

# Analiza przeżycia pacjentek u których zdiagnozowano raka piersi

Ewelina Kowalewska  
Zuzanna Waluś

# Opis zbioru df.metabric

Profile kliniczne 2509 pacjentek z rakiem piersi.

W zbiorze danych są wymienione dwa zdarzenia. Natomiast Analiza przeżycia została dokonana tylko dla Overall Survival Status. Okres trwania związany ze zdarzeniem to Overall Survival Status (Months).

Każda obserwacja ma swój PatientID.

Liczba zmiennych niezależnych: 26

Źródło danych:

<https://www.kaggle.com/datasets/gunesevitan/breast-cancer-metabric>

# Zakodowanie zmiennych

**Column: Oncotree Code**

Original BRCA => Encoded 0  
Original BREAST => Encoded 1  
Original IDC => Encoded 2  
Original ILC => Encoded 3  
Original IMMC => Encoded 4  
Original MBC => Encoded 5  
Original MDLC => Encoded 6  
Original PBS => Encoded 7

**Column: PR Status**

Original Negative => Encoded 0  
Original Positive => Encoded 1

**Column: Radio Therapy**

Original No => Encoded 0  
Original Yes => Encoded 1

**Column: Sex**

Original Female => Encoded 0

**Column: 3-Gene classifier subtype**

Original ER+/HER2- High Prolif => Encoded 0  
Original ER+/HER2- Low Prolif => Encoded 1  
Original ER-/HER2- => Encoded 2  
Original HER2+ => Encoded 3

**Column: HER2 Status**

Original Negative => Encoded 0  
Original Positive => Encoded 1

**Column: Tumor Other Histologic Subtype**

Original Ductal/NST => Encoded 0  
Original Lobular => Encoded 1  
Original Medullary => Encoded 2  
Original Metaplastic => Encoded 3  
Original Mixed => Encoded 4  
Original Mucinous => Encoded 5  
Original Other => Encoded 6  
Original Tubular/ cribriform => Encoded 7

**Column: Hormone Therapy**

Original No => Encoded 0  
Original Yes => Encoded 1

**Column: Inferred Menopausal State**

Original Post => Encoded 0  
Original Pre => Encoded 1

**Column: Primary Tumor Laterality**

Original Left => Encoded 0  
Original Right => Encoded 1

**Column: Cellularity**

Original High => Encoded 0  
Original Low => Encoded 1  
Original Moderate => Encoded 2

**Column: Chemotherapy**

Original No => Encoded 0  
Original Yes => Encoded 1

**Column: Pam50Claudinlowsubtype**

Original Basal => Encoded 0  
Original Her2 => Encoded 1  
Original LumA => Encoded 2  
Original LumB => Encoded 3  
Original NC => Encoded 4  
Original Normal => Encoded 5  
Original claudin-low => Encoded 6

**Column: ER status measured by IHC**

Original Negative => Encoded 0  
Original Positive => Encoded 1

**Column: ER Status**

Original Negative => Encoded 0  
Original Positive => Encoded 1

**Column: Type of Breast Surgery**

Original Breast Conserving => Encoded 0  
Original Mastectomy => Encoded 1

**Column: Cancer Type**

Original Breast Cancer => Encoded 0  
Original Breast Sarcoma => Encoded 1

**Column: Cancer Type Detailed**

Original Breast => Encoded 0  
Original Breast Angiosarcoma => Encoded 1  
Original Breast Invasive Ductal Carcinoma => Encoded 2  
Original Breast Invasive Lobular Carcinoma => Encoded 3  
Original Breast Invasive Mixed Mucinous Carcinoma => Encoded 4  
Original Breast Mixed Ductal and Lobular Carcinoma => Encoded 5  
Original Invasive Breast Carcinoma => Encoded 6  
Original Metaplastic Breast Cancer => Encoded 7

# Cel i zakres projektu

Celem projektu jest obserwacja przeżycia pacjentek z rakiem piersi w czasie oraz ustalenie które zmienne mają największy wpływ na przeżycie.

W tym celu wykorzystano metody:

- nieparametryczne
- semiparametryczne
- parametryczne

# Modele nieparametryczne

# Modele nieparametryczne

**Informacje uzyskane z analizy nieparametrycznej mogą pomóc w decydowaniu o dalszych krokach analizy. Metoda Kaplana-Meiera i tablice trwania życia są używane do estymacji funkcji przeżycia, co daje wizualizację i kwantyfikację czasu do zdarzenia.**

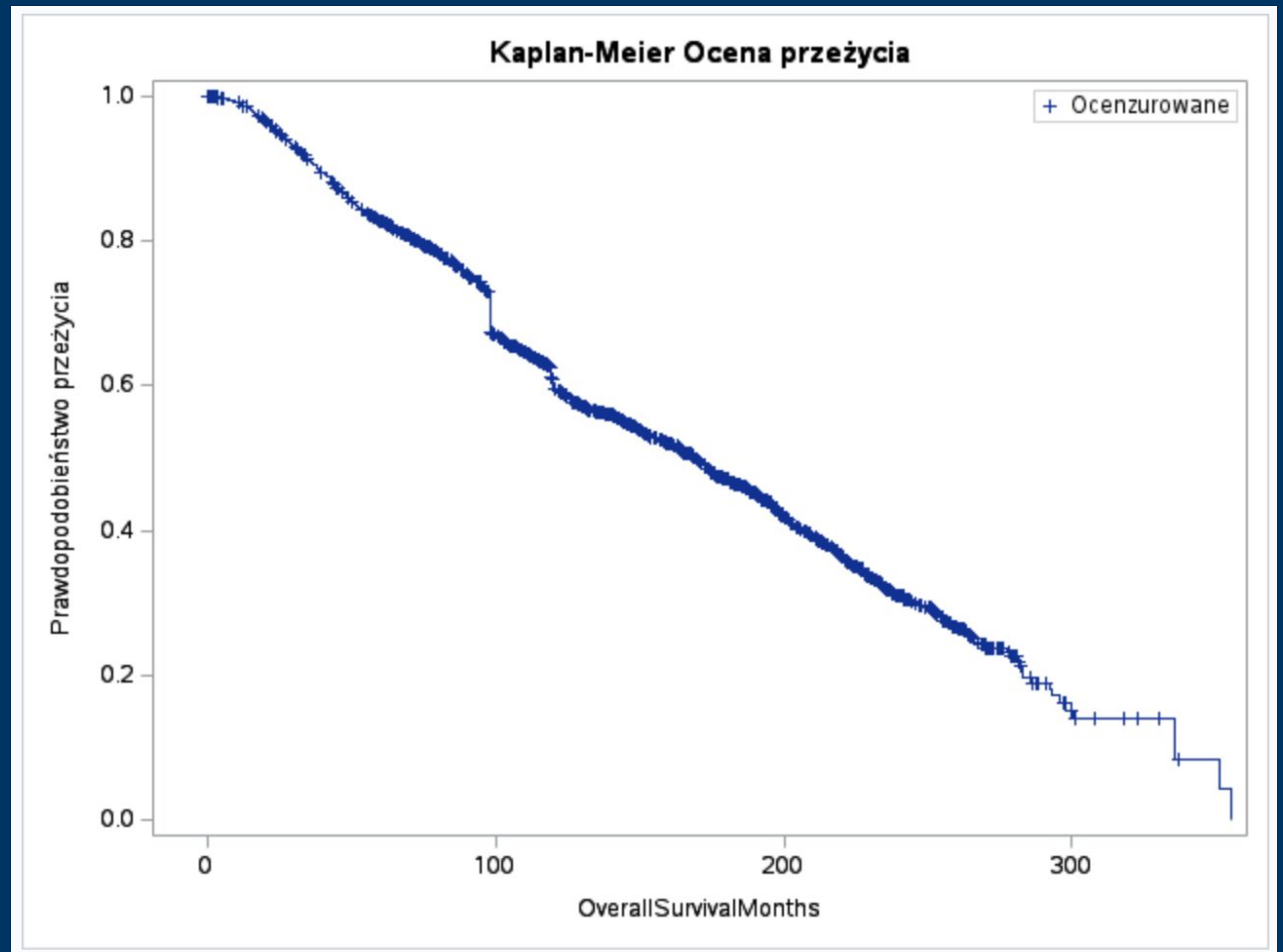
**Techniki te pomogą w zrozumieniu struktury danych, zidentyfikowaniu trendów, anomalii, wzorców i zależności między zmiennymi. W tym kontekście, mogą pomóc w identyfikacji istotnych zmiennych, które mogą wpływać na czas przeżycia.**



# Ocena przeżycia Kaplana Meiera

Analiza została zrobiona na podstawie dwóch zmiennych:  
*OverallSurvivalMonths* - zmienna czasu oraz  
*OverallSurvivalStatus* - opisującej wystąpienie zdarzenia (śmierć).

Przedstawiony wykres funkcji przeżycia kończy się na ostatniej obserwacji nieodciętej. W punkcie, gdzie funkcja przeżycia osiąga wartość 0.



# Statystyki agregujące

Wartość “Nie powiodło” mówi nam o tym ile obserwacji doświadczyło zdarzenia. W naszym przypadku zdarzeniem jest śmierć pacjentki.

Liczba obserwacji obciętych wynosi 1166.

Statystyki agregujące zmiennej czasowej OverallSurvivalMonths				
Oceny kwartylowe				
Procent	Ocena punktowa	Przedział ufności 95%		
		Przekształcenie	[Dolna	Górna)
75	265.567	LINEAR	255.000	282.300
50	169.233	LINEAR	159.700	176.033
25	91.100	LINEAR	86.067	96.833

Średnia	Błąd standardowy
176.237	2.863

Sumarycznie wartości ocenzone i nieocenzone			
Suma	Nie powiodło się	Ocenzurowane	Procent ocenzone
2509	1343	1166	46.47

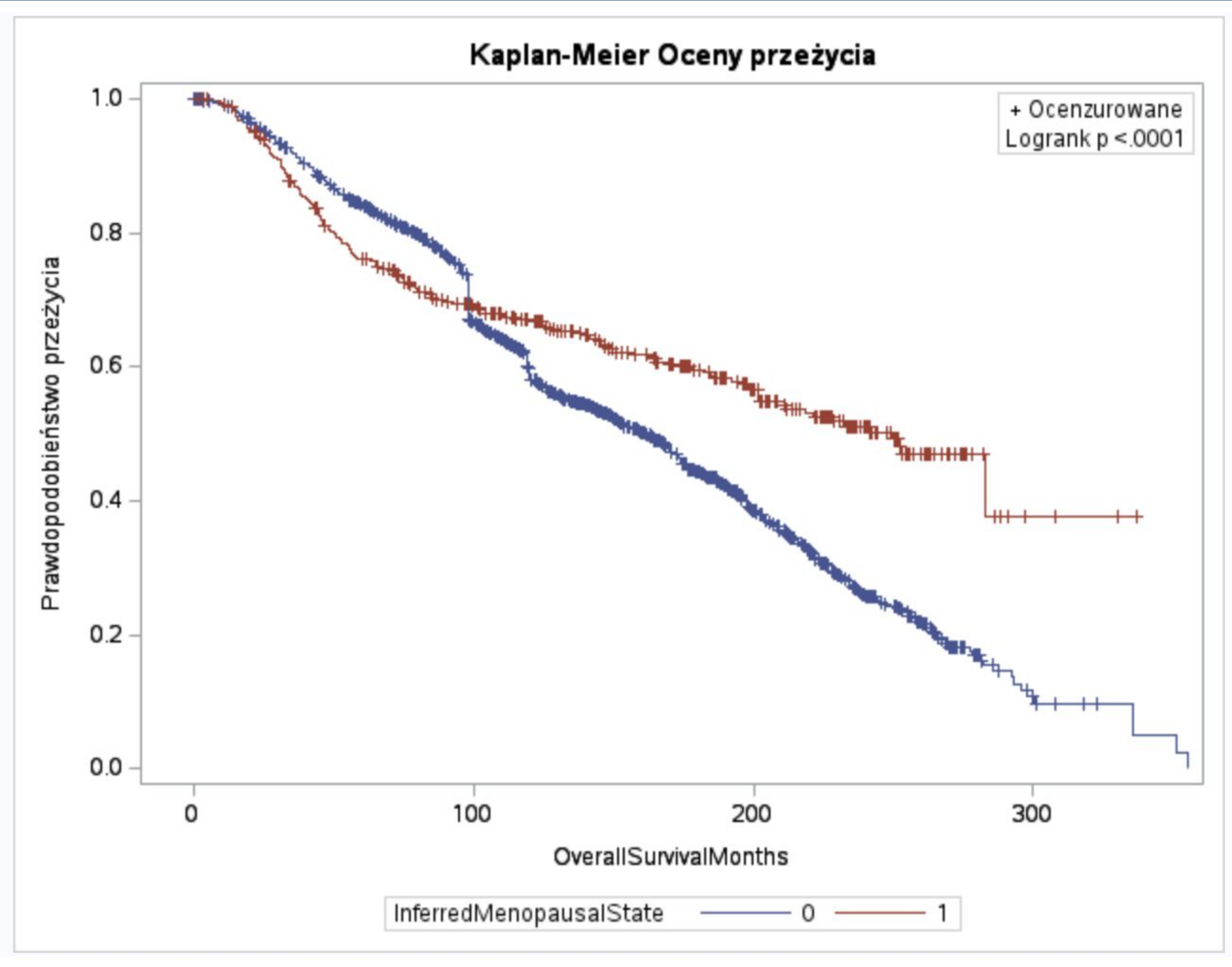


Test różnic funkcji przeżycia pomiędzy grupami dla:

- InferredMenopausalState
- Cellularity

# InferredMenopausalState

Procedura LIFETEST			
Testowanie homogeniczności krzywych przeżycia dla OverallSurvivalMonths w warstwach			
Statystyki rangowe			
InferredMenopausalState	Log. rang	Wilcoxon	
0	70.943	34437	
1	-70.943	-34437	
Macierz kowariancji dla statystyk log. rang			
InferredMenopausalState	0	1	
0	196.923	-196.923	
1	-196.923	196.923	
Macierz kowariancji dla statystyk Wilcoxona			
InferredMenopausalState	0	1	
0	5.2197E8	-5.22E8	
1	-5.22E8	5.2197E8	
Test równości w warstwach			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kwadrat
Log. rang	25.5578	1	<.0001
Wilcoxon	2.2720	1	0.1317
-2Log(LR)	20.6462	1	<.0001



# Cellularity

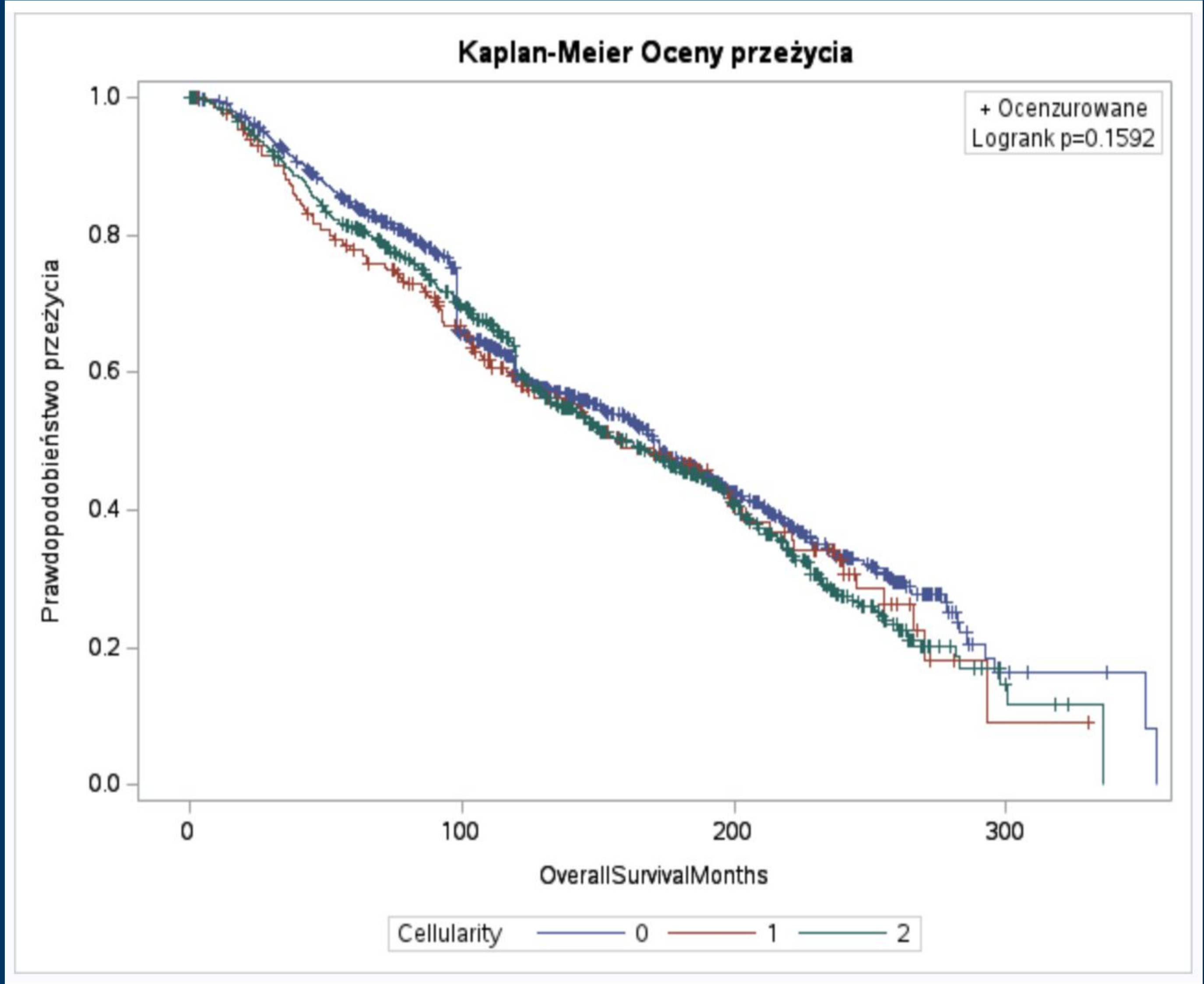
Testowanie homogeniczności krzywych przeżycia dla OverallSurvivalMonths w warstwach

Statystyki rangowe		
Cellularity	Log. rang	Wilcoxon
0	-34.145	-47564
1	8.199	20921
2	25.946	26643

Macierz kowariancji dla statystyk log. rang			
Cellularity	0	1	2
0	319.606	-62.511	-257.096
1	-62.511	98.264	-35.753
2	-257.096	-35.753	292.849

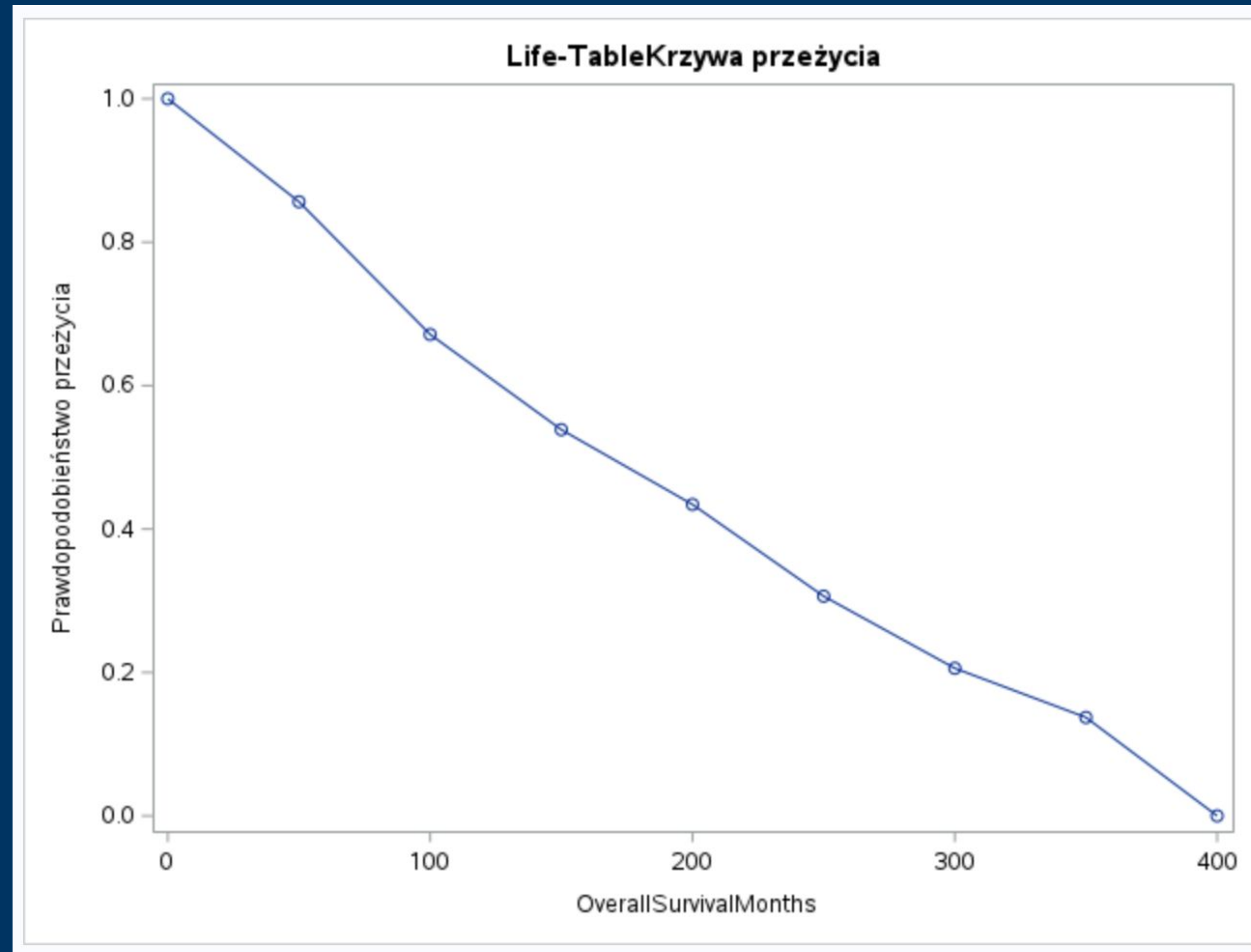
Macierz kowariancji dla statystyk Wilcoxona			
Cellularity	0	1	2
0	9.2378E8	-1.834E8	-7.404E8
1	-1.834E8	2.7968E8	-9.631E7
2	-7.404E8	-9.631E7	8.3672E8

Test równości w warstwach			
Testowanie	Chi-kwadrat	DF	Pr. > chi-kwadrat
Log. rang	3.6747	2	0.1592
Wilcoxon	2.9907	2	0.2242
-2Log(LR)	4.5993	2	0.1003

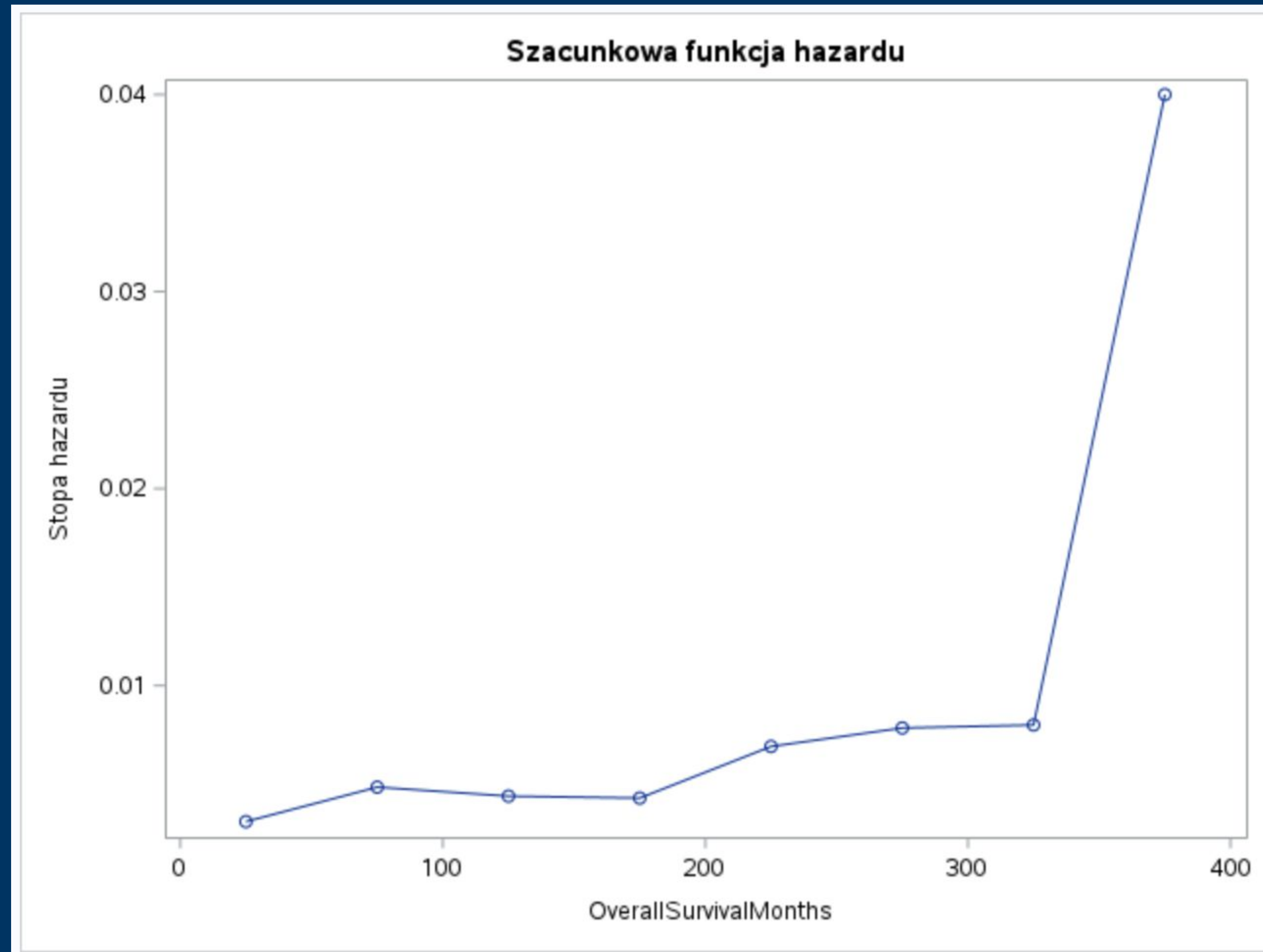




# Metoda aktuarialna - prawdopodobieństwo przeżycia



# Metoda aktuarialna - funkcja hazardu



Wraz ze wzrostem OverallSurvivalMonths  
ryzyko śmierci zwiększa się.



# Jednowymiarowy test Wilcoxona

Jednowymiarowy test Wilcoxona został przeprowadzony na modelu ze wszystkimi zmiennymi. Test ten sprawdza istotność zmiennych w modelu dla wszystkich jednocześnie.

Na podstawie wyników jednowymiarowego testu Wilcoxona możemy odrzucić z modelu zmienne dla których p-value przyjmuje wartość powyżej 0,05.

Postępująco-krokowa sekwencja chi-kwadratów dla testu Wilcoxona						
Zmienna	DF	Chi-kwadrat	Pr. > chi-kwadrat	Przyrost chi-kwadrat	Pr. > przyrost	Etykieta
Lymphnodesexaminedpositive	1	207.3	<.0001	207.3	<.0001	Lymphnodesexaminedpositive
AgeatDiagnosis	2	296.9	<.0001	89.6689	<.0001	AgeatDiagnosis
Chemotherapy	3	352.5	<.0001	55.5773	<.0001	Chemotherapy
TumorStage	4	390.8	<.0001	38.3283	<.0001	TumorStage
HER2statusmeasuredbySNP6	5	428.1	<.0001	37.2822	<.0001	HER2statusmeasuredbySNP6
PRStatus	6	451.7	<.0001	23.5416	<.0001	PRStatus
TumorSize	7	467.9	<.0001	16.2414	<.0001	TumorSize
RadioTherapy	8	481.3	<.0001	13.4228	0.0002	RadioTherapy
ERStatus	9	497.4	<.0001	16.0619	<.0001	ERStatus
Nottinghamprognosticindex	10	507.7	<.0001	10.3092	0.0013	Nottinghamprognosticindex
Gene classifiers subtype	11	516.1	<.0001	8.4142	0.0037	Gene classifiers subtype
HormoneTherapy	12	522.1	<.0001	5.9661	0.0146	HormoneTherapy
HER2Status	13	527.7	<.0001	5.6261	0.0177	HER2Status
OncotreeCode	14	531.6	<.0001	3.8692	0.0492	OncotreeCode
TumorOtherHistologicSubtype	15	535.6	<.0001	4.0217	0.0449	TumorOtherHistologicSubtype
PrimaryTumorLaterality	16	537.3	<.0001	1.7552	0.1852	PrimaryTumorLaterality
NeoplasmHistologicGrade	17	538.8	<.0001	1.4396	0.2302	NeoplasmHistologicGrade
Cellularity	18	540.0	<.0001	1.2428	0.2649	Cellularity
Pam50Claudinlowsubtype	19	541.3	<.0001	1.3107	0.2523	Pam50Claudinlowsubtype
CancerType	20	542.2	<.0001	0.8491	0.3568	CancerType
ERstatusmeasuredbyIHC	21	543.0	<.0001	0.7910	0.3738	ERstatusmeasuredbyIHC
Cohort	22	543.2	<.0001	0.1895	0.6634	Cohort
InferredMenopausalState	23	543.4	<.0001	0.2132	0.6443	InferredMenopausalState
TypeofBreastSurgery	24	543.4	<.0001	0.0727	0.7875	TypeofBreastSurgery
CancerTypeDetailed	25	543.5	<.0001	0.0683	0.7939	CancerTypeDetailed
MutationCount	26	543.5	<.0001	0.00560	0.9403	MutationCount



# Postępująco-krokowa sekwencja chi-kwadratów dla testu Wilcoxona

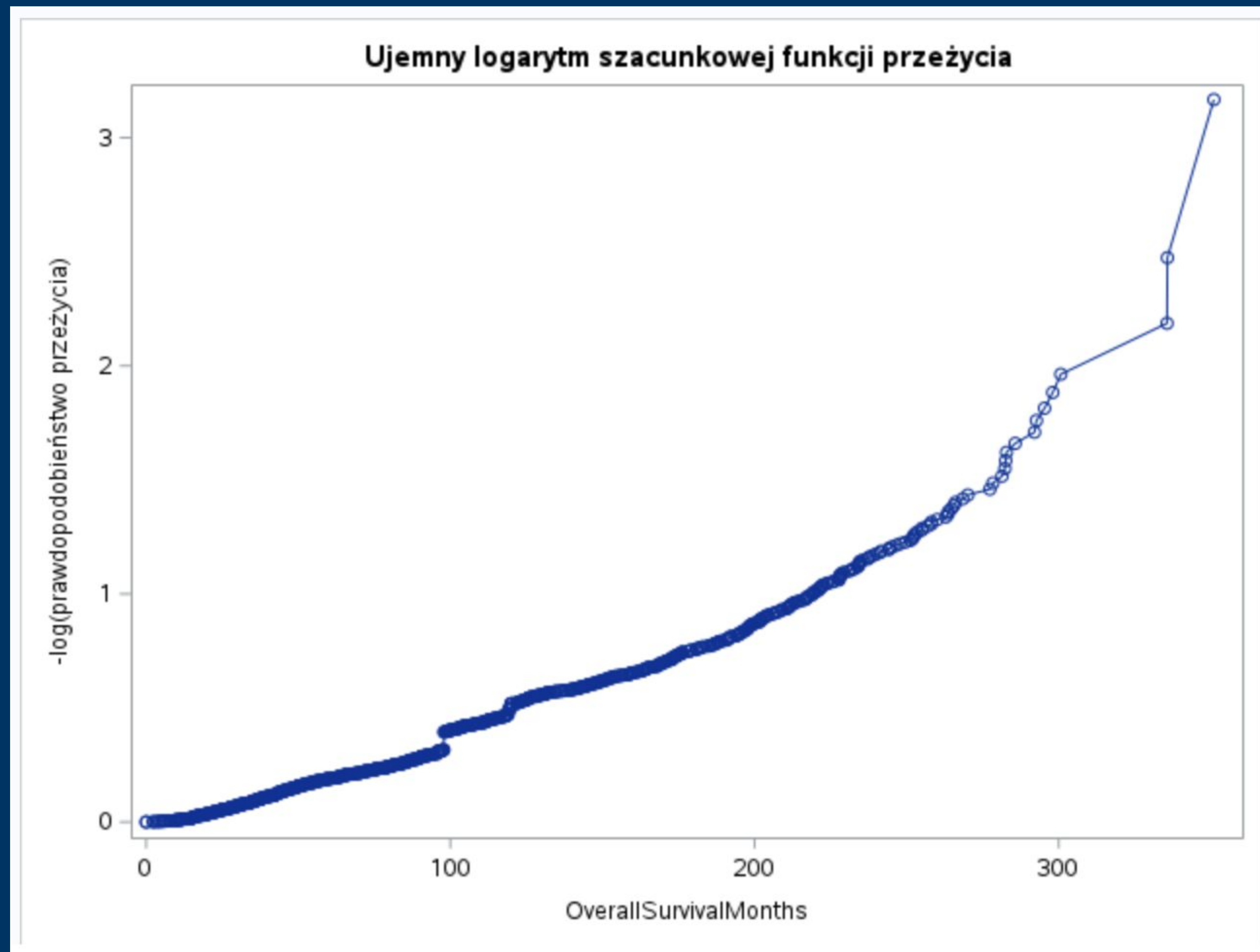
Test Wilcoxona z postępująco-krokową sekwencją chi-kwadratów wprowadza zmienne do modelu stopniowo.

Na podstawie wyników tego testu możemy odrzucić zmienne, dla których p-value przyjmuje wartość powyżej 0,05

Postępująco-krokowa sekwencja chi-kwadratów dla testu Wilcoxona						
Zmienna	DF	Chi-kwadrat	Pr. > chi-kwadrat	Przyrost chi-kwadrat	Pr. > przyrost	Etykieta
Lymphnodesexaminedpositive	1	207.3	<.0001	207.3	<.0001	Lymphnodesexaminedpositive
AgeatDiagnosis	2	296.9	<.0001	89.6689	<.0001	AgeatDiagnosis
Chemotherapy	3	352.5	<.0001	55.5773	<.0001	Chemotherapy
TumorStage	4	390.8	<.0001	38.3283	<.0001	TumorStage
HER2statusmeasuredbySNP6	5	428.1	<.0001	37.2822	<.0001	HER2statusmeasuredbySNP6
PRStatus	6	451.7	<.0001	23.5416	<.0001	PRStatus
TumorSize	7	467.9	<.0001	16.2414	<.0001	TumorSize
RadioTherapy	8	481.3	<.0001	13.4228	0.0002	RadioTherapy
ERStatus	9	497.4	<.0001	16.0619	<.0001	ERStatus
Nottinghamprognosticindex	10	507.7	<.0001	10.3092	0.0013	Nottinghamprognosticindex
Gene classifiers subtype	11	516.1	<.0001	8.4142	0.0037	Gene classifiers subtype
HormoneTherapy	12	522.1	<.0001	5.9661	0.0146	HormoneTherapy
HER2Status	13	527.7	<.0001	5.6261	0.0177	HER2Status
OncotreeCode	14	531.6	<.0001	3.8692	0.0492	OncotreeCode
TumorOtherHistologicSubtype	15	535.6	<.0001	4.0217	0.0449	TumorOtherHistologicSubtype
PrimaryTumorLaterality	16	537.3	<.0001	1.7552	0.1852	PrimaryTumorLaterality
NeoplasmHistologicGrade	17	538.8	<.0001	1.4396	0.2302	NeoplasmHistologicGrade
Cellularity	18	540.0	<.0001	1.2428	0.2649	Cellularity
Pam50Claudinlowsubtype	19	541.3	<.0001	1.3107	0.2523	Pam50Claudinlowsubtype
CancerType	20	542.2	<.0001	0.8491	0.3568	CancerType
ERstatusmeasuredbyIHC	21	543.0	<.0001	0.7910	0.3738	ERstatusmeasuredbyIHC
Cohort	22	543.2	<.0001	0.1895	0.6634	Cohort
InferredMenopausalState	23	543.4	<.0001	0.2132	0.6443	InferredMenopausalState
TypeofBreastSurgery	24	543.4	<.0001	0.0727	0.7875	TypeofBreastSurgery
CancerTypeDetailed	25	543.5	<.0001	0.0683	0.7939	CancerTypeDetailed
MutationCount	26	543.5	<.0001	0.00560	0.9403	MutationCount

# Wybór modelu

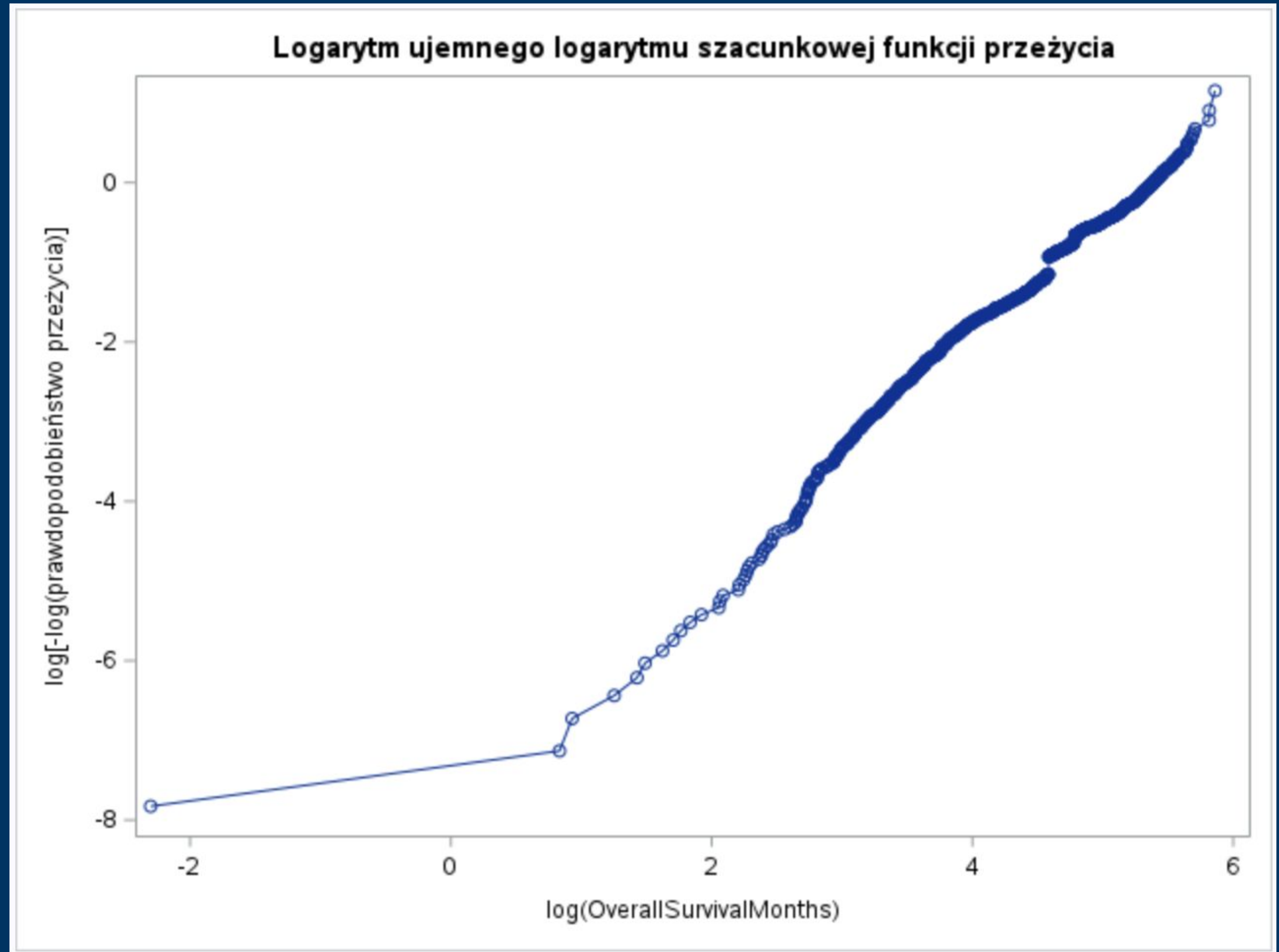
Wykres jest zbliżony do postaci liniowej oraz wychodzi z punktu 0 co mogłoby sugerować wykres wykładniczy natomiast funkcja hazardu nie jest stała w czasie.





# Wybór modelu

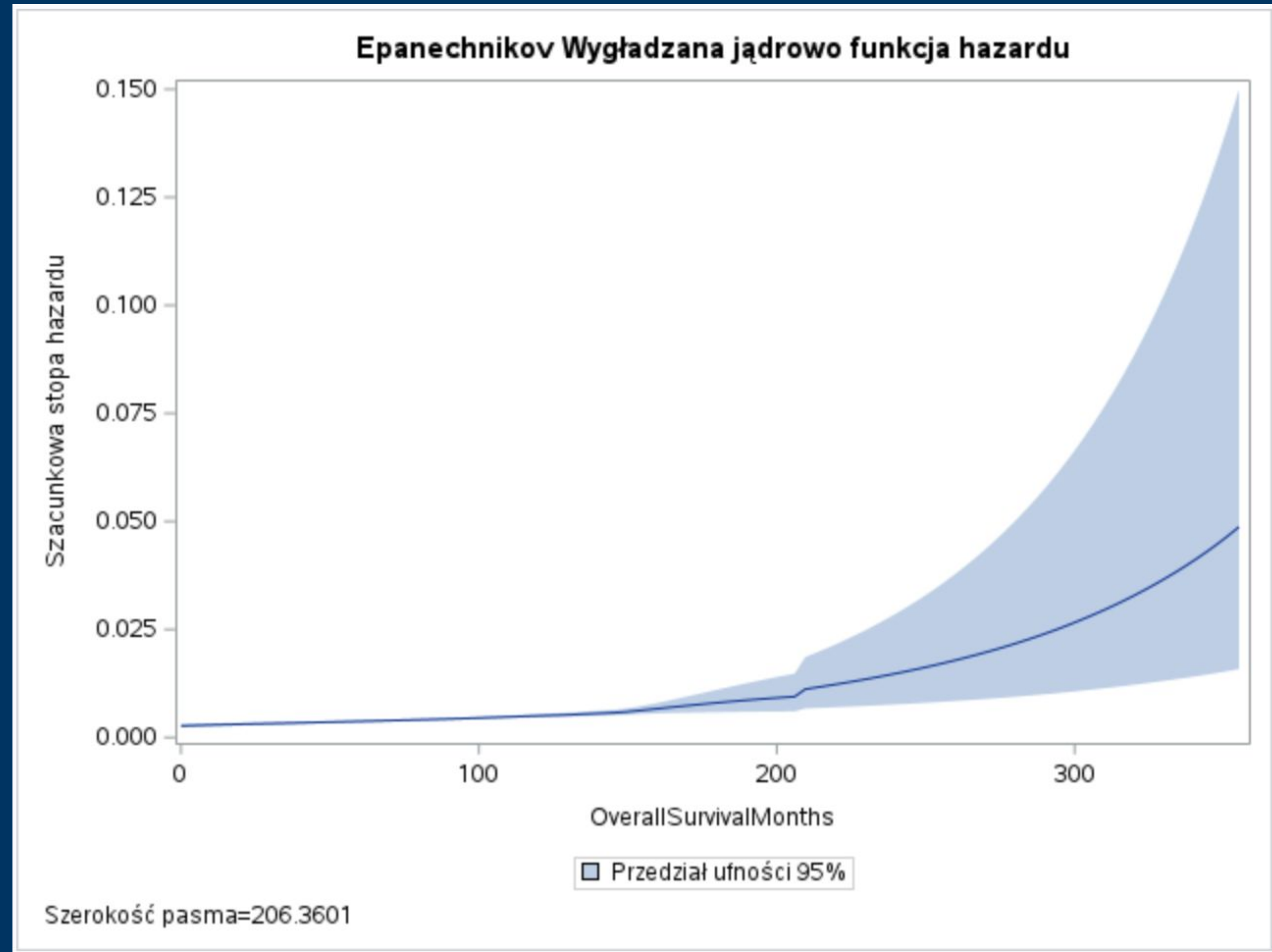
Wykres  $\log(-\log S(t))$  nie jest liniowy  
co sugeruje, że model wykładniczy  
może nie być odpowiedni.



# Wybór modelu

Funkcja hazardu jest rosnąca co jest zgodne z modelem Weibulla. W przypadku modelu wykładniczego funkcja byłaby stała w czasie.

**Wobec tego wybór modelu wykładniczego byłby nieodpowiedni. Możemy wstępnie odrzucić model wykładniczy**





# Modele parametryczne

Modele parametryczne i nieparametryczne różnią się od siebie założeniem dotyczącym rozkładu czasu a zdarzeniami. W podejściu parametrycznym przyjmuje się założenie, że czas między zdarzeniami jest zmienną losową należącą do konkretnej rodziny rozkładów.

# Modele parametryczne

Do wykonania modeli parametrycznych, z uwagi na wysoki współczynnik p-value, odrzucone zostały poniższe zmienne:

PrimaryTumorLaterality

MutationCount

TumorSize

TypeofBreastSurgery

CancerType

CancerTypeDetailed

Cellularity

Pam50Claudinlowsubtype

Cohort

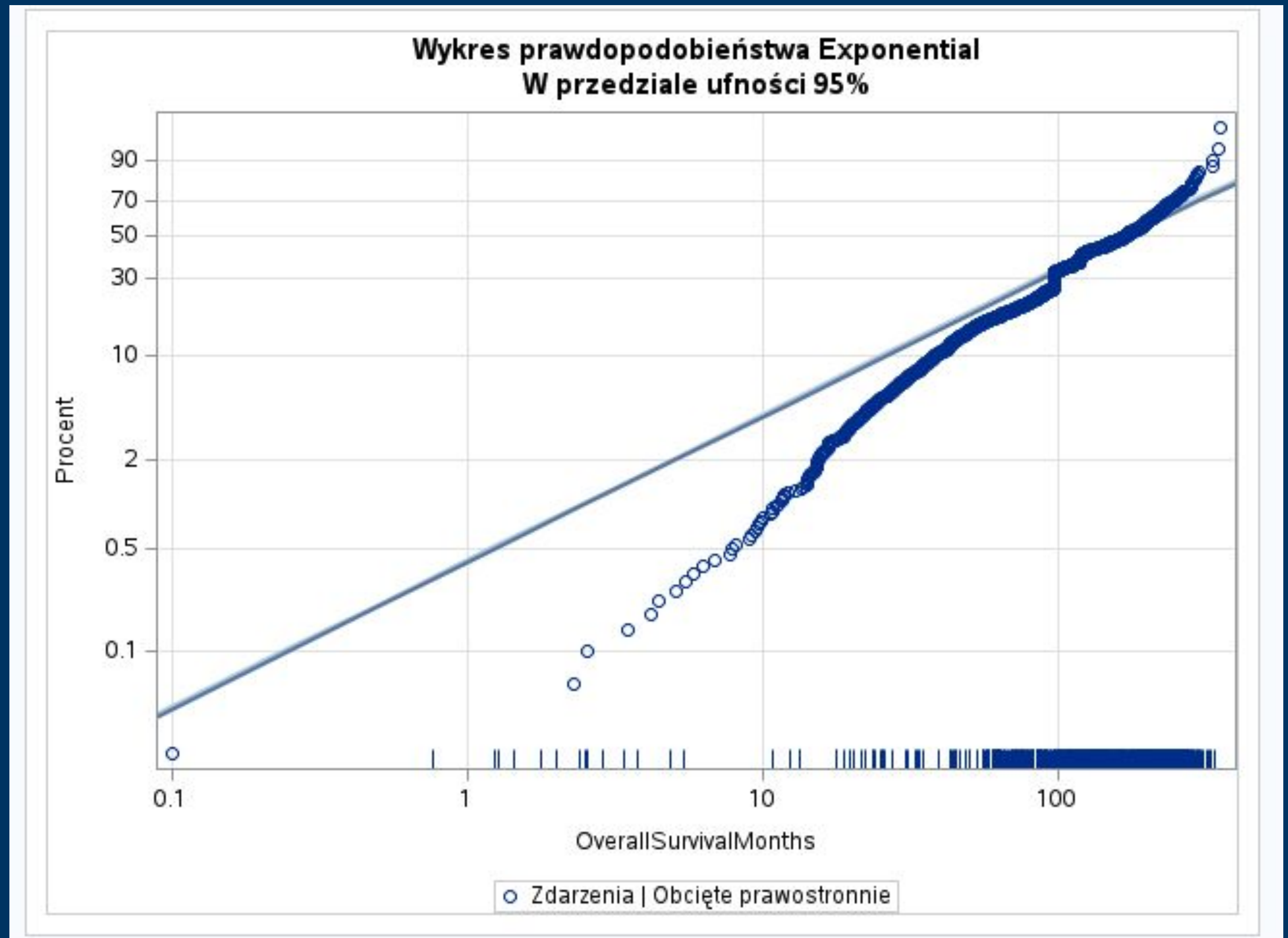
ERstatusmeasuredbyIHC

NeoplasmHistologicGrade

InferredMenopausalState

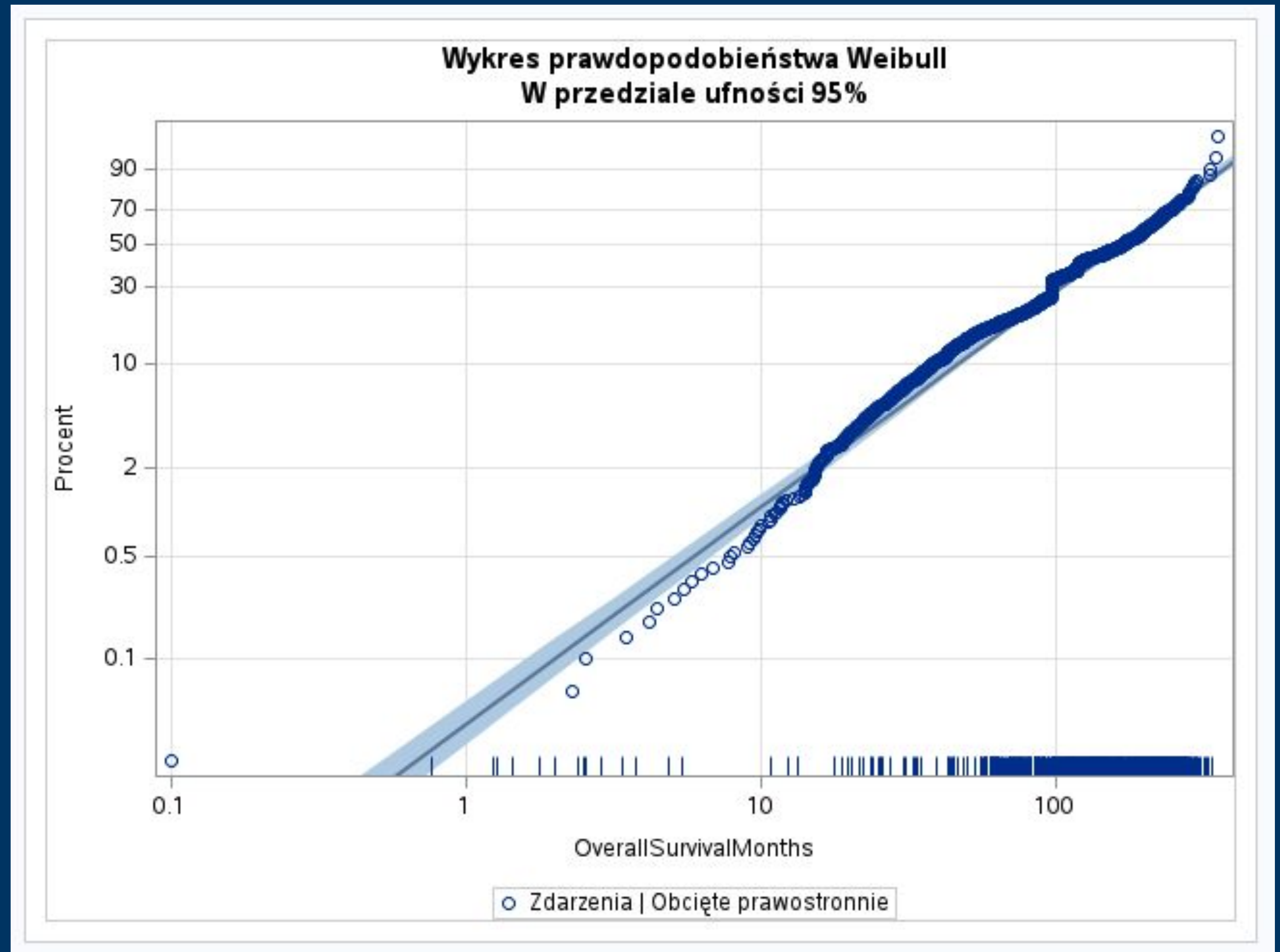
# Graficzna ocena - model wykładniczy

Punkty empiryczne leżą poza przedziałem ufności z czego wynika, że model wykładniczy jest nieodpowiedni dla analizowanych danych.



# Graficzna ocena - model Weibulla

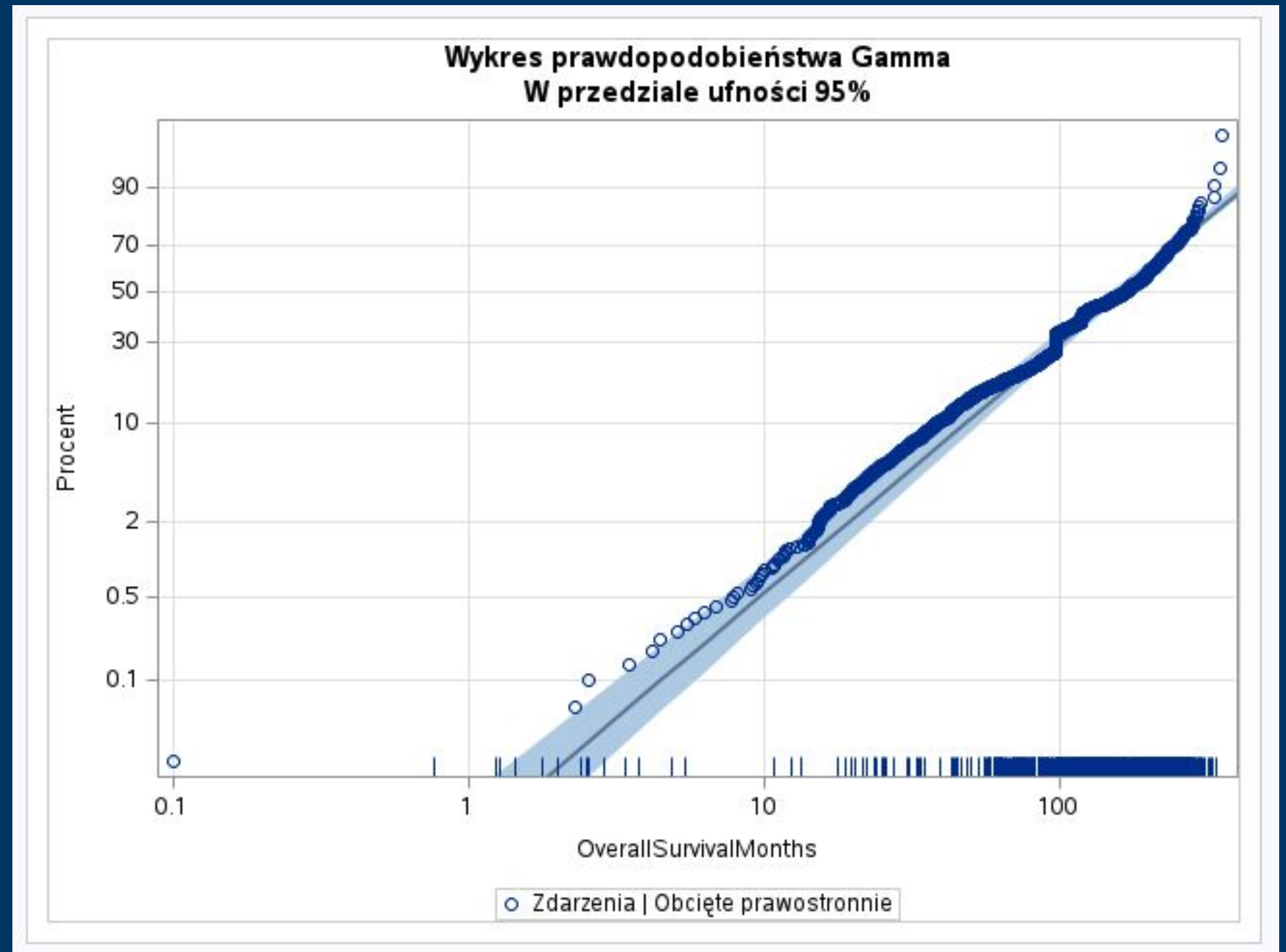
Punkty empiryczne leżą poza przedziałem ufności z czego wynika, że model Weibulla jest nieodpowiedni dla analizowanych danych.





# Graficzna ocena - model Gamma

Model Gamma w porównaniu do uprzednio analizowanych modeli pasuje najlepiej do danych. Potwierdzają to również statystyki dopasowania pokazane na następnym slajdzie.





# Podsumowanie - wybór modelu parametrycznego na podstawie statystyk dopasowania

## 1. Model Wykładniczy

Statystyki dopasowania (odpowiedź nierejestrowana)	
-2 log. wiarygodności	17008.74
ExponentialAIC (jak najmniejsze)	17038.74
ExponentialAICC (jak najmniejsze)	17038.94
ExponentialBIC (jak najmniejsze)	17126.15

## 2. Model Weibulla

Statystyki dopasowania (odpowiedź nierejestrowana)	
-2 log. wiarygodności	16760.59
WeibullAIC (jak najmniejsze)	16792.59
WeibullAICC (jak najmniejsze)	16792.81
WeibullBIC (jak najmniejsze)	16885.83

## 3. Model Gamma

Statystyki dopasowania (odpowiedź nierejestrowana)	
-2 log. wiarygodności	16738.15
GammaAIC (jak najmniejsze)	16772.15
GammaAICC (jak najmniejsze)	16772.39
GammaBIC (jak najmniejsze)	16871.21

# Model wykładniczy

P-value dla statystyki mnożnika Lagrange'a jest mniejsze od 0,05, zatem możemy odrzucić hipotezę o stałości hazardu. Brak wskazań do przyjęcia modelu wykładniczego.

Analiza efektów dla rozkładu wykładniczego wskazuje na brak istotności przy zmiennych HER2Status oraz ERStatus

Statystyki mnożnika Lagrange'a		
Parametr	Chi-kwadrat	Pr. > chi-kw.
Skala	635.0663	<.0001

Analiza efektów typu III			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
AgeatDiagnosis	1	108.3547	<.0001
Chemotherapy	1	4.3581	0.0368
ERStatus	1	3.4226	0.0643
HER2statusmeasuredby	1	5.0337	0.0249
HER2Status	1	2.7330	0.0983
TumorOtherHistologic	1	3.8852	0.0487
HormoneTherapy	1	5.2896	0.0215
Lymphnodesexaminedpo	1	33.5781	<.0001
Nottinghamprognostic	1	14.7257	0.0001
OncotreeCode	1	8.0144	0.0046
PRStatus	1	3.9062	0.0481
RadioTherapy	1	28.2958	<.0001
Gene classifiersubtyp	1	8.8942	0.0029
TumorStage	1	18.4626	<.0001



# Model Weibulla

Parametr kształtu Weibulla osiąga wartość 1,4815, co oznacza, że: 1% wzrostu czasu jest związany z 1,4815% spadku w ryzyku śmierci  
Wartość parametru kształtu Weibulla świadczy również o kształcie funkcji hazardu.

W przypadku, gdy jest większy od 1 wskazuje to, że hazard jest malejący w czasie.

Analiza ocen parametrów maksymalnej wiarygodności							
Parametr	DF	Ocena	Błąd standardowy	Przedział ufności 95%		Chi-kwadrat	Pr. > chi-kw.
Intercept	1	6.9699	0.1681	6.6404	7.2993	1719.19	<.0001
AgeatDiagnosis	1	-0.0205	0.0018	-0.0239	-0.0170	136.51	<.0001
Chemotherapy	1	-0.1868	0.0648	-0.3138	-0.0599	8.32	0.0039
ERStatus	1	0.1201	0.0549	0.0125	0.2277	4.79	0.0287
HER2statusmeasuredby	1	0.0845	0.0315	0.0227	0.1463	7.18	0.0074
HER2Status	1	-0.1618	0.0858	-0.3300	0.0065	3.55	0.0595
TumorOtherHistologic	1	0.0480	0.0226	0.0038	0.0923	4.53	0.0334
HormoneTherapy	1	0.0617	0.0462	-0.0289	0.1522	1.78	0.1821
Lymphnodesexaminedpo	1	-0.0298	0.0043	-0.0381	-0.0214	48.93	<.0001
Nottinghamprognostic	1	-0.0928	0.0219	-0.1357	-0.0498	17.88	<.0001
OncotreeCode	1	-0.0702	0.0240	-0.1172	-0.0232	8.57	0.0034
PRStatus	1	0.0920	0.0443	0.0051	0.1788	4.30	0.0380
RadioTherapy	1	0.2330	0.0411	0.1524	0.3136	32.11	<.0001
Geneclassifiersubtyp	1	0.0813	0.0253	0.0316	0.1309	10.30	0.0013
TumorStage	1	-0.1750	0.0382	-0.2499	-0.1002	20.99	<.0001
Skala	1	0.6750	0.0155	0.6453	0.7060		
Kształt Weibulla	1	1.4815	0.0340	1.4164	1.5496		



# Gamma

Analiza efektów dla rozkładu Gamma wskazuje na brak istotności przy zmiennych HER2Status

Analiza efektów typu III			
Efekt	DF	Chi-kwadrat Walda	Pr. > chi-kw.
AgeatDiagnosis	1	121.0084	<.0001
Chemotherapy	1	8.7123	0.0032
ERStatus	1	7.6833	0.0056
HER2statusmeasuredby	1	7.6818	0.0056
HER2Status	1	3.4215	0.0644
TumorOtherHistologic	1	4.6720	0.0307
HormoneTherapy	1	5.4596	0.0195
Lymphnodesexaminedpo	1	44.3037	<.0001
Nottinghamprognostic	1	17.6512	<.0001
OncotreeCode	1	8.5769	0.0034
PRStatus	1	7.8886	0.0050
RadioTherapy	1	30.1405	<.0001
Gene classifiersubtyp	1	11.0191	0.0009
TumorStage	1	25.2960	<.0001

# Metody semiparametryczne



# Metody semiparametryczne

Na podstawie analizy maksymalnej wiarygodności w modelu semi-parametrycznym otrzymujemy współczynnik Hazard Ratio, który określa jak zmienia się hazard danej zmiennej w czasie.

Przykładowa interpretacja zmiennej istotnej statystycznie (Chemotherapy)

Współczynnik hazardu dla parametru Chemotherapy - 1,322 oznacza, że ryzyko śmierci dla osób po chemioterapii jest o 32,2% wyższe od ryzyka dla tych, którzy nie mieli chemioterapii.

Analiza ocen maksymalnej wiarygodności							
Parametr	DF	Ocena parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi-kw.	Współczynnik hazardu	Etykieta
AgeatDiagnosis	1	0.03067	0.00259	140.1864	<.0001	1.031	AgeatDiagnosis
Chemotherapy	1	0.27903	0.09600	8.4481	0.0037	1.322	Chemotherapy
ERStatus	1	-0.19128	0.08149	5.5096	0.0189	0.826	ERStatus
HER2statusmeasuredby	1	-0.12445	0.04681	7.0667	0.0079	0.883	HER2statusmeasuredbySNP6
HER2Status	1	0.22802	0.12745	3.2011	0.0736	1.256	HER2Status
TumorOtherHistologic	1	-0.07403	0.03363	4.8457	0.0277	0.929	TumorOtherHistologicSubtype
HormoneTherapy	1	-0.07363	0.06922	1.1315	0.2875	0.929	HormoneTherapy
Lymphnodesexaminedpo	1	0.04431	0.00634	48.8955	<.0001	1.045	Lymphnodesexaminedpositive
Nottinghamprognostic	1	0.13536	0.03236	17.4970	<.0001	1.145	Nottinghamprognosticindex
OncotreeCode	1	0.10644	0.03559	8.9443	0.0028	1.112	OncotreeCode
PRStatus	1	-0.12848	0.06576	3.8173	0.0507	0.879	PRStatus
RadioTherapy	1	-0.34107	0.06071	31.5629	<.0001	0.711	RadioTherapy
Gene classifiersubtyp	1	-0.11608	0.03751	9.5764	0.0020	0.890	Gene classifiersubtype
TumorStage	1	0.25259	0.05634	20.0974	<.0001	1.287	TumorStage



# Metody semiparametryczne

W modelu Coxa hazardy muszą być proporcjonalne, aby można było interpretować wpływ danej zmiennej na zajście zdarzenia. Jeżeli założenie to nie jest spełnione, oznacza to, że pewne zmienne wchodzą w interakcję z czasem, czyli hazardy nie są równoległe. Postanowiono sprawdzić, czy w badanym zbiorze znajdują się zmienne zależne od czasu.

Zmienne zależne od czasu:

- AgeatDiagnosis\_t
- ERStatus\_t
- HER2statusmeasuredbySNP6\_t
- TumorOtherHistologicSubtype\_t
- Lymphnodesexaminedpositive\_t
- Nottinghamprognosticindex\_t
- OncotreeCode\_t
- PRStatus\_t
- RadioTherapy\_t
- Geneclifierssubtype\_t
- TumorStage\_t
- Chemotherapy\_t

Parametr	DF	Ocena parametru	Błąd standardowy	Chi-kwadrat	Pr. > chi-kw.	Współczynnik hazardu	Etykieta
AgeatDiagnosis_t	1	0.0001996	0.0000438	20.7747	<.0001	1.000	
ERStatus_t	1	0.00427	0.00157	7.3677	0.0066	1.004	
HER2statusmeasuredby	1	0.00158	0.0006355	6.2047	0.0127	1.002	
TumorOtherHistologic	1	0.0003161	0.0004554	0.4819	0.4876	1.000	
Lymphnodesexaminedpo	1	-0.0000353	0.0001359	0.0676	0.7949	1.000	
Nottinghamprognostic	1	-0.0007690	0.0004952	2.4109	0.1205	0.999	
OncotreeCode_t	1	-0.0002537	0.0005078	0.2496	0.6173	1.000	
PRStatus_t	1	0.00350	0.00108	10.4509	0.0012	1.004	
RadioTherapy_t	1	0.00207	0.0009395	4.8517	0.0276	1.002	
Geneclifierssubtype	1	0.0004247	0.0005988	0.5030	0.4782	1.000	
TumorStage_t	1	-0.00220	0.0008941	6.0408	0.0140	0.998	
Chemotherapy_t	1	-0.00706	0.00199	12.6374	0.0004	0.993	
AgeatDiagnosis	1	0.01139	0.00483	5.5635	0.0183	1.011	AgeatDiagnosis
Chemotherapy	1	0.77273	0.17364	19.8037	<.0001	2.166	Chemotherapy
ERStatus	1	-0.61211	0.15519	15.5577	<.0001	0.542	ERStatus
HER2statusmeasuredby	1	-0.30234	0.06728	20.1923	<.0001	0.739	HER2statusmeasuredbySNP6
TumorOtherHistologic	1	-0.11895	0.06826	3.0371	0.0814	0.888	TumorOtherHistologicSubtype
Lymphnodesexaminedpo	1	0.04122	0.01225	11.3202	0.0008	1.042	Lymphnodesexaminedpositive
Nottinghamprognostic	1	0.22291	0.06022	13.7009	0.0002	1.250	Nottinghamprognosticindex
OncotreeCode	1	0.14445	0.07116	4.1201	0.0424	1.155	OncotreeCode
PRStatus	1	-0.52084	0.12520	17.3051	<.0001	0.594	PRStatus
RadioTherapy	1	-0.54707	0.11298	23.4484	<.0001	0.579	RadioTherapy
Geneclifierssubtype	1	-0.13845	0.06479	4.5663	0.0326	0.871	Geneclifierssubtype
TumorStage	1	0.47845	0.10404	21.1481	<.0001	1.614	TumorStage

# Podsumowanie

Celem projektu było sprawdzenie jakie czynniki mają wpływ na śmiertelność pacjentek onkologicznych (rak piersi), już we wstępnej analizie danych przy pomocy modeli nieparametrycznych możemy zauważyć, że istotny wpływ na to zdarzenie ma między innymi: wiek pacjentki; stopień zaawansowania nowotworu;

Według statystyk dopasowania poszczególnych modeli parametrycznych najlepiej dopasowany do danych jest model Gamma.