W4 – Eksperyment niezawodnościowy

Henryk Maciejewski

Marek Woda

Badania niezawodnościowe i analiza statystyczna wyników

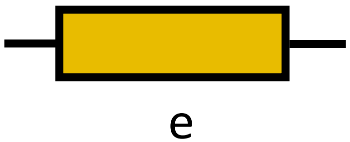
1. Co to są badania niezawodnościowe i jak się je prowadzi

2. Sposoby prezentacji danych z eksperymentu

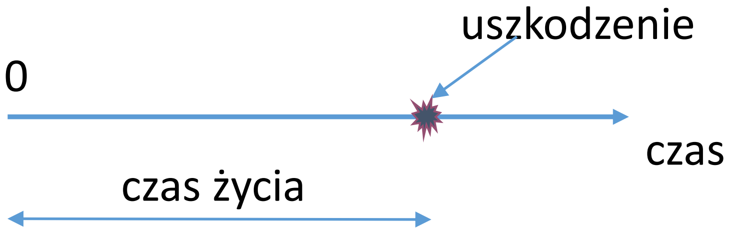
3. Sposoby wyznaczanie rozkładu zmiennej losowej na podstawie danych z eksperymentu

Co to są badania niezawodnościowe i jak się je przeprowadza

Celem badań jest określenie rozkładu zmiennej losowej opisującej czas życia elementu.



Czas życia określamy zwykle na podstawie eksperymentu.



zmienna losowa T
– określająca
model niezawodnościowy
elementu

Pojęcia podstawowe i stosowana terminologia

Populacja – zbiór wszystkich elementów danego rodzaju

Próba losowa – wybrane w pewien sposób niektóre elementy populacji, opisane przez wektor:

$$T = T_1, T_2, \ldots, T_n$$

T_i - zmienna losowa, czas życia i-tego elementu ($1 \leq i \leq n$)

Badanie – sposób uzyskania wiedzy o elementach z próby losowej, wynikiem badania jest

$$t_1, t_2, \ldots, t_n$$

t_i – realizacja zmiennej losowej T_i

Statystyka – funkcja wektora losowego

$$S(T) = S(T_1, T_2, \ldots, T_n)$$

Przeprowadzanie badań niezawodnościowych

- Pobranie próby losowej
- Notowanie chwil uszkodzeń kolejnych elementów, z jednoczesnym sprawdzaniem czy nie jest spełnione kryterium zakończenia badania

Kryteria zakończenia badania:

- Uszkodzenie wszystkich n elementów próby losowej (badanie pełne)
- Przekroczenie dopuszczalnego czasu badania T_{test} (próba obcięta, *Type I censoring, time-censored test*)
- Uszkodzenie zadanej liczby (k<n) elementów (próba obcięta, *Type II censoring, failure-censored test*)
- Uszkodzenie zadanej liczby elementów k lub przekroczenie dopuszczalnego czasu badania T_{test}

Wynik badania niezawodnościowego

- Dla próby pełnej – zbiór realizacji czasów życia:

$$t_1 < t_2 < \ldots < t_n$$

- Dla próby obciętej:

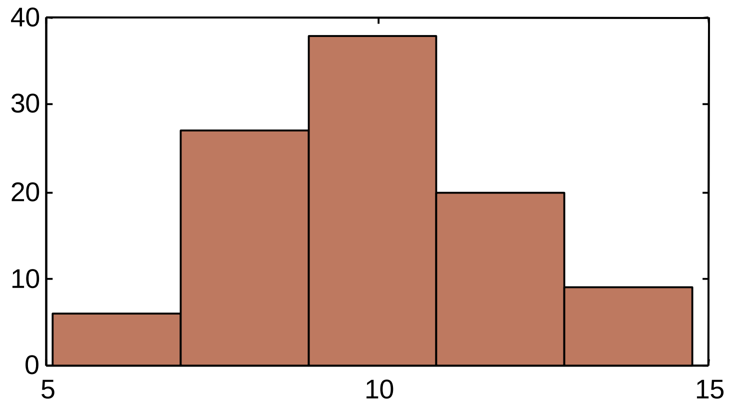
$t_1 < t_2 < \ldots < t_k$ – czasy życia zmierzone dla k elementów, które uległy uszkodzeniu

$T_1, \ldots, T_{n-k}$ – czasy pracy n-k elementów, które do końca testu nie uległy uszkodzeniu

<u>Przyspieszone badania niezawodności</u>: jeśli w warunkach nominalnych nie jest możliwe w realnym czasie przeprowadzenie eksperymentu, wówczas wymusza się szybsze wystąpienie uszkodzeń przez zwiększenie poziomu stresu, a następnie odwzorowuje się wyniki badań na warunki nominalne.

Sposoby prezentacji danych z eksperymentu

Histogram

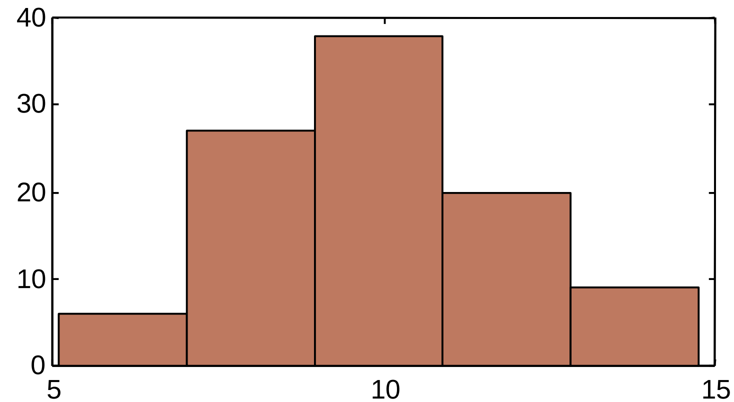


Budowa histogramu:

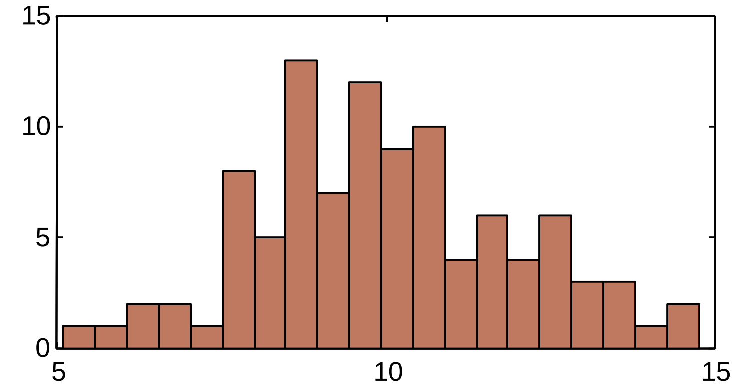
- Wyznaczamy zakres rysowania
- Podział zakresu rysowania na równe przedziały
- Wyznaczenie dla każdego przedziału liczby należących do niego danych (lub częstości)
- Narysowanie wykresu słupkowego

Przykład

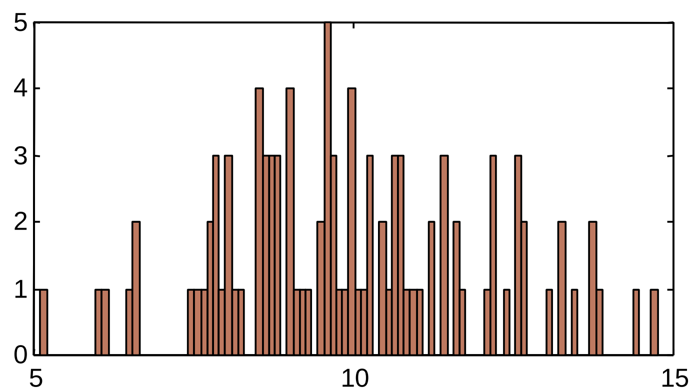
Z rozkładu normalnego o parametrach $\mu = 10$ i $\sigma = 2$, wylosowano $n = 100$ liczb. Wpływ liczby m przedziałów na wygląd histogramu.



m=5



m=20

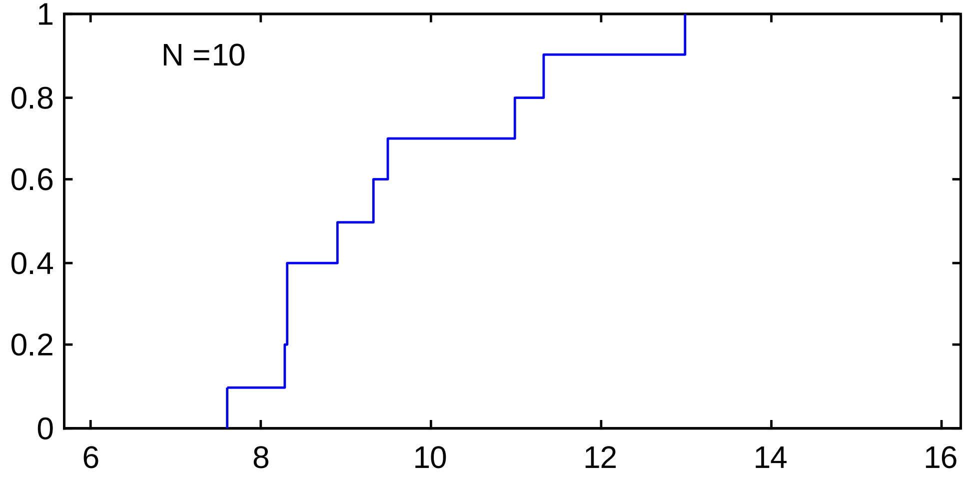


m=100

Dystrybuanta empiryczna

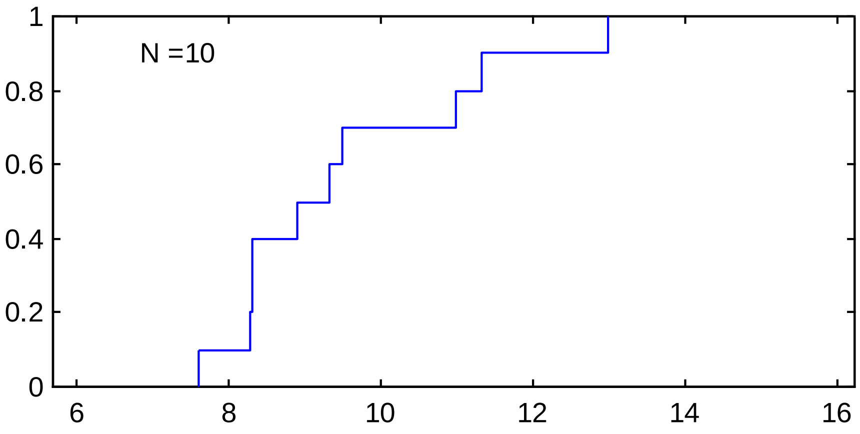
Funkcja określona następująco:

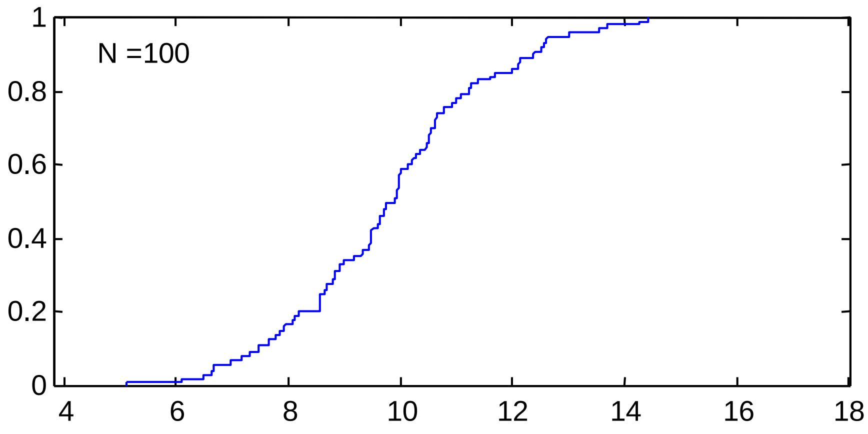
$$\hat{F}_n(t) = \begin{cases} 0 & t \le t_1 \\ \dfrac{k}{n} & t_k < t \le t_{k+1} \\ 1 & t > t_n \end{cases}$$

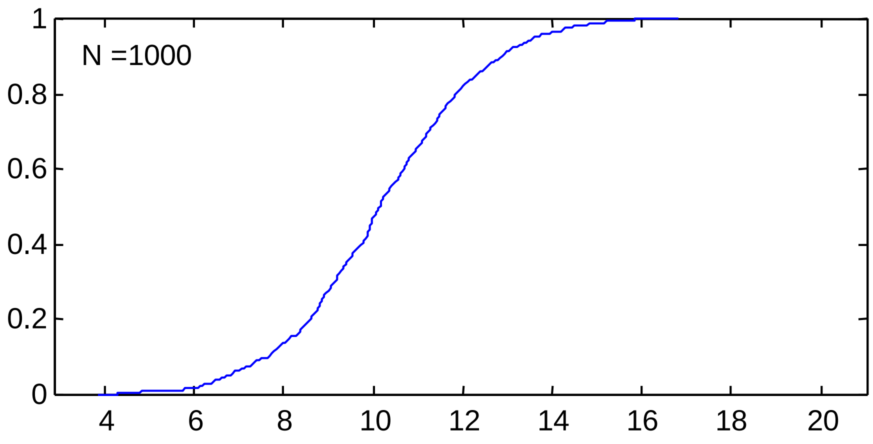


Przykład

Z rozkładu normalnego o parametrach $\mu = 10$ i $\sigma = 2$, wylosowano N liczb. Jak wygląda dystrybuanta empiryczna dla różnych N?







Twierdzenie (Gliwienko-Cantelli)

Jeśli F(t) jest dystrybuantą, z której pobrano próbę losową i niech

$$D_n = \sup_{-\infty<t<+\infty} \left|\hat{F}_n(t) - F(t)\right|$$

Wówczas

$$P\left(\lim_{n\to\infty} D_n = 0\right) = 1$$

Dystrybuanta empiryczna zbiega się do teoretycznej z prawdopodobieństwem 1.

Wyznaczanie rozkładu zmiennej losowej na podstawie danych z eksperymentu

Sformułowanie zadania estymacji parametrycznej

Zakładamy że wektor $T = T_1, T_2, \ldots, T_n$ opisujący próbę losową został pobrany z parametrycznej rodziny rozkładów

$$\{F_\theta : \theta \in \Theta\}$$

przy czym wartość parametru (parametrów) θ identyfikującego rozkład nie jest znana.

Jak wyznaczyć nieznany parametr θ na podstawie próby losowej?

Estymacja θ metodą największej wiarogodności

Metodę podał R.A. Fisher

Jako wartość parametru θ przyjąć wartość $\hat{\theta}$, która maksymalizuje funkcję wiarogodności:

$$L(t_1, t_2, \dots, t_n, \theta) = \prod_{i=1}^{n} f(t_i, \theta)$$

gdzie f(t,θ) – gęstość rozkładu populacji.

Estymacja θ metodą największej wiarogodności

W przypadku <u>prób obciętych</u>, parametr $\hat{\theta}$ wyznaczamy poprzez maksymalizację wyrażenia:

$$L(t_1, t_2, \dots, t_k, T_1, T_2, \dots, T_{n-k}, \theta) =$$

$$= \prod_{i=1}^{k} f(t_i, \theta) \cdot \prod_{j=1}^{n-k} \left(1 - F\left(T_j, \theta\right)\right)$$

gdzie f(t,θ) – gęstość, F(t,θ) - dystrybuanta rozkładu populacji.

Komputerowe metody analizy wyników badań niezawodnościowych

Procedurę zilustrujemy przykładem (czasy życia transformatorów):

- zarejestrowano następujące czasy życia t
- censor = 0 oznacza czas do awarii,
- censor = 1 oznacza czas pracy do zakończenia badania, element nie uległ awarii.

t	censor
14219	1
14251	1
14277	1
14277	1
14277	1
14277	1
14323	1
14324	1
14383	1
14403	1
14961	1
15029	1
15129	1
15129	1
15129	1
15174	1
15187	1
15187	1
15192	1
14130	0
10795	0
8558	0
10667	0
6153	0
11795	0
9265	0
5685	0
8151	0
11599	0
5403	0
8618	0
10468	0
8711	0
12132	0
9917	0
9664	0
5372	0
3565	0
9141	0
7753	0
9327	0
5834	0

Jaki jest rozkład zmiennej losowej opisującej czas życie elementu?

Narzędzia:

- Uniwersalne oprogramowanie matematyczne (np. Matlab, R)
- Specjalistyczne oprogramowanie do analizy danych (np. system SAS firmy SAS Institute)

Przykładowa procedura (dane obcięte)

1. Sporządzić wykres probabilistyczny dla wybranych rozkładów prawdopodobieństwa (np. Weibulla, wykładniczego, …)

2. Jeśli dane pochodzą z danego rozkładu, wówczas naniesione na wykres probabilistyczny powinny ułożyć się mniej więcej wzdłuż linii prostej.

3. Wyznaczamy parametry rozkładu (np. metodą największej wiarogodności).

Wykres probabilistyczny dla danego rozkładu – uzyskujemy przez przeskalowanie osi t i F(t) tak żeby dystrybuanta dla tego rozkładu była linią prostą.

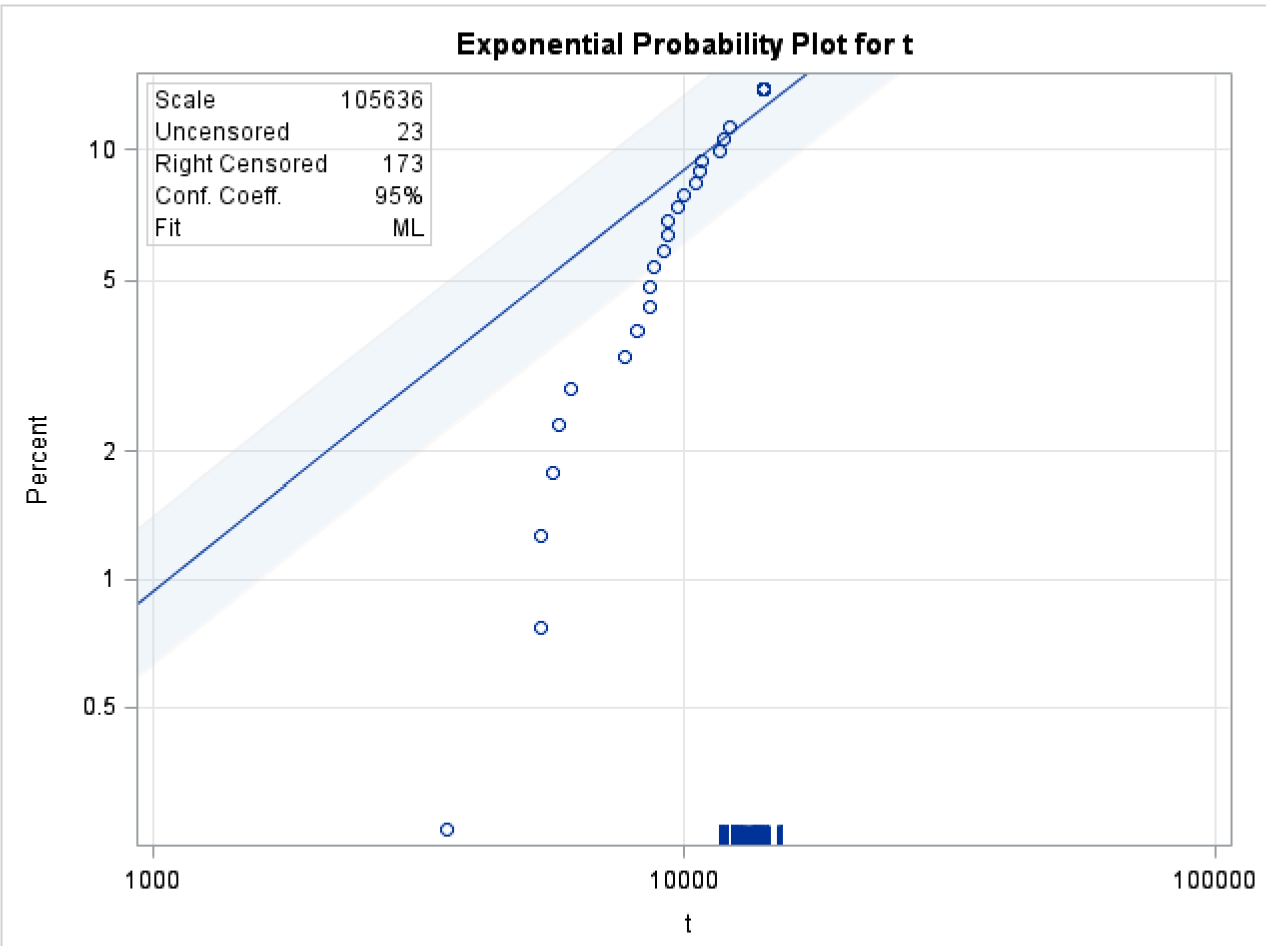
Np. dla rozkładu wykładniczego $F(t)=1-\exp(-\lambda t)$, zachodzi $\log(1-F(t)) = -\lambda t$ $\rightarrow$ stąd w układzie współrzędnych $(t, \log(1-F(t)))$ dystrybuanta dla tego rozkładu jest linią.

W praktyce często stosujemy log przy podstawie 10.

Przykład – analiza czasu życia transformatorów

Zakładamy rozkład wykładniczy

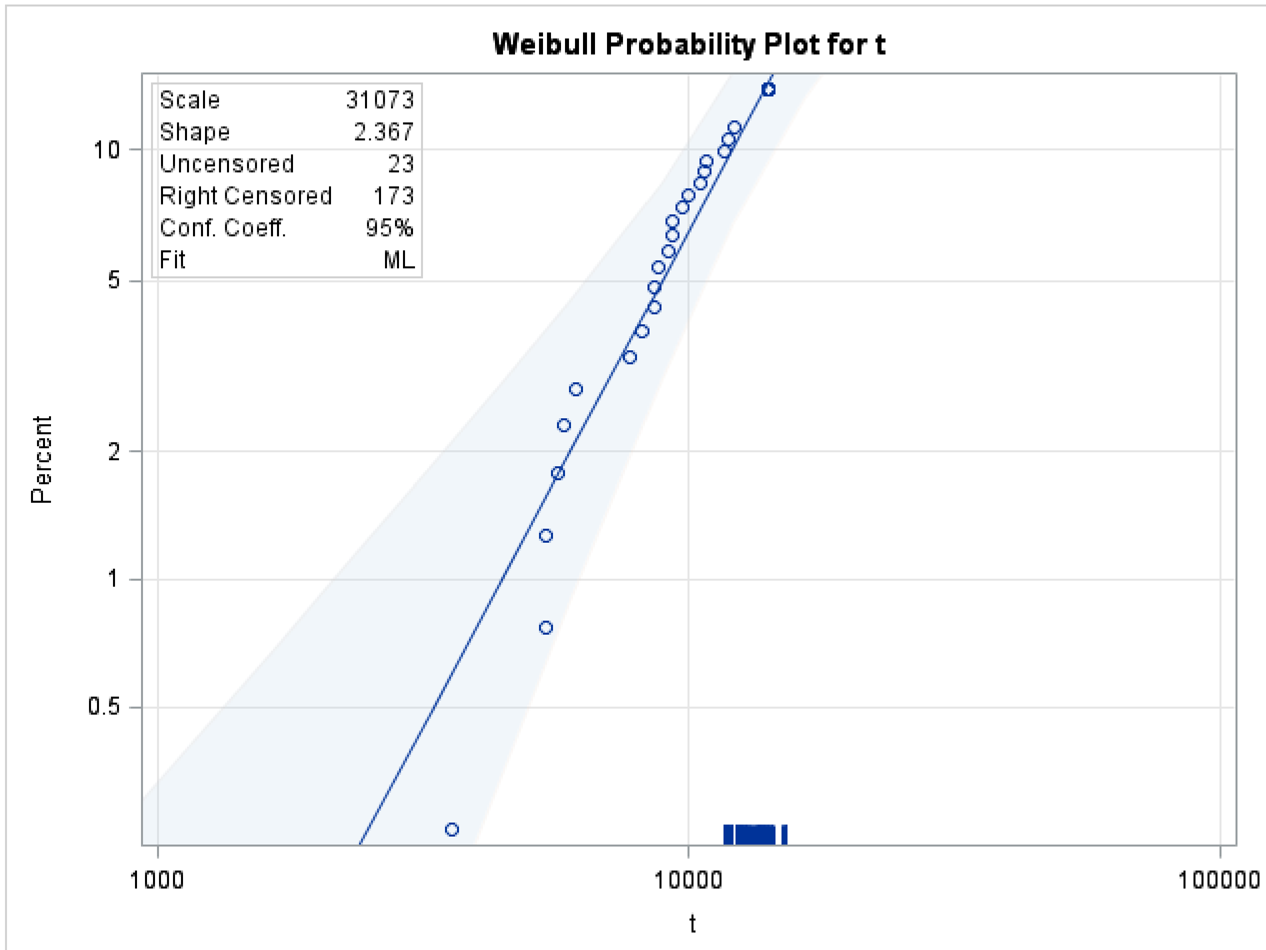
```
proc reliability data=trans;
     distribution exponential;
     pplot t*censor( 1 );
run;
```



t	censor
14219	1
14251	1
14277	1
14277	1
14277	1
14277	1
14323	1
14324	1
14383	1
14403	1
14961	1
15029	1
15129	1
15129	1
15129	1
15174	1
15187	1
15187	1
15192	1
14130	0
10795	0
8558	0
10667	0
6153	0
11795	0
9265	0
5685	0
8151	0
11599	0
5403	0
8618	0
10468	0
8711	0
12132	0
9917	0
9664	0
5372	0
3565	0
9141	0
7753	0
9327	0
5834	0

Zakładamy rozkład Weibulla

```
proc reliability data=trans;
      distribution weibull;
      pplot t*censor( 1 );
run;
```

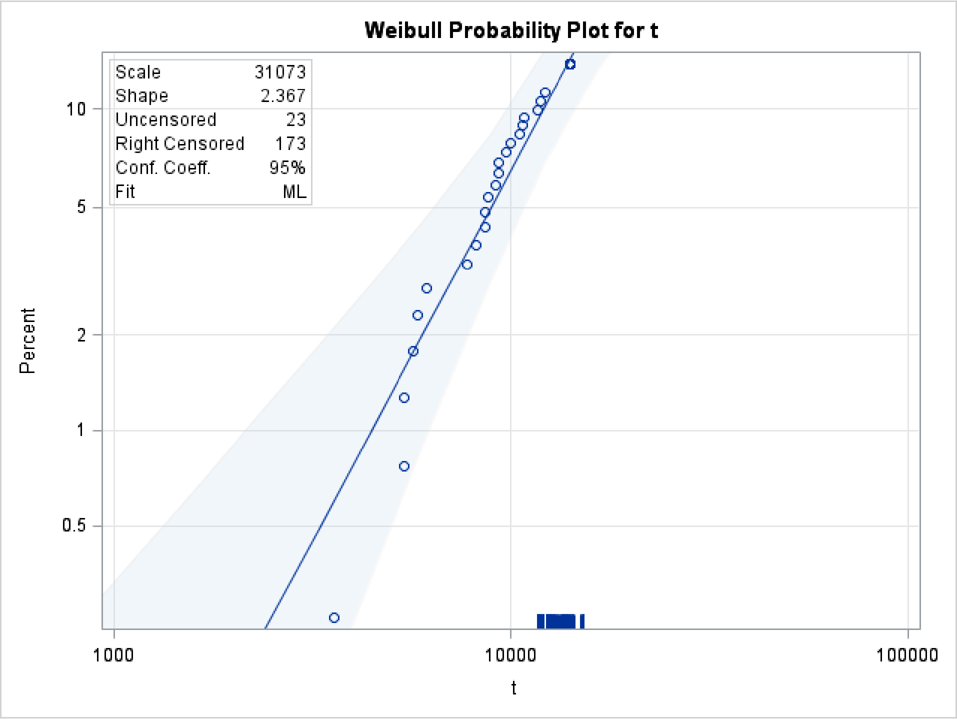


Analiza wyników

- Stwierdzamy, że dla rozkładu Weibulla dane „układają się” wzdłuż linii prostej – wybieramy więc ten rozkład
- Wyznaczamy parametry rozkładu (metoda największej wiarogodności):

  Scale = 31073

  Shape = 2.367



- Wartość średnia dla tego rozkładu:

  E(T) = 27539

Przykładowa procedura (dane pełne)

- Sporządzić histogram dla zarejestrowanych danych.
- Na podstawie kształtu histogramu założyć możliwy rozkład prawdopodobieństwa czasu życia.
- Dla wytypowanego rozkładu wyznaczyć parametry (np. metodą największej wiarogodności).
- Sprawdzić jakość dopasowania (*goodness-of-fit*), poprzez weryfikację hipotezy H0 o zgodności danych z założonym rozkładem.

  Procedura testowa:
  jeśli p-value < 0.05 – wówczas odrzucamy hipotezę o zgodności przyjętego rozkładu z danymi. Należy wtedy poszukać innego rozkładu.

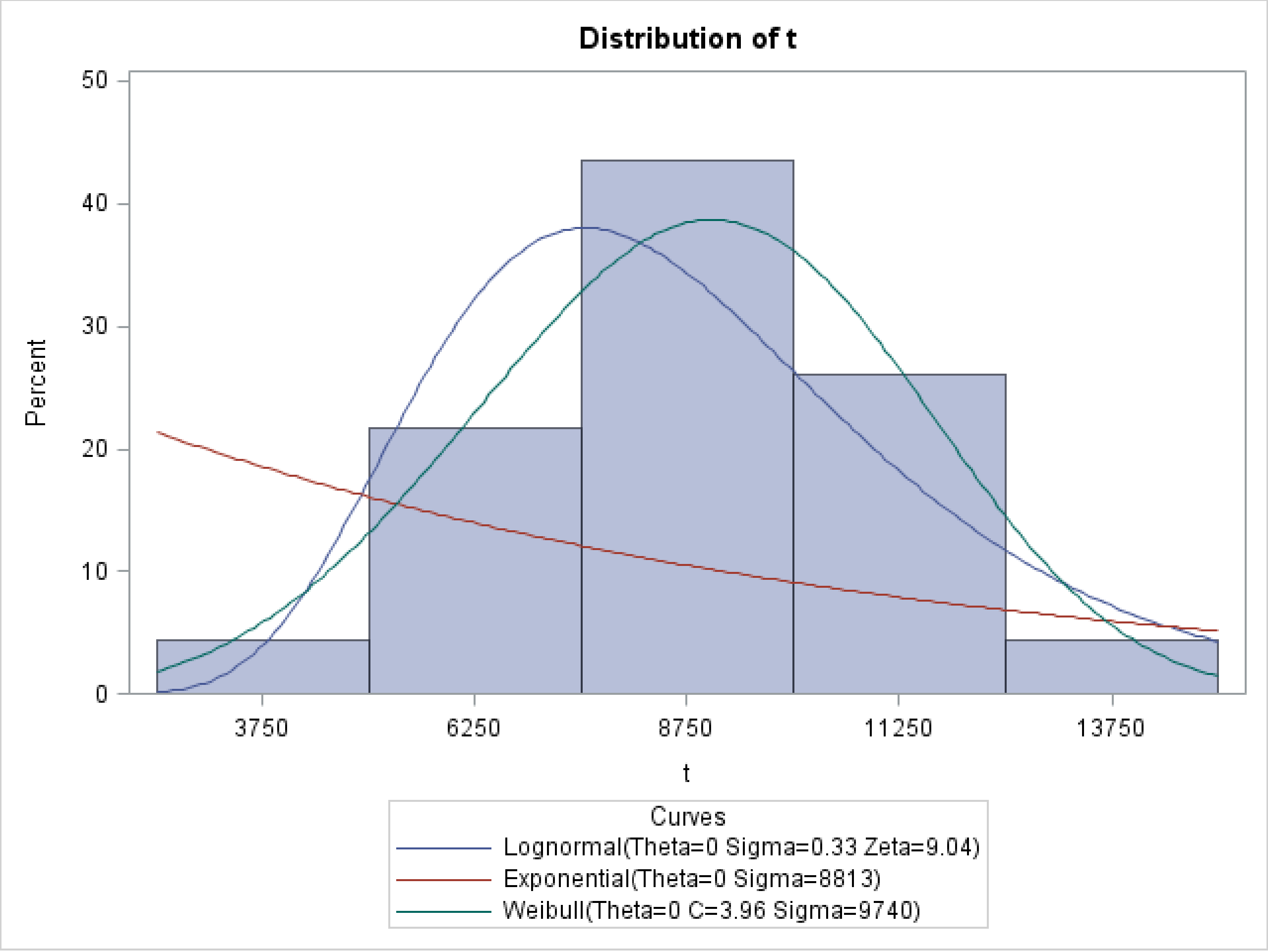
- Możemy również posłużyć się wykresami probabilistycznymi – jak w przykładzie dla prób obciętych.

Przykład – analiza próby pełnej

```
proc capability data=trans2 noprint;
      var t;
      histogram /   weibull
                    lognormal
                    exponential;
run;
```

Zakładamy rozkłady:

- Wykładniczy
- Weibulla
- Lognormalny



Goodness-of-fit testing

test zgodności rozkładu

H_0: zgodne
H_1: niezgodne

Goodness-of-Fit Tests for Lognormal Distribution					
Test	Statistic		DF	p Value	
Kolmogorov-Smirnov	D	0.17409624		Pr > D	0.070
Cramer-von Mises	W-Sq	0.11987306		Pr > W-Sq	0.058
Anderson-Darling	A-Sq	0.65180068		Pr > A-Sq	0.081
Chi-Square	Chi-Sq	2.37114531	2	Pr > Chi-Sq	0.306

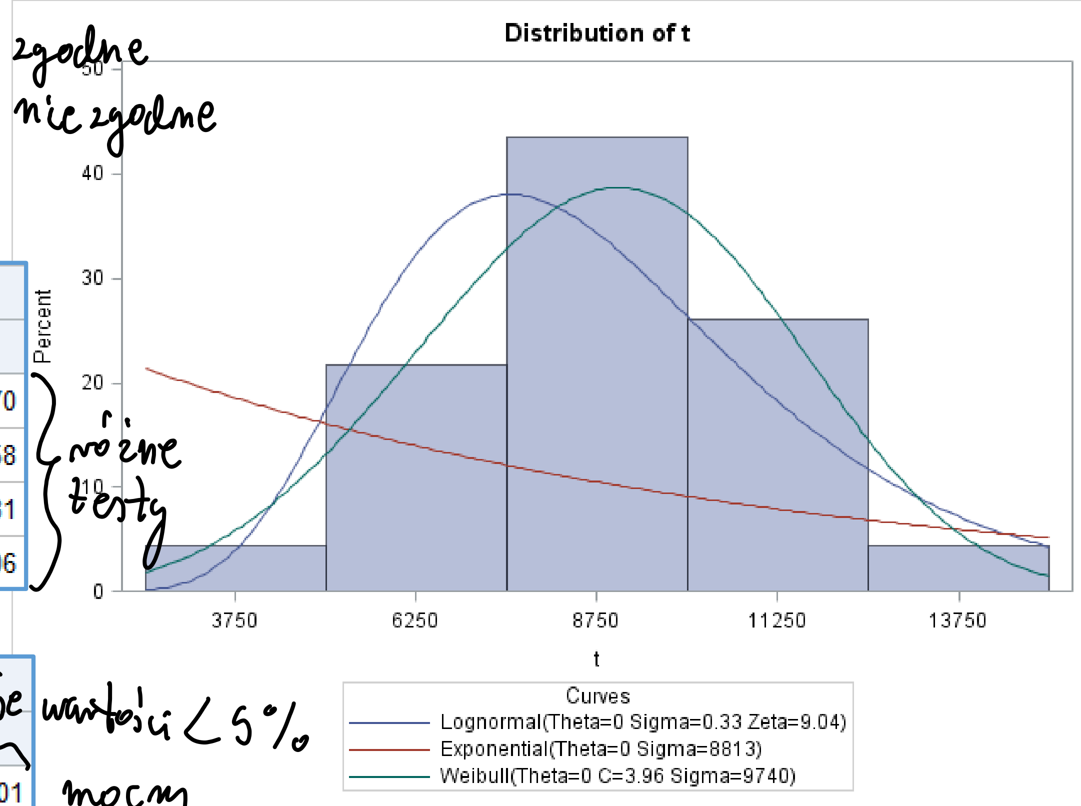
różne testy

jeżeli p jest małe np. 5%

Goodness-of-Fit Tests for Exponential Distribution					
Test	Statistic		DF	p Value	
Kolmogorov-Smirnov	D	0.4129220		Pr > D	<0.001
Cramer-von Mises	W-Sq	1.0943972		Pr > W-Sq	<0.001
Anderson-Darling	A-Sq	5.3663322		Pr > A-Sq	<0.001
Chi-Square	Chi-Sq	36.7861790	3	Pr > Chi-Sq	<0.001

małe wartości < 5%

mocny sygnał by odrzucić



wzorek/technika

Goodness-of-Fit Tests for Weibull Distribution					
Test	Statistic		DF	p Value	
Cramer-von Mises	W-Sq	0.03991042		Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.26384750		Pr > A-Sq	>0.250
Chi-Square	Chi-Sq	0.58073678	2	Pr > Chi-Sq	0.748

sens „odległości" 1 od 1

duże wartości / brak podstaw do odrzucenia

Analiza wartość p testu (p-value) wskazuje że czas życia ma rozkład Weibulla.

Parametry tego rozkładu:
Scale = 9740
Shape = 3.96

czy się często zdarza? :p

Wartość średnia = 8823

Analiza nieparametryczna

- Nie zakładamy, że czas życia należy do parametrycznej rodziny rozkładów
- Wyznaczamy estymatory nieparametryczne funkcji niezawodności, intensywności uszkodzeń itd. bezpośrednio z próby (obciętej).
- Narzędzie komputerowe – proc lifetest (system SAS)

```
title "Analiza nieparametryczna";
proc lifetest   data=trans
                method=lt
                ninterval=10
                plots=(s,p,h);
    time t*censor( 1 );
run;
```

t	censor
14219	1
14251	1
14277	1
14277	1
14277	1
14277	1
14323	1
14324	1
14383	1
14403	1
14961	1
15029	1
15129	1
15129	1
15129	1
15174	1
15187	1
15187	1
15192	1
14130	0
10795	0
8558	0
10667	0
6153	0
11795	0
9265	0
5685	0
8151	0
11599	0
5403	0
8618	0
10468	0
8711	0
12132	0
9917	0
9664	0
5372	0
3565	0
9141	0
7753	0
9327	0
5834	0

Analiza nieparametryczna

Przykład dla transformatorów – cd

```
title "Analiza nieparametryczna";
proc lifetest   data=trans
                method=lt
                ninterval=10
                plots=(s,p,h);
    time t*censor( 1 );
run;
```

Wyznacz:

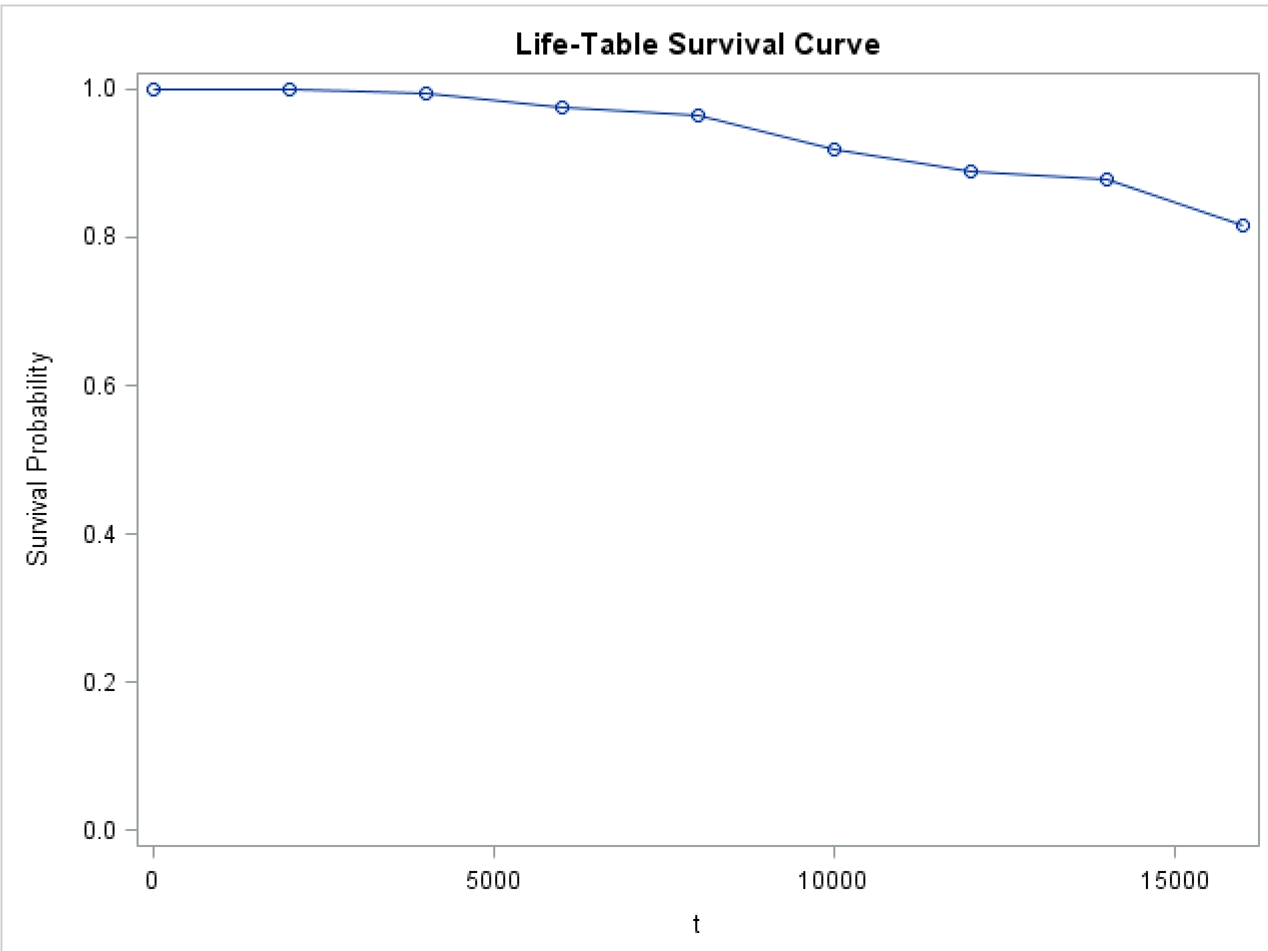
s – survival (funkcję niezawodności)

p – pdf (gęstość prawdopodobieństwa)

h – hazard (intensywność uszkodzeń)

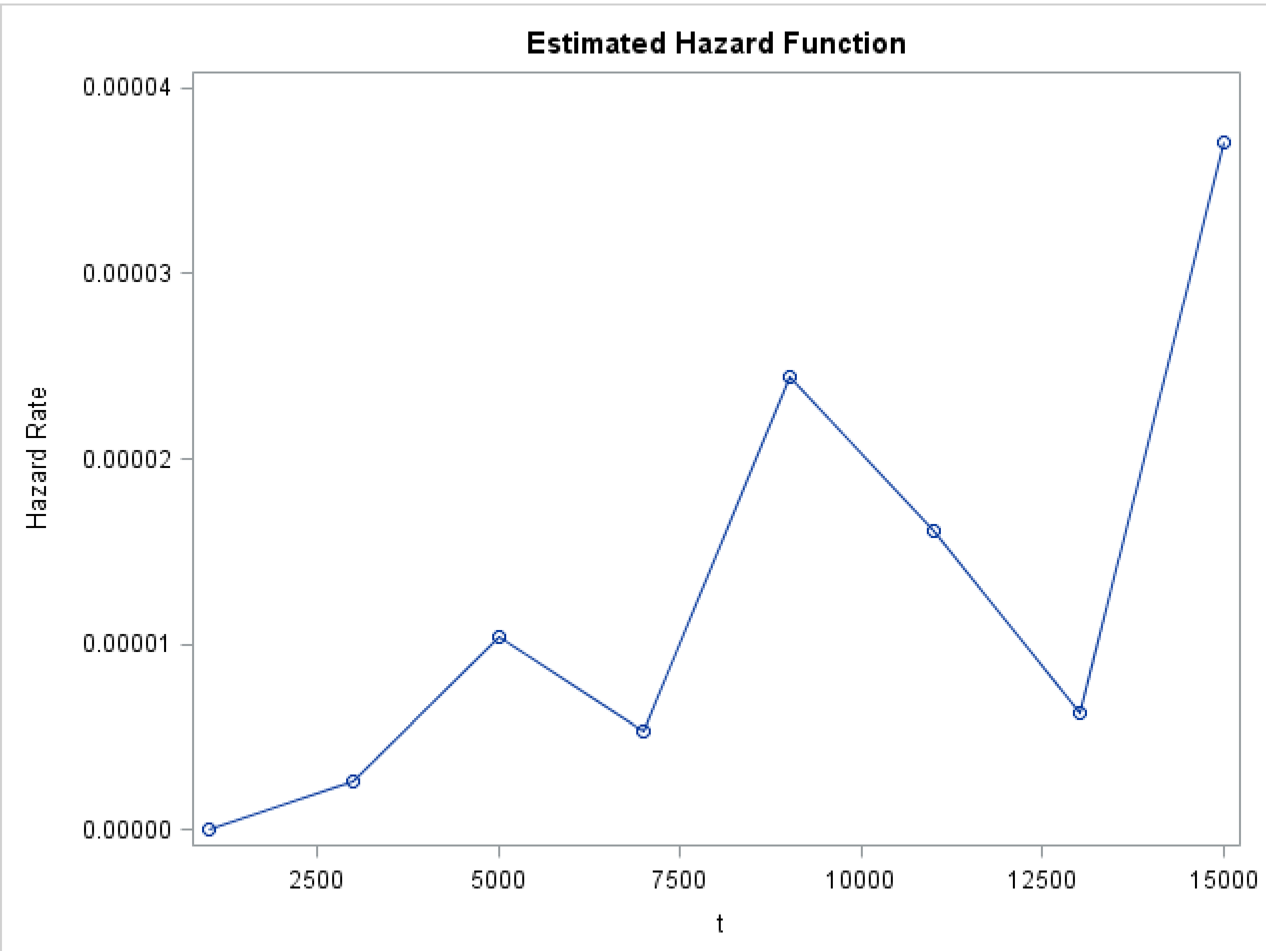
t	censor
14219	1
14251	1
14277	1
14277	1
14277	1
14277	1
14323	1
14324	1
14383	1
14403	1
14961	1
15029	1
15129	1
15129	1
15129	1
15174	1
15187	1
15187	1
15192	1
14130	0
10795	0
8558	0
10667	0
6153	0
11795	0
9265	0
5685	0
8151	0
11599	0
5403	0
8618	0
10468	0
8711	0
12132	0
9917	0
9664	0
5372	0
3565	0
9141	0
7753	0
9327	0
5834	0

```
title "Analiza nieparametryczna";
proc lifetest   data=trans
                method=lt
                ninterval=10
                plots=(s,p,h);
    time t*censor( 1 );
run;
```



t	censor
14219	1
14251	1
14277	1
14277	1
14277	1
14277	1
14323	1
14324	1
14383	1
14403	1
14961	1
15029	1
15129	1
15129	1
15129	1
15174	1
15187	1
15187	1
15192	1
14130	0
10795	0
8558	0
10667	0
6153	0
11795	0
9265	0
5685	0
8151	0
11599	0
5403	0
8618	0
10468	0
8711	0
12132	0
9917	0
9664	0
5372	0
3565	0
9141	0
7753	0
9327	0
5834	0

```
title "Analiza nieparametryczna";
proc lifetest   data=trans
                method=lt
                ninterval=10
                plots=(s,p,h);
    time t*censor( 1 );
run;
```



t	censor
14219	1
14251	1
14277	1
14277	1
14277	1
14277	1
14323	1
14324	1
14383	1
14403	1
14961	1
15029	1
15129	1
15129	1
15129	1
15174	1
15187	1
15187	1
15192	1
14130	0
10795	0
8558	0
10667	0
6153	0
11795	0
9265	0
5685	0
8151	0
11599	0
5403	0
8618	0
10468	0
8711	0
12132	0
9917	0
9664	0
5372	0
3565	0
9141	0
7753	0
9327	0
5834	0