

**ANALIZA PRZEŻYCIA**

# **SPRAWOZDANIE**

**OPRACOWAŁA:**  
**ALEKSANDRA GRZESZCZUK**  
**NUMER ALBUMU: 255707**

# SPIS TREŚCI

|   |            |    |
|---|------------|----|
| 1 | WSTĘP      | 2  |
| 2 | ZADANIE 1A | 4  |
| 3 | ZADANIE 1B | 9  |
| 4 | ZADANIE 2  | 11 |
| 5 | ZADANIE 3  | 14 |
| 6 | ZADANIE 4  | 15 |

# 1 WSTEP

```
surv_diff <- survdiff(Surv(time, status) ~ sex, data = lung)
surv_diff

## Call:
## survdiff(formula = Surv(time, status) ~ sex, data = lung)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 138      112      91.6      4.55      10.3
## sex=2  90       53      73.4      5.68      10.3
##
##  Chisq= 10.3  on 1 degrees of freedom, p= 0.001
```

```
x <- rnorm(50)
y <- runif(30)
ks.test(x, y)

##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  x and y
## D = 0.54, p-value = 1.598e-05
## alternative hypothesis: two-sided
```

Wszystkie zadania do sprawozdania oparte są na danych `lung` z biblioteki `survival`. Dane te dotyczą pacjentów z zaawansowanym rakiem płuc. Nasz zbiór zawiera:

```
nrow(dane)

## [1] 228
```

informacji o pacjentach, których zbiór charakterystyk obejmuje:

```
ncol(dane)

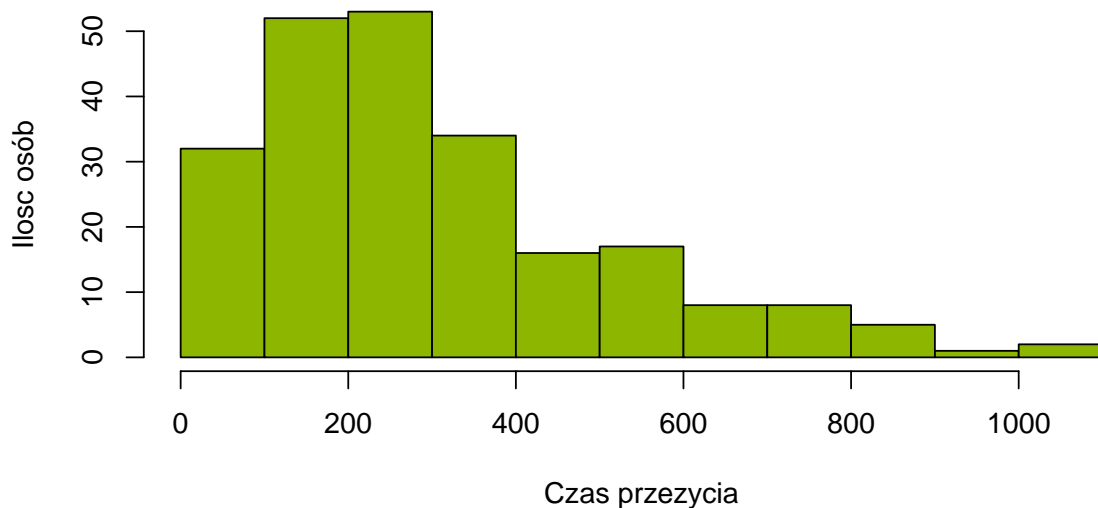
## [1] 10
```

następujących zmiennych:

- *inst* - kod instytucji
- *time* - czas przeżycia w dniach
- *status* - czy otrzymaliśmy zdarzenie (2 - śmierć), czy też nasze dane zostały ocen-zurowane (1 - cenzura)
- *age* - wiek w latach

- *sex* - płeć, gdzie 1 oznacza mężczyznę, zaś 2 - kobietę
- *ph.ecog* - skala sprawności ECPG według lekarza (0 oznacza sprawność prawidłową, 5 - zgon)
- *ph.karno* - skala sprawności Karnofsky'ego według lekarza (100 oznacza sprawność prawidłową, 0 - zgon)
- *pat.karno* - skala sprawności Karnofsky'ego według pacjenta
- *meal.cal* - kalorie spożywane podczas posiłków
- *wt.loss* - utrata masy ciała w ciągu ostatnich sześciu miesięcy

W celu sprawdzenia, ile najczęściej wynosi czas przeżycia pacjentów w dniach (ocenzurowany czy też nie) narysujemy histogram.



Rysunek 1: Histogram czasu przeżycia w dniach (zmiennej time)

Widzimy, że najczęściej pacjenci przeżywają od 100 do 300 dni, czyli nie przeżywają nawet roku.

W celu zilustrowania danych liczbowych i przedstawienia częstości występowania poszczególnych wartości stworzymy tablice liczebności różnych cech.

|   | Płeć      | Ocenzurowanie | Śmierć | Suma osób |
|---|-----------|---------------|--------|-----------|
| 1 | Mężczyzna | 26            | 112    | 138       |
| 2 | Kobieta   | 37            | 53     | 90        |

Tabela 1: Zestawienie danych dla sex i status

Widzimy, że z 90 kobiet biorących udział w badaniu, aż 41% danych dotyczących ich przeżycia zostało ocenzone podczas gdy u mężczyzn (ze 138 w sumie) jest to jedynie 19%.

|   | ph.ecog | Liczba mężczyzn | Liczba kobiet |
|---|---------|-----------------|---------------|
| 1 | 0       | 27              | 36            |
| 2 | 1       | 42              | 71            |
| 3 | 2       | 21              | 29            |
| 4 | 3       | 0               | 1             |
| 5 | 4       | 0               | 0             |
| 6 | 5       | 0               | 0             |
| 7 | NA      | 0               | 1             |

Tabela 2: Zestawienie danych dla sex i ph.ecog

|    | ph.karno | Liczba mężczyzn | Liczba kobiet |
|----|----------|-----------------|---------------|
| 1  | 0        | 0               | 0             |
| 2  | 10       | 0               | 0             |
| 3  | 20       | 0               | 0             |
| 4  | 30       | 0               | 0             |
| 5  | 40       | 0               | 0             |
| 6  | 50       | 2               | 4             |
| 7  | 60       | 8               | 11            |
| 8  | 70       | 12              | 20            |
| 9  | 80       | 27              | 40            |
| 10 | 90       | 29              | 45            |
| 11 | 100      | 12              | 17            |
| 12 | NA       | 0               | 1             |

Tabela 3: Zestawienie danych dla sex i ph.karno

W powyższych tabelach widzimy, że występują jedynie 2 brakujące wartości - dla danych ph.ecog oraz ph.karno dla kobiet.

## 2 ZADANIE 1A

W podpunkcie A zadania 1 mamy naszkicować estymator Kaplana-Meiera funkcji przeżycia (zmiennej time) w całej badanej grupie oraz w podgrupach ze względu na płeć, bez zaznaczonych realizacji przedziałów ufności. Wykorzystamy w tym celu funkcję `Surv` z biblioteki `survival`, która tworzy obiekt przeżycia, zwykle używany jako zmienna odpowiedzi w formule modelu.

Za pomocą funkcji `survfit` tworzymy prostą krzywą przeżycia estymatora Kaplana-Meiera, która nie uwzględnia żadnych podgrup.

```
fit_overall = survfit(Surv(time, status) ~ 1, data = lung,
                      type = "kaplan-meier")
summary(fit_overall)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, data = lung, type = "kaplan-meier")
##
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##      5      228      1  0.9956 0.00438      0.9871      1.000
##     11      227      3  0.9825 0.00869      0.9656      1.000
##     12      224      1  0.9781 0.00970      0.9592      0.997
##     13      223      2  0.9693 0.01142      0.9472      0.992
##     15      221      1  0.9649 0.01219      0.9413      0.989
##     26      220      1  0.9605 0.01290      0.9356      0.986
##     30      219      1  0.9561 0.01356      0.9299      0.983
##     31      218      1  0.9518 0.01419      0.9243      0.980
##     53      217      2  0.9430 0.01536      0.9134      0.974
##     54      215      1  0.9386 0.01590      0.9079      0.970
##     59      214      1  0.9342 0.01642      0.9026      0.967
##     60      213      2  0.9254 0.01740      0.8920      0.960
##     61      211      1  0.9211 0.01786      0.8867      0.957
##     62      210      1  0.9167 0.01830      0.8815      0.953
##     65      209      2  0.9079 0.01915      0.8711      0.946
##     71      207      1  0.9035 0.01955      0.8660      0.943
##     79      206      1  0.8991 0.01995      0.8609      0.939
##     81      205      2  0.8904 0.02069      0.8507      0.932
##     88      203      2  0.8816 0.02140      0.8406      0.925
##     92      201      1  0.8772 0.02174      0.8356      0.921
##     93      199      1  0.8728 0.02207      0.8306      0.917
##     95      198      2  0.8640 0.02271      0.8206      0.910
##    105      196      1  0.8596 0.02302      0.8156      0.906
##    107      194      2  0.8507 0.02362      0.8056      0.898
##    110      192      1  0.8463 0.02391      0.8007      0.894
##    116      191      1  0.8418 0.02419      0.7957      0.891
##    118      190      1  0.8374 0.02446      0.7908      0.887
##    122      189      1  0.8330 0.02473      0.7859      0.883
##    131      188      1  0.8285 0.02500      0.7810      0.879
##    132      187      2  0.8197 0.02550      0.7712      0.871
##    135      185      1  0.8153 0.02575      0.7663      0.867
##    142      184      1  0.8108 0.02598      0.7615      0.863
##    144      183      1  0.8064 0.02622      0.7566      0.859
##    145      182      2  0.7975 0.02667      0.7469      0.852
##    147      180      1  0.7931 0.02688      0.7421      0.848
##    153      179      1  0.7887 0.02710      0.7373      0.844
##    156      178      2  0.7798 0.02751      0.7277      0.836
##    163      176      3  0.7665 0.02809      0.7134      0.824
##    166      173      2  0.7577 0.02845      0.7039      0.816
##    167      171      1  0.7532 0.02863      0.6991      0.811
##    170      170      1  0.7488 0.02880      0.6944      0.807
##    175      167      1  0.7443 0.02898      0.6896      0.803
##    176      165      1  0.7398 0.02915      0.6848      0.799
##    177      164      1  0.7353 0.02932      0.6800      0.795
##    179      162      2  0.7262 0.02965      0.6704      0.787
```

|    |     |     |   |        |         |        |       |
|----|-----|-----|---|--------|---------|--------|-------|
| ## | 180 | 160 | 1 | 0.7217 | 0.02981 | 0.6655 | 0.783 |
| ## | 181 | 159 | 2 | 0.7126 | 0.03012 | 0.6559 | 0.774 |
| ## | 182 | 157 | 1 | 0.7081 | 0.03027 | 0.6511 | 0.770 |
| ## | 183 | 156 | 1 | 0.7035 | 0.03041 | 0.6464 | 0.766 |
| ## | 186 | 154 | 1 | 0.6989 | 0.03056 | 0.6416 | 0.761 |
| ## | 189 | 152 | 1 | 0.6943 | 0.03070 | 0.6367 | 0.757 |
| ## | 194 | 149 | 1 | 0.6897 | 0.03085 | 0.6318 | 0.753 |
| ## | 197 | 147 | 1 | 0.6850 | 0.03099 | 0.6269 | 0.749 |
| ## | 199 | 145 | 1 | 0.6803 | 0.03113 | 0.6219 | 0.744 |
| ## | 201 | 144 | 2 | 0.6708 | 0.03141 | 0.6120 | 0.735 |
| ## | 202 | 142 | 1 | 0.6661 | 0.03154 | 0.6071 | 0.731 |
| ## | 207 | 139 | 1 | 0.6613 | 0.03168 | 0.6020 | 0.726 |
| ## | 208 | 138 | 1 | 0.6565 | 0.03181 | 0.5970 | 0.722 |
| ## | 210 | 137 | 1 | 0.6517 | 0.03194 | 0.5920 | 0.717 |
| ## | 212 | 135 | 1 | 0.6469 | 0.03206 | 0.5870 | 0.713 |
| ## | 218 | 134 | 1 | 0.6421 | 0.03218 | 0.5820 | 0.708 |
| ## | 222 | 132 | 1 | 0.6372 | 0.03231 | 0.5769 | 0.704 |
| ## | 223 | 130 | 1 | 0.6323 | 0.03243 | 0.5718 | 0.699 |
| ## | 226 | 126 | 1 | 0.6273 | 0.03256 | 0.5666 | 0.694 |
| ## | 229 | 125 | 1 | 0.6223 | 0.03268 | 0.5614 | 0.690 |
| ## | 230 | 124 | 1 | 0.6172 | 0.03280 | 0.5562 | 0.685 |
| ## | 239 | 121 | 2 | 0.6070 | 0.03304 | 0.5456 | 0.675 |
| ## | 245 | 117 | 1 | 0.6019 | 0.03316 | 0.5402 | 0.670 |
| ## | 246 | 116 | 1 | 0.5967 | 0.03328 | 0.5349 | 0.666 |
| ## | 267 | 112 | 1 | 0.5913 | 0.03341 | 0.5294 | 0.661 |
| ## | 268 | 111 | 1 | 0.5860 | 0.03353 | 0.5239 | 0.656 |
| ## | 269 | 110 | 1 | 0.5807 | 0.03364 | 0.5184 | 0.651 |
| ## | 270 | 108 | 1 | 0.5753 | 0.03376 | 0.5128 | 0.645 |
| ## | 283 | 104 | 1 | 0.5698 | 0.03388 | 0.5071 | 0.640 |
| ## | 284 | 103 | 1 | 0.5642 | 0.03400 | 0.5014 | 0.635 |
| ## | 285 | 101 | 2 | 0.5531 | 0.03424 | 0.4899 | 0.624 |
| ## | 286 | 99  | 1 | 0.5475 | 0.03434 | 0.4841 | 0.619 |
| ## | 288 | 98  | 1 | 0.5419 | 0.03444 | 0.4784 | 0.614 |
| ## | 291 | 97  | 1 | 0.5363 | 0.03454 | 0.4727 | 0.608 |
| ## | 293 | 94  | 1 | 0.5306 | 0.03464 | 0.4669 | 0.603 |
| ## | 301 | 91  | 1 | 0.5248 | 0.03475 | 0.4609 | 0.597 |
| ## | 303 | 89  | 1 | 0.5189 | 0.03485 | 0.4549 | 0.592 |
| ## | 305 | 87  | 1 | 0.5129 | 0.03496 | 0.4488 | 0.586 |
| ## | 306 | 86  | 1 | 0.5070 | 0.03506 | 0.4427 | 0.581 |
| ## | 310 | 85  | 2 | 0.4950 | 0.03523 | 0.4306 | 0.569 |
| ## | 320 | 82  | 1 | 0.4890 | 0.03532 | 0.4244 | 0.563 |
| ## | 329 | 81  | 1 | 0.4830 | 0.03539 | 0.4183 | 0.558 |
| ## | 337 | 79  | 1 | 0.4768 | 0.03547 | 0.4121 | 0.552 |
| ## | 340 | 78  | 1 | 0.4707 | 0.03554 | 0.4060 | 0.546 |
| ## | 345 | 77  | 1 | 0.4646 | 0.03560 | 0.3998 | 0.540 |
| ## | 348 | 76  | 1 | 0.4585 | 0.03565 | 0.3937 | 0.534 |
| ## | 350 | 75  | 1 | 0.4524 | 0.03569 | 0.3876 | 0.528 |
| ## | 351 | 74  | 1 | 0.4463 | 0.03573 | 0.3815 | 0.522 |

|    |     |    |   |        |         |        |       |
|----|-----|----|---|--------|---------|--------|-------|
| ## | 353 | 73 | 2 | 0.4340 | 0.03578 | 0.3693 | 0.510 |
| ## | 361 | 70 | 1 | 0.4278 | 0.03581 | 0.3631 | 0.504 |
| ## | 363 | 69 | 2 | 0.4154 | 0.03583 | 0.3508 | 0.492 |
| ## | 364 | 67 | 1 | 0.4092 | 0.03582 | 0.3447 | 0.486 |
| ## | 371 | 65 | 2 | 0.3966 | 0.03581 | 0.3323 | 0.473 |
| ## | 387 | 60 | 1 | 0.3900 | 0.03582 | 0.3258 | 0.467 |
| ## | 390 | 59 | 1 | 0.3834 | 0.03582 | 0.3193 | 0.460 |
| ## | 394 | 58 | 1 | 0.3768 | 0.03580 | 0.3128 | 0.454 |
| ## | 426 | 55 | 1 | 0.3700 | 0.03580 | 0.3060 | 0.447 |
| ## | 428 | 54 | 1 | 0.3631 | 0.03579 | 0.2993 | 0.440 |
| ## | 429 | 53 | 1 | 0.3563 | 0.03576 | 0.2926 | 0.434 |
| ## | 433 | 52 | 1 | 0.3494 | 0.03573 | 0.2860 | 0.427 |
| ## | 442 | 51 | 1 | 0.3426 | 0.03568 | 0.2793 | 0.420 |
| ## | 444 | 50 | 1 | 0.3357 | 0.03561 | 0.2727 | 0.413 |
| ## | 450 | 48 | 1 | 0.3287 | 0.03555 | 0.2659 | 0.406 |
| ## | 455 | 47 | 1 | 0.3217 | 0.03548 | 0.2592 | 0.399 |
| ## | 457 | 46 | 1 | 0.3147 | 0.03539 | 0.2525 | 0.392 |
| ## | 460 | 44 | 1 | 0.3076 | 0.03530 | 0.2456 | 0.385 |
| ## | 473 | 43 | 1 | 0.3004 | 0.03520 | 0.2388 | 0.378 |
| ## | 477 | 42 | 1 | 0.2933 | 0.03508 | 0.2320 | 0.371 |
| ## | 519 | 39 | 1 | 0.2857 | 0.03498 | 0.2248 | 0.363 |
| ## | 520 | 38 | 1 | 0.2782 | 0.03485 | 0.2177 | 0.356 |
| ## | 524 | 37 | 2 | 0.2632 | 0.03455 | 0.2035 | 0.340 |
| ## | 533 | 34 | 1 | 0.2554 | 0.03439 | 0.1962 | 0.333 |
| ## | 550 | 32 | 1 | 0.2475 | 0.03423 | 0.1887 | 0.325 |
| ## | 558 | 30 | 1 | 0.2392 | 0.03407 | 0.1810 | 0.316 |
| ## | 567 | 28 | 1 | 0.2307 | 0.03391 | 0.1729 | 0.308 |
| ## | 574 | 27 | 1 | 0.2221 | 0.03371 | 0.1650 | 0.299 |
| ## | 583 | 26 | 1 | 0.2136 | 0.03348 | 0.1571 | 0.290 |
| ## | 613 | 24 | 1 | 0.2047 | 0.03325 | 0.1489 | 0.281 |
| ## | 624 | 23 | 1 | 0.1958 | 0.03297 | 0.1407 | 0.272 |
| ## | 641 | 22 | 1 | 0.1869 | 0.03265 | 0.1327 | 0.263 |
| ## | 643 | 21 | 1 | 0.1780 | 0.03229 | 0.1247 | 0.254 |
| ## | 654 | 20 | 1 | 0.1691 | 0.03188 | 0.1169 | 0.245 |
| ## | 655 | 19 | 1 | 0.1602 | 0.03142 | 0.1091 | 0.235 |
| ## | 687 | 18 | 1 | 0.1513 | 0.03090 | 0.1014 | 0.226 |
| ## | 689 | 17 | 1 | 0.1424 | 0.03034 | 0.0938 | 0.216 |
| ## | 705 | 16 | 1 | 0.1335 | 0.02972 | 0.0863 | 0.207 |
| ## | 707 | 15 | 1 | 0.1246 | 0.02904 | 0.0789 | 0.197 |
| ## | 728 | 14 | 1 | 0.1157 | 0.02830 | 0.0716 | 0.187 |
| ## | 731 | 13 | 1 | 0.1068 | 0.02749 | 0.0645 | 0.177 |
| ## | 735 | 12 | 1 | 0.0979 | 0.02660 | 0.0575 | 0.167 |
| ## | 765 | 10 | 1 | 0.0881 | 0.02568 | 0.0498 | 0.156 |
| ## | 791 | 9  | 1 | 0.0783 | 0.02462 | 0.0423 | 0.145 |
| ## | 814 | 7  | 1 | 0.0671 | 0.02351 | 0.0338 | 0.133 |
| ## | 883 | 4  | 1 | 0.0503 | 0.02285 | 0.0207 | 0.123 |

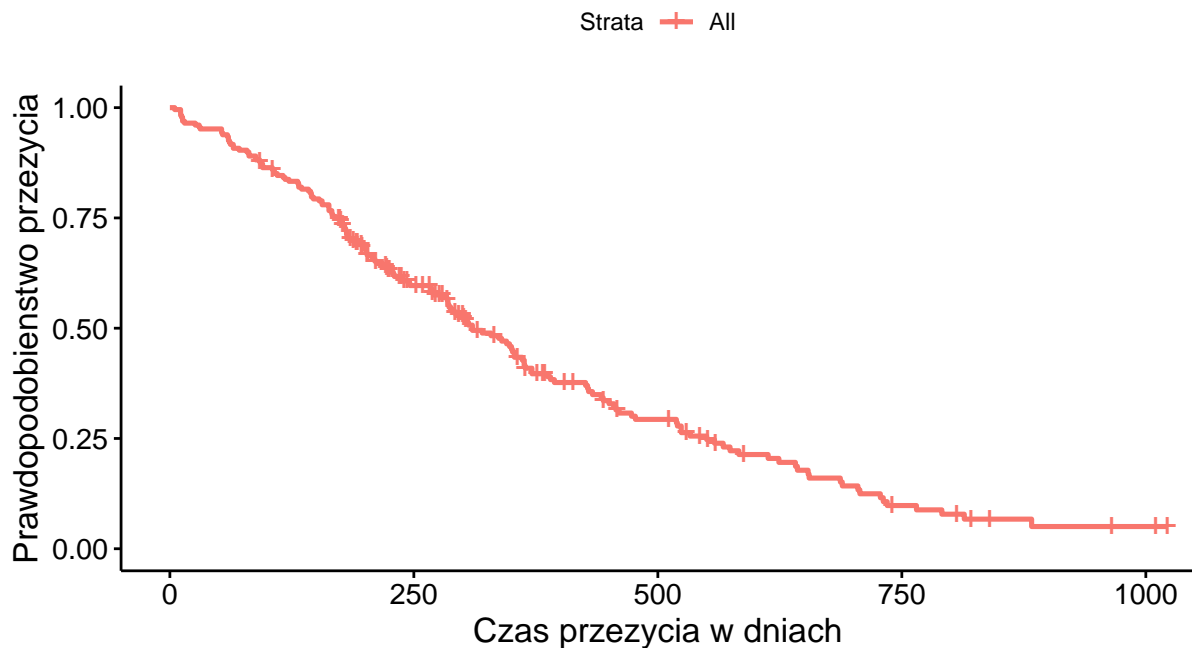
Powyższa funkcja pokazuje nam liczbę obserwacji, liczbę zdarzeń, medianę przeżycia oraz



granice ufności dla mediany. Zauważmy, że mediana (czyli wartość środkowa) czasu przeżycia wynosi 310 dni.

Teraz zaprezentujemy to na wykresie Kaplana-Meiera.

```
ggsurvplot(fit_overall, xlab = "Czas przeżycia w dniach",  
            ylab = "Prawdopodobieństwo przeżycia", conf.int = FALSE)
```



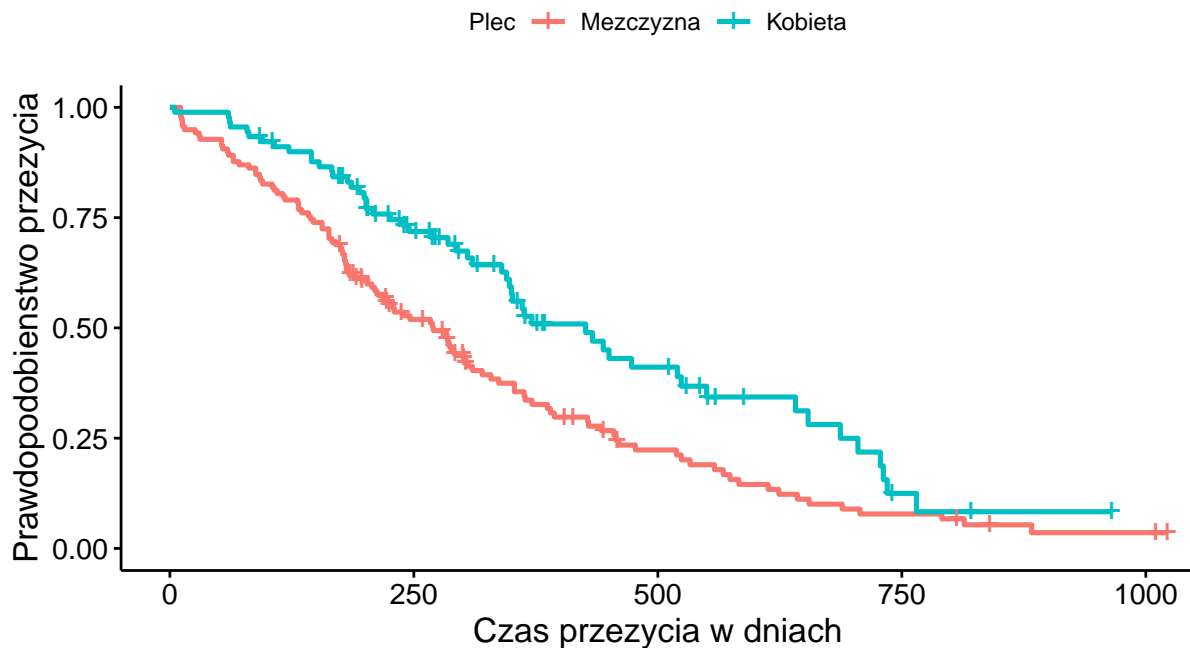
Rysunek 2: Estymator KM w całej badanej grupie

Ponownie, za pomocą funkcji `survfit` tworzymy prostą krzywą przeżycia estymatora Kaplana-Meiera w pogrupie ze względu na płeć.

```
fit_overall1 = survfit(Surv(time, status) ~ sex, data = lung,  
                       type = "kaplan-meier")
```

Zauważmy, że mediana przeżycia kobiet prawie o połowę większa, niż mediana przeżycia mężczyzn, mimo, że badanych kobiet jest o 48 mniej.

```
ggsurvplot(fit_overall1, legend.title = "Płeć",  
            legend.labs = c("Mężczyzna", "Kobieta"),  
            xlab = "Czas przeżycia w dniach",  
            ylab = "Prawdopodobieństwo przeżycia")
```



Rysunek 3: Estymator KM w podgrupach ze względu na płeć

Powyższy wykres potwierdza wcześniejsze wnioski. Krzywa niebieska będą estymatorem KM kobiet przez cały okres badania jest ponad czerwoną krzywą estymatora KM mężczyzn. To znaczy, że w całym okresie badania prawdopodobieństwo przeżycia kobiet jest większe.

### 3 ZADANIE 1B

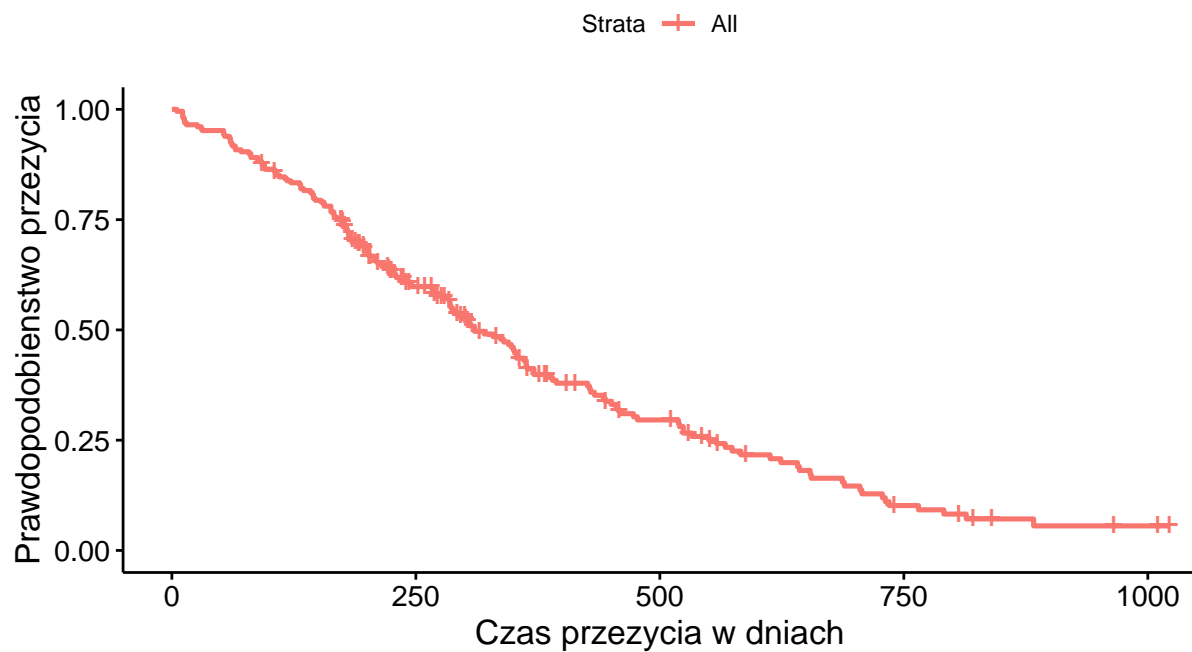
W podpunkcie B zadania 1 mamy naszkicować estymator Fleminga-Harringtona funkcji przeżycia (zmiennej `time`) w całej badanej grupie i w podgrupach ze względu na płeć, ponownie bez zaznaczonych realizacji przedziałów ufności.

Za pomocą funkcji `survfit` tworzymy prostą krzywą przeżycia estymatora Fleminga-Harringtona, która nie uwzględnia żadnych podgrup.

```
fit_overall2 = survfit(Surv(time, status) ~ 1, data = lung,
                      type = "fleming-harrington")
print(fit_overall2)

## Call: survfit(formula = Surv(time, status) ~ 1, data = lung, type = "fleming-harrington")
##
##           n events median 0.95LCL 0.95UCL
## [1,] 228     165     310     285     363

ggsurvplot(fit_overall2, xlab = "Czas przeżycia w dniach",
            ylab = "Prawdopodobieństwo przeżycia", conf.type = "none",
            conf.int = FALSE)
```

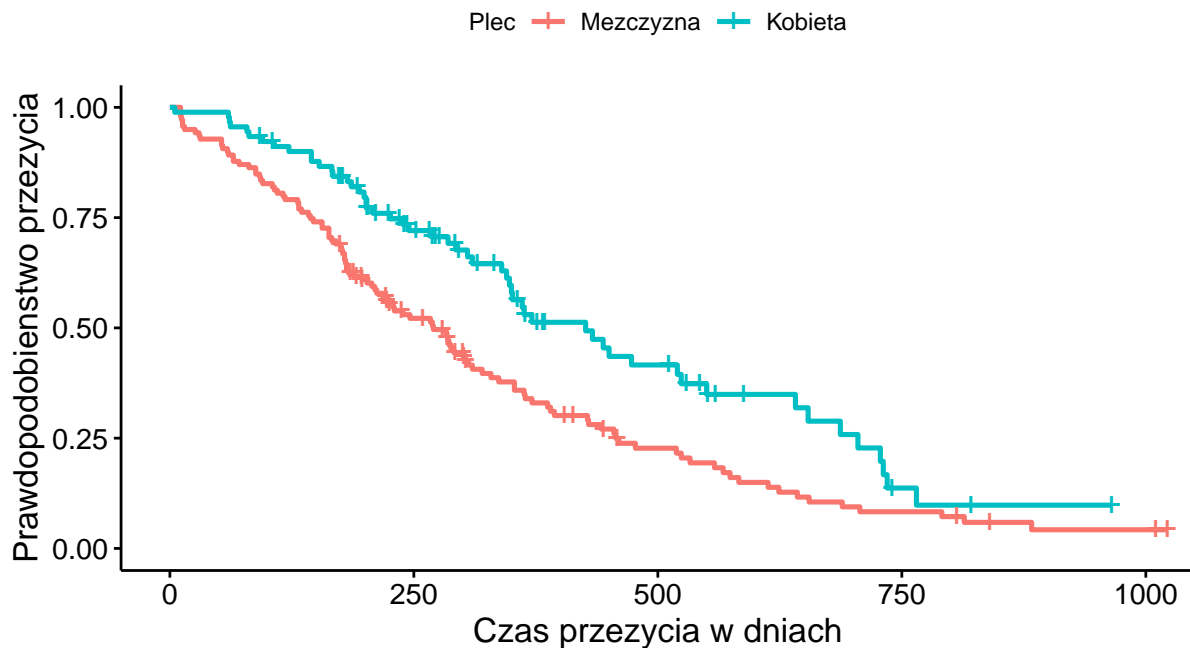


Rysunek 4: Estymator FH w całej badanej grupie

```
fit_overall3 = survfit(Surv(time, status) ~ sex, data = lung,
                      type = "fleming-harrington")
print(fit_overall3)

## Call: survfit(formula = Surv(time, status) ~ sex, data = lung, type = "fleming-harrington")
##
##           n events median 0.95LCL 0.95UCL
## sex=1 138    112    270    218    320
## sex=2  90     53    426    348    550
```

```
ggsurvplot(fit_overall3, legend.title = "Płeć",
            legend.labs = c("Mężczyzna", "Kobieta"),
            xlab = "Czas przeżycia w dniach",
            ylab = "Prawdopodobieństwo przeżycia")
```



Rysunek 5: Estymator FH w podgrupach ze względu na płeć

```
print(fit_overall2)

## Call: survfit(formula = Surv(time, status) ~ 1, data = lung, type = "fleming-harrin
##
##          n events median 0.95LCL 0.95UCL
## [1,] 228    165    310    285    363
```

Widzimy, że powyższe wykresy nie różnią się od siebie praktycznie wogóle.

## 4 ZADANIE 2

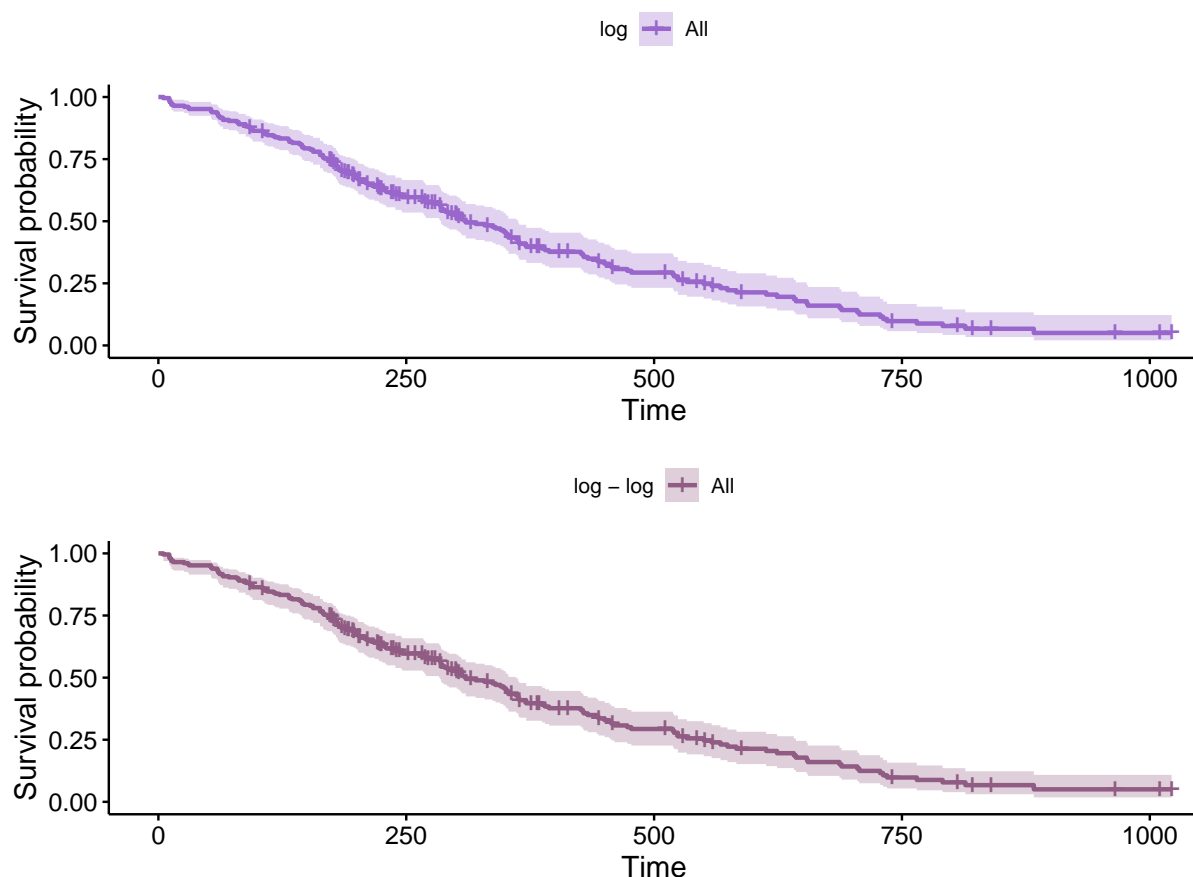
Mamy naszkicować wykres estymatora Kaplana-Meiera w funkcji przeżycia w całej badanej grupie oraz podgrupach, ze względu na wartość zmiennej `ph.ecog` wraz z realizacjami przedziałów ufności (asymptotycznych punktowo) wybierając różne typy tych przedziałów. Następnie mamy zastanowić się nad wyborem typu przedziału ufności i uzasadnieniem tego wyboru oraz sformułować wnioski wynikające z porównania uzyskanych wykresów w podgrupach ze względu na wartość zmiennej `ph.ecog`.

|   | typ       | n   | events | median | 0.95 LCL | 0.95 UCL |
|---|-----------|-----|--------|--------|----------|----------|
| 1 | log       | 228 | 165    | 310    | 285      | 363      |
| 2 | log - log | 228 | 165    | 310    | 284      | 361      |
| 3 | plain     | 228 | 165    | 310    | 284      | 361      |

Tabela 4: Zestawienie danych dla całej badanej grupy

Powyżej zaprezentowana została tabela podająca nam liczbę obserwacji, zdarzeń, medianę przeżycia oraz granice ufności dla mediany. Analizując ją widzimy, że typ przedziału

log-log oraz plain zwracają nam te same wartości, dlatego w dalszej analizie będziemy pisać tylko jeden z nich. Różnica przedziału ufności dla typu log wynosi 78 podczas gdy dla typu log - log jest równa 77. Ponieważ wiemy, że lepszym przedziałem ufności jest ten krótszy, to mogłoby się wydawać, że typ przedziału log - log jest lepszy od log. Poniżej prezentujemy dwa wykresy estymatora Kaplana - Meiera funkcji przeżycia w całej badanej grupie.



Dalej przejdziemy do analizowania funkcji przeżycia w podgrupach ze względu na wartość zmiennej `ph.ecog`.

- Dla `ph.ecog = 0`, różnica przedziałów ufności dla typu przedziału log wynosi 228, podczas gdy dla typu przedziału log - log jest ona równa 218.
- Dla `ph.ecog = 1`, różnica przedziałów ufności dla typu przedziału log wynosi 161, podczas gdy dla typu przedziału log - log jest ona równa 181.
- Dla `ph.ecog = 2`, różnica przedziałów ufności dla typu przedziału log wynosi 132, podczas gdy dla typu przedziału log - log jest ona równa 132.
- Dla `ph.ecog = 3` mamy jedynie 1 obserwację, zatem jest to za mało, żeby wyznaczyć granice przedziałów ufności.

|   | typ       | n  | events | median | 0.95 LCL | 0.95 UCL |
|---|-----------|----|--------|--------|----------|----------|
| 1 | log       | 63 | 37     | 394    | 348      | 574      |
| 2 | log - log | 63 | 37     | 394    | 340      | 558      |
| 3 | plain     | 63 | 37     | 394    | 340      | 558      |

Tabela 5: Zestawienie danych dla  $ph.ecog = 0$

|   | typ       | n   | events | median | 0.95 LCL | 0.95 UCL |
|---|-----------|-----|--------|--------|----------|----------|
| 1 | log       | 113 | 82     | 306    | 268      | 429      |
| 2 | log - log | 113 | 82     | 306    | 245      | 426      |
| 3 | plain     | 113 | 82     | 306    | 245      | 426      |

Tabela 6: Zestawienie danych dla  $ph.ecog = 1$

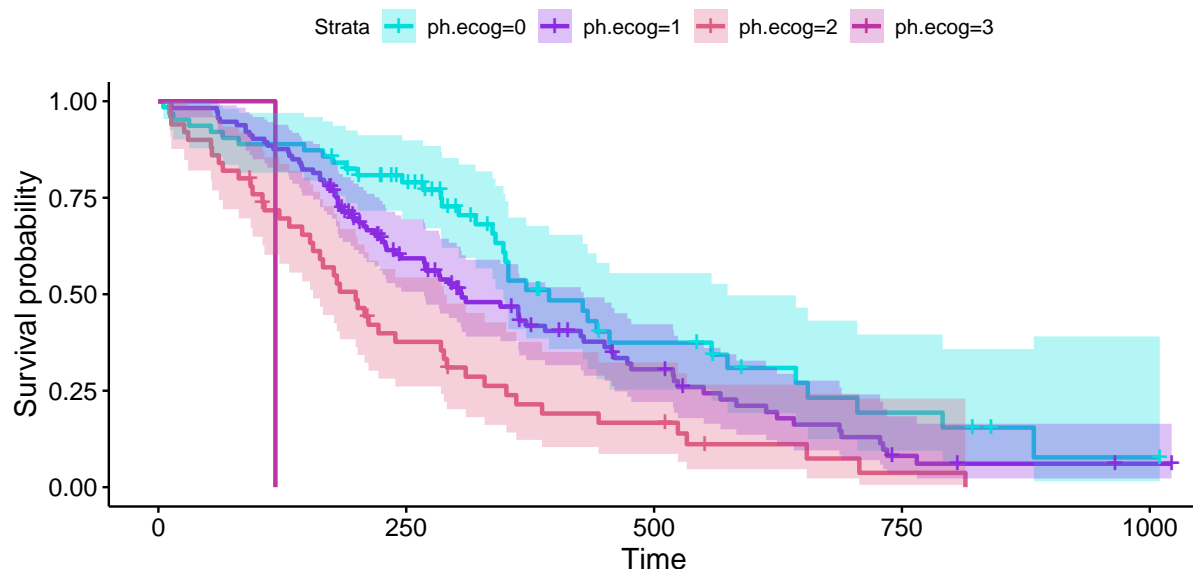
|   | typ       | n  | events | median | 0.95 LCL | 0.95 UCL |
|---|-----------|----|--------|--------|----------|----------|
| 1 | log       | 50 | 44     | 199    | 156      | 288      |
| 2 | log - log | 50 | 44     | 199    | 153      | 285      |
| 3 | plain     | 50 | 44     | 199    | 153      | 285      |

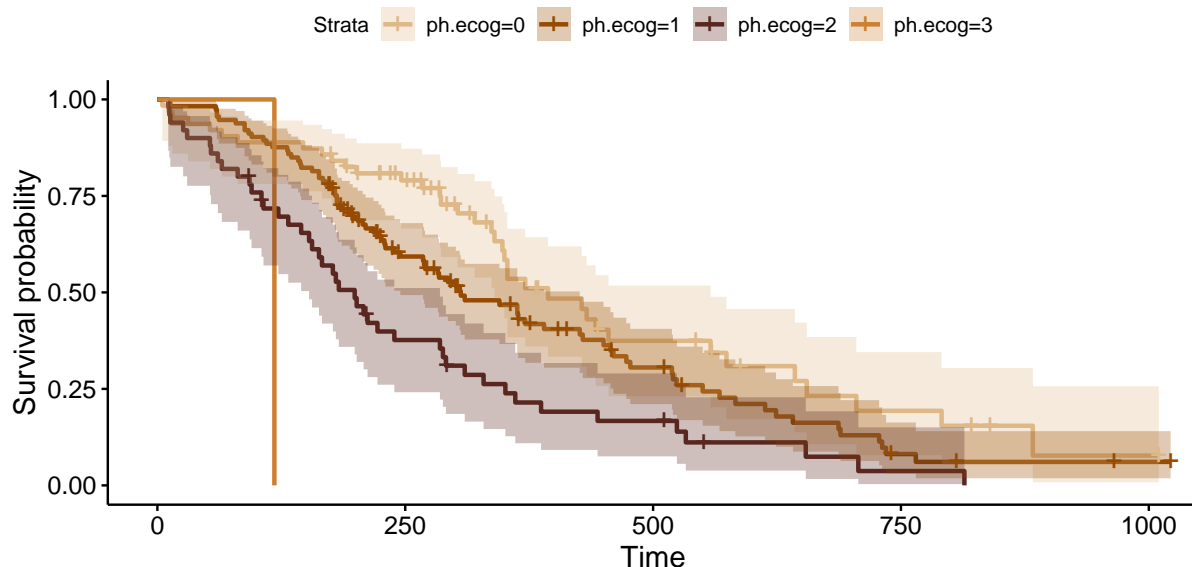
Tabela 7: Zestawienie danych dla  $ph.ecog = 2$

|   | typ       | n | events | median | 0.95 LCL | 0.95 UCL |
|---|-----------|---|--------|--------|----------|----------|
| 1 | log       | 1 | 1      | 118    | NA       | NA       |
| 2 | log - log | 1 | 1      | 118    | NA       | NA       |
| 3 | plain     | 1 | 1      | 118    | NA       | NA       |

Tabela 8: Zestawienie danych dla  $ph.ecog = 3$

Czyli ponownie, prawie za każdym razem, mniejsze przedziały ufności wyznacza nam typ przedziału  $\log - \log$ . Poniżej rysujemy wykresy estymatora funkcji Kaplana - Meiera funkcji przeżycia w podgrupach, ze względu na wartość zmiennej  $ph.ecog$ .





Najbardziej odpowiednim typem przedziału ufności wydaje się być  $\log - \log$ , ponieważ zawęży on nasze pole, zmniejszy wahania i daje dokładniejsze przewidywania. Przyglądając się powyższemu wykresom KM widzimy, że typ  $\log - \log$  bardziej przypomina funkcję schodkową, zaś typ  $\log$  ma większe wahania i szersze zakresy przewidywań. Ponadto widzimy, że pacjenci oznaczeni skalą sprawności ECOG jako 0 według lekarza, mają znacznie dłuższą krzywą przeżycia niż pozostałe osoby.

## 5 ZADANIE 3

Zadanie 3 polega na oszacowaniu punktowym i przedziałowym wartości oczekiwanej i mediany rozkładu czasu przeżycia w całej badanej grupie i w podgrupach ze względu na wartość zmiennej `ph.ecog`.

Korzystając z funkcji `survmean` (która oblicza średnie czasy przeżycia w oparciu o oszacowanie przeżycia do pewnego punktu w czasie obserwacji) z biblioteki `survival` oszacujemy najpierw wartości średniej i mediany w całej badanej grupie.

|                                | Średnia   | Mediana   |
|--------------------------------|-----------|-----------|
| Oszacowanie punktowe           | 375.16714 | 310.00000 |
| Oszacowanie przedziałowe dolne | 355.72828 | 285.00000 |
| Oszacowanie przedziałowe górne | 394.60600 | 363.00000 |

Tabela 9: Oszacowanie punktowe i przedziałowe mediany i wartości średniej w całej grupie

Oraz w podgrupach najpierw wartość średniej ze względu na wartość zmiennej `ph.ecog`.

|                                | <code>ph.ecog = 0</code> | <code>ph.ecog = 1</code> | <code>ph.ecog = 2</code> | <code>ph.ecog = 3</code> |
|--------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Oszacowanie punktowe           | 463.61368                | 383.50235                | 258.73793                | 118.00000                |
| Oszacowanie przedziałowe dolne | 421.45458                | 356.64388                | 227.49531                | 118.00000                |
| Oszacowanie przedziałowe górne | 505.77278                | 410.36083                | 289.98055                | 118.00000                |

Tabela 10: Oszacowanie wartości średniej ze względu na wartość zmiennej `ph.ecog`

Oraz w podgrupach wartość mediany ze względu na wartość zmiennej `ph.ecog`.

|                                | ph.ecog = 0 | ph.ecog = 1 | ph.ecog = 2 | ph.ecog = 3 |
|--------------------------------|-------------|-------------|-------------|-------------|
| Oszacowanie punktowe           | 394.00000   | 306.00000   | 199.00000   | 118         |
| Oszacowanie przedziałowe dolne | 348.00000   | 268.00000   | 156.00000   | NA          |
| Oszacowanie przedziałowe górne | 574.00000   | 429.00000   | 288.00000   | NA          |

Tabela 11: Oszacowanie wartości mediany ze względu na wartość zmiennej `ph.ecog`

## 6 ZADANIE 4

Dalej, w oparciu o realizacje przedziałów ufności, na poziomie ufności  $1 - \alpha = 0.95$ , dla średnich w podgrupach ze względu na wartość zmiennej `ph.ecog`, mamy sformułować wniosek dotyczący weryfikacji hipotezy o równości średnich w tych podgrupach, na poziomie istotności  $\alpha = 0.05$ . Wykorzystamy w tym celu funkcję `survdif` z biblioteki `survival`, która testuje, czy istnieje różnica między dwiema lub więcej krzywymi przeżycia. Wartość `ph.ecog` występuje w 4 różnych wartościach, zatem otrzymujemy cztery podpopulacje.

```
survdif(Surv(time, status) ~ ph.ecog, data = dane)

## Call:
## survdiff(formula = Surv(time, status) ~ ph.ecog, data = dane)
##
## n=227, 1 obserwacja została skasowana z uwagi na braki w niej zawarte.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## ph.ecog=0  63      37    54.153    5.4331    8.2119
## ph.ecog=1 113      82    83.528    0.0279    0.0573
## ph.ecog=2  50      44    26.147   12.1893   14.6491
## ph.ecog=3   1       1     0.172    3.9733    4.0040
##
##  Chisq= 22  on 3 degrees of freedom, p= 7e-05
```

```
survdif(Surv(time, status) ~ ph.ecog, rho = 0.1, data = dane)

## Call:
## survdiff(formula = Surv(time, status) ~ ph.ecog, data = dane,
##          rho = 0.1)
##
## n=227, 1 obserwacja została skasowana z uwagi na braki w niej zawarte.
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## ph.ecog=0  63   34.253   50.547    5.2529    8.418
## ph.ecog=1 113   76.782   78.470    0.0363    0.079
## ph.ecog=2  50   41.946   24.776   11.8985   15.093
## ph.ecog=3   1    0.983    0.171    3.8603    3.922
```



```
##
## Chisq= 22.4 on 3 degrees of freedom, p= 5e-05
```

```
survdif(Surv(time, status) ~ ph.ecog, rho = 0.5, data = dane)

## Call:
## survdiff(formula = Surv(time, status) ~ ph.ecog, data = dane,
##          rho = 0.5)
##
## n=227, 1 obserwacja została skasowana z uwagi na braki w niej zawarte.
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## ph.ecog=0  63    26.11   39.726    4.6657    8.770
## ph.ecog=1 113    60.78   62.957    0.0751    0.193
## ph.ecog=2  50    35.51   20.473   11.0400   16.317
## ph.ecog=3   1     0.92    0.165    3.4368    3.606
##
## Chisq= 23.3 on 3 degrees of freedom, p= 3e-05
```

```
survdif(Surv(time, status) ~ ph.ecog, rho = 1, data = dane)

## Call:
## survdiff(formula = Surv(time, status) ~ ph.ecog, data = dane,
##          rho = 1)
##
## n=227, 1 obserwacja została skasowana z uwagi na braki w niej zawarte.
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## ph.ecog=0  63    19.877   31.190    4.103    8.605
## ph.ecog=1 113    47.814   50.346    0.127    0.367
## ph.ecog=2  50    30.002   16.843   10.279   16.964
## ph.ecog=3   1     0.846    0.159    2.967    3.236
##
## Chisq= 23.4 on 3 degrees of freedom, p= 3e-05
```

|   | Wartość RHO | Wartość p |
|---|-------------|-----------|
| 1 | rho = 0     | p = 7e-05 |
| 2 | rho = 0.1   | p = 5e-05 |
| 3 | rho = 0.5   | p = 3e-05 |
| 4 | rho = 1     | p = 3e-05 |

Tabela 12: Różne wartości wartości poziomów krytycznych dla różnych wartości RHO

Widzimy, że dla funkcja `survdif` dla różnych wartości  $\rho$  daje różne wartości poziomów krytycznych, w każdym przypadku są one bardzo małe. Na poziomie istotności  $\alpha = 0.05$  odrzucamy hipotezę o równości średnich we wskazanych podgrupach ze względu na wartość zmiennej `ph.ecog`.