Raport 3

Analiza przeżycia

Olga Foriasz, Tomasz Warzecha

2025-12-18

Spis treści

1 Lista 9

W zadaniu 9 wykorzystywane są dane lung z pakietu survival, dotyczące pacjentów z zaawansowanym rakiem płuc.

1.1 Zadanie 1 - Oszacowanie parametrów

Najpierw oszacowano parametry przyspieszonego modelu awarii. Za zmienną zależną przyjęto zmienną time, natomiast za charakterystyki zmienne: age, sex, ph.ecog, ph.karno.

Tabela 1: Oszacowania parametrów modelu Weibulla

Parametr	Wartosc
age	-0.0086
as.factor(sex)2	0.4083
as.factor(ph.ecog)1	-0.4216
as.factor(ph.ecog)2	-0.9261
as.factor(ph.ecog)3	-1.6872
ph.karno	-0.0101
mi	6.0431
sigma	0.7543

Powyższa tabela przedstawia oszacowane parametry modelu. Pierwszych sześć wierszy zawiera oszacowania współczynników wektora β. Kolejny wiersz odpowiada parametrowi μ, który określa bazowy logarytm czasu przeżycia. Ostatni wiersz zawiera parametr σ, opisujący zmienność czasu przeżycia w modelu.

1.2 Zadanie 2 - Interpretacja parametrów modelu

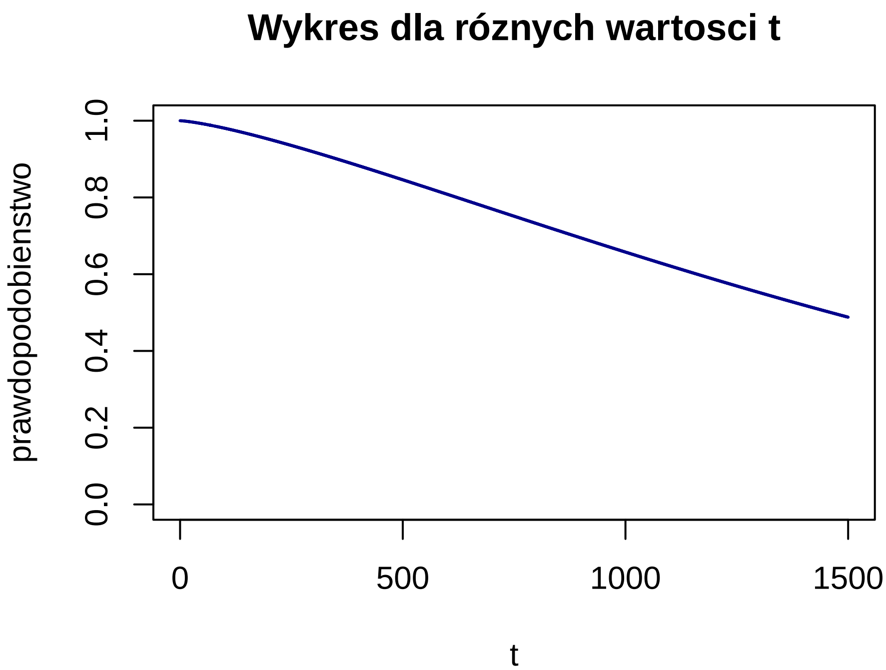
- Współczynnik μ wynosi 6{,}04. Jest to parametr określający bazowy logarytm czasu przeżycia pacjenta, tj. przy założeniu, że zmienna age przyjmuje wartość równą średniej z danych, ph.ecog = 0, ph.karno jest równe swojej średniej w próbie oraz sex = 1 (mężczyzna).
- Ponieważ μ = 6,04, otrzymujemy $\log(T) = 6.04$. Po zastosowaniu funkcji wykładniczej uzyskujemy bazowy oczekiwany czas przeżycia pacjenta o „bazowych" parametrach, wyrażony w dniach. U nas wynosi on w przybliżeniu 421.
- Następny parametr σ, kontrolujący zmienność czasu przeżycia, wynosi 0.75. Im większa jest wartość parametru, tym większe zróżnicowanie czasu przeżycia wśród badanych danych. Wartość 0.75 wskazuje zatem na niezbyt duże zróżnicowanie badanych danych.
- Parametr β składa się z sześciu współczynników. Pierwszy – age = -0.86 – wskazuje, że wzrost wieku pacjenta powoduje skrócenie przewidywanego czasu przeżycia. Współczynnik as.factor(sex)2 wynosi 0.41 i oznacza, że pacjenci ze zmienną sex=2 (kobiety) mają większą oczekiwaną wartość funkcji przeżycia. Następnie współczynniki as.factor(ph.ecog)1, as.factor(ph.ecog)2 oraz as.factor(ph.ecog)3 są coraz niższe. Wskazuje to, że im większy wskaźnik ph.ecog (gdzie 0 oznacza sprawność prawidłową, a 5 – zgon), tym krótszy przewidywany czas przeżycia, co jest zgodne z intuicją. Ostatnim współczynnikiem β jest ph.karno wynoszące -0.01. Wskazuje to, że lepsza sprawność fizyczna powoduje bardzo niewielki spadek przewidywanego czasu życia, co w praktyce prawdopodobnie nie odzwierciedla rzeczywistego wpływu sprawności fizycznej na długość życia.

1.3 Zadanie 3 - Oszacowanie funkcji przeżycia

W zadaniu wyznaczono oszacowanie funkcji przeżycia, odpowiadającej rozkładowi czasu życia kobiety w wieku 70 lat, przy wartościach ph.karno=90 oraz ph.ecog=1, a następnie na jej podstawie obliczono prawdopodobieństwo, że badana kobieta przeżyje ponad 300 dni. Z wyliczeń wyszło, że to prawdopodobieństwo wynoi 0.918547.

1.4 Zadanie 4 - Wykres

Poniższy wykres przedstawia prawdopodobieństwo przeżycia t dni, dla t z przedziału [0,1500], przy parametrach jak w zdaniu 3.



1.5 Zadanie 5 - Oszacowanie parametry modelu proporcjonalnych hazardów

W zadaniu oszacowano parametry modelu proporcjonalnych hazardów, przyjmując za zmienną zależną zmienną time, a za charakterystyki zmienne: age, sex, ph.ecog, ph.karno.

Tabela 2: Oszacowania parametrów modelu Weibulla

Parametr	Wartosc
age	0.0119
as.factor(sex)2	-0.5665
as.factor(ph.ecog)1	0.5850
as.factor(ph.ecog)2	1.2851
as.factor(ph.ecog)3	2.3413
ph.karno	0.0140
mi	6.3012
sigma	1.3876

Powyższa tabela przedstawia oszacowane parametry modelu. Pierwszych sześć wierszy zawiera oszacowania współczynników wektora β. Kolejny wiersz odpowiada parametrowi μ, który określa bazowy logarytm czasu przeżycia. Ostatni wiersz zawiera parametr σ, opisujący zmienność czasu przeżycia w modelu.

1.6 Zadanie 6 - Interpretacja współczynników modelu

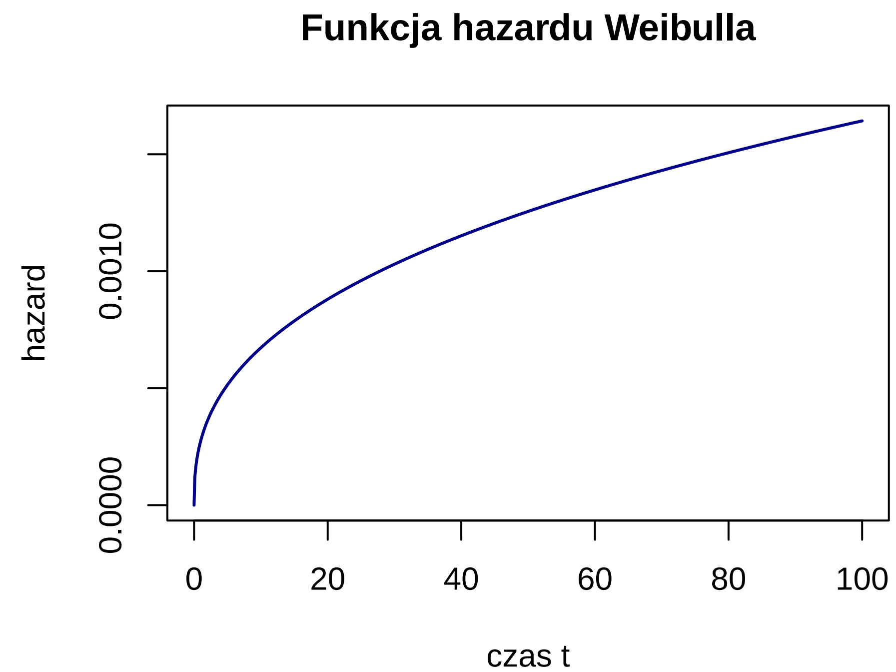
Oszacowane parametry modelu Weibulla zostały uzyskane metodą największej wiarygodności i odnoszą się do modelu liniowego logarytmu czasu przeżycia (AFT).

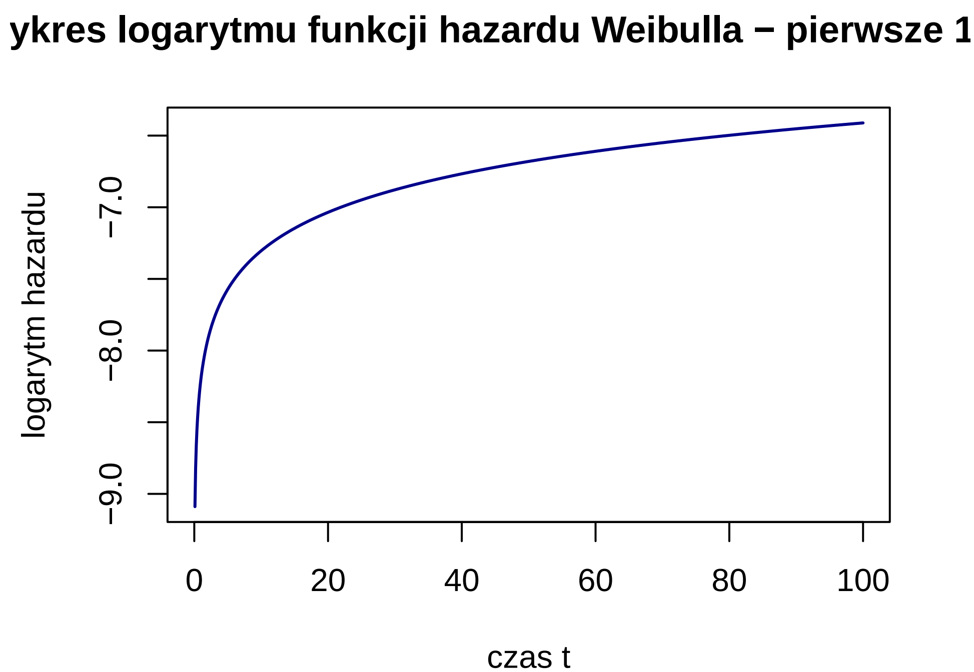
- Parametr μ=6,30 opisuje bazowy logarytm czasu przeżycia pacjenta o wartościach referencyjnych zmiennych objaśniających.Oznacza to, że dla pacjenta z grupy referencyjnej (mężczyzny, o najlepszej sprawności według skali ph.ecog oraz przy wartościach referencyjnych pozostałych zmiennych) model przewiduje przeciętny czas przeżycia wynoszący około 2129 dni, gdzie dodatkowo w zależności od pozostałych cech, szacowany czas może się wydłużyć lub skrócić. wydłużeniu.
- Parametr σ=1,3876 kontroluje zmienność czasu przeżycia i wskazuje na dość duże zróżnicowanie obserwowanych czasów przeżycia w badanej populacji.
- Wzrost wieku pacjenta nieznacznie wydłuża przewidywany czas przeżycia (age=0,0119).
- Ujemny współczynnik natomiast dla zmiennej sex=2 (dla kobiet) , wynoszący- 0,5665 oznacza, że kobiety mają krótszy oczekiwany czas przeżycia niż grupa referencyjna (mężczyźni).
- Współczynniki przy zmiennej ph.ecog rosną wraz z jej poziomem, co wskazuje, że pogarszająca się sprawność chorego znacznie skraca przewidywany czas przeżycia.
- Zmienna ph.karno ma bardzo niewielki dodatni wpływ na czas przeżycia (0,0140), co sugeruje słaby wpływ tej zmiennej w obecnym modelu.

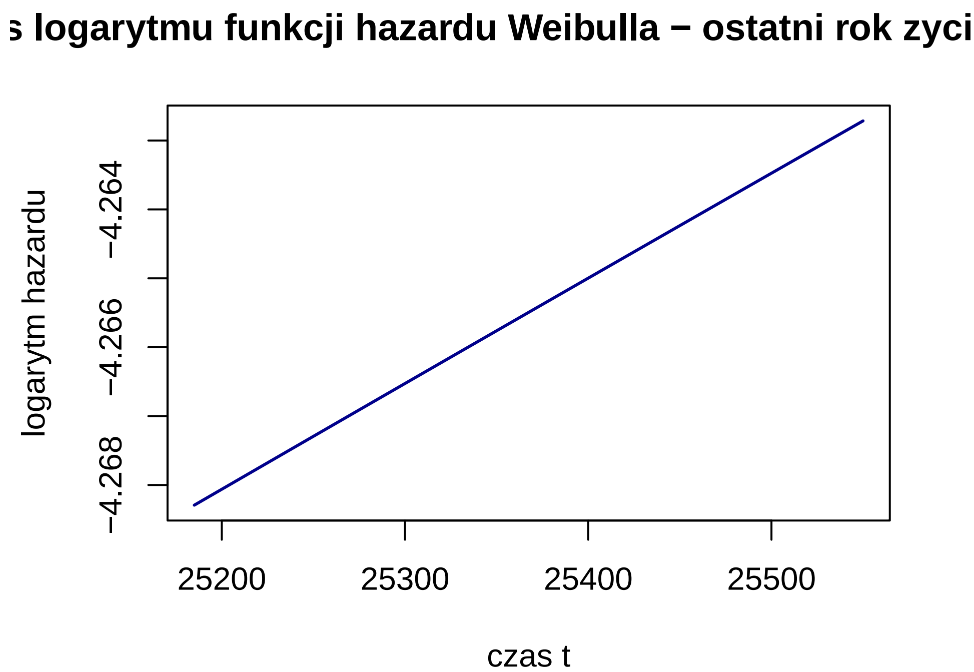
1.7 Zadanie 7 - Oszacowanie funkcji hazardu

Dla kobiety w wieku 70 lat o charakterystyce ph.ecog=1 i ph.karno=90 oszacowanie funkcji hazardu wyniosło 0.00271, gdy przyjęliśmy za t - 70 lat podane w dniach.

```
##  log(scale)
## 0.002713376
```

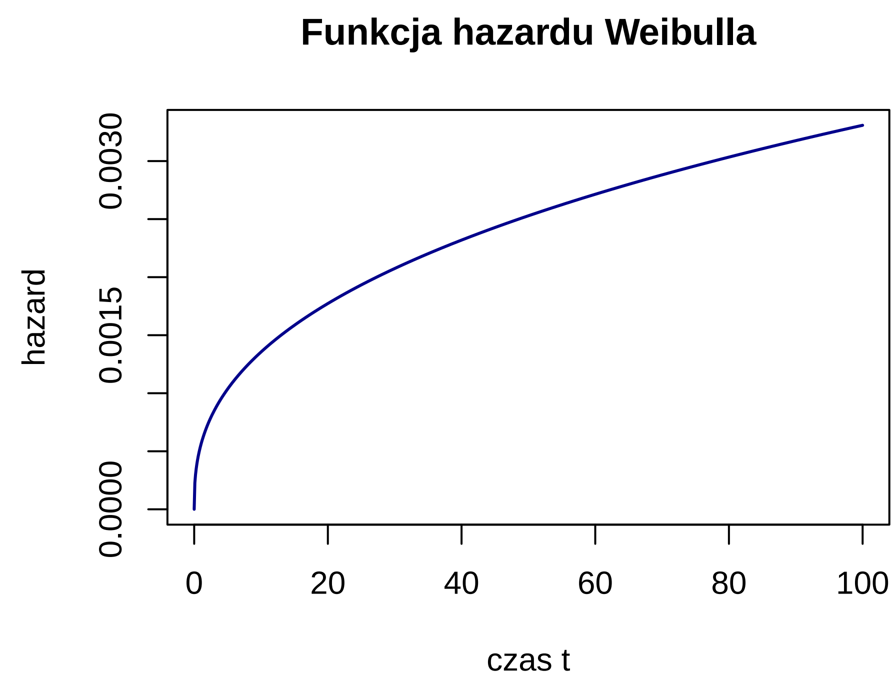

Funkcja hazardu Weibulla
hazard
0.0000
0.0010
0
20
40
60
80
100
czas t


ykres logarytmu funkcji hazardu Weibulla − pierwsze 1
logarytm hazardu
−9.0
−8.0
−7.0
0
20
40
60
80
100
czas t

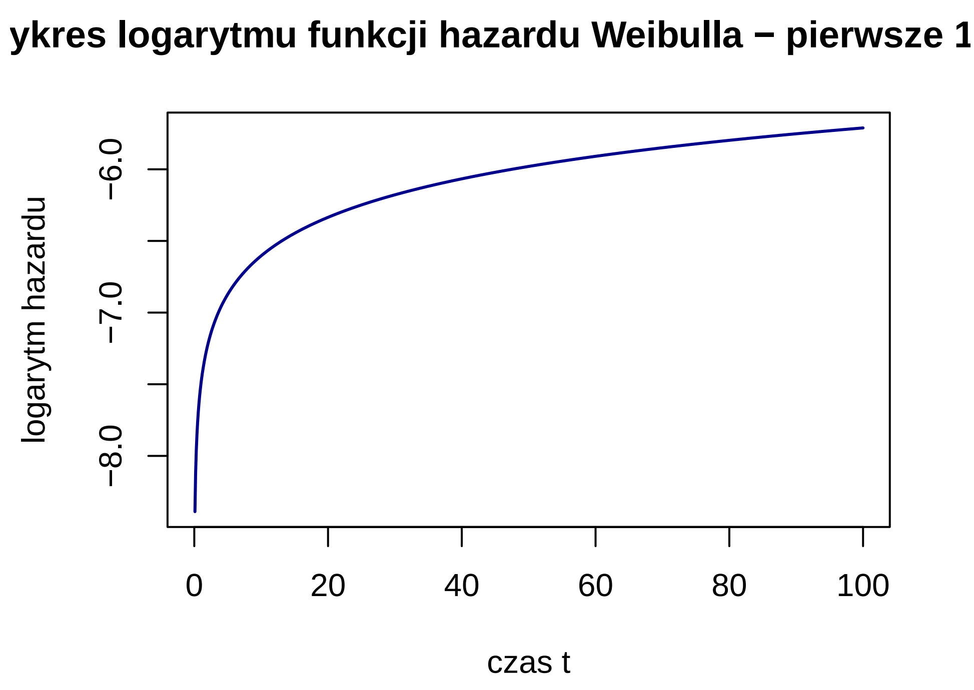


Dla kobiety w wieku 70 lat o charakterystyce ph.ecog=2 i ph.karno=90 oszacowanie funkcji hazardu wyniosło 0.00546, gdy przyjęliśmy za t - 70 lat podane w dniach.

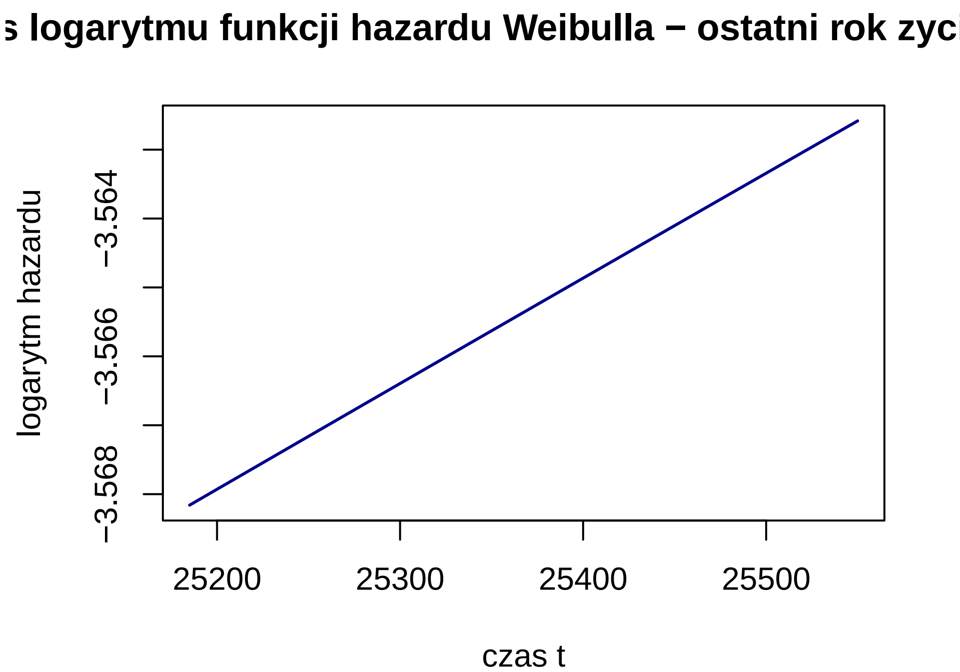
```
##  log(scale)
## 0.005464792
```



Funkcja hazardu Weibulla
hazard
0.0000
0.0015
0.0030
0
20
40
60
80
100
czas t



ykres logarytmu funkcji hazardu Weibulla − pierwsze 1
logarytm hazardu
−8.0
−7.0
−6.0
0
20
40
60
80
100
czas t



Na przyjętym przedziale czasowym dla t od 0 do 100 wykresy logarytmm hazardów nie są liniowe. Aby nie mieć wątpliowści co do przyjętego modelu proporcjonalnych hazardów, oczekiwalibyśmy, że wykresy będą prawie liniowe oraz wykresy logarytmów będą do siebie równoległe. Jak wcześniej napisano, nie są to wykresy liniowe na zadanym przedziale, zatem można mieć wątpliowści co do przyjętego modelu.

Jeśli natomiast ustalimy t - zmienną czasu, tylko na ostatni rok życia pacjentek, to wykresu logarytmów funkcji hazardów są liniowe oraz można uznać, że równoległe do siebie. Na takim przedziale czasowym nie powinno się mieć wątpliwości co do przyjętego modelu.

Stąd można wysnuć wnioski, że na pełnym przedziale czasowym od 0 do 70 lat, można mieć wątpliwości co do przyjętego modelu, jednak biorać krótszy przedział, np. ostatni rok życia, można uznać przyjęty model za akceptowalny.

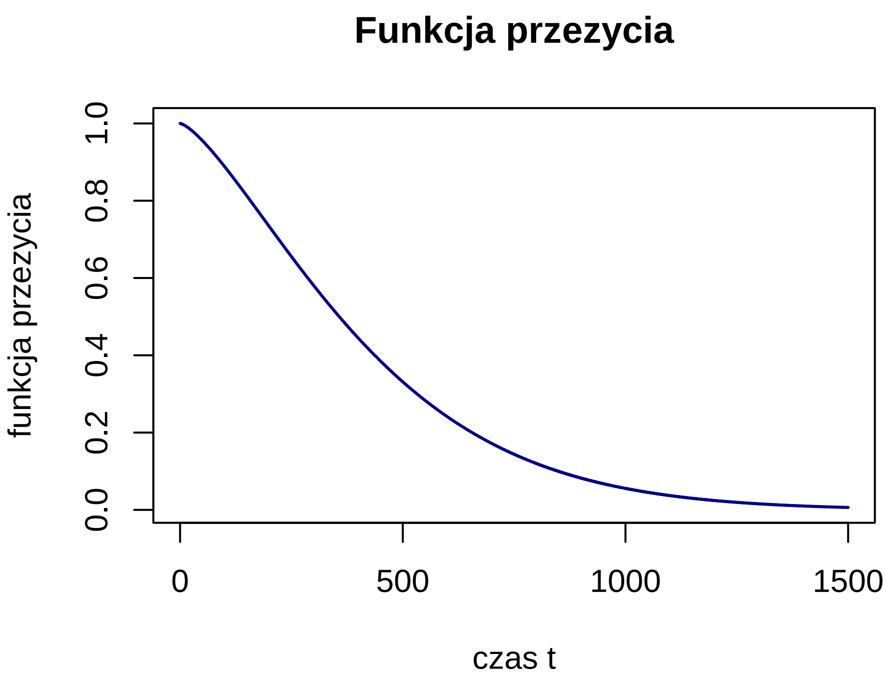
1.8 Zadanie 8 - Oszacowanie funkcji przeżycia

Dla kobiety w wieku 70 lat o charakterystyce ph.ecog=1 i ph.karno=90 oszacowane prawdopodobieństwo, że czas życia będzie większy niż 300 dni, wynosi 0,58. W zadaniu 3, przy wcześniejszej metodzie obliczeń, prawdopodobieństwo to wynosiło 0,51, czyli było o 0,07 niższe. Różnica wynika z użycia parametrów w różnych skalach: w zadaniu 3 z modelu AFT (log-czas), obecnie z modelu proporcjonalnych hazardów (hazard), co daje nieco inne prawdopodobieństwo przeżycia.

Dla kobiety w wieku 70 lat o charakterystyce ph.ecog=2 i ph.karno=90 szacowane prawdopodobieństwo, że czas życia będzie większy niż 300 dni wynosi 0.33.

1.9 Zadanie 9 - Wykresy oszacowanych funkcji przeżycia

Poniższy wykres przedstawia oszacowanie funkcji przeżycia odpowiadającej rozkładowi czasu życia kobiety w wieku 70 lat o charakterystyce ph.ecog=1 i ph.karno=90.



Porównując wyniki z obecnego oszacowania funkcji przeżycia z wykresem z zadania 4, gdzie również liczono funkcję przeżycia (w dniach) dla kobiety o tej samej charakterystyce, można zauważyć, że nowa funkcja przeżycia spada znacznie łagodniej. Spadek prawdopodobieństwa jest stopniowy, podczas gdy w funkcji z zadania 4 spadek był gwałtowny i niemal od razu zbliżał się do zera. Obecnie funkcja przeżycia osiąga wartość bliską 0 dopiero w okolicach 1500 dni, natomiast w zadaniu 4 wartość ta była osiągana już około 500 dnia.