

Wnioskowanie statystyczne w modelach czasu życia na podstawie danych klinicznych

Analiza przeżycia

Zuzanna Klaman

2026-02-17

Spis treści

1	Estymacja funkcji przeżycia i analiza asymptotyczna	2
1.1	Porównanie estymatorów Kaplana-Meiera i Fleminga-Harringtona	2
1.2	Implementacja estymatora Kaplana-Meiera z ogonem BHK	4
1.3	Analiza asymptotycznej normalności estymatora Kaplana-Meiera	6
2	Estymacja średniego czasu życia	7
2.1	Algorytmy estymacji średniego czasu życia	7
2.2	Porównawcza analiza średniego czasu do remisji	8
3	Wnioskowanie przedziałowe	9
3.1	Konstrukcja przedziałów ufności	9
3.2	Estymacja przedziałowa średniego czasu do progresji choroby	10
4	Testowanie hipotez o równości rozkładów w grupach o różnym stopniu zaawansowania choroby	11
4.1	Weryfikacja hipotezy o zgodności rozkładów czasu przeżycia	11
4.2	Wizualizacja funkcji przeżycia i wag statystyk testowych	13

1 Estymacja funkcji przeżycia i analiza asymptotyczna

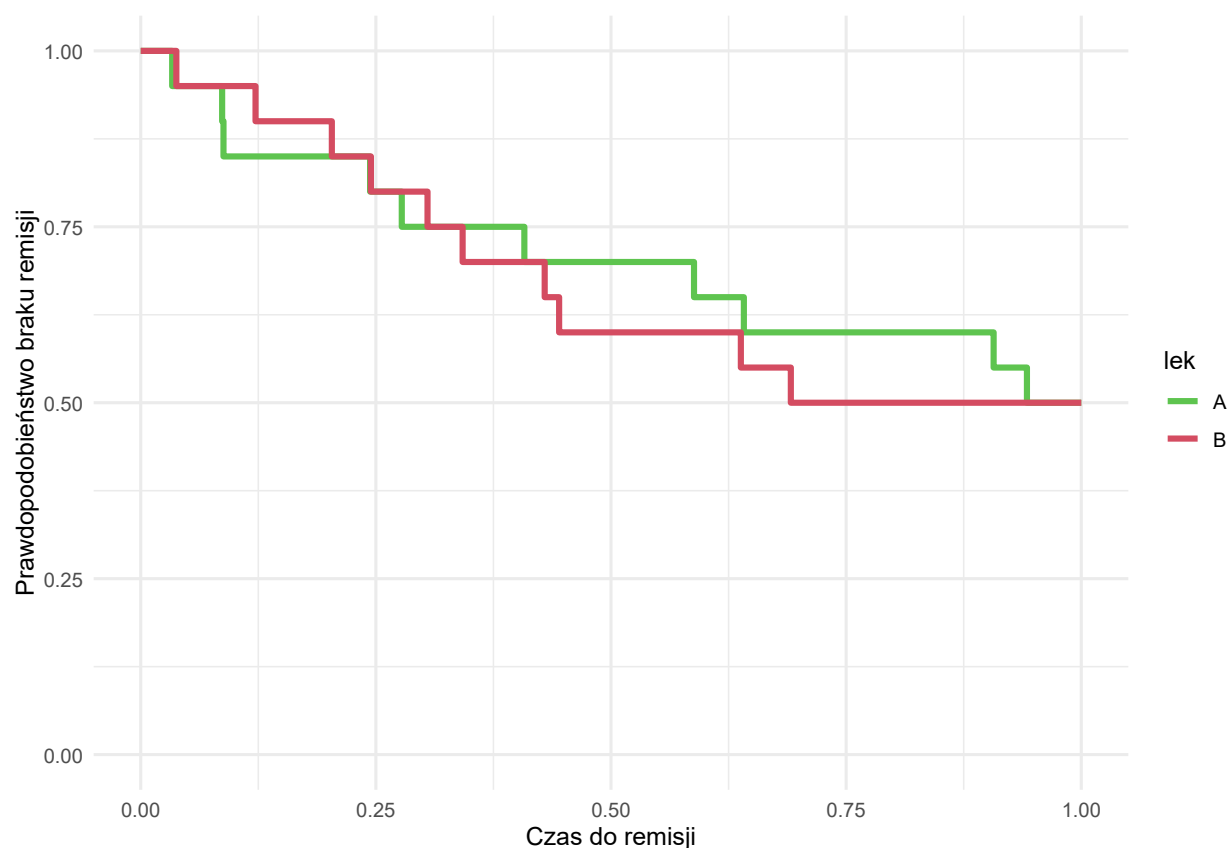
1.1 Porównanie estymatorów Kaplana-Meiera i Fleminga-Harringtona

Celem jest zaprezentowanie wykresów dwóch różnych estymatorów przeżycia: Kaplana-Meiera oraz Fleminga-Harringtona. Wykorzystamy dane używane wcześniej, opisujące czasy do remisji choroby w dwóch grupach pacjentów leczonych lekami A i B.

1.1.1 Estymatory Kaplana-Meiera

Wykres przedstawiający estymatory Kaplana-Meiera:

```
km.fit.A <- survfit(Surv(time, status)~1, data=dane.A, type="kaplan-meier")
km.fit.B <- survfit(Surv(time, status)~1, data=dane.B, type="kaplan-meier")
```



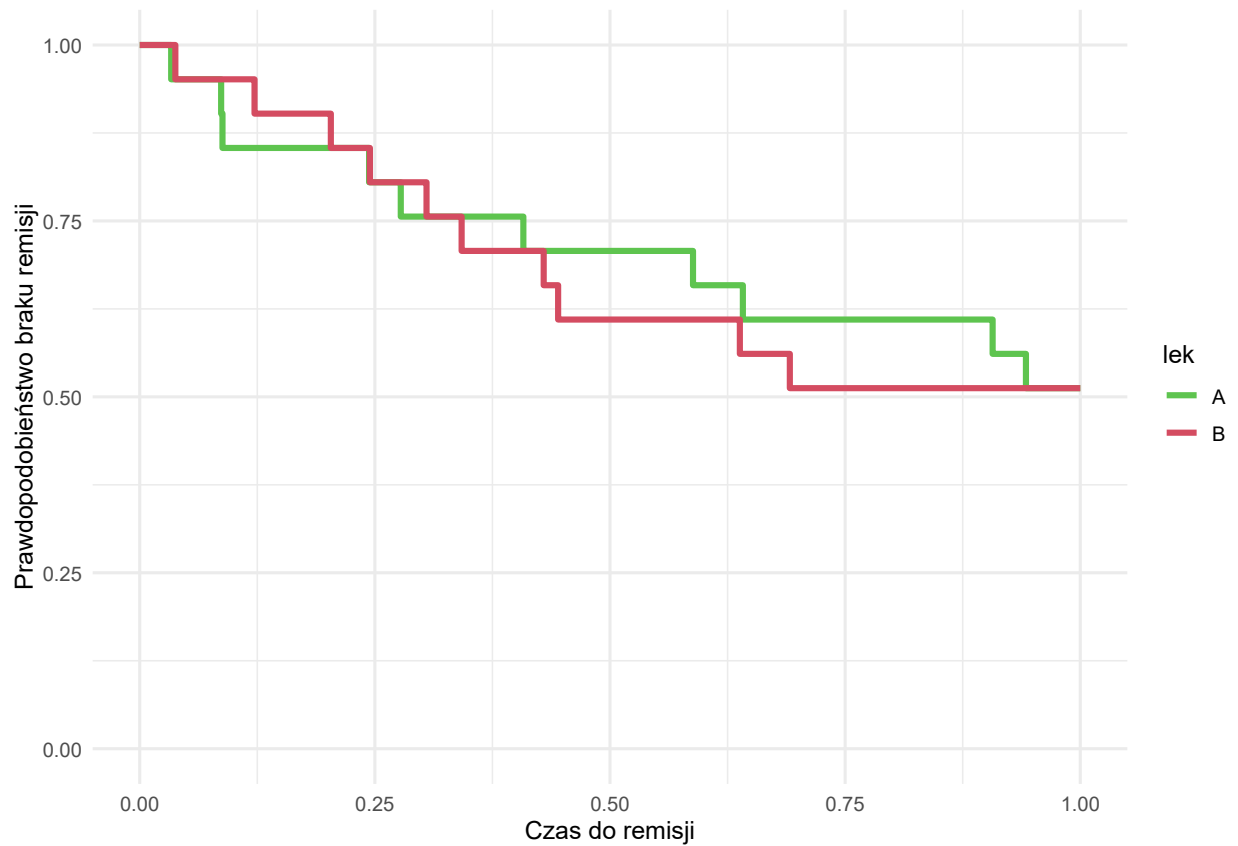
Rysunek 1: Estymatory Kaplana-Meiera dla leku A i B

Na podstawie wykresu 1 można stwierdzić, że lek B wskazuje lepszą skuteczność. Mimo że w początkowej fazie krzywe przeżycia przebiegają blisko siebie i wielokrotnie się przecinają, to w końcowym etapie obserwacji krzywa dla leku B opada szybciej. Oznacza to, że wszyscy pacjenci z grupy B osiągnęli remisję w czasie krótszym niż pacjenci z grupy A.

1.1.2 Estymatory Fleminga-Harringtona

Wykres przedstawiający estymatory Fleminga-Harringtona:

```
fh.fit.A <- survfit(Surv(time, status)~1, data=dane.A,  
                    type = "fleming-harrington")  
fh.fit.B <- survfit(Surv(time, status)~1, data=dane.B,  
                    type = "fleming-harrington")
```



Rysunek 2: Estymatory Fleminga-Harringtona dla leku A i B

Analiza wykresu 2 wykazuje spójność z wynikami uzyskanymi metodą Kaplana-Meiera. Wybór innego estymatora nie wpłynął na interpretację danych, krzywe zachowują podobny przebieg, uwydatniając końcową przewagę leku B w szybkości leczenia.

1.2 Implementacja estymatora Kaplana-Meiera z ogonem BHK

Zaimplementujemy algorytm rysujący estymator Kaplana-Meiera z dorysowanym ogonem według propozycji Browna, Hollandera i Kowara. Stworzony program posłużył do wygenerowania wykresów funkcji przeżycia dla pacjentów przyjmujących lek A oraz lek B.

```
library(survival)
km_bhk <- function(time, status, label, t_max = 4.0){
  model_km <- survfit(Surv(time, status)~1)

  df_km <- data.frame(time = model_km$time,
    surv = model_km$surv, type = "Kaplan-Meier")

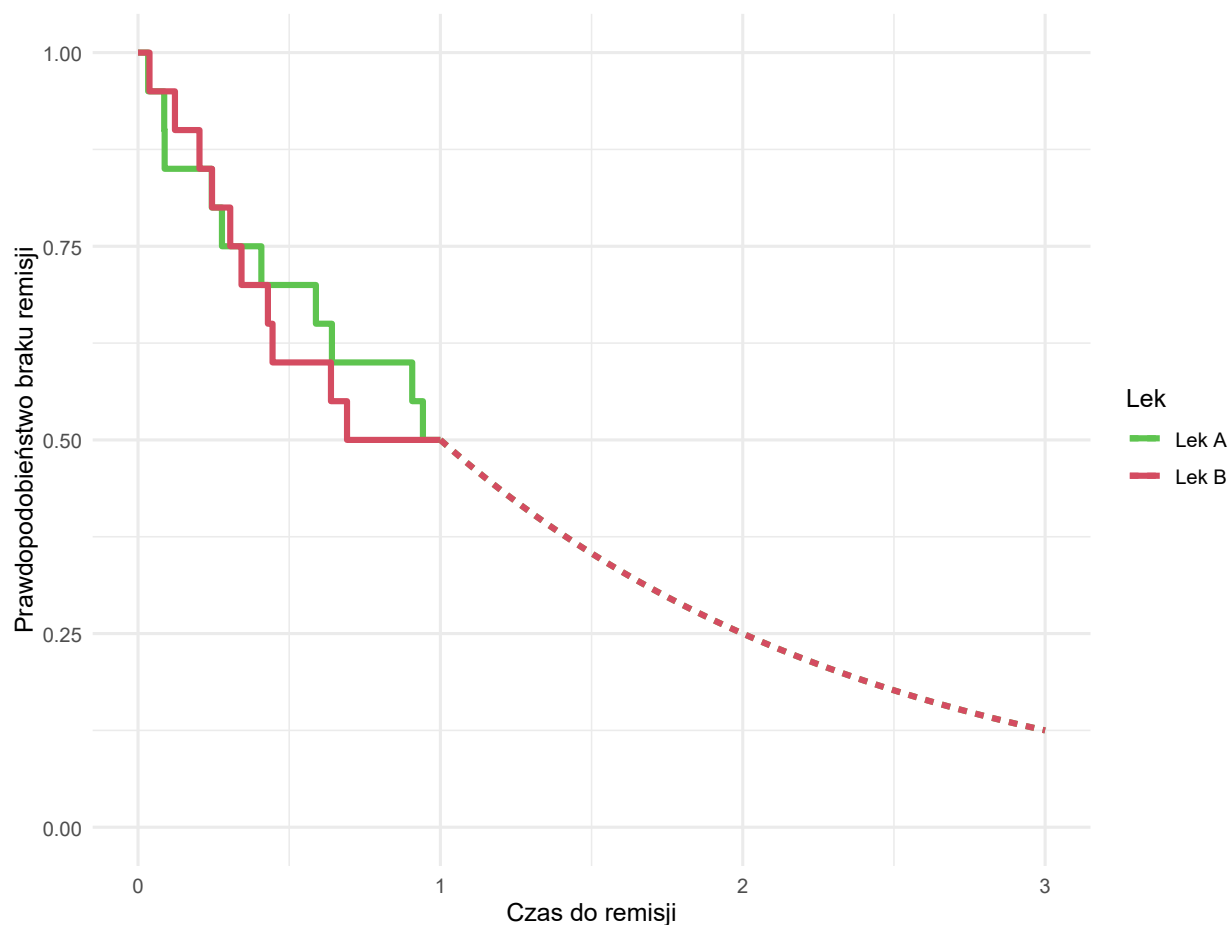
  df_km <- rbind(data.frame(time = 0, surv = 1,
    type = "Kaplan-Meier"), df_km)

  ostatnie_t <- length(model_km$time)
  t_plus <- max(model_km$time)
  s_t_plus <- model_km$surv[ostatnie_t]
  df_ogon <- NULL

  if (s_t_plus > 0){
    t <- seq(from = t_plus, to = t_max, length.out = 200)
    estymator <- exp((log(s_t_plus) / t_plus) * t)

    df_ogon <- data.frame(time = t, surv = estymator,
      type = "Ogon BHK")
  }
  else
    df_ogon <- data.frame(time = c(t_plus, t_max),
      surv = c(0, 0), type = "Ogon BHK")

  df <- rbind(df_km, df_ogon)
  df$label <- label
  return(df)
}
```



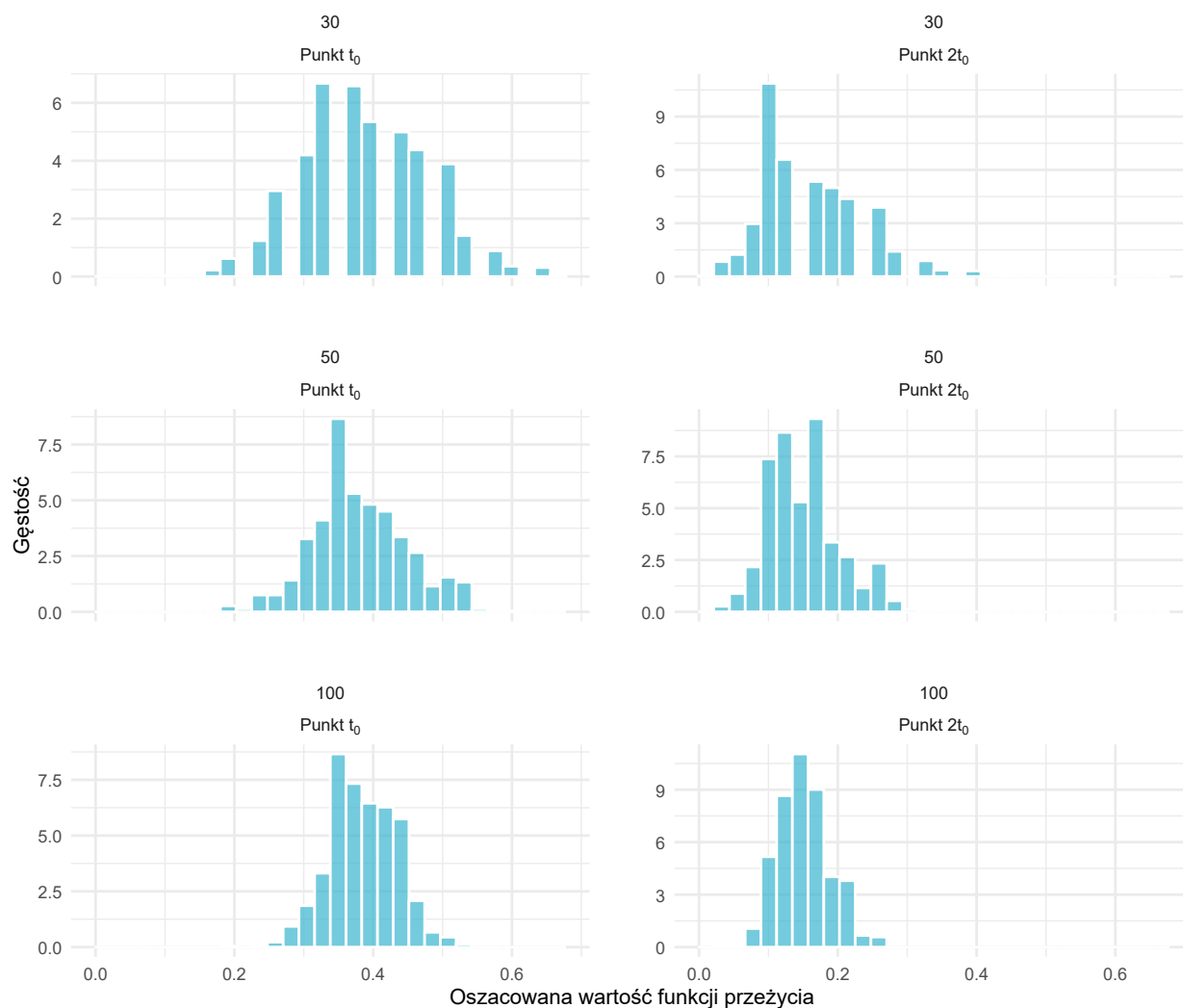
Rysunek 3: Estymator Fleminga-Harringtona - BHK dla leku A i B

Zastosowanie estymatora Browna, Hollandera i Kowara (zobacz rysunek 3) pozwoliło na prognozę przebiegu remisji poza okres obserwacji. Dorysowany ogon dąży do zera, co sugeruje, że przy wystarczająco długim czasie obserwacji remisja wystąpiłaby u całej populacji pacjentów. Proces ten jest jednak rozciągnięty w czasie. Model przewiduje, że po upływie dwóch lat odsetek pacjentów bez remisji spadłby do poziomu 0.25. Wskazuje to na konieczność dłuższego monitorowania pacjentów w celu potwierdzenia skuteczności leków.

1.3 Analiza asymptotycznej normalności estymatora Kaplana-Meiera

Wyznaczymy oszacowania funkcji przeżycia dla danych cenzurowanych I-go typu. Na podstawie wygenerowanych zbiorów obliczono wartości funkcji przeżycia w punktach t_0 i $2t_0$, korzystając z estymatora Kaplana-Meiera z ogonem estymowanym zgodnie z propozycją Browna, Hollandera i Kowara. Otrzymane wyniki przedstawiono w formie histogramów.

Histogramy dla $n = 30, 50, 100$ w punktach t_0 i $2t_0$ dla parametrów $\lambda = 0.5$, $\alpha = 1.2$, $t_0 = 2.2$



Rysunek 4: Oszacowania funkcji przeżycia dla danych cenzurowanych I-go typu

Analiza otrzymanych histogramów (zobacz rysunek 4) nie potwierdza przypuszczenia o normalności rozkładu estymatora w badanych punktach. Otrzymane rozkłady są nieregularne i niesymetryczne, co świadczy o tym, że przy takich liczebnościach prób przybliżenie rozkładem normalnym jest niewystarczające.

2 Estymacja średniego czasu życia

2.1 Algorytmy estymacji średniego czasu życia

Celem było przygotowanie programu służącego do estymacji średniego czasu życia z wykorzystaniem dwóch estymatorów funkcji przeżycia: Kaplana-Meiera oraz Fleminga-Harringtona. W obu przypadkach należało zastosować metodę przedłużenia ogona zaproponowaną przez Browna, Hollandera i Kowara. Różnica między estymatorami sprowadza się jedynie do sposobu wyznaczania funkcji przeżycia, kontrolowanego w funkcji `survfit` parametrem `type`. Pozostałe elementy algorytmu są identyczne dla obu podejść. Z tego powodu zamiast tworzyć dwa odrębne programy przygotowano jedno uniwersalne rozwiązanie, w którym użytkownik może wybrać estymator za pomocą parametru `type`.

```
library(survival)

srednia_z <- function(czas, status, type){
  model <- survfit(Surv(czas, status) ~ 1, type = type)

  czasy <- c(0, model$time)
  przezycie <- c(1, model$surv)
  punkty <- length(model$time)
  pole <- 0

  for (i in 1:punkty){
    szerokosc <- czasy[i+1] - czasy[i]
    wysokosc <- przezycie[i]
    pole <- pole + (szerokosc * wysokosc)
  }

  t_plus <- max(model$time)
  s_t_plus <- model$surv[punkty]
  pole_ogon <- 0

  if (s_t_plus > 0){
    lambda <- -log(s_t_plus) / t_plus
    pole_ogon <- s_t_plus / lambda
  }

  sredni_czas <- pole + pole_ogon

  return(sredni_czas)
}
```

2.2 Porównawcza analiza średniego czasu do remisji

Korzystając z napisanego wyżej programu, wyznaczono oszacowane średnie czasy do remisji choroby pacjentów leczonych lekiem A i lekiem B.

```
wynik_km_A <- srednia_z(dane.A$time, dane.A$status, "kaplan-meier")
wynik_fh_A <- srednia_z(dane.A$time, dane.A$status, "fleming-harrington")

wynik_km_B <- srednia_z(dane.B$time, dane.B$status, "kaplan-meier")
wynik_fh_B <- srednia_z(dane.B$time, dane.B$status, "fleming-harrington")
```

Wyniki oszacowań widoczne są poniżej.

Tabela 1: Oszacowanie średniego czasu do remisji

Lek	Estymator Kaplana-Meiera	Estymator Fleminga-Harringtona
Lek A	1.4321	1.4840
Lek B	1.3943	1.4471

Analiza oszacowań średniego czasu do remisji (Tabela 1) wskazuje na przewagę leku B. Niezależnie od zastosowanego estymatora, średni czas oczekiwania na ustąpienie objawów jest krótszy w grupie pacjentów przyjmujących lek B w porównaniu do grupy leku A. Wartości te, przekraczające czas trwania badania (1 rok), wynikają z zastosowania estymacji ogona metodą Browna, Hollandera i Kowara, która uwzględnia prognozowany czas leczenia pacjentów, u których remisja nie nastąpiła przed końcem obserwacji.

3 Wnioskowanie przedziałowe

3.1 Konstrukcja przedziałów ufności

Zaimplementowana poniżej funkcja służy do wyznaczania realizacji przedziałów ufności dla średniego czasu do progresji choroby. Funkcja przyjmuje jako argumenty dane cenzurowane losowo, poziom ufności oraz wartość graniczną τ , zwracając dolną i górną granicę przedziału ufności oraz oszacowania punktowe.

```
library(survival)

sredni_czas_przezycia <- function(czas, status, poziom_ufnosci, tau){
  model <- survfit(Surv(czas, status) ~ 1)
  czasy <- c(0, model$time)
  przezycie <- c(1, model$surv)
  r <- model$n.risk
  d <- model$n.event
  punkty <- length(model$time)
  mu_tau <- 0
  for (i in 1:length(czasy)){
    t_start <- czasy[i]
    if (t_start >= tau) break
    if (i < length(czasy)) t_end <- min(czasy[i+1], tau)
    else t_end <- tau
    mu_tau <- mu_tau + ((t_end - t_start) * przezycie[i])
  }
  suma <- 0
  pole <- 0
  for (i in 1:punkty){
    t_biezacy <- model$time[i]
    pole <- pole + ((t_biezacy - czasy[i]) * przezycie[i])
    if (t_biezacy > tau) break
    if (d[i] > 0 && (r[i] - d[i]) > 0){
      calka_reszta <- mu_tau - pole
      if (calka_reszta < 0)
        calka_reszta <- 0
      suma <- suma + (calka_reszta^2) * (d[i] / (r[i] * (r[i] - d[i])))
    }
  }
  alpha <- 1 - poziom_ufnosci
  z <- qnorm(1 - alpha/2)
  dolna_granica <- mu_tau - z * sqrt(suma)
  gorna_granica <- mu_tau + z * sqrt(suma)
  return(c(P = mu_tau, L = dolna_granica, U = gorna_granica))
}
```

3.2 Estymacja przedziałowa średniego czasu do progresji choroby

Analizowano dane z badania klinicznego, którego celem było sprawdzenie, czy stopień zaawansowania choroby ma związek z czasem do jej progresji. Dane dotyczyły dwóch stadiów: niskiego stopnia zaawansowania (II) oraz wysokiego stopnia zaawansowania (IIIA). Na podstawie uzyskanych danych wyznaczono realizację przedziałów ufności na poziomie ufności 0.95, dla średniego czasu do progresji choroby w dwóch badanych grupach pacjentek. Obliczenia przeprowadzono dla trzech wybranych wartości τ , z których dwie odpowiadają maksymalnym czasom obserwacji w poszczególnych grupach.

Tabela 2: Wyniki estymacji średniego czasu do progresji choroby

Stopień zaawansowania	τ	Oszacowanie punktowe	Przedział ufności
niski (II)	1500	937.733	[589.91; 1285.557]
wysoki (IIIA)	1500	413.204	[232.733; 593.675]
niski (II)	1206	774.400	[507.705; 1041.095]
wysoki (IIIA)	1206	387.479	[249.103; 525.856]
niski (II)	1119	726.067	[483.181; 968.952]
wysoki (IIIA)	1119	379.867	[253.659; 506.074]

Analiza uzyskanych wyników (zobacz Tabela 2) wskazuje na to, że wyższy stopień zaawansowania raka jajnika (IIIA) wiąże się ze znacznie gorszym rokowaniem. Dla $\tau = 1206$ średni czas do progresji w grupie o niskim stopniu zaawansowania wynosi około 774 dni, podczas gdy w grupie o wysokim stopniu jest on dwukrotnie krótszy (387 dni). Podobna zależność występuje również dla $\tau = 1119$, gdzie wartości te wynoszą około 726 i 380 dni. Niewielkie nakładanie się przedziałów ufności nie zmienia ogólnego wniosku, pacjentki z mniej zaawansowaną chorobą cechują się lepszym rokowaniem niż te ze stadium IIIA.

4 Testowanie hipotez o równości rozkładów w grupach o różnym stopniu zaawansowania choroby

4.1 Weryfikacja hipotezy o zgodności rozkładów czasu przeżycia

Przeprowadzono weryfikację hipotezy o jednakowym rozkładzie czasu do progresji choroby w dwóch analizowanych grupach pacjentek. W tym celu zastosowano kilka rodzajów testów statystycznych, przyjmując poziom istotności równy 0.05.

```
model <- survfit(Surv(czas, status) ~ grupa, data = dane)
```

4.1.1 Test log-rank

```
test_logrank <- surv_pvalue(model, method = "log-rank")
```

Tabela 3: Wynik testu Log-rank

Metoda	Wartość p
Log-rank	0.0183

Wniosek: Ponieważ wartość p jest mniejsza od przyjętego poziomu istotności $\alpha = 0.05$ ($0.0183 < 0.05$), odrzucamy hipotezę zerową.

4.1.2 Test Gehana-Breslowa

```
test_GB <- surv_pvalue(model, method = "gehan-breslow")
```

Tabela 4: Wynik testu Gehana-Breslowa

Metoda	Wartość p
Gehan-Breslow	0.135

Wniosek: Ponieważ $p > 0.05$, nie ma podstaw do odrzucenia hipotezy zerowej.

4.1.3 Test Tarone'a-Warego

```
test_TW <- surv_pvalue(model, method = "tarone-ware")
```

Tabela 5: Wynik testu Tarone-Ware

Metoda	Wartość p
Tarone-Ware	0.0555

Wniosek: Wynik jest na granicy istotności ($p \approx 0.05$), ale formalnie $p > 0.05$, więc nie ma podstaw do odrzucenia hipotezy zerowej.

4.1.4 Test Peto-Peto

```
test_PP <- surv_pvalue(model, method = "peto-peto")
```

Tabela 6: Wynik testu Peto-Peto

Metoda	Wartość p
Peto-Peto	0.1021

Wniosek: Ponieważ $p > 0.05$, nie ma podstaw do odrzucenia hipotezy zerowej.

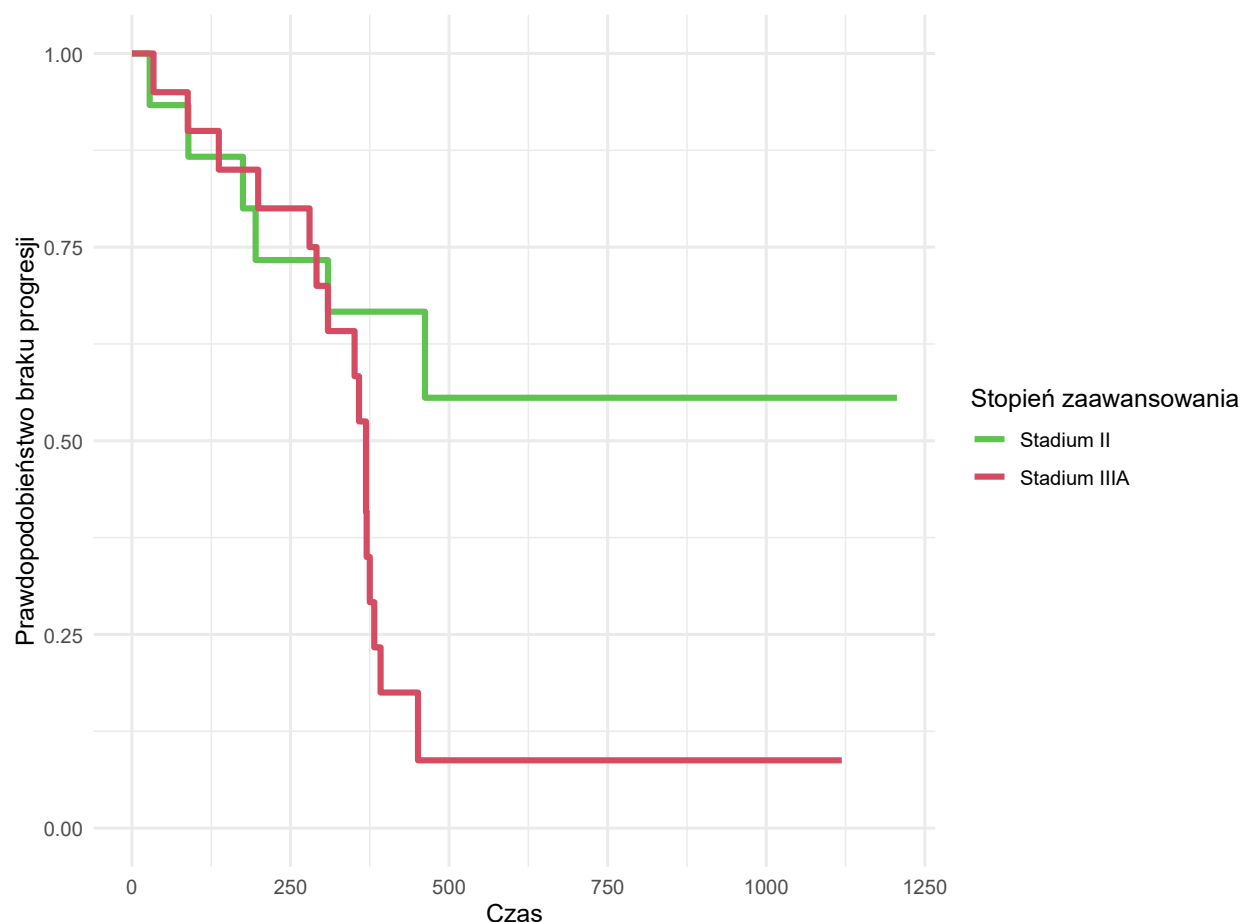
Ogólne wnioski

- Różnice między wynikami testów wynikają z tego, że każdy z nich inaczej waży zdarzenia w czasie trwania badania. Testy Gehana-Breslowa i Peto-Peto kładą największy nacisk na zdarzenia z początku obserwacji, dlatego ich wyniki sugerują brak istotnych różnic między grupami we wczesnym okresie. Test Log-rank traktuje wszystkie zdarzenia jednakowo, dzięki czemu lepiej wykrywa rozbieżności pojawiające się w późniejszym czasie. Test Tarone-Ware, stosujący wagi pośrednie, dał wynik na granicy istotności ($p \approx 0.055$), co wskazuje na stopniowe narastanie różnic w czasie.
- Testy Gehana-Breslowa czy Peto-Peto nie odrzuciły hipotezy ($p > 0.05$), co oznacza, że na początku obserwacji przebieg choroby w obu grupach wygląda podobnie.
- Wynik testu Log-rank prowadzi do odrzucenia hipotezy o jednakowym przebiegu choroby. Otrzymany efekt sugeruje, że stadium zaawansowania nowotworu stanowi istotny czynnik różnicujący pacjentki.
- Korzystniejsza sytuacja pacjentek w stadium II ujawnia się dopiero wraz z upływem czasu. W dalszych fazach obserwacji progresja choroby występuje u nich znacznie rzadziej niż w grupie IIIA.

4.2 Wizualizacja funkcji przeżycia i wag statystyk testowych

Wyznamy wykresy estymatorów Kaplana-Meiera funkcji przeżycia dla czasu do progresji choroby w dwóch badanych grupach oraz funkcji wagowych w statystykach testowych rozważanych wyżej.

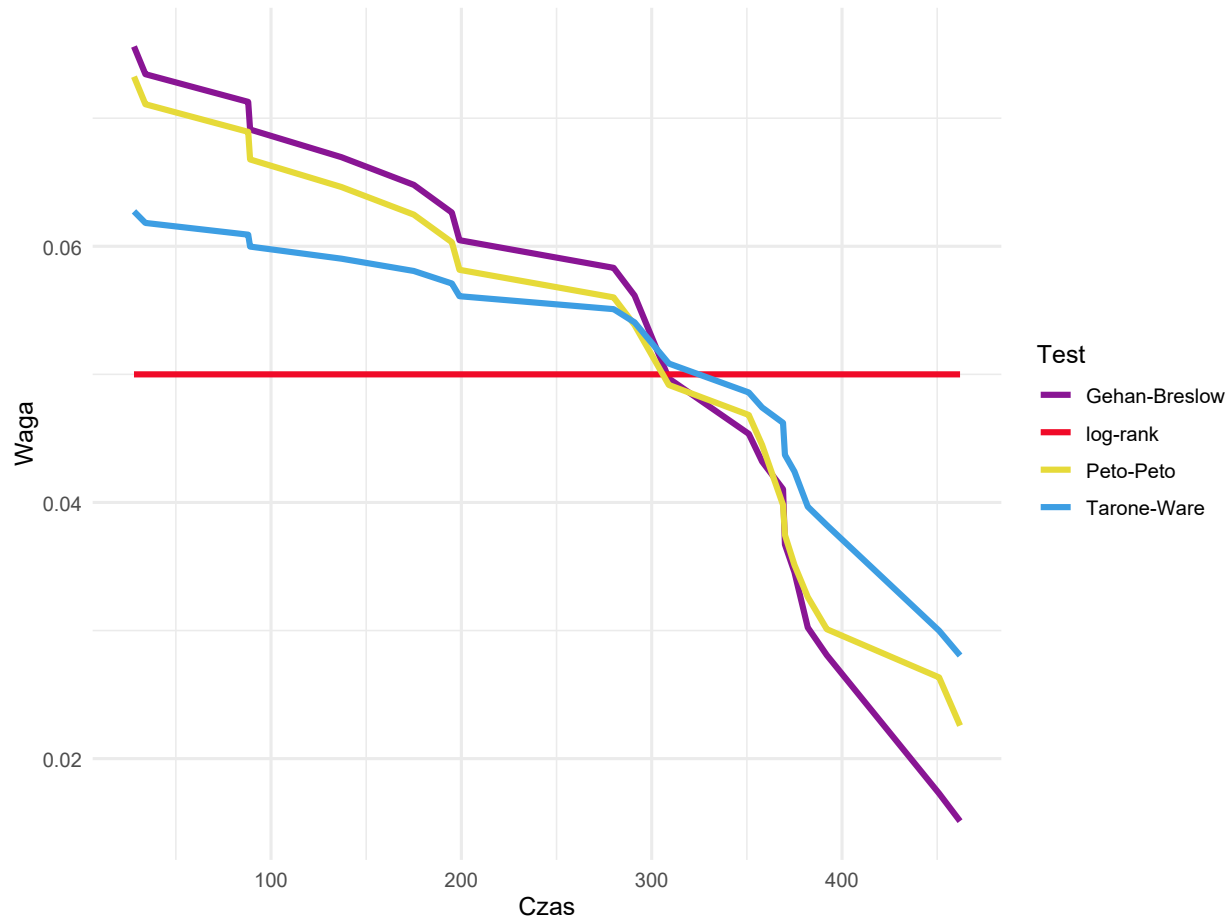
- Wykres przedstawiający estymatory Kaplana-Meiera:



Rysunek 5: Wykresy estymatorów Kaplana-Meiera funkcji przeżycia dla czasu do progresji choroby

Wykres 5 wizualnie potwierdza, że pacjentki z niższym stopniem zaawansowania mają większe szanse na długotrwałe przeżycie bez progresji choroby w porównaniu do pacjentek ze stadium zaawansowanego. Różnica między grupami jest bardzo wyraźna. Linia dla stadium II stabilizuje się na poziomie powyżej 0.50, podczas gdy linia dla stadium IIIA spada poniżej 0.25.

- Wykres przedstawiający wykresy funkcji wagowych w statystykach testowych:



Rysunek 6: Wykresy funkcji wagowych w statystykach testowych

Na wykresie 6 można zauważyć, że test log-rank stosuje stałe wagi, traktując jednakowo zdarzenia w całym czasie obserwacji, co sprzyja wykrywaniu różnic o stałym charakterze. W przeciwieństwie do niego, testy Gehan-Breslow, Peto-Peto i Tarone-Ware używają wag malejących, co czyni je bardziej czułymi na różnice pojawiające się we wczesnej fazie badania. Zatem w przypadku występowania wczesnych efektów testy ważone wykażą wyższą moc i niższe wartości p niż test log-rank.