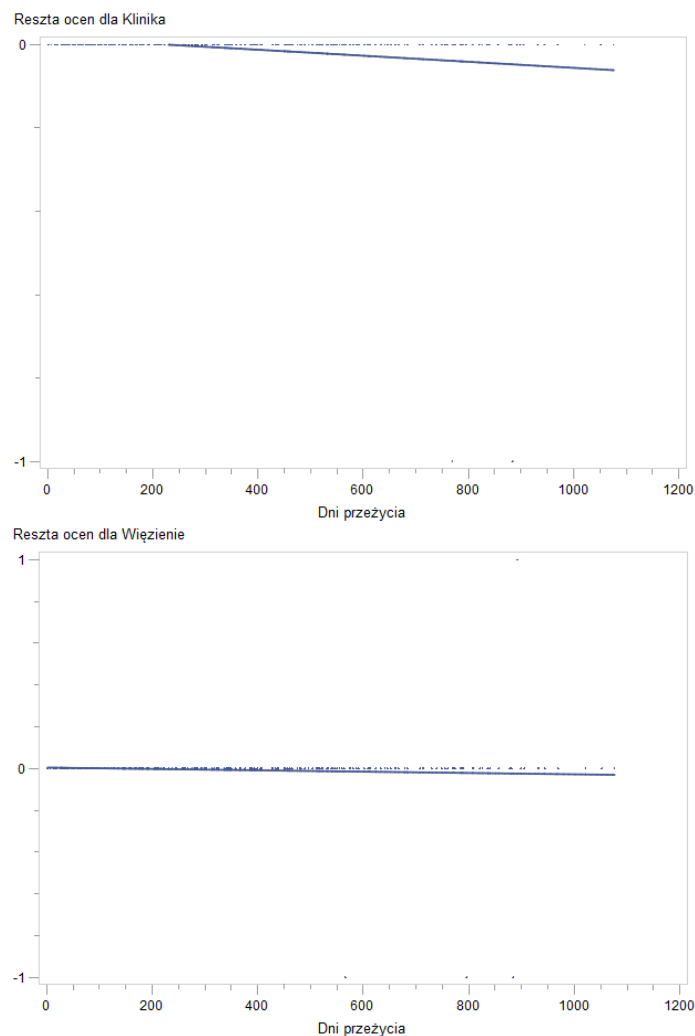
Wykres 2. Wykres reszt martyngałowych dla zmiennej DAWKA

Na podstawie wykresu autor stwierdza, że założenie o liniowości zostało spełnione. Wartości reszt martyngałowych dla zmiennej DAWKA oscylują wokół zera i są reprezentowane przez krzywą charakteryzującą się trendem liniowym.

2.4. Usunięcie obserwacji odstających

Identyfikacja: autor zdecydował, aby obserwacje odstające zidentyfikować za pomocą graficznej analizy reszt ocen dla każdej ze zmiennych. Poniżej przedstawiono wykresy reszt ocen dla zmiennych DAWKA, KLINIKA i WIĘZIENIE.





Wykres 3. Reszta ocen dla zmiennych DAWKA, KLINIKA i WIĘZIENIE

Po przeanalizowaniu powyższych wykresów, autor zdecydował się na następujące ograniczenia reszt ocen:

- dla DAWKA- od -20 do +20
- dla KLINIKA- więcej niż -0,5
- dla WIĘZIENIE- od -0,5 do 0,5

3. Remodelowanie

Następnie, autor przeprowadził ponowną analizę hazardów proporcjonalnych. Nowy zbiór charakteryzował się o 100 mniejszą liczebnością obserwacji i podobnym odsetkiem obserwacji odciętych.

3.1. Model hazardów proporcjonalnych po usunięciu obserwacji odstających

Informacje o modelu	
Zbiór	WORK.SCORE1
Zmienna zależna	Dni przeżycia
Zmienna obciążenia	Status
Wartości obciążenia	0
Rangi wiązane	EFRON

Wczytano obserwacji	186
Użyto obserwacji	186

Podsumowanie liczby zdarzeń i wartości obciążonych			
Suma	Zdarzenie	Obciążone	Procent obciążonych
186	117	69	37.10

Statystyki dopasowania		
Kryterium	Bez zmiennych towarzyszących	Ze zmiennymi towarzyszącymi
-2 LOG L	1027.627	900.409
AIC	1027.627	906.409
SBC	1027.627	914.696

Testowanie globalnej hipotezy zerowej: BETA=0			
Test	Chi-kwadrat	DF	Pr. > chi-kw.
Iloraz wiaryg.	127.2175	3	<.0001
Ocena	94.0518	3	<.0001
Wald	78.8520	3	<.0001

Analiza ocen maksymalnej wiarygodności							
Parametr	DF	Ocena parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi-kw.	Współczynnik hazardu	Przedział ufności współczynnika hazardu 95%
Klinika	1	-2.58285	0.51494	25.1590	<.0001	0.076	0.028 0.207
Dawka	1	-0.06042	0.00853	50.1378	<.0001	0.941	0.926 0.957
Więzienie	1	0.56276	0.19079	8.6997	0.0032	1.756	1.208 2.552

Tabela 3. Wyniki remodelowania modelu hazardów proporcjonalnych

Statystyki dopasowania są wyraźnie niższe niż w modelu pierwotnym, zatem ten model jest znacząco lepiej dopasowany. Testy globalnej hipotezy zerowej pokazują że na poziomie istotności 5% co najmniej jedna zmienna jest istotna statystycznie. Analiza ocen parametrów ukazuje, że na poziomie istotności 5% każda ze zmiennych jest istotna statystycznie. Szczególną poprawę widać dla zmiennej WIĘZIENIE.

Interpretacja współczynnika hazardu:

- KLINIKA- 0,076 czyli u pacjentów z kliniki 2 ryzyko zajścia zdarzenia jest o 93% mniejsze niż u pacjentów z kliniki 1
- DAWKA- 0,941 czyli ryzyko wystąpienia zdarzenia wraz ze zwiększeniem dawki metadonu o 1 mg/dzień jest o 6% mniejsze
- WIĘZIENIE- 1,756 czyli u pacjentów, którzy w przeszłości przebywali w więzieniu ryzyko zajścia zdarzenia rośnie o 76%, w porównaniu z tymi, którzy w więzieniu nigdy nie byli

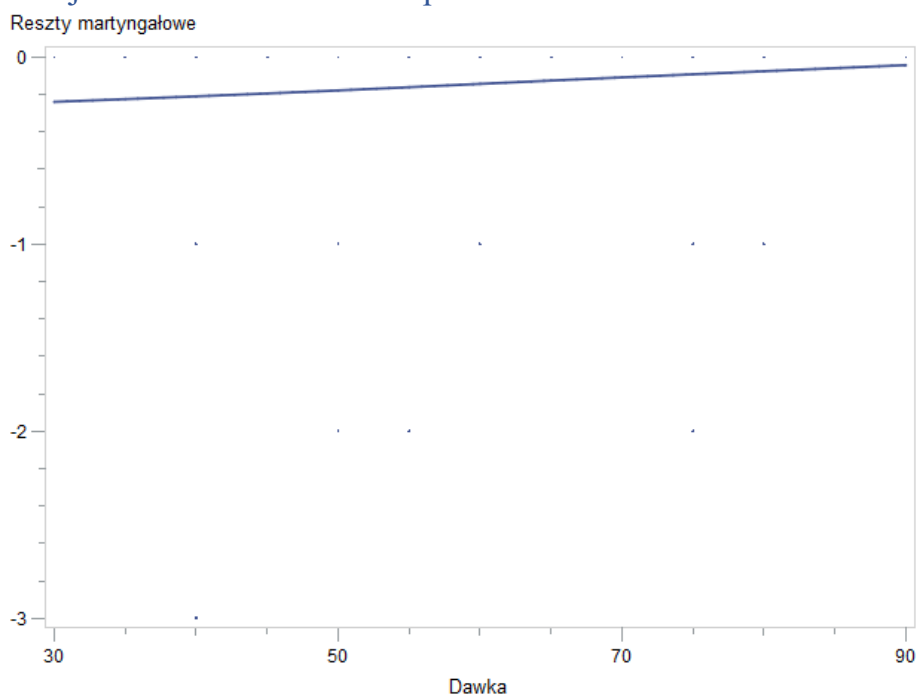
3.2. Weryfikacja założenia proporcjonalnych hazardów po remodelowaniu

Analiza ocen maksymalnej wiarygodności						
Parametr	DF	Ocena parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi-kw.	Współczynnik hazardu
Klinika	1	-2.95832	1.26720	5.4500	0.0196	0.052
Klinika_t	1	0.0007013	0.00228	0.0949	0.7581	1.001
Dawka	1	-0.05143	0.01466	12.3075	0.0005	0.950
Dawka_t	1	-0.0000296	0.0000403	0.5397	0.4625	1.000
Więzenie	1	0.45178	0.34277	1.7371	0.1875	1.571
Więzenie_t	1	0.0003240	0.0008660	0.1400	0.7083	1.000

Tabela 4. Analiza ocen parametrów modelu ze zmiennymi zależnymi od czasu po remodelowaniu

Powyższa tabela ukazuje, że żadna ze zmiennych zależnych od czasu nie jest istotna statystycznie na poziomie istotności 5%. Zatem, w tym wypadku globalne założenie proporcjonalności hazardów jest spełnione.

3.3. Weryfikacja założenia o liniowości po remodelowaniu



Wykres 4. Reszty martynałowe po remodelowaniu

Na podstawie powyższego wykresu autor stwierdza, że założenie o liniowości zostało spełnione. Wartości reszt martynałowych oscylują wokół zera i są reprezentowane przez krzywą charakteryzującą się trendem liniowym.

4. 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>

5. Kody SAS

```
TITLE1 "Analiza hazardów proporcjonalnych";
```

```
/* Uruchamia PROC PHREG i wykonuje analizę.*/
```

```
PROC PHREG DATA=WORK.ADDICTS
```

```
    PLOTS=SURVIVAL
```

```
    PLOTS=CUMHAZ
```

```
;
```

```
    MODEL "Dni przeżycia"n * Status (0) = Klinika Dawka "Więzenie"n /
```

```
        TIES=EFRON
```

```
        RISKLIMITS ALPHA=0.05
```

```
        SELECTION=STEPWISE
```

```
        SLENTY=0.05
```

```
        SLSTAY=0.05
```

```
        INCLUDE=3
```

```
;
```

```
RUN;TITLE;
```

```
TITLE1 "Weryfikacja modelu hazardów proporcjonalnych";
```

```
PROC PHREG DATA=WORK.ADDICTS;
```

```
    MODEL "Dni przeżycia"n * Status (0) = Klinika Klinika_t Dawka Dawka_t "Więzenie"n
```

```
    Więzenie t /
```

```
        TIES=EFRON;
```

```
        Klinika_t = Klinika*"Dni przeżycia"n;
```

```
        Dawka_t = Dawka*"Dni przeżycia"n;
```

```
        Więzenie_t = "Więzenie"n*"Dni przeżycia"n;
```

```
RUN;TITLE;
```

```
proc phreg data=WORK.ADDICTS;
```

```
/*weryfikacja założenia o liniowości dla zmiennej dawka metodą reszt martyngałowych*/
```

```
model "Dni przeżycia"n * Status (0)=dawka klinika "Więzenie"n/risklimits ties=efron;
```

```
output out=dane_resz_m RESMART=reszty_m;
```

```
id dawka;
```

```
run;
```

```
proc gplot data=dane_resz_m;
```

```
/*wykres reszt martyngałowych*/
```

```
plot reszty_m*dawka/vaxis=axis1 haxis=axis2;
```

```
symbol1 v=point width=2 i=sm90s;
```

```
symbol2 v=point;
```

```
axis1 label=('Reszty martyngałowe');
```

```
axis2 label=('Dawka') order=(10 30 50 70 90 110) offset=(1,1);
```

```
run;
```

```
/*Estymacja modelu z wykresem reszt ocen */
```

```
proc phreg data = WORK.ADDICTS;
```

```
model "Dni przeżycia"n * Status (0)=dawka klinika "Więzenie"n / ties = efron;
```

```
output out = score
```

```
ressco = sc_dawka sc_klinika sc_wiez /* reszty ocen */
```

```
resmart = resmart; /* reszty martyngałowe */
```

```
id dawka;
```

```
run;
```

```
proc gplot data = score;
```

```
plot sc_dawka*"Dni przeżycia"n;
```

```

plot sc_klinika*"Dni przeżycia"n;
plot sc_wiez*"Dni przeżycia"n;
run;

/*Usuwanie obserwacji odstających */
PROC SQL;
    CREATE VIEW WORK.score1 AS
        SELECT "Dni przeżycia"n, Status, Klinika, Dawka, "Więzenie"n, sc_dawka,
sc_klinika, sc_wiez
    FROM score
    where sc_dawka between -20 and 20
        AND sc_klinika > -0.5
        AND sc_wiez between -0.5 and 0.5
;
TITLE1 "Analiza hazardów proporcjonalnych";
PROC PHREG DATA=WORK.score1;
    MODEL "Dni przeżycia"n * Status (0) = Klinika Dawka "Więzenie"n /
        TIES=EFRON
        RISKLIMITS ALPHA=0.05
        SELECTION=STEPWISE
        SLENTY=0.05
        SLSTAY=0.05
        INCLUDE=3;
RUN;TITLE;

TITLE1 "Weryfikacja modelu hazardów proporcjonalnych";
PROC PHREG DATA=WORK.score1;
    MODEL "Dni przeżycia"n * Status (0) = Klinika Klinika_t Dawka Dawka_t "Więzenie"n
Więzenie_t /
        TIES=EFRON;
    Klinika_t = Klinika*"Dni przeżycia"n;
    Dawka_t = Dawka*"Dni przeżycia"n;
    Więzenie_t ="Więzenie"n*"Dni przeżycia"n;
RUN;TITLE;

proc phreg data=WORK.score1;
/*weryfikacja założenia o liniowości dla zmiennej dawka metodą reszt martyngałowych*/
model "Dni przeżycia"n * Status (0)=dawka klinika "Więzenie"n/risklimits ties=efron;
output out=dane_resz_m2 RESMART=reszty_m2;
id dawka;
run;
proc gplot data=dane_resz_m2;
/*wykres reszt martyngałowych*/
plot reszty_m2*dawka/vaxis=axis1 haxis=axis2;
symbol1 v=point width=2 i=sm90s mode=include;
symbol2 v=point;
axis1 label=('Reszty martyngałowe');
axis2 label=('Dawka') order=(30 50 70 90) offset=(1,1);
run;
QUIT;

```