

Raport 2

Analiza przeżycia

Jan Solarz
Aleksander Kaczmarek

7 grudnia 2020

Spis treści

1	Lista 1	2
1.1	Zadanie 1	2
1.2	Zadanie 2	3
1.3	Zadanie 3	4
2	Lista 2	7
2.1	Zadanie 1	7
2.2	Zadanie 2	13
2.3	Zadanie 3.	15
3	Lista 3.	16
3.1	Zadanie 1a.	16
3.2	Zadanie 1b.	17
3.3	Zadanie 2.	17
3.4	Zadanie 3.	17

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
##
## Attaching package: 'tidyr'
## The following object is masked from 'package:reshape2':
##
##   smiths
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:readr':
##
##   col_factor
## Registered S3 methods overwritten by 'broom':
##   method          from
##   tidy.glht        jtools
##   tidy.summary.glht jtools
##
## Attaching package: 'data.table'
## The following objects are masked from 'package:reshape2':
##
##   dcast, melt
## The following objects are masked from 'package:dplyr':
##
##   between, first, last
```

1 Lista 1

1.1 Zadanie 1

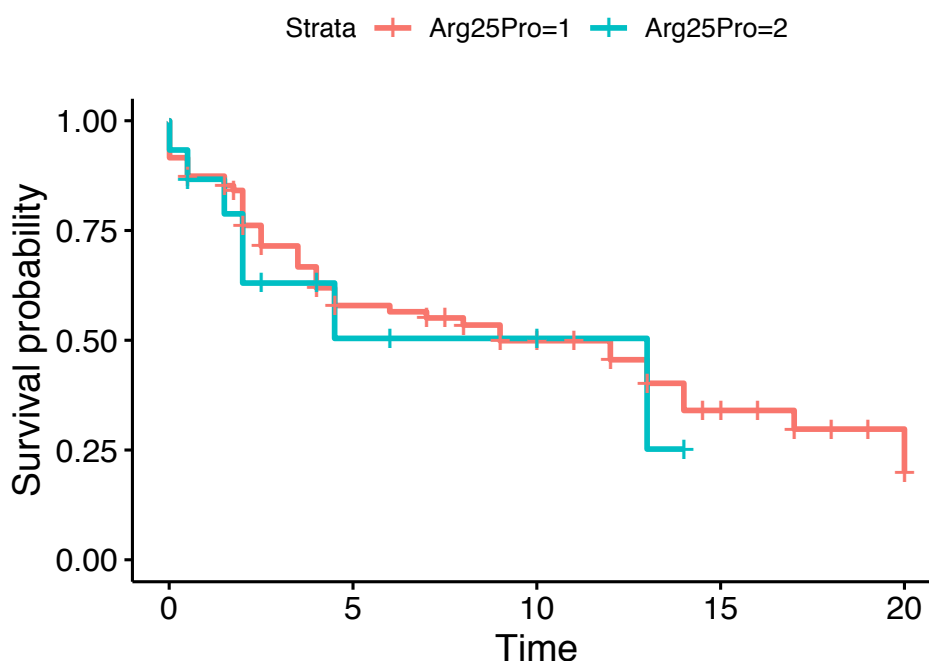
Szkicujemy wykres estymatora Kaplana-Meiera funkcji przeżycia w całej grupie pacjentów i w podgrupach ze względu na zmienną Arg25Pro bez zaznaczonych realizacji przedziałów ufności

```
attach(CzasDoDializy)
```

```
model11<-survfit(Surv(Czas,Cenzura)~Arg25Pro, data= CzasDoDializy)
```

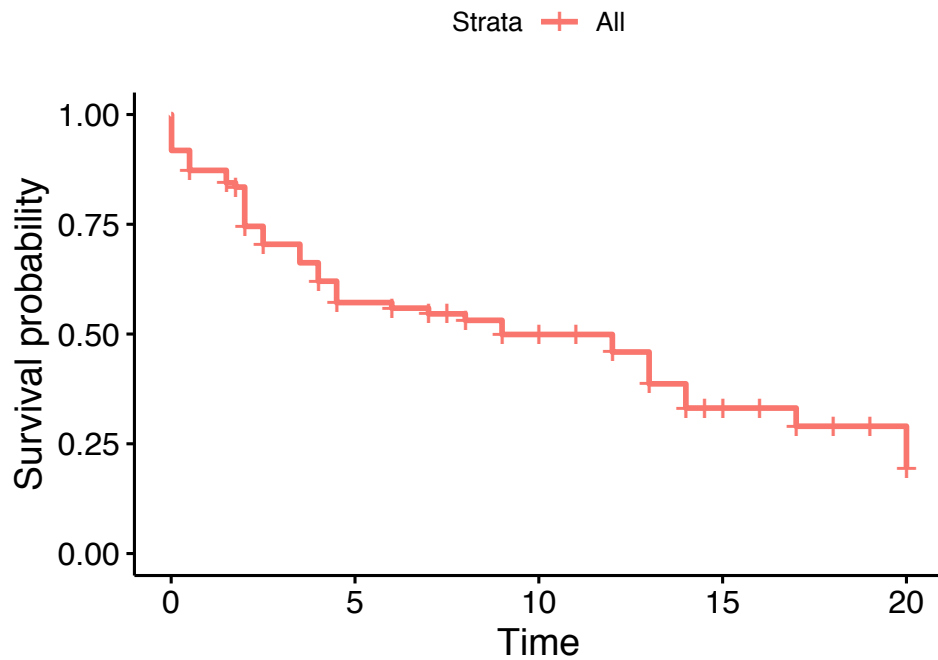
```
ggsurvplot(model11, data=CzasDoDializy,conf.int = FALSE,title="Wykres Kaplan-Meier dla p
```

Wykres Kaplan–Meier dla podgrup Arg25Pr



```
model12<-survfit(Surv(Czas,Cenzura)~1, data= CzasDoDializy)
ggsurvplot(model12, data=CzasDoDializy,conf.int = FALSE,title="Wykres Kaplan-Meier dla c
```

Wykres Kaplan–Meier dla całej grupy

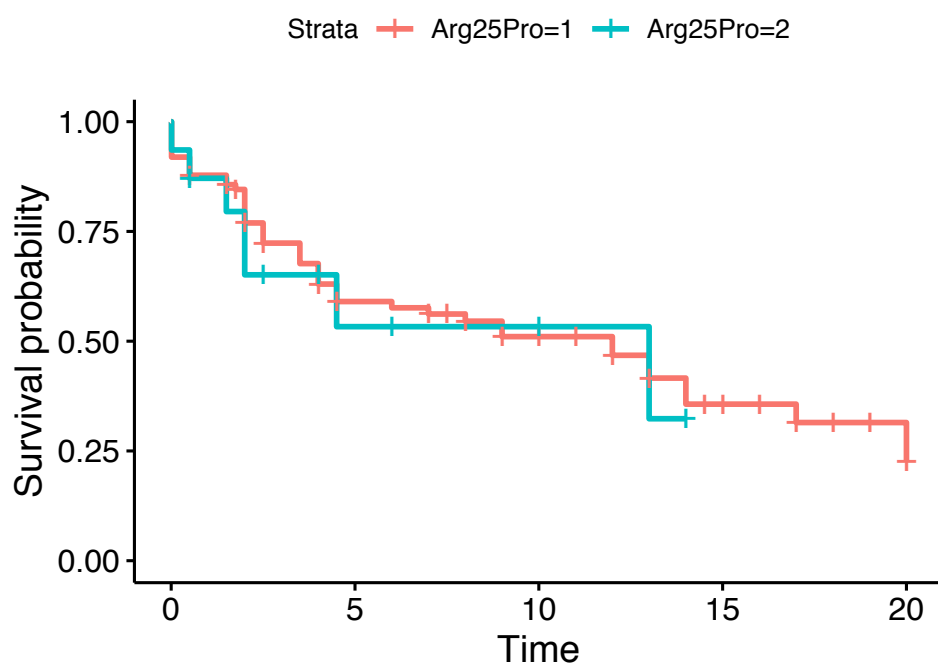


1.2 Zadanie 2

Szkicujemy wykres estymatora Kaplana-Meiera funkcji przeżycia w całej grupie pacjentów i w podgrupach ze względu na zmienną Arg25Pro bez zaznaczonych realizacji przedziałów ufności

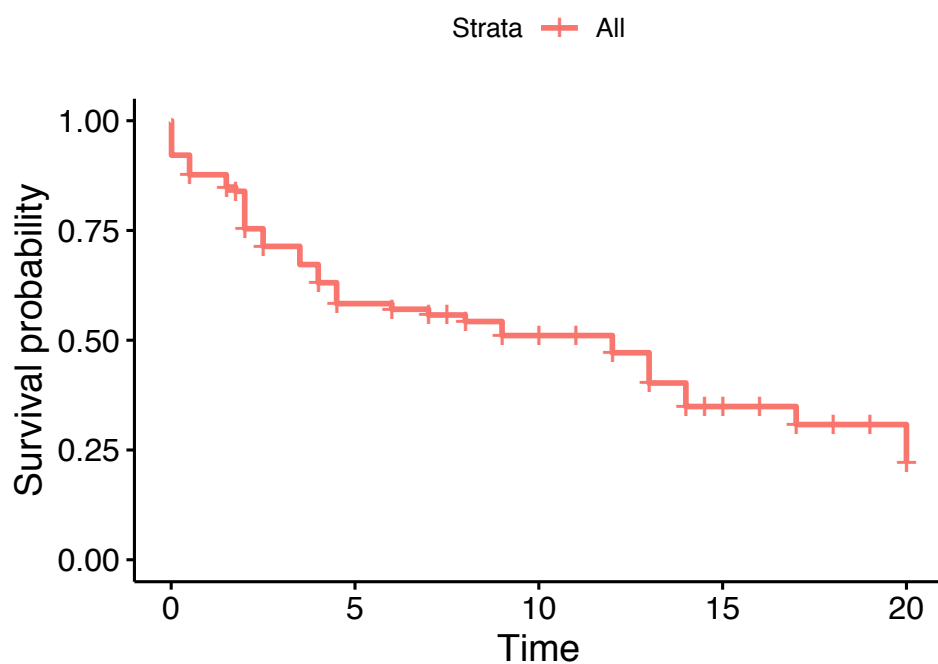
```
model21<-survfit(Surv(Czas,Cenzura)~Arg25Pro, data= CzasDoDializy,type="fleming-harrington")
ggsurvplot(model21, data=CzasDoDializy,conf.int = FALSE, title="Wykres Fleminga- Harringtona")
```

Wykres Fleminga– Harringtona dla podgrup



```
model22<-survfit(Surv(Czas,Cenzura)~1, data= CzasDoDializy,type="fleming-harrington")
ggsurvplot(model22, data=CzasDoDializy,conf.int = FALSE,title="Wykres Fleminga- Harringtona dla podgrup")
```

Wykres Fleminga– Harringtona dla całej gru



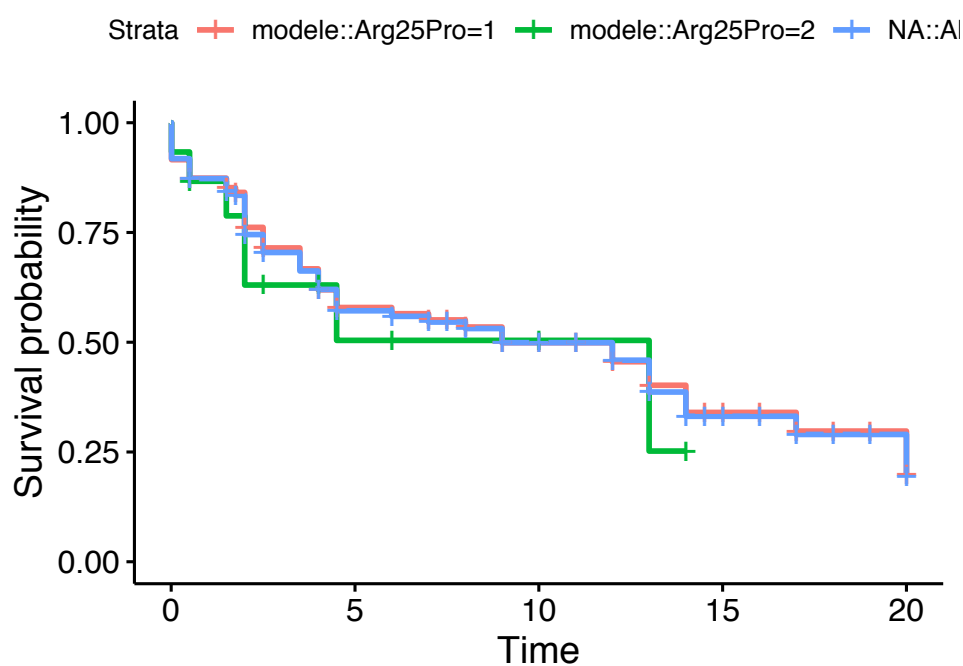
1.3 Zadanie 3

```
count(CzasDoDializy,Arg25Pro==1)

## # A tibble: 2 x 2
##   `Arg25Pro == 1`      n
##   <lgl>           <int>
## 1 FALSE             15
## 2 TRUE              95

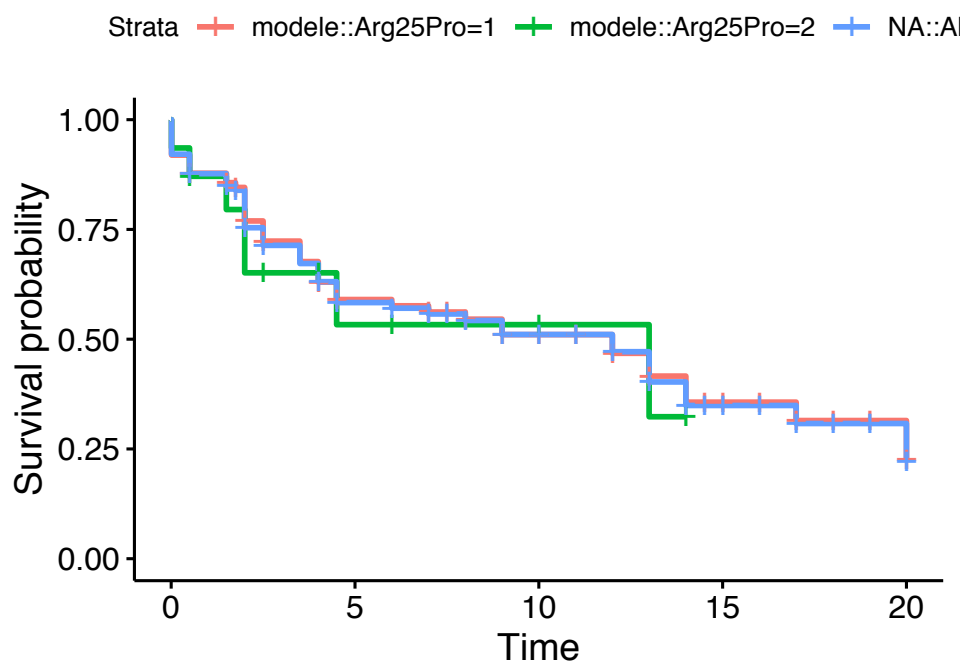
modele<-list(model11,model12)
ggsurvplot_combine(modele, data=CzasDoDializy, title="Wykres 1.")
```

Wykres 1.



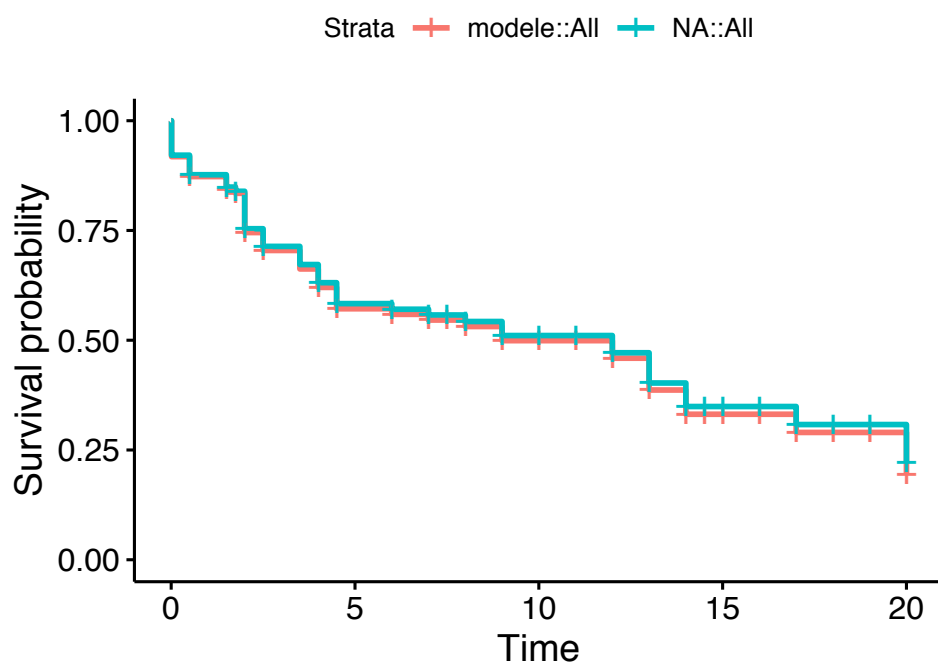
```
modele<-list(model21,model22)
ggsurvplot_combine(modele, data=CzasDoDializy, title="Wykres 2.")
```

Wykres 2.



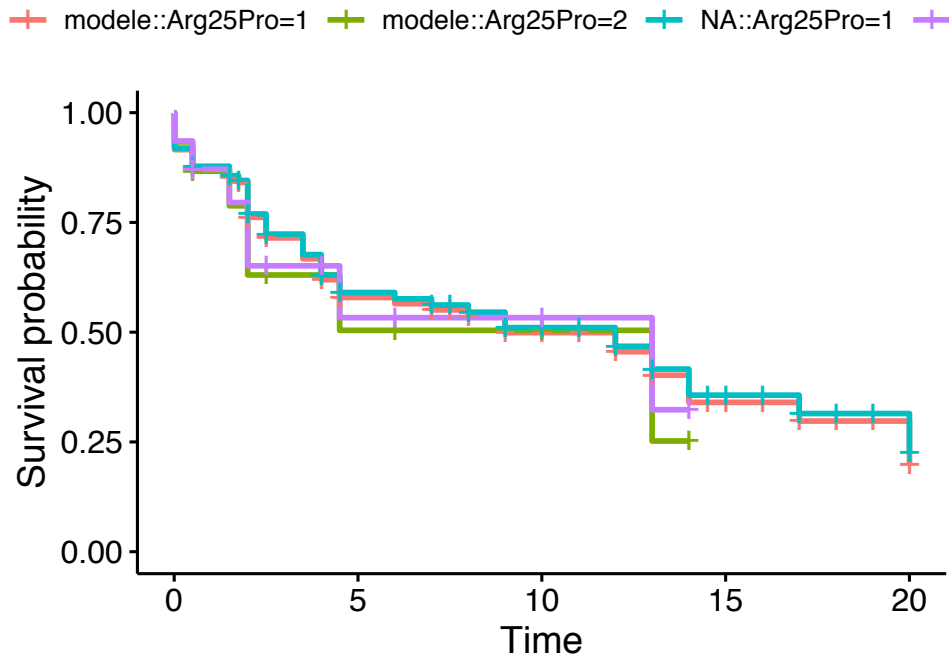
```
modele<-list(model12,model22)
ggsurvplot_combine(modele, data=CzasDoDializy, title="Wykres 3.")
```

Wykres 3.



```
modele<-list(model11,model21)
ggsurvplot_combine(modele, data=CzasDoDializy, title="Wykres 4.")
```

Wykres 4.



- Badane przeżycie występuje w okresie 20 lat, najniższe prawdopodobieństwo jest równe ok. 0.25
- Wśród 110 pacjentów, 95 ma genotyp Arg25Pro o typie=1, 15 o typie=2.
- Wykres funkcji dla Arg25Pro o typie 2 jest w większości czasu pod wykresem funkcji dla typu 1. Wniosek jest taki że "przeżycie" dla Arg25Pro=1 jest bardziej optymistyczne, wykres ten jest też bardziej rozciągnięty w czasie.
- W obu typach estymatorów(wykres 1. i 2.) widzimy pokrywanie się wykresu dla podgrupy Arg25Pro=1 i całej grupy (wykres dla Arg25Pro=2 odbiega i kończy się w okolicach czasu równego 15). Wynika to z faktu, iż znacząca ilość pacjentów ma genotyp Arg25Pro=1 (95 do 110)
- Oba wykresy różnych estymatorów dla całych grup (wykres 3.) niemal się pokrywają w całym okresie czasu- z czasem widzimy większe prawdopodobieństwo przeżycia dla estymatora Fleminga-Harringtona.

2 Lista 2

2.1 Zadanie 1

Szkicujemy wykres estymatora Kaplana-Meiera funkcji przeżycia w całej grupie pacjentów i w podgrupach ze względu na zmienną Arg25Pro z realizacją przedziałów ufności

```
model31<-survfit(Surv(Czas,Cenzura)~Arg25Pro, data= CzasDoDializy, type="kaplan-meier",
model32<-survfit(Surv(Czas,Cenzura)~Arg25Pro, data= CzasDoDializy, type="kaplan-meier",
model33<-survfit(Surv(Czas,Cenzura)~Arg25Pro, data= CzasDoDializy, type="kaplan-meier",
```

```

model34<-survfit(Surv(Czas,Cenzura)~Arg25Pro, data= CzasDoDializy, type="kaplan-meier",
model35<-survfit(Surv(Czas,Cenzura)~Arg25Pro, data= CzasDoDializy, type="kaplan-meier",

model312<-survfit(Surv(Czas,Cenzura)~1, data= CzasDoDializy, type="kaplan-meier", conf.int=TRUE,
model322<-survfit(Surv(Czas,Cenzura)~1, data= CzasDoDializy, type="kaplan-meier", conf.int=TRUE,
model332<-survfit(Surv(Czas,Cenzura)~1, data= CzasDoDializy, type="kaplan-meier", conf.int=TRUE,
model342<-survfit(Surv(Czas,Cenzura)~1, data= CzasDoDializy, type="kaplan-meier", conf.int=TRUE,
model352<-survfit(Surv(Czas,Cenzura)~1, data= CzasDoDializy, type="kaplan-meier", conf.int=TRUE,

```

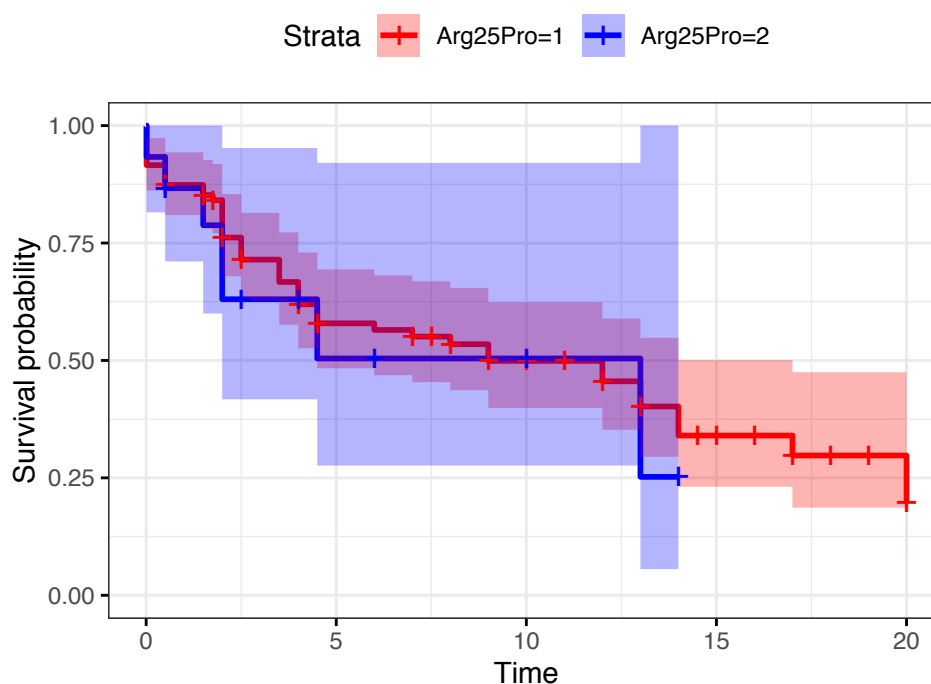
Wykresy estymatora Kaplana-Meiera dla różnych typów przedziałów ufności. Estymacja względem podgrup Arg25Pro

```

ggsurvplot(model31, data=CzasDoDializy,conf.int = TRUE, title="Typ log-log", palette = c("red", "blue"),

```

Typ log-log

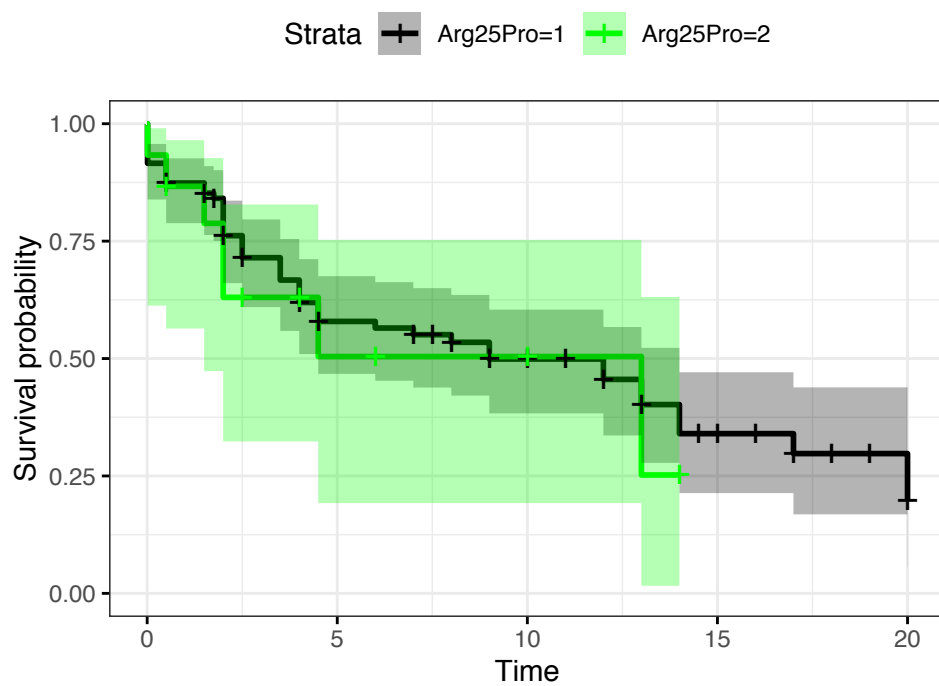


```

ggsurvplot(model32, data=CzasDoDializy,conf.int = TRUE, title="Typ log-log", palette = c("red", "blue"),

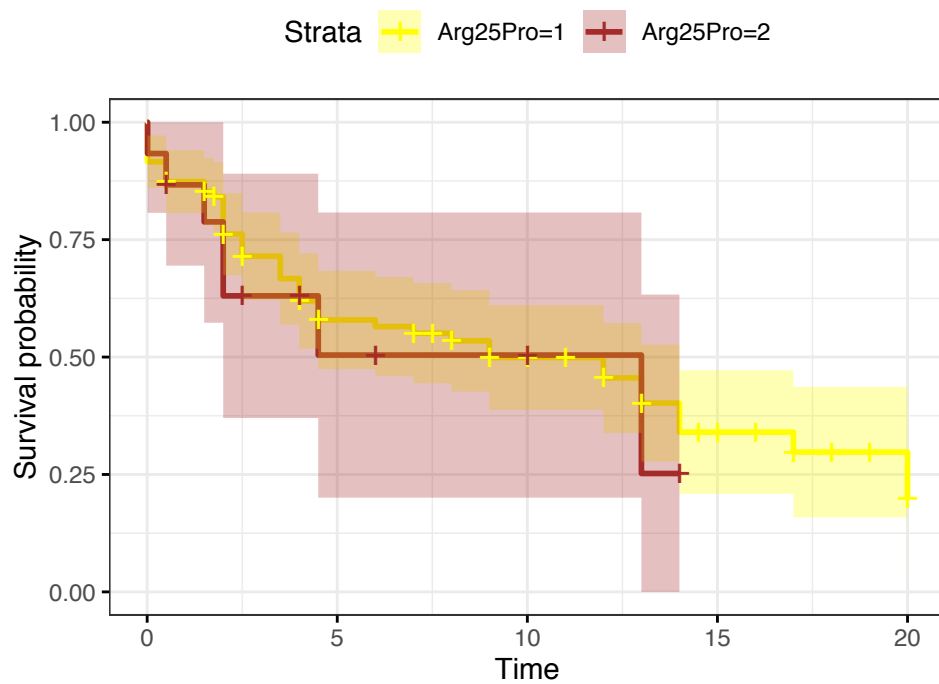
```


Typ log-log



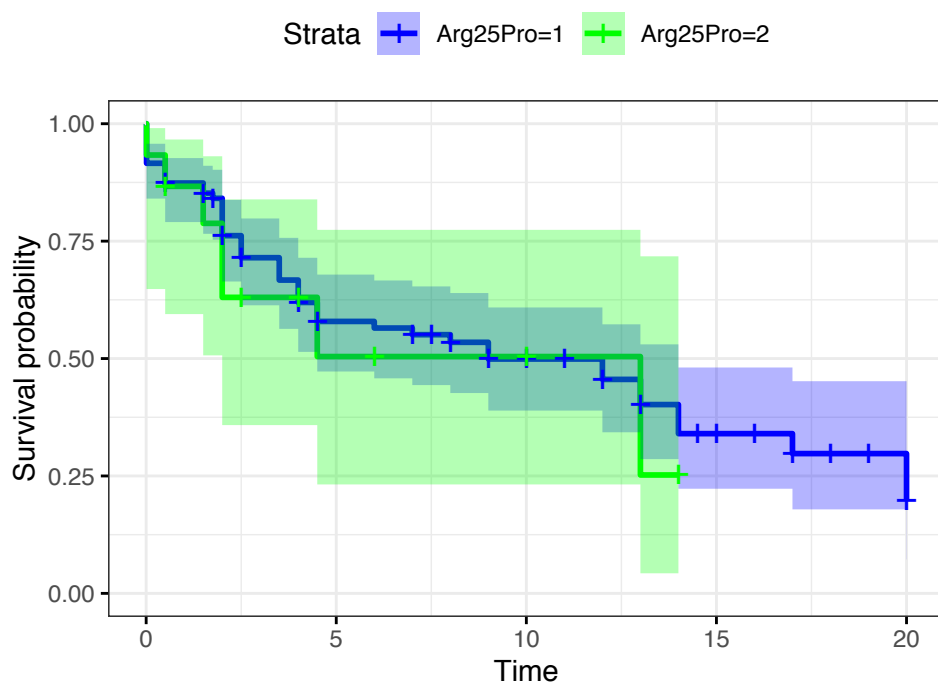
```
ggsurvplot(model133, data=CzasDoDializy, conf.int = TRUE, title="Typ plain", palette = c('
```

Typ plain



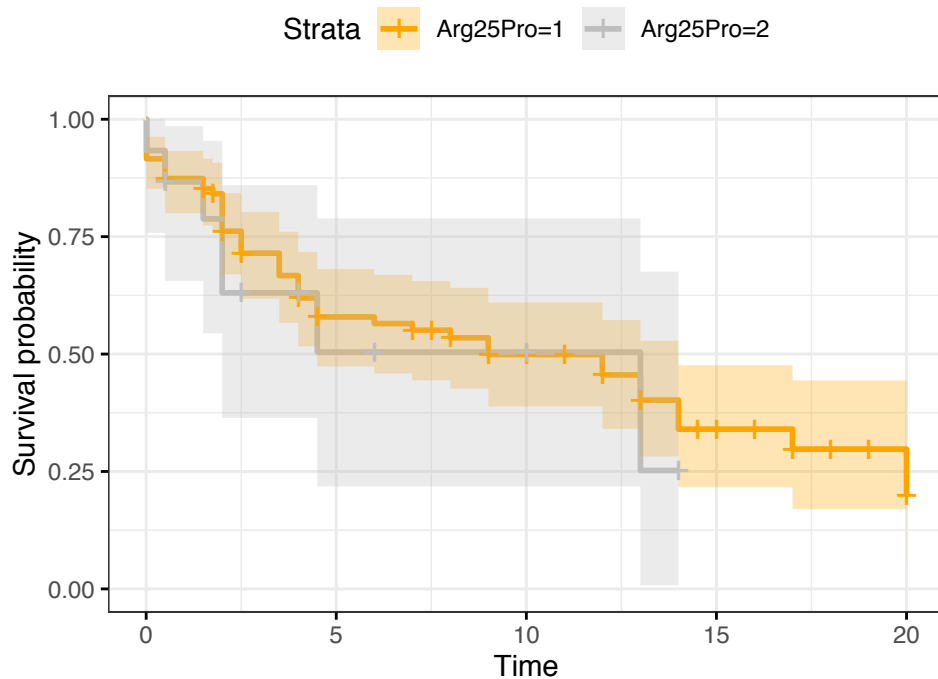
```
ggsurvplot(model134, data=CzasDoDializy, conf.int = TRUE, title="Typ logit", palette = c('
```

Typ logit



```
ggsurvplot(model35, data=CzasDoDializy, conf.int = TRUE, title="Typ arcsin", palette = c("blue", "green"))
```

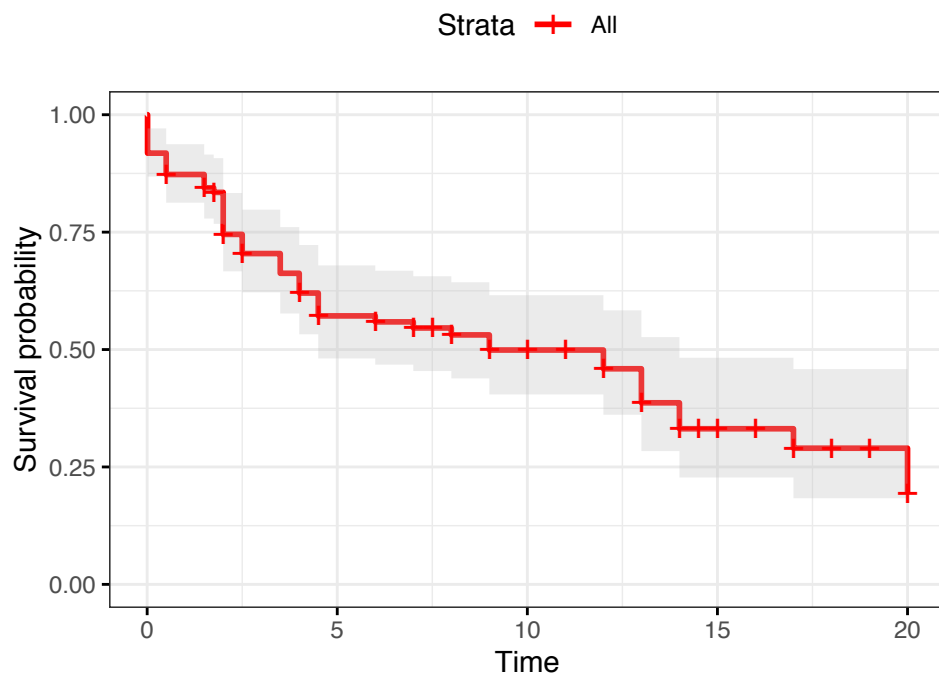
Typ arcsin



Wykresy estymatora Kaplana-Meiera dla dla różnych typów przedziałów ufności. Estymacja względem całej grupy

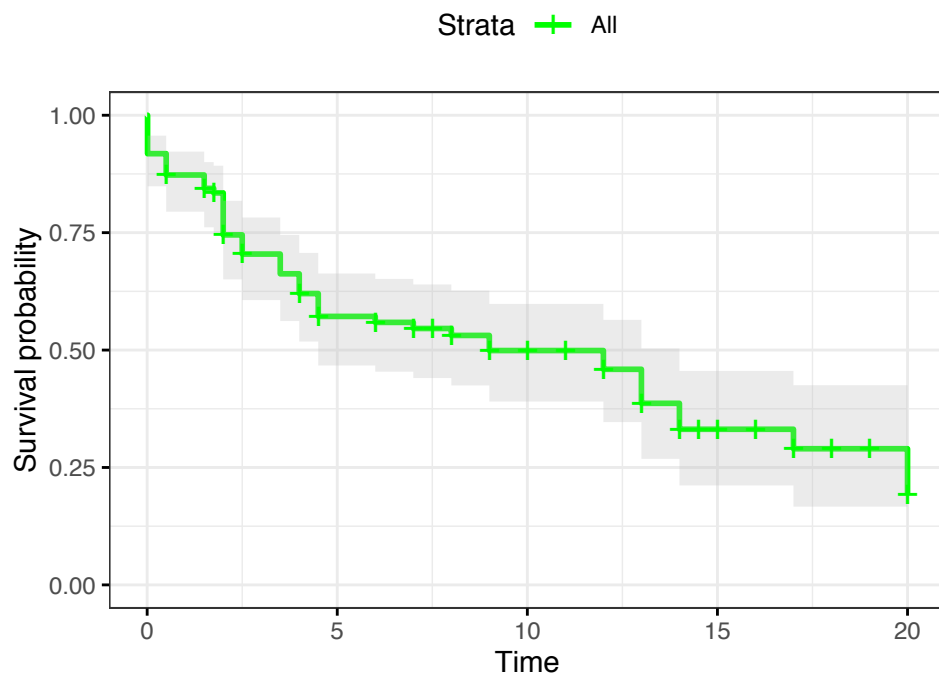
```
ggsurvplot(model312, data=CzasDoDializy, conf.int = TRUE, title="Typ log", palette = c("blue", "green"))
```

Typ log



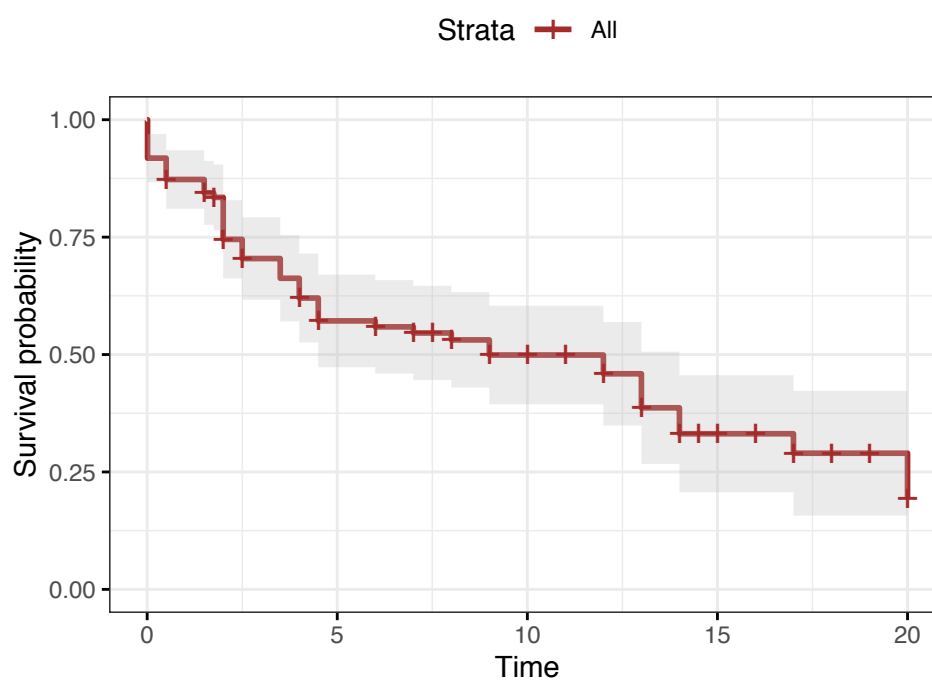
```
ggsurvplot(model322, data=CzasDoDializy, conf.int = TRUE, title="Typ log-log", palette =
```

Typ log-log



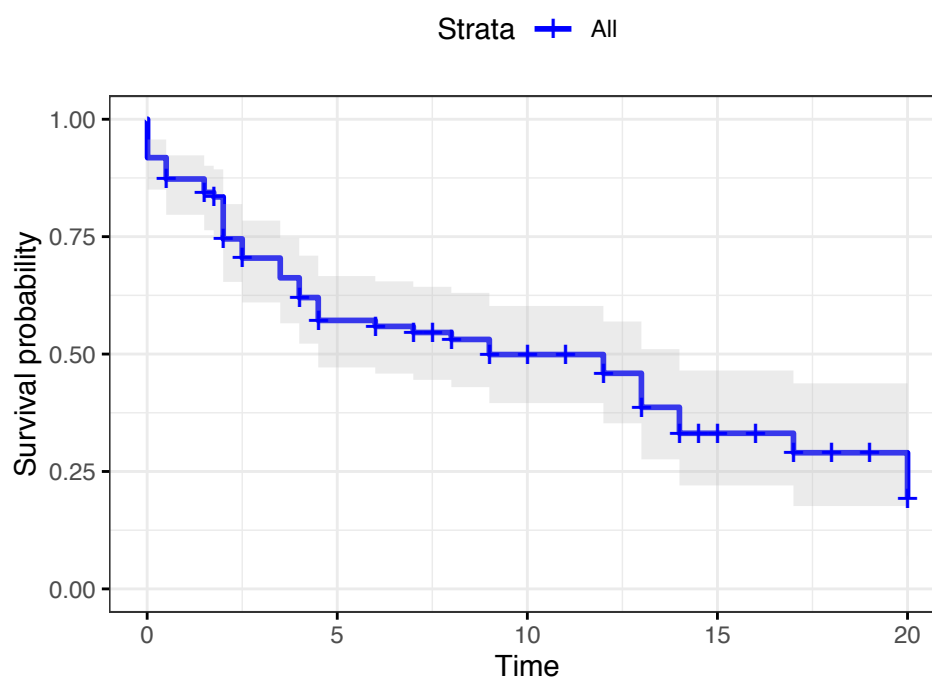
```
ggsurvplot(model332, data=CzasDoDializy, conf.int = TRUE, title="Typ plain", palette = c(
```

Typ plain



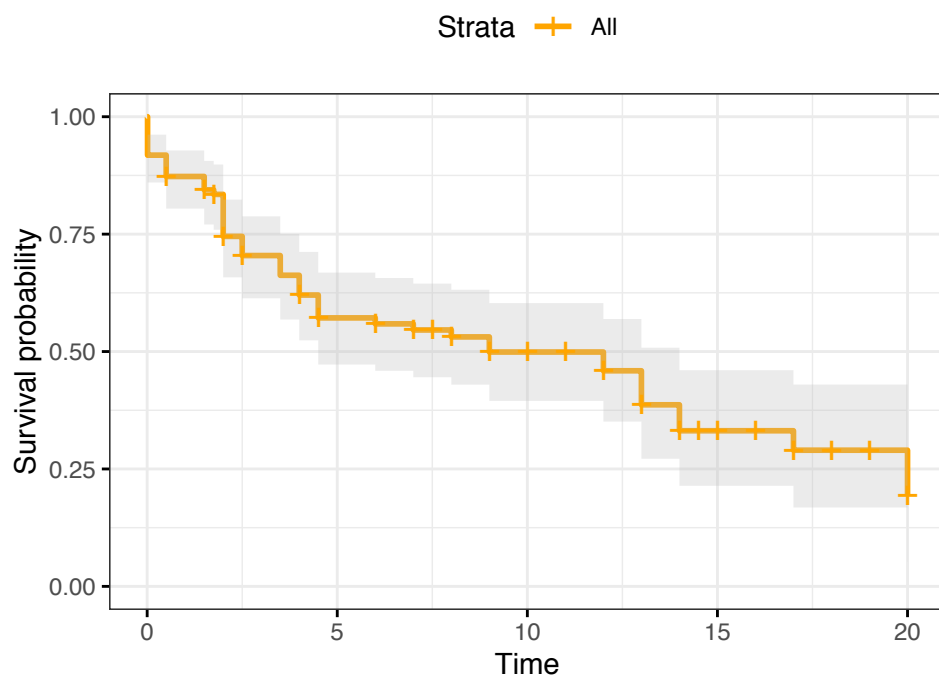
```
ggsurvplot(model342, data=CzasDoDializy, conf.int = TRUE, title="Typ logit", palette = c("red"))
```

Typ logit



```
ggsurvplot(model352, data=CzasDoDializy, conf.int = TRUE, title="Typ arcsin", palette = c("blue"))
```

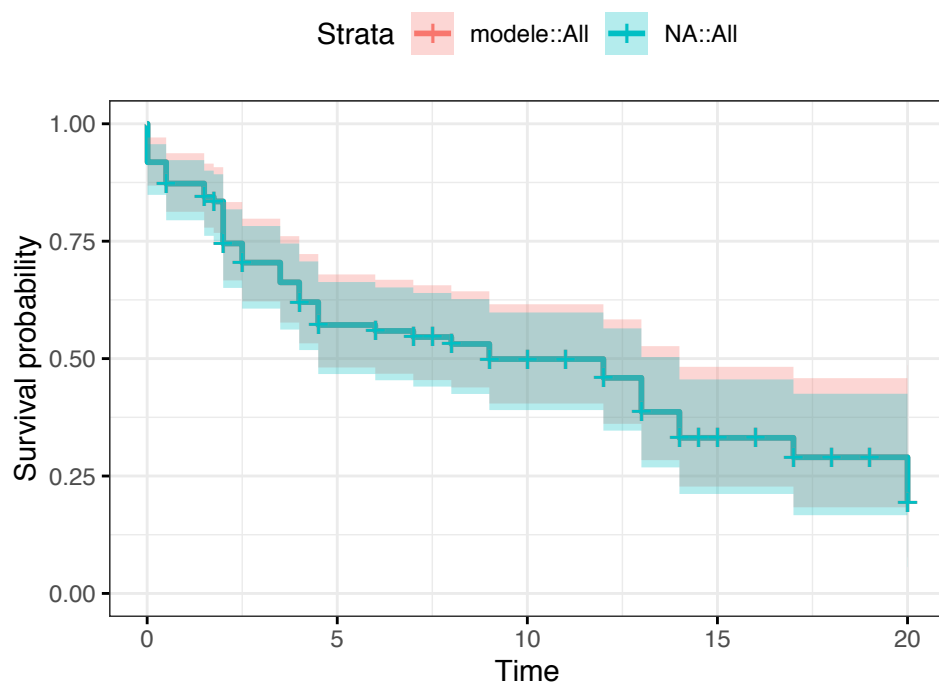
Typ arcsin



2.2 Zadanie 2

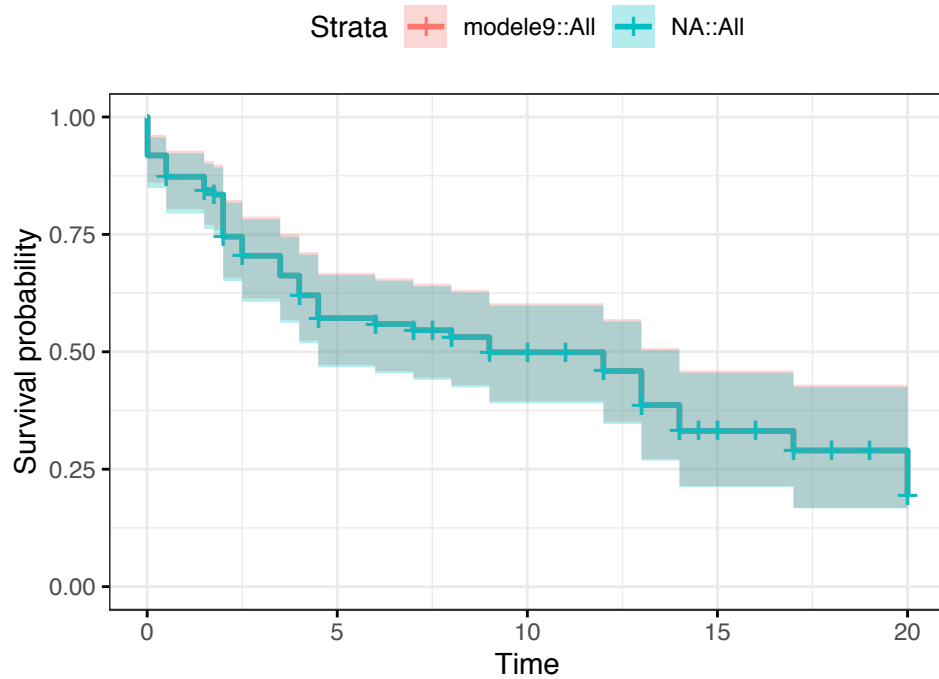
```
modele<-list(model312,model322)
ggsurvplot_combine(modele, data=CzasDoDializy,conf.int = TRUE, title="Typ log vs typ log-log")
```

Typ log vs typ log-log



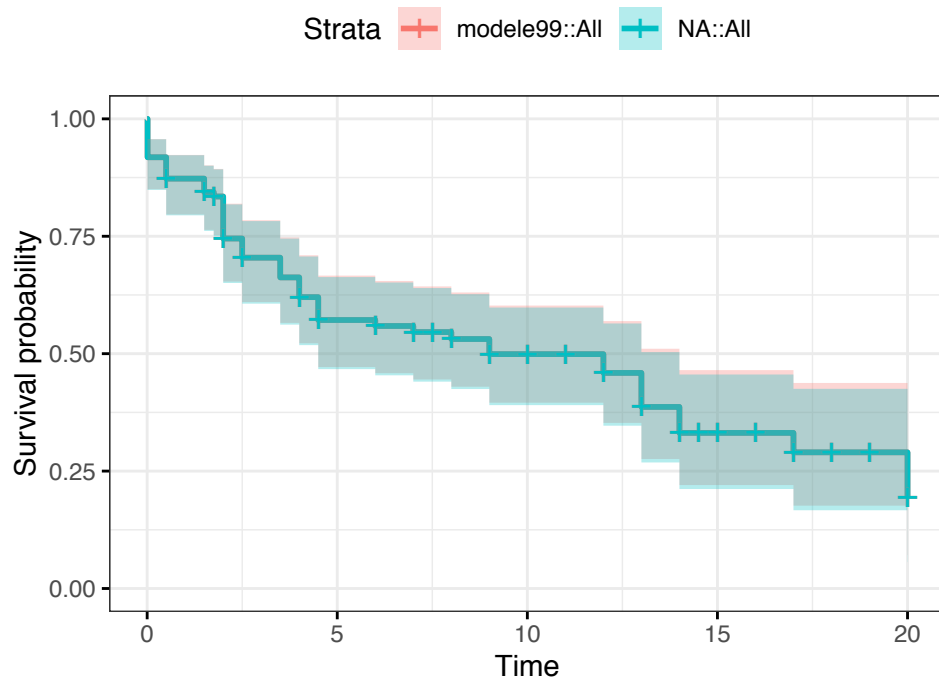
```
modele9<-list(model352,model322)
ggsurvplot_combine(modele9, data=CzasDoDializy,conf.int = TRUE, title="Typ arcsin vs log-log")
```

Typ arcsin vs log-log



```
modele99<-list(model342,model322)
ggsurvplot_combine(modele99, data=CzasDoDializy,conf.int = TRUE, title="Typ logit vs log-log")
```

Typ logit vs log-log



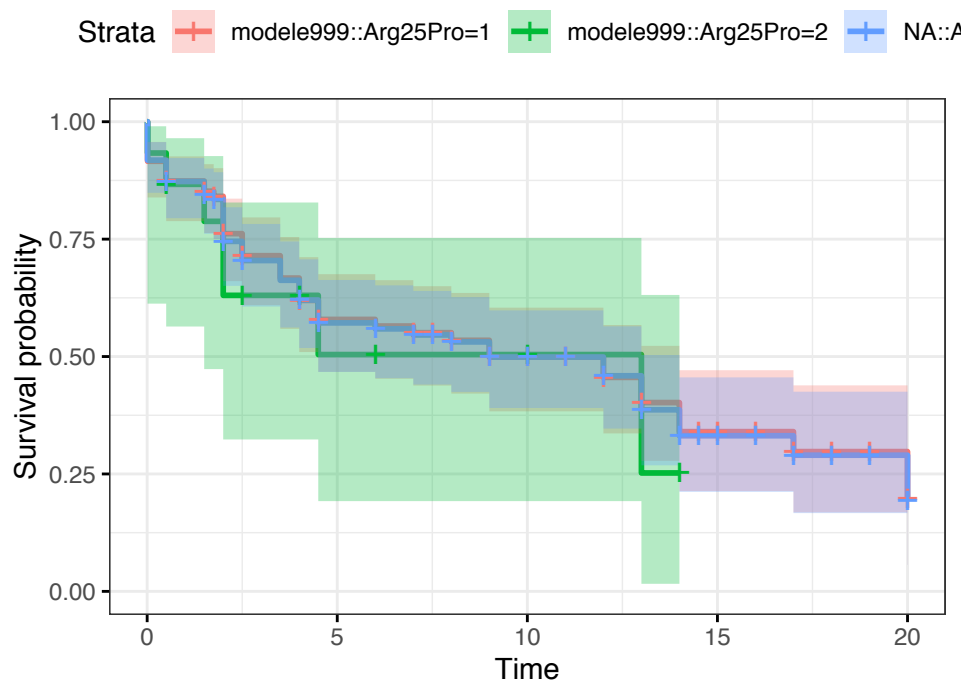
- Widzimy, że typ log-log posiada węższy przedział ufności niż typ log

- Typy log-log, plain, logit i arcsin wykazują podobne przedziały ufności, pokrywają się one. Można wskazać że typ arcsin i log-log generuje nieco węższy przedział niż pozostałe. Najszerszy (najgorszy) typ log.

2.3 Zadanie 3.

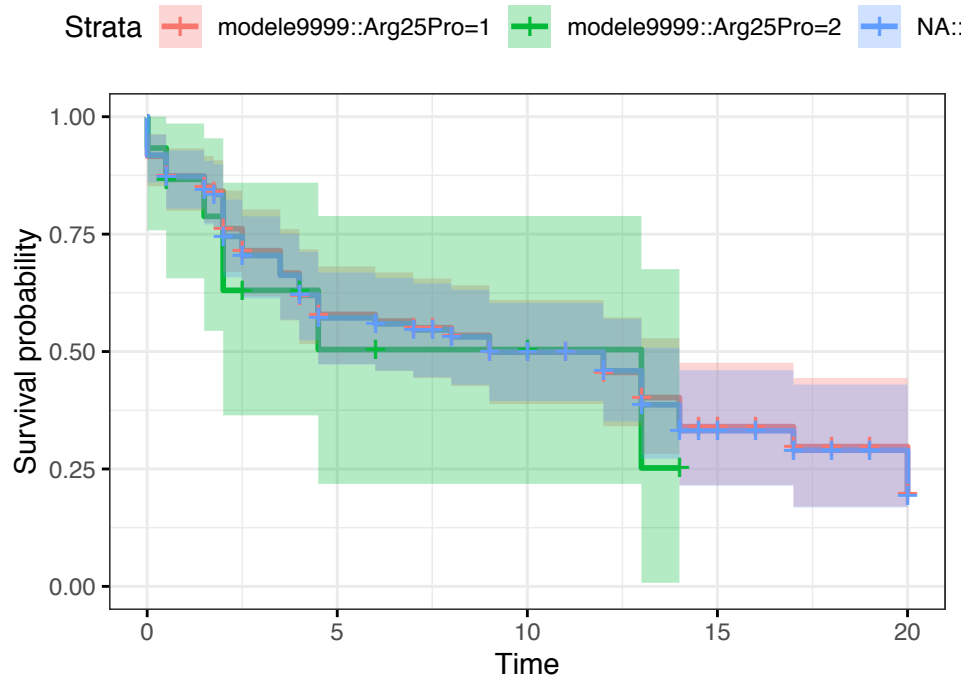
```
modele999<-list(model32,model322)
ggsurvplot_combine(model999, data=CzasDoDializy,conf.int = TRUE, title="Typ log-log dla
```

Typ log-log dla podgrup i całej grupy



```
modele9999<-list(model35,model352)
ggsurvplot_combine(model9999, data=CzasDoDializy,conf.int = TRUE, title="Typ arcsin dla
```

Typ arcsin dla podgrup i całej grupy



- Najlepszą skuteczność mają typy przedziału ufności "arcsin" i "log-log"- po porównaniu pokrywania się przedziałów, im węższy tym lepsza skuteczność
- Typy przedziałów ufności generalnie mają podobną skuteczność. Najgorszy jest typ log
- Tak jak w poprzednim zadaniu zauważamy prawie całkowite pokrywanie się wykresu dla całej grupy i Arg25Pro=1
- Zdecydowanie najszerze przedziały ufności generuje podgrupa o Arg25Pro typu 2, bez względu na typ przedziału. Trzeba pamiętać, że grupa ta jest mało liczna.

3 Lista 3.

3.1 Zadanie 1a.

```
CzasDoDializy <- as.data.table(CzasDoDializy)

#cala grupa

km <- survfit(Surv(CzasDoDializy$Czas, event=CzasDoDializy$Cenzura)~1,stype=1,ctype=1)
print(km, print.rmean = TRUE)

## Call: survfit(formula = Surv(CzasDoDializy$Czas, event = CzasDoDializy$Cenzura) ~
##      1, stype = 1, ctype = 1)
##
##              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
```


3.2 Zadanie 1b.

```
#podgrupy

km2 <- survfit(Surv(CzasDoDializy$Czas, event=CzasDoDializy$Cenzura) ~ CzasDoDializy$Arg
print(km2, print.rmean = TRUE)

## Call: survfit(formula = Surv(CzasDoDializy$Czas, event = CzasDoDializy$Cenzura) ~
##      CzasDoDializy$Arg25Pro, stype = 1, ctype = 1)
##
##              n events *rmean *se(rmean) median 0.95LCL 0.95UCL
## CzasDoDializy$Arg25Pro=1 95      50   9.39      0.752      9      4.5      17
## CzasDoDializy$Arg25Pro=2 15       7   8.60      2.070     13      2.0      NA
##      * restricted mean with upper limit = 17
```

3.3 Zadanie 2.

- Tabelka dla całej grupy oraz dla podgrup ze względu na zmienną *Arg25Pro* reprezentuje się następująco.

```
table(print(km, print.rmean=TRUE), print(km2, print.rmean=TRUE))

## Call: survfit(formula = Surv(CzasDoDializy$Czas, event = CzasDoDializy$Cenzura) ~
##      1, stype = 1, ctype = 1)
##
##           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
## Call: survfit(formula = Surv(CzasDoDializy$Czas, event = CzasDoDializy$Cenzura) ~
##      CzasDoDializy$Arg25Pro, stype = 1, ctype = 1)
##
##              n events *rmean *se(rmean) median 0.95LCL 0.95UCL
## CzasDoDializy$Arg25Pro=1 95      50   9.39      0.752      9      4.5      17
## CzasDoDializy$Arg25Pro=2 15       7   8.60      2.070     13      2.0      NA
##      * restricted mean with upper limit = 17

## Error in order(y): niezaimplementowany typ 'list' w 'orderVector1'
```

3.4 Zadanie 3.

```
mean_s <- 10.28
sd_s <- 0.9
alpha <- 0.05
lb_i <- mean_s - qnorm(1 - alpha/2)*sd_s
```

```

ub_i <- mean_s + qnorm(1 - alpha/2)*sd_s

mean_s <- 9.35
sd_s <- 2.51
alpha <- 0.05
lb_ii <- mean_s - qnorm(1 - alpha/2)*sd_s
ub_ii <- mean_s + qnorm(1 - alpha/2)*sd_s

c(lb_i, ub_i)

## [1] 8.516032 12.043968

c(lb_ii, ub_ii)

## [1] 4.43049 14.26951

```

- Widzimy, że jeden przedział ufności całkowicie zawiera się w drugim, dlatego nie mamy podstaw do odrzucenia hipotezy H_0 .