

Analiza przeżycia

Raport 2

Romana Żmuda

13 grudnia 2020

Spis treści

1	Zadanie do sprawozdania - Część 1	3
1.1	Zadanie 1	3
1.2	Zadanie 2	6
1.3	Zadanie 3	7
2	Zadanie do sprawozdania - Część 2	8
2.1	Zadanie 1	8
2.2	Zadanie 2	12
2.3	Zadanie 3	12
3	Zadanie do sprawozdania - Część 3	13
3.1	Zadanie 1,2	13
3.2	Zadanie 3	15

1 Zadanie do sprawozdania - Część 1

1.1 Zadanie 1

Moim zadaniem jest narysowanie estymatora nieparametrycznego Kaplana - Meiera funkcji przeżycia na danych dotyczących pacjentów leczonych na niewydolność nerek. Zmienna *Czas* przyjmuje wartości 1 lub 0, gdzie pierwsza zmienna oznacza czas, gdy zaczęto dializować pacjenta od momentu rozpoznania choroby (w latach), natomiast 0 określa czas obserwacji pacjenta od momentu rozpoznania choroby (w latach). Pozostałe cztery zmienne dotyczą genotypu pacjenta. Interesuje nas czas od momentu rozpoznania choroby do momentu, gdy dializa jest konieczna do przeprowadzenia. W poniższych zadaniach czas od momentu rozpoznania choroby do dializy pełni rolę “czasu życia”. Oczywiście niektóre dane są cenzurowane (Zmienna *Cenzura*) i zakładamy, że czas cenzurowania jest niezależny od czasu do wystąpienia zdarzenia, co oznacza że możemy estymować te zmienne i nie będzie problemu z określeniem funkcji przeżycia. Nie przyjmujemy dodatkowych założeń dotyczących “czasu życia”, tzn. nie przyjmujemy postaci rozkładu tego czasu.

Estymator kaplana - meiera w podgrupie Arg25Pro

Poniżej wgrywamy te dane :

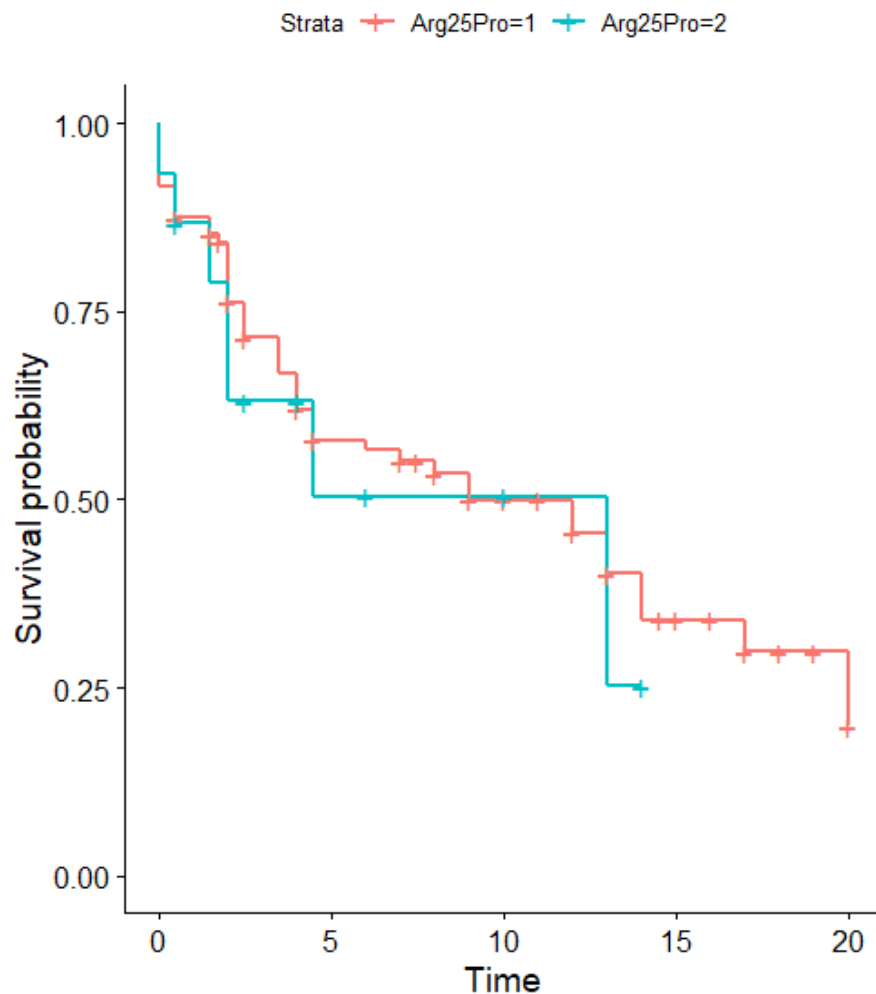
```
> dane3<-read.delim("CzasDoDializy.csv", header = TRUE, ";")
> dane3$Czas <- as.numeric(gsub(",", ".", as.character(dane3$Czas)))
>
```

Tworzymy ten estymator za pomocą bibliotek survival i survminer:

```
> surv_object1<-Surv(dane3$Czas, dane3$Cenzura)
> fit1 <- survfit(surv_object1 ~ dane3$Arg25Pro, data = dane3,
+               type = "kaplan-meier")
> g1<-ggsurvplot(fit1, data = dane3, conf.int = FALSE) + ggtitle("Estymator kaplana -
```

Dany estymator ukazujemy z podziałem w podgrupie Arg25Pro na wykresie [1](#).

Estymator kaplana - meiera dla zmiennych Arg



Rysunek 1: Estymator Kaplana - Meiera funkcji przeżycia w podgrupie Arg25Pro

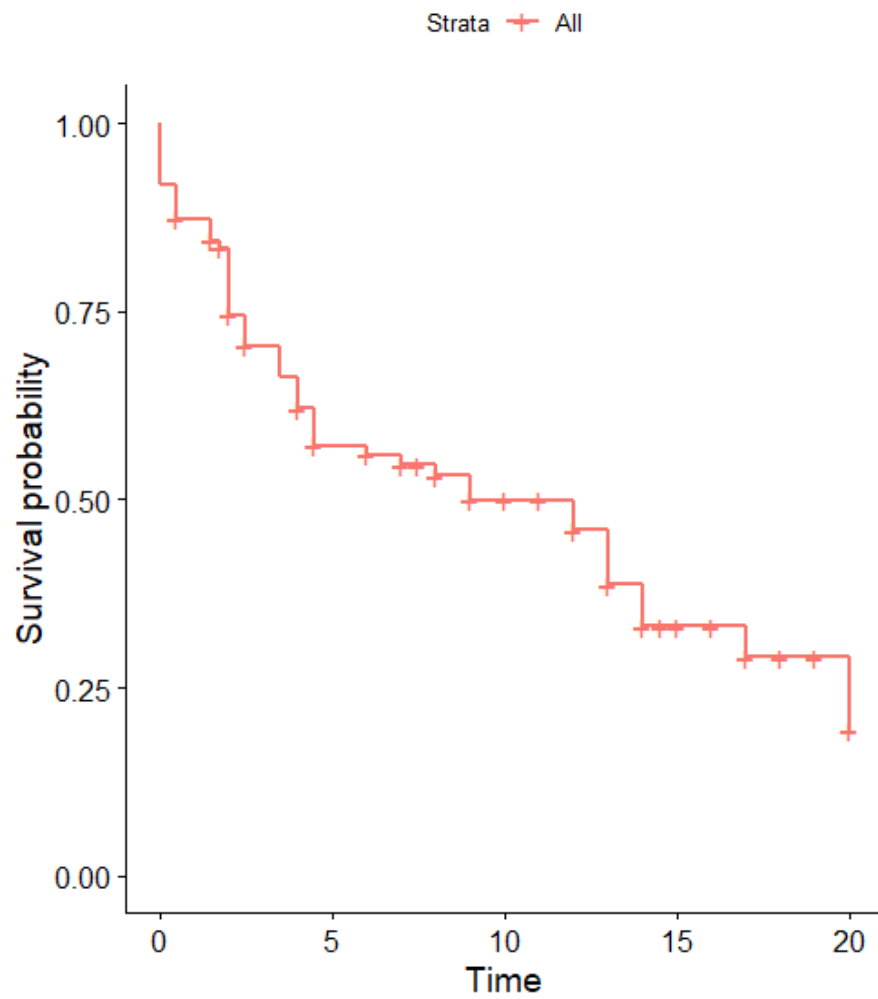
Estymator kaplana - meiera w całej bazie danych

Postępujemy analogicznie jak w przypadku podgrupy, jednak w funkcji *survfit* wpisujemy 1, gdyż to oznacza wszystkie zmienne.

```
> fit1.2 <- survfit(surv_object1 ~ 1, data = dane3,  
+                   type = "kaplan-meier")  
> g1.2 <- ggsurvplot(fit1.2, data = dane3, conf.int = FALSE) +  
+   ggtitle("Estymator kaplana - meiera w dla całej bazy danych")
```

Dany estymator ukazuje na wykresie 2 dla wszystkich danych.

Estymator kaplana - meiera w dla całych dany



Rysunek 2: Estymator Kaplana - Meiera funkcji przeżycia w całej bazie

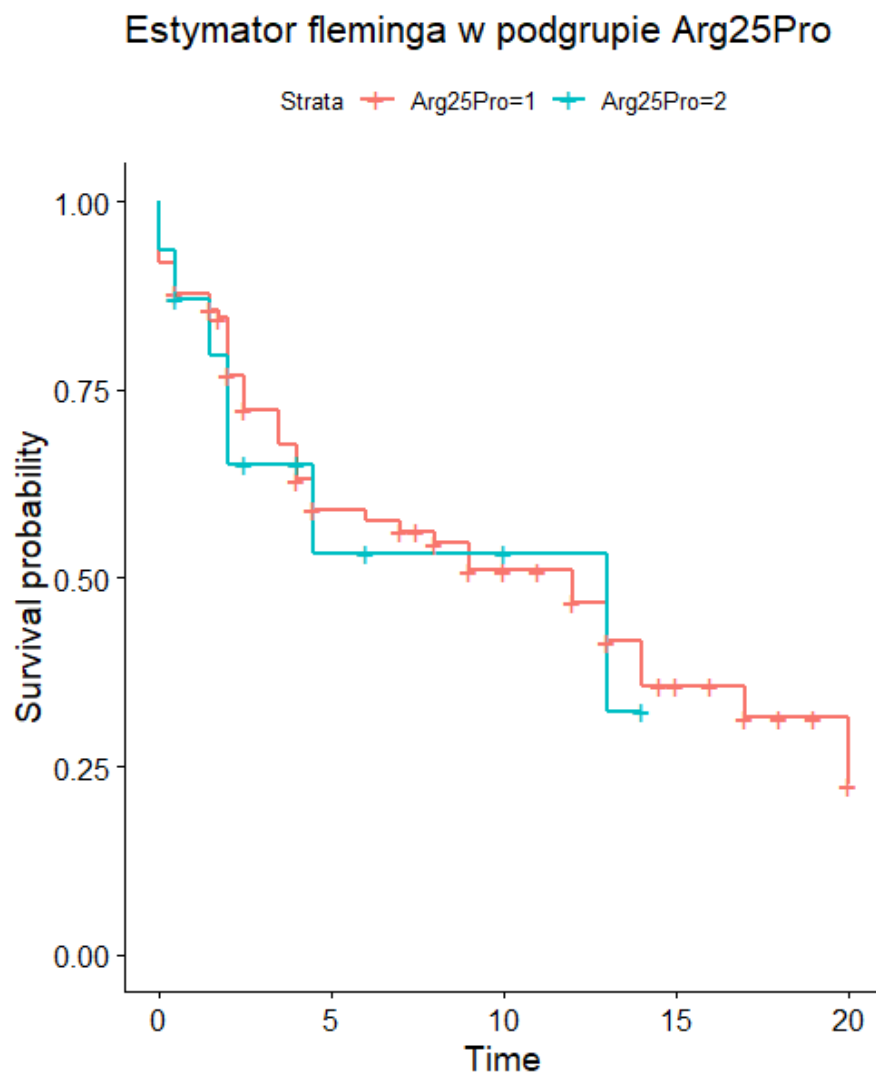
1.2 Zadanie 2

Estymator fleminga - harringtona w podgrupie Arg25Pro

W tym zadaniu zbudujemy estymator Fleminga - Harringtona i zaprezentujemy tą samą zależność.

```
> fit2 <- survfit(surv_object1 ~ dane3$Arg25Pro, data = dane3,  
+               type = "fleming-harrington")  
> g2 <- ggsurvplot(fit2, data = dane3, conf.int = FALSE) +  
+   ggtitle("Estymator fleminga w podgrupie Arg25Pro")
```

Na wykresie 3 prezentuje powstały estymator funkcji przeżycia w podgrupach Arg25Pro.



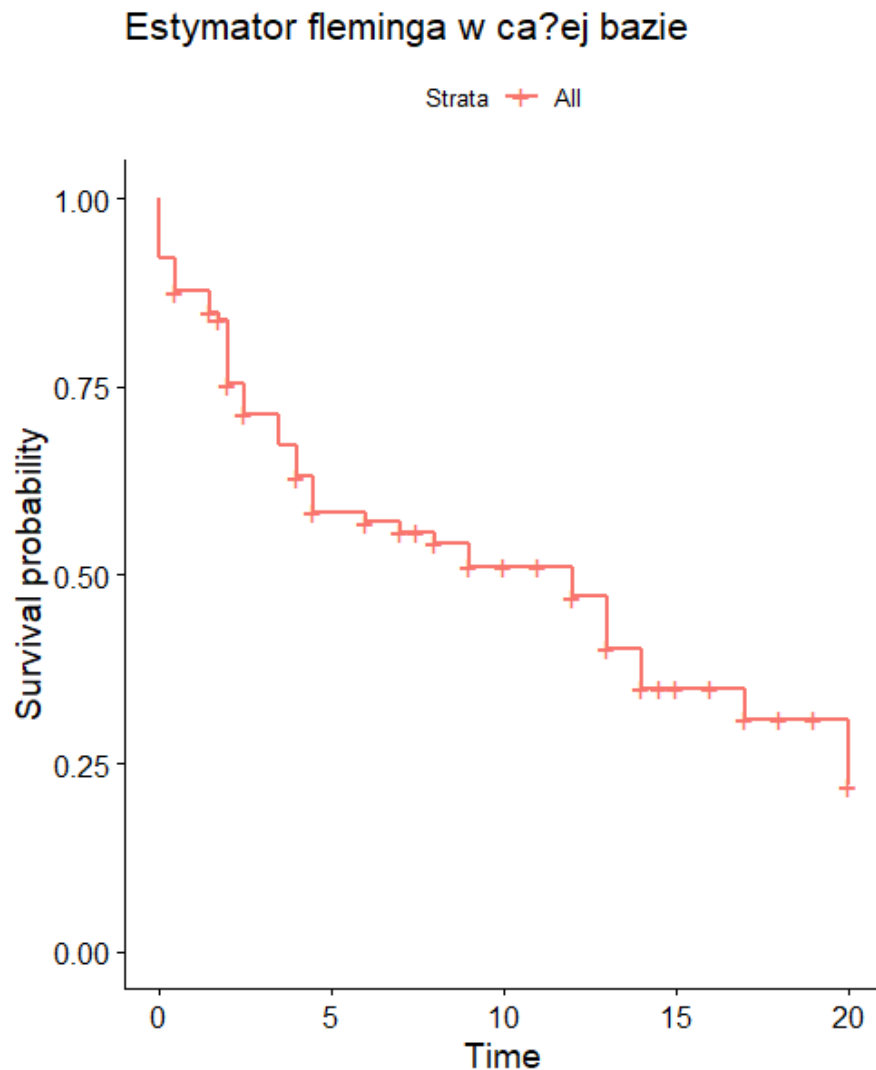
Rysunek 3: Estymator Fleminga - Harringtona funkcji przeżycia w podgrupie Arg25Pro

Estymator fleminga - harringtona w całej bazie danych

Stworzenie modelu :

```
> fit2.1 <- survfit(surv_object1 ~ 1, data = dane3,  
+                   type = "fleming-harrington")  
> g2 <- ggsurvplot(fit2.1, data = dane3, conf.int = FALSE) +  
+   ggtitle("Estymator fleminga w bazie")
```

Na wykresie 4 prezentuje powstały estymator funkcji przeżycia w całej bazie danych



Rysunek 4: Estymator Fleminga - Harringtona funkcji przeżycia w całej bazie

1.3 Zadanie 3

Warto zauważyć, że wykresy między sobą nie różnią się wiele, co oznacza że oba estymatory dobrze przybliżają funkcję przeżycia stworzoną na zmiennych o niewydolności nerek. Nawet podział na podgrupę Arg25Pro nadal zachowuje kształt wykresu funkcji przeżycia, również dla odpowiednich estymatorów.

2 Zadanie do sprawozdania - Część 2

2.1 Zadanie 1

Kolejnym zadaniem dotyczącym tego samego pliku jest narysowanie przedziałów ufności w przypadku całego pliku oraz z podziałem w podgrupie Arg25Pro, rysując przedziały mamy uwzględnić 3 typy, wszystkie z podanych typów korzystają z metody delty poniżej same typy i odpowiadają im podane przekształcenia, z których korzystamy w metodzie delty:

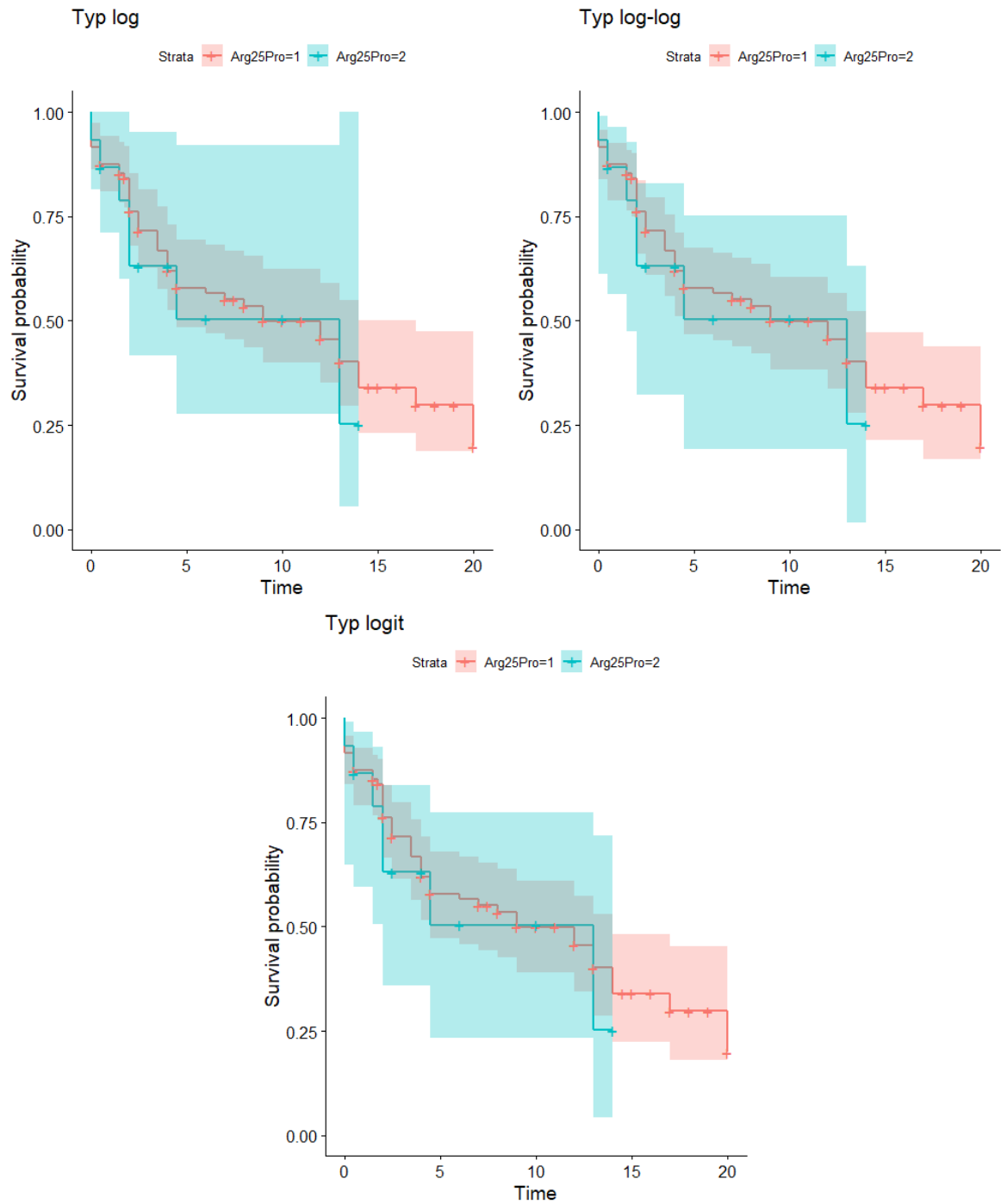
- \log ————— $g(x) = \log(x)$,
- $\log - \log$ — $g(x) = \log(-\log(x))$,
- logit ————— $g(x) = \log\left(\frac{x}{1-x}\right)$

Przedziały ufności w podgrupie Arg25Pro

Postępujemy podobnie jak w części pierwszej, jednak teraz do naszych modeli dodajemy warunek *conf.type*, który przyjmuje odpowiednie typy przedziałów.

```
> surv_object1<-Surv(dane3$Czas, dane3$Cenzura)
> model1 <- survfit(surv_object1 ~ dane3$Arg25Pro, data = dane3,
+                  type = c("kaplan-meier"), conf.type = c("log"))
> model2 <- survfit(surv_object1 ~ dane3$Arg25Pro, data = dane3,
+                  type = c("kaplan-meier"), conf.type = c("log-log"))
> model3 <- survfit(surv_object1 ~ dane3$Arg25Pro, data = dane3,
+                  type = c("kaplan-meier"), conf.type = c("logit"))
> wykres1 <- ggsurvplot(model1, data = dane3, conf.int = TRUE) + ggtitle("Typ log")
> wykres2 <- ggsurvplot(model2, data = dane3, conf.int = TRUE) + ggtitle("Typ log-log")
> wykres3 <- ggsurvplot(model3, data = dane3, conf.int = TRUE) + ggtitle("Typ logit")
```

Poniżej wykresy 5 funkcji przeżycia wraz z przedziałami ufności dla różnych przekształceń w metodzie delty:



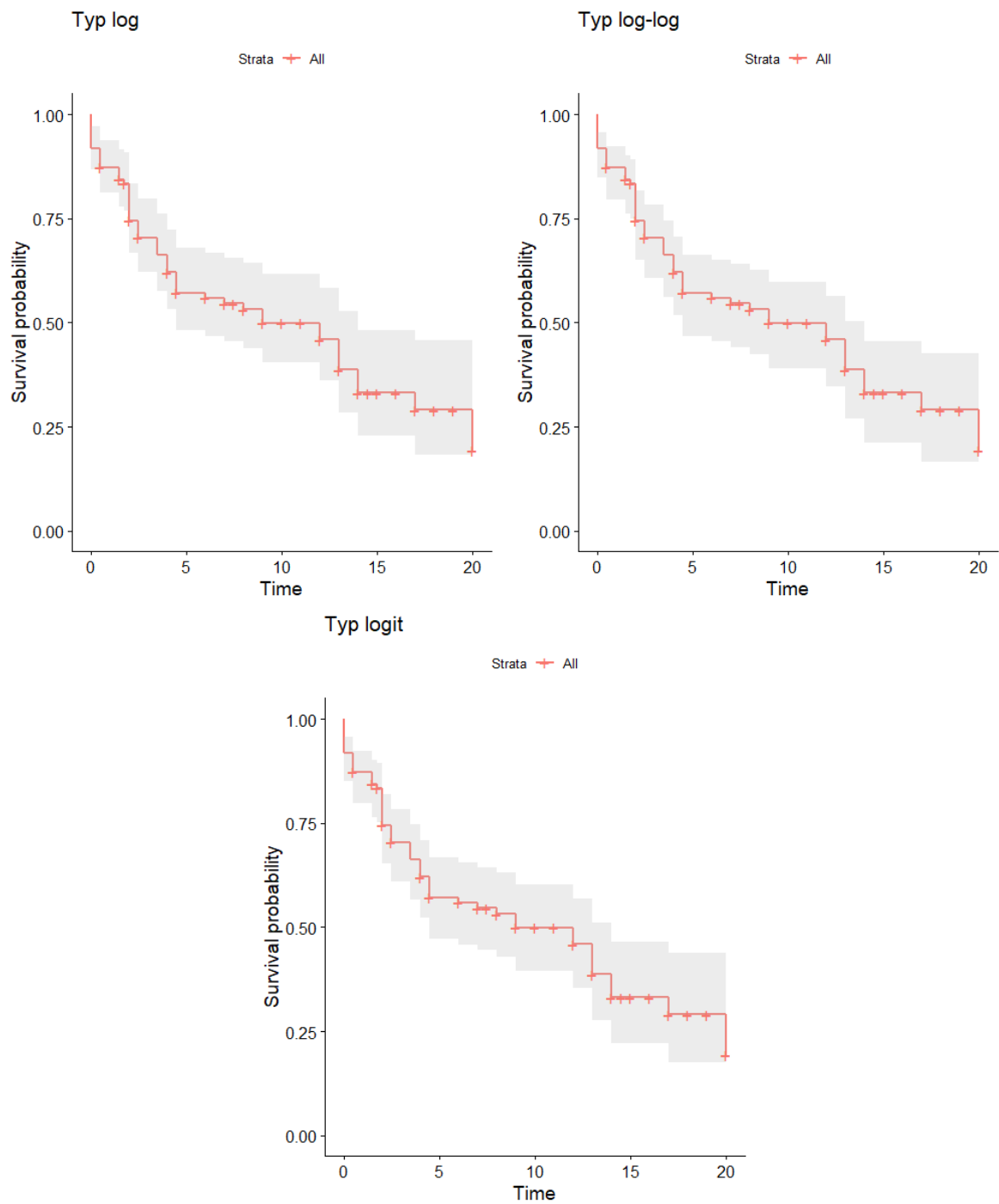
Rysunek 5: (a) log (b) log - log (c) logit

Przedziały ufności dla całej bazy

Analogiczne postępowanie:

```
> surv_object1<-Surv(dane3$Czas, dane3$Cenzura)
> model1 <- survfit(surv_object1 ~ 1, data = dane3,
+                   type = c("kaplan-meier"), conf.type = c("log"))
> model2 <- survfit(surv_object1 ~ 1, data = dane3,
+                   type = c("kaplan-meier"), conf.type = c("log-log"))
> model3 <- survfit(surv_object1 ~ 1, data = dane3,
+                   type = c("kaplan-meier"), conf.type = c("logit"))
> wykres1 <- ggsurvplot(model1, data = dane3, conf.int = TRUE) + ggtitle("Typ log")
> wykres2 <- ggsurvplot(model2, data = dane3, conf.int = TRUE) + ggtitle("Typ log-log")
> wykres3 <- ggsurvplot(model3, data = dane3, conf.int = TRUE) + ggtitle("Typ logit")
```

Wykres 6 estymatora Kaplana-Meiera funkcji przeżycia w całej badanej grupie wraz z realizacjami przedziałów ufności:



Rysunek 6: (a) log (b) log - log (c) logit

2.2 Zadanie 2

Jeśli chodzi o wybór, który typ przedziału wybrać to po analizie samych wykresów dla całej bazy nie możemy jednoznacznie zdecydować, który przedział jest tym najlepszym, a same zmiany szerokości przedziałów są niewielkie w poszczególnych punktach czasu.

Tabela 1: Przedziały ufności

	log_lower	log_upper	log.log_lower	log.log_upper	logit_lower	logit_upper
1	0.87	0.97	0.85	0.96	0.85	0.96
2	0.81	0.94	0.79	0.92	0.80	0.92
3	0.78	0.92	0.76	0.90	0.76	0.90
4	0.77	0.91	0.75	0.89	0.75	0.89
5	0.67	0.83	0.65	0.82	0.65	0.82
6	0.62	0.80	0.61	0.78	0.61	0.78
7	0.58	0.76	0.56	0.75	0.57	0.75
8	0.53	0.72	0.52	0.71	0.52	0.71
9	0.48	0.68	0.47	0.66	0.47	0.67
10	0.47	0.67	0.45	0.65	0.46	0.65
11	0.45	0.66	0.44	0.64	0.45	0.64
12	0.45	0.66	0.44	0.64	0.45	0.64
13	0.44	0.64	0.42	0.63	0.43	0.63
14	0.40	0.62	0.39	0.60	0.40	0.60
15	0.40	0.62	0.39	0.60	0.40	0.60
16	0.40	0.62	0.39	0.60	0.40	0.60
17	0.36	0.58	0.35	0.56	0.35	0.57
18	0.28	0.53	0.27	0.50	0.28	0.51
19	0.23	0.48	0.21	0.46	0.22	0.47
20	0.23	0.48	0.21	0.46	0.22	0.47
21	0.23	0.48	0.21	0.46	0.22	0.47
22	0.23	0.48	0.21	0.46	0.22	0.47
23	0.18	0.46	0.17	0.43	0.18	0.44
24	0.18	0.46	0.17	0.43	0.18	0.44
25	0.18	0.46	0.17	0.43	0.18	0.44
26	0.08	0.49	0.06	0.39	0.07	0.43

2.3 Zadanie 3

Bardzo duże przedziały są w przypadku zmiennej $Arg25Pro = 2$ i to tutaj widać największe różnice w wyliczaniu przedziałów, w przypadku typu log znacznie większe przedziały są podane dla górnej granicy, odwrotny przypadek dzieje się dla $log - log$. Najbardziej wypośrodkowany wydaje się typ przedziału ufności dla $logit$. W przypadku wartości 1 to badając jej przedziały dla różnych typów nie widzimy znaczących różnic w wykresach. Warto dodać, że zmiennych równych 2 jest sporo mniej niż 1 temu też te przedziały dają takie, a nie inne wartości. Warto dodać, że zmiennych $Arg25Pro = 2$ jest znacznie mniej niż $Arg25Pro = 1$, co negatywnie wpływa na tworzone przedziały, dlatego są takie szerokie i uzależnione od typu wyliczenia przedziałów.

3 Zadanie do sprawozdania - Część 3

3.1 Zadanie 1,2

W ostatniej części sprawozdania musimy oszacować punktowo i przedziałowo wartość oczekiwaną i medianę rozkładu czasu od momentu rozpoznania choroby do konieczności przeprowadzenia dializy w całej badanej grupie i w podgrupie *Arg25Pro*.

Estymacja mediany i średniej w podgrupie Arg25Pro

Tak jak w poprzednich zadaniach tworzymy model estymatora funkcji przeżycia. Z modelu odczytamy wartość średniej, a z danych z modelu jesteśmy w stanie podać ich przedziały ufności. Oczywiście korzystając z obliczonych na wykładzie przedziałów ufności tzn $[T_l, T_u]$:

```
> surv_object1<-Surv(dane3$Czas, dane3$Cenzura)
> fit1 <- survfit(surv_object1 ~ dane3$Arg25Pro, type = c("kaplan-meier"), data = dane3)
> print(fit1, print.rmean = TRUE)
```

```
Call: survfit(formula = surv_object1 ~ dane3$Arg25Pro, data = dane3,
               stype = 1, ctype = 1, type = c("kaplan-meier"))
```

	n	events	*rmean	*se(rmean)	median	0.95LCL	0.95UCL
dane3\$Arg25Pro=1	95	50	10.28	0.90	9	4.5	17
dane3\$Arg25Pro=2	15	7	9.35	2.51	13	2.0	NA

* restricted mean with upper limit = 20

```
> x <- qnorm(1-0.05/2)
> srednia.1 <- 10.28
> srednia.1.low <- srednia.1 - x * 0.92
> srednia.1.upper <- srednia.1 + x * 0.92
> srednia.2 <- 9.35
> srednia.2.low <- srednia.2 - x * 2.51
> srednia.2.upper <- srednia.2 + x * 2.51
>
```

To samo zdefiniujemy sobie dla mediany, jednak tutaj model sam podaje nam przedziały ufności na poziomie ufności 0.95:

```
> mediana.1 <- 9
> mediana.1.low <- 4.5
> mediana.1.upper <- 17
> mediana.2 <- 13
> mediana.2.low <- 2
> mediana.2.upper <- "NA"
```

Estymacja mediany i średniej dla całej bazy

Postępujemy analogicznie w wyliczeniach, jedyną różnicą jest konstrukcja modelu, ale to już zostało omówione parę zadań wyżej. Zaczynamy od wyliczenia odpowiednich wartości dla estymatora średniej.

```
> surv_object1<-Surv(dane3$Czas, dane3$Cenzura)
> fit1 <- survfit(surv_object1 ~ 1, type = c("kaplan-meier"), data = dane3, stype = 1,
> print(fit1, print.rmean = TRUE)
```

```
Call: survfit(formula = surv_object1 ~ 1, data = dane3, stype = 1,
             ctype = 1, type = c("kaplan-meier"))
```

n	events	*rmean	*se(rmean)	median	0.95LCL	0.95UCL
110.000	57.000	10.171	0.847	9.000	4.500	14.000
* restricted mean with upper limit = 20						

```
> x <- qnorm(1-0.05/2)
> srednia.all <- 10.171
> srednia.all.low <- srednia.all - x * 0.847
> srednia.all.upper <- srednia.all + x * 0.847
```

Odczytujemy wartości estymatora mediany dla całej bazy:

```
> mediana.all <- 9
> mediana.all.low <- 4.5
> mediana.all.upper <- 14
```

Ostatecznie prezentuje wnioski w dwóch tabelach

- Estymacja Średniej (2)
- Estymacja Mediany (3) w której każda z nich ma 3 kolumny, odpowiadające kolejno:
 - wyliczenia dla wszystkich danych,
 - podgrupa $Arg25Pro = 1$,
 - podgrupa $Arg25Pro = 2$.

```
> All <- data.frame(c(srednia.all, srednia.all.low, srednia.all.upper))
> Arg.1 <- data.frame(c(srednia.1, srednia.1.low, srednia.1.upper))
> Arg.2 <- data.frame(c(srednia.2, srednia.2.low, srednia.2.upper))
> srednia <- data.frame(All, Arg.1, Arg.2)
> rownames(srednia) <- c("Estymacja średniej", " T_1", " T_u")
> colnames(srednia) <- c("All", "Arg = 1", "Arg = 2")

> All.med <- data.frame(c(mediana.all, mediana.all.low, mediana.all.upper))
> Arg.1.med <- data.frame(c(mediana.1, mediana.1.low, mediana.1.upper))
> Arg.2.med <- data.frame(c(mediana.2, mediana.2.low, mediana.2.upper))
> mediana <- data.frame(All.med, Arg.1.med, Arg.2.med)
> rownames(mediana) <- c("Estymacja mediany", " T_1", " T_u")
> colnames(mediana) <- c("All", "Arg = 1", "Arg = 2")
```

Tabela 2: Średnia

	All	Arg = 1	Arg = 2
Estymacja średniej	10.17	10.28	9.35
T_l	8.51	8.48	4.43
T_u	11.83	12.04	14.27

Tabela 3: Mediana

	All	Arg = 1	Arg = 2
Estymacja mediany	9.00	9.00	13
T_l	4.50	4.50	2
T_u	14.00	17.00	NA

3.2 Zadanie 3

Porównując wartość wyestymowanej średniej wraz z przedziałami dla danych podgrup $Arg25Pro$ widzimy, że nie są one równe na takim poziomie istotności, przypuszczam, że podgrupa $Arg25Pro = 2$ jest tą odstającą od właściwej wartości, wiemy że do niej wpadało $n = 15$ elementów ze 110. Myślę, że odrzucilibyśmy hipotezę o ich równości, ale oczywiście mogą się mylić. Dla $Arg25Pro = 1$ widzimy zbliżoność wartości do wartości średniej z estymatora ze wszystkich danych.