**Temat projektu:** Modele semiparametryczne

Autor: Sebastian Boruch

Spis treści

[1. Wstęp 2](#_Toc511043979)

[2. Analiza zbioru 3](#_Toc511043980)

[2.1. Model Coxa hazardów proporcjonalnych 3](#_Toc511043981)

[2.2. Weryfikacja założenia proporcjonalnych hazardów modelu Coxa 5](#_Toc511043982)

[2.3. Weryfikacja założenia o liniowości 5](#_Toc511043983)

[2.4. Usunięcie obserwacji odstających 6](#_Toc511043984)

[3. Remodelowanie 7](#_Toc511043985)

[3.1. Model hazardów proporcjonalnych po usunięciu obserwacji odstających 7](#_Toc511043986)

[3.2. Weryfikacja założenia proporcjonalnych hazardów po remodelowaniu 8](#_Toc511043987)

[3.3. Weryfikacja założenia o liniowości po remodelowaniu 8](#_Toc511043988)

[4. Bibliografia 9](#_Toc511043989)

[5. Kody SAS 9](#_Toc511043990)

# 1. Wstęp

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

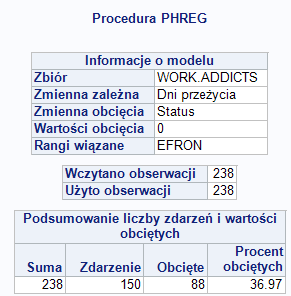
W tym referacie autor przeprowadzi najpierw analizę hazardów proporcjonalnych za pomocą modelu Coxa, zinterpretuje wyniki oraz zweryfikuje założenia o liniowości i proporcjonalności hazardów. Następnie, za pomocą reszt ocen, usunie obserwacje odstające i wykona ponowną analizę zbioru danych bez obserwacji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

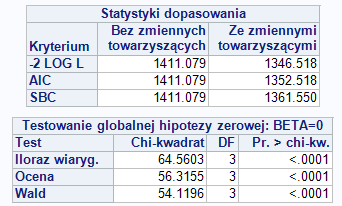
Analiza zostanie wykonana za pomocą procedury PHREG w programie SAS. Procedura PHREG służy do estymacji modeli semiparametrycznych – modeli regresji Coxa, nazywanych najczęściej modelami proporcjonalnego hazardu. Metoda zaproponowana przez Coxa nie wymaga wyboru konkretnego rozkładu prawdopodobieństwa 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: -2ln(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.

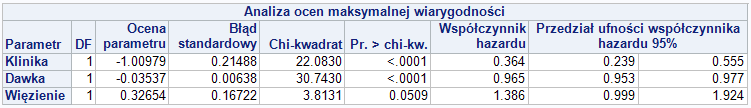
Ze względu na występowanie danych połączonych, autor zdecydował się na użycie metody Efrona konstrukcji funkcji częściowej wiarygodności.

# 2. Analiza zbioru

## 2.1. Model Coxa hazardów proporcjonalnych







*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.

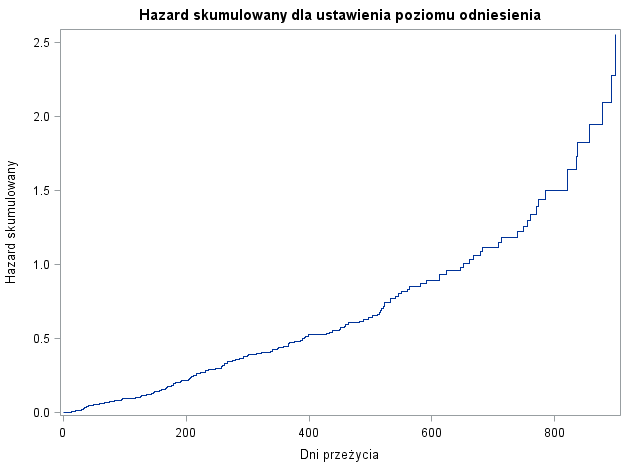
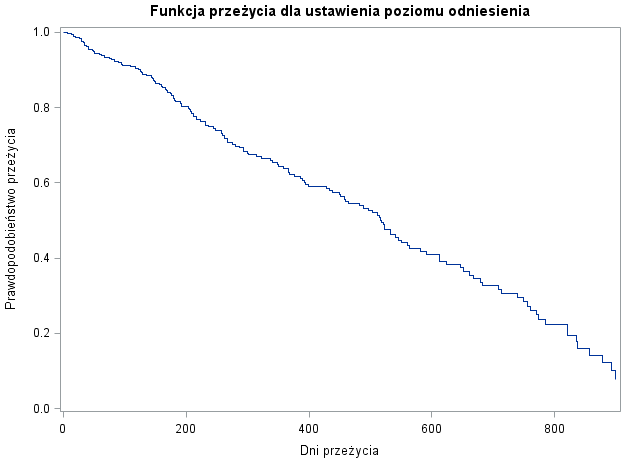
Tabela zatytułowana ,,Testowanie globalnej hipotezy zerowej BETA = 0” przedstawia wartości statystyki Chi-kwadrat, liczbę stopni swobody i wartości p-value dla testów: ilorazu wiarygodności, oceny oraz Walda. Hipoteza zerowa w każdym teście mówi o tym, że wszystkie zmienne są nieistotne statystycznie. Hipoteza alternatywna, że minimum jedna zmienna jest istotna statystycznie. Na przyjętym poziomie istotności 𝛼=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 𝛼=0,05 i wartości p-value <0,0001 należy odrzucić hipotezę zerowa mówiącą o braku istotności statystycznej zmiennej na rzecz hipotezy alternatywnej mówiącej, że zmienna jest statystycznie istotna.

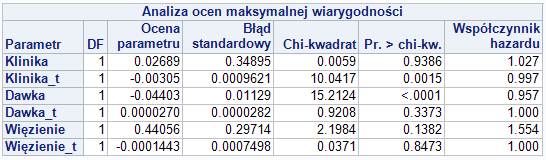
Wartość współczynnika hazardu przy zmiennej DAWKA wynosi 0,965, zatem ryzyko wystąpienia zdarzenia wraz ze zwiększeniem dawki metadonu o 1 mg/dzień jest o 3,5% mniejsze. Zmienna jest istotna statystycznie na poziomie istotności 𝛼=0,05.

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



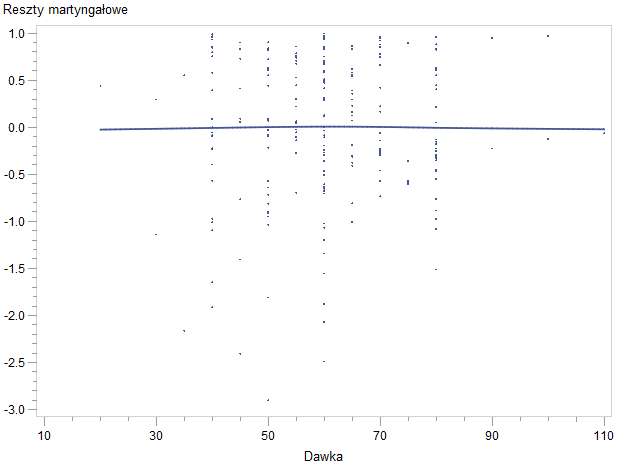
*Tabela 2. Analiza ocen parametrów modelu ze zmiennymi zależnymi od czasu*

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 Więzienie\_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

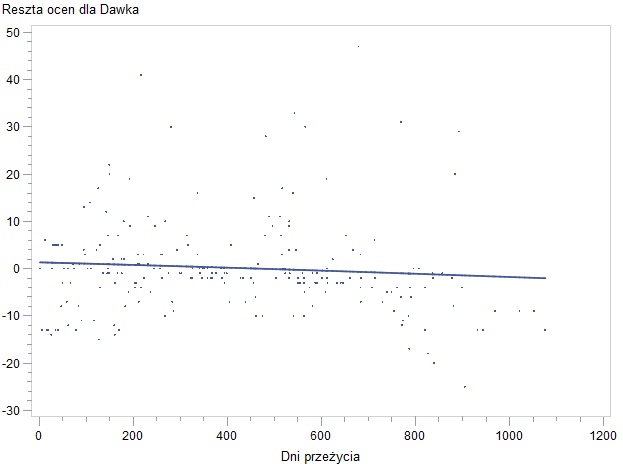


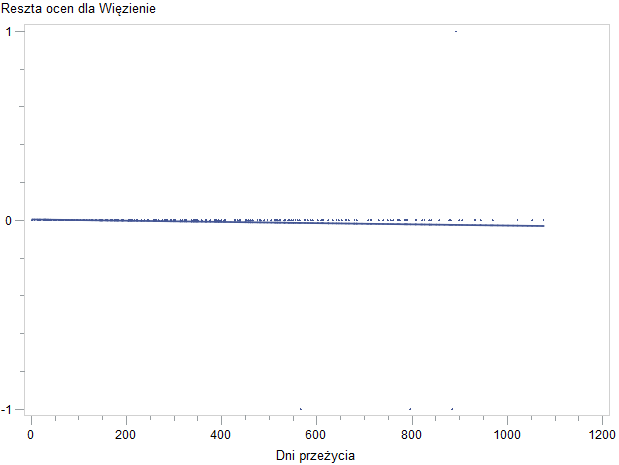
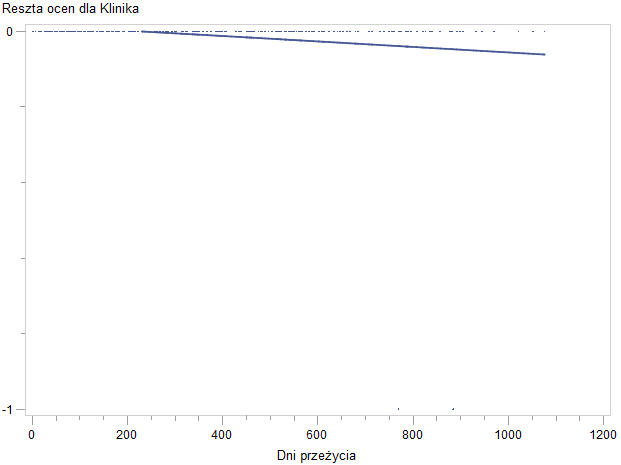
*Wykres 2. Wykres reszt martyngałowych dla zmiennej DAWKA*

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*

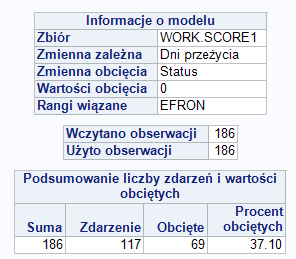
Po przeanalizowaniu powyższych wykresów, autor zdecydował się na następujące ograniczenia reszt ocen:

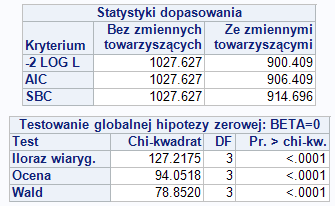
* dla DAWKA- od -20 do +20
* dla KLINIKA- więcej niż -0,5
* dla WIĘZIENIE- od -0,5 do 0,5

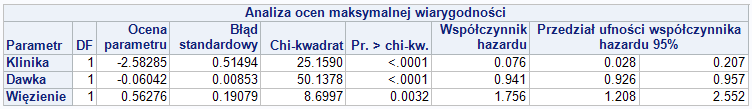
# 3. Remodelowanie

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







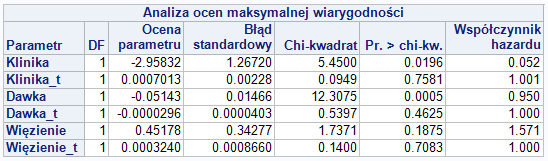
*Tabela 3. Wyniki remodelowania modelu hazardów proporcjonalnych*

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

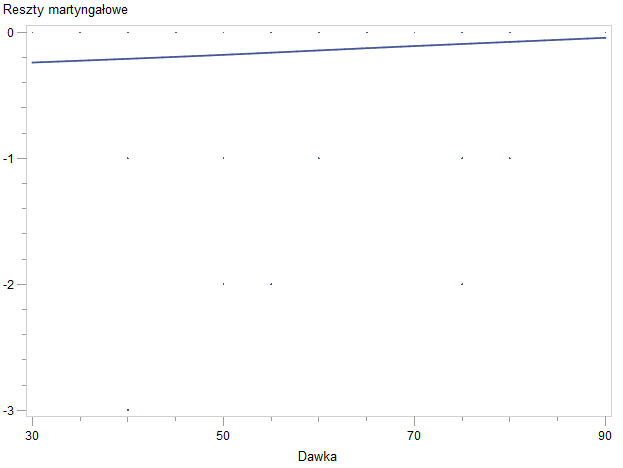
## 3.2. Weryfikacja założenia proporcjonalnych hazardów po remodelowaniu



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

Powyższa tabela ukazuje, że żadna za 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 martyngałowe po remodelowaniu*

Na podstawie powyższego wykresu autor stwierdza, że założenie o liniowości zostało spełnione. Wartości reszt martyngałowych 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”; Szkołą 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 "Więzienie"n /

TIES=EFRON

RISKLIMITS ALPHA=**0.05**

SELECTION=STEPWISE

SLENTRY=**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 "Więzienie"n Więzienie\_t /

TIES=EFRON;

Klinika\_t = Klinika\*"Dni przeżycia"n;

Dawka\_t = Dawka\*"Dni przeżycia"n;

Więzienie\_t ="Więzienie"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 "Więzienie"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 "Więzienie"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, "Więzienie"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 "Więzienie"n /

TIES=EFRON

RISKLIMITS ALPHA=**0.05**

SELECTION=STEPWISE

SLENTRY=**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 "Więzienie"n Więzienie\_t /

TIES=EFRON;

Klinika\_t = Klinika\*"Dni przeżycia"n;

Dawka\_t = Dawka\*"Dni przeżycia"n;

Więzienie\_t ="Więzienie"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 "Więzienie"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**;