Raport 3
Analiza przeżycia

Jan Solarz
Aleksander Kaczmarek

11 stycznia 2021

Spis treści


```
## Loading required package: ggplot2
## Loading required package: ggpubr
```

1 Lista 1

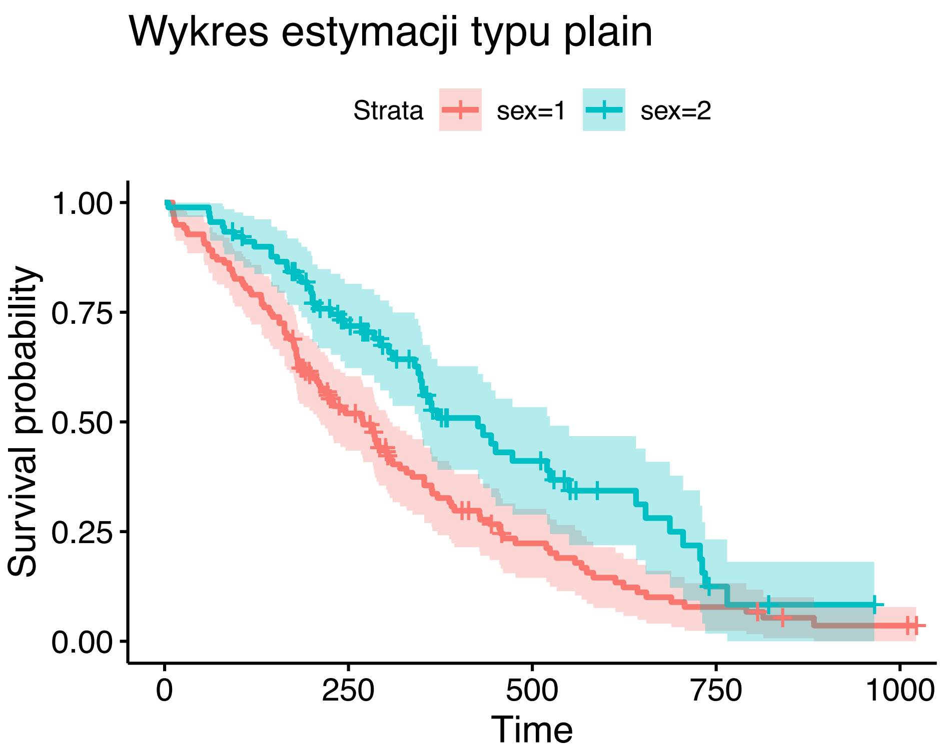
1.1 Zadanie 1

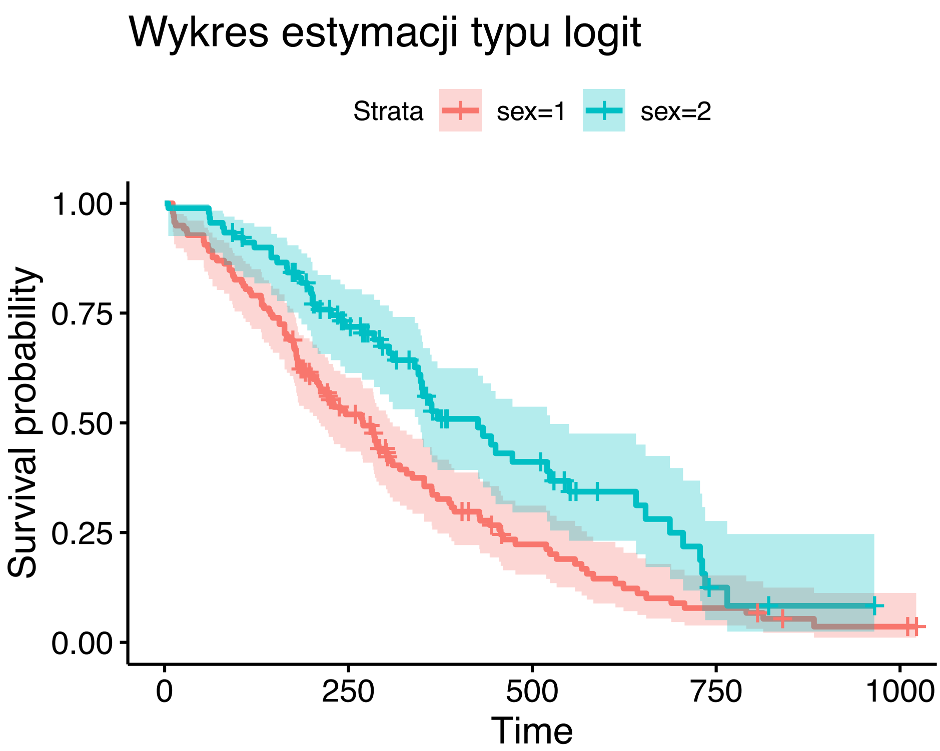
Wykonując odpowiednia testy chcemy zweryfikować hipotezę o równości rozkładów czasu przeżycia w grupie kobiet i mężczyzn na poziomie istotności alpha = 0.05

```
## Call:
## survdiff(formula = Surv(time, status) ~ sex, data = lung, rho = 0)
##
```

```
##         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
## Call:
## survdiff(formula = Surv(time, status) ~ sex, data = lung, rho = 1)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 138     70.4     55.6      3.95      12.7
## sex=2  90     28.7     43.5      5.04      12.7
##
##  Chisq= 12.7  on 1 degrees of freedom, p= 4e-04
##
##  Asymptotic Two-Sample Gehan-Breslow Test
##
## data:  Surv(time, status) by sex (1, 2)
## Z = -3.5745, p-value = 0.0003509
## alternative hypothesis: true theta is not equal to 1
##
##  Asymptotic Two-Sample Prentice-Marek Test
##
## data:  Surv(time, status) by sex (1, 2)
## Z = -3.6131, p-value = 0.0003026
## alternative hypothesis: true theta is not equal to 1
##
##  Asymptotic Two-Sample Prentice Test
##
## data:  Surv(time, status) by sex (1, 2)
## Z = -3.6133, p-value = 0.0003024
## alternative hypothesis: true theta is not equal to 1
```

Na podstawie wykonanych testów możemy jednoznacznie odrzucić naszą hipotezę.
Wizualizacja funkcji przeżycia z zadania ze względu na płeć



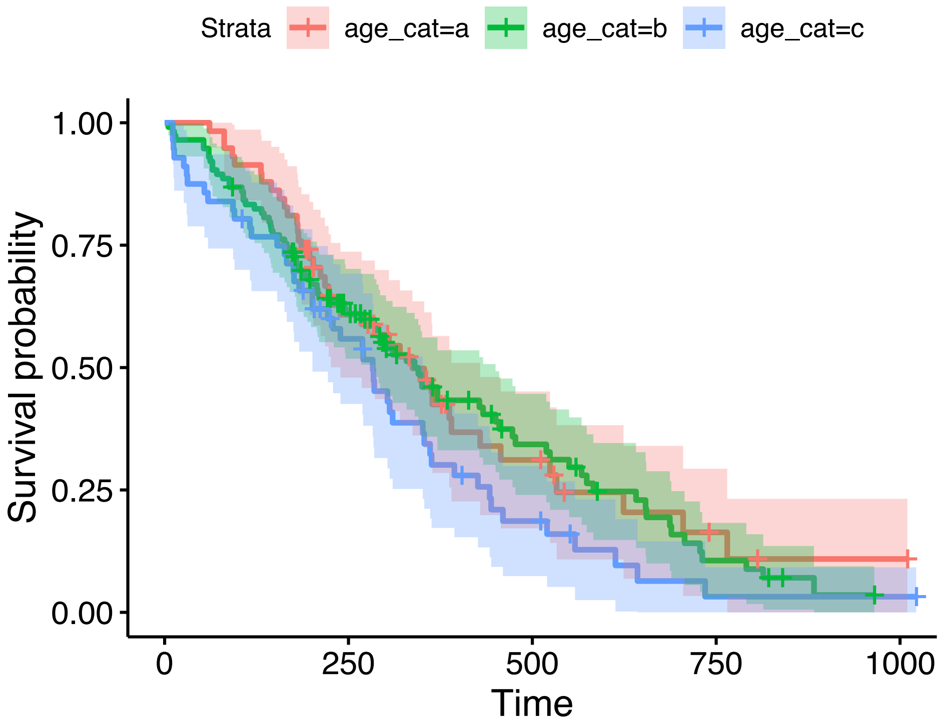


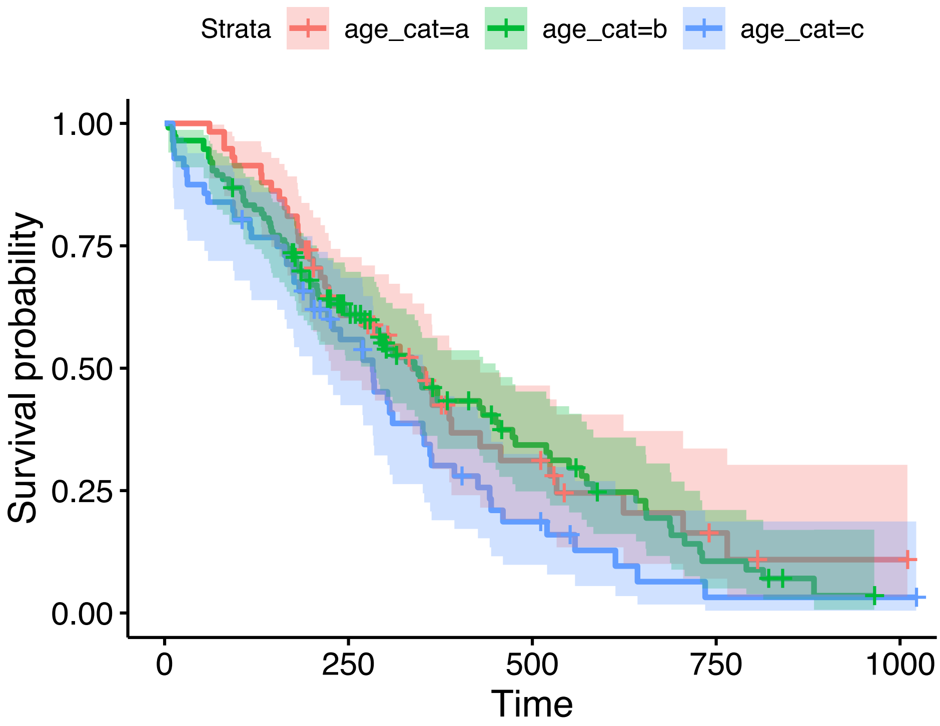
1.2 Zadanie 2

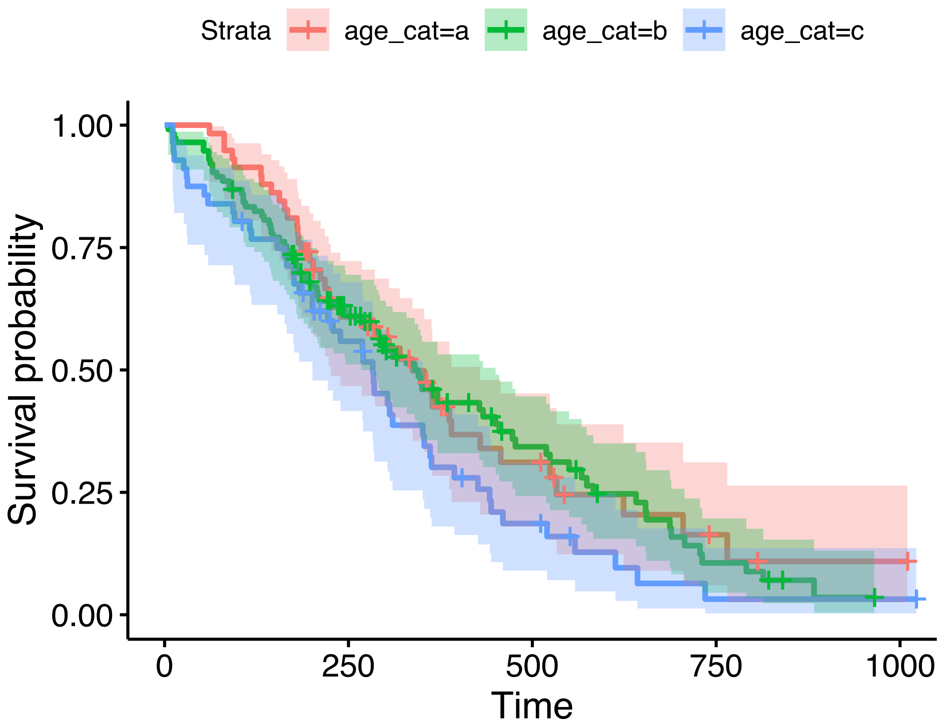
```
#Zad2
#kategoryzacja na rozne przedzialy wiekowe
lung[, age_cat := cut(age, breaks = c(-Inf, quantile(age, 0.25),quantile(age, 0.75), Inf
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ age, data = lung, rho = 0)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## age=39  2        0   0.9330  9.33e-01  9.45e-01
## age=40  1        1   0.1981  3.25e+00  3.27e+00
## age=41  1        0   0.3633  3.63e-01  3.66e-01
## age=42  1        0   0.3700  3.70e-01  3.73e-01
## age=43  1        0   0.9191  9.19e-01  9.30e-01
## age=44  5        3   5.9334  1.45e+00  1.54e+00
## age=45  1        0   0.7243  7.24e-01  7.31e-01
## age=46  1        1   0.7119  1.17e-01  1.18e-01
## age=47  1        1   0.8301  3.48e-02  3.51e-02
## age=48  4        3   3.8279  1.79e-01  1.85e-01
## age=49  2        2   0.5494  3.83e+00  3.87e+00
## age=50  6        5   5.2192  9.20e-03  9.61e-03
## age=51  2        1   2.6072  9.91e-01  1.02e+00
## age=52  2        1   0.6662  1.67e-01  1.69e-01
## age=53  9        7   4.9150  8.85e-01  9.26e-01
## age=54  4        3   4.0610  2.77e-01  2.88e-01
## age=55  6        4   3.4501  8.77e-02  9.02e-02
## age=56  9        7   8.0263  1.31e-01  1.40e-01
## age=57  9        6   4.9538  2.21e-01  2.30e-01
## age=58  8        5   7.2323  6.89e-01  7.26e-01
## age=59  8        6   8.0714  5.32e-01  5.64e-01
## age=60 11        8   8.0541  3.63e-04  3.86e-04
## age=61  5        4   1.5450  3.90e+00  3.98e+00
## age=62  7        5   4.3279  1.04e-01  1.08e-01
## age=63 11        7  10.2729  1.04e+00  1.13e+00
## age=64 11        6   9.2788  1.16e+00  1.24e+00
## age=65  8        6   4.4274  5.59e-01  5.82e-01
## age=66  7        5   5.3616  2.44e-02  2.57e-02
## age=67  8        6   3.1144  2.67e+00  2.77e+00
## age=68 10        9  10.6116  2.45e-01  2.65e-01
## age=69 11        7   8.0073  1.27e-01  1.35e-01
## age=70 10        8   7.7352  9.07e-03  9.61e-03
## age=71  7        6   4.8780  2.58e-01  2.70e-01
## age=72  7        7   4.7679  1.04e+00  1.08e+00
## age=73  6        6   1.3362  1.63e+01  1.67e+01
## age=74 10        6   8.8303  9.07e-01  9.70e-01
## age=75  5        4   3.4998  7.15e-02  7.37e-02
## age=76  5        5   1.1472  1.29e+01  1.32e+01
## age=77  2        0   1.7414  1.74e+00  1.77e+00
## age=80  2        2   1.4335  2.24e-01  2.27e-01
## age=81  1        1   0.0176  5.48e+01  5.54e+01
## age=82  1        1   0.0493  1.84e+01  1.85e+01
##
##  Chisq= 138  on 41 degrees of freedom, p= 2e-12
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ age, data = lung, rho = 1)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## age=39  2    0.000   0.7414  7.41e-01  9.25e-01
## age=40  1    0.829   0.1803  2.33e+00  2.57e+00
## age=41  1    0.000   0.3057  3.06e-01  3.62e-01
## age=42  1    0.000   0.3103  3.10e-01  3.69e-01
## age=43  1    0.000   0.6034  6.03e-01  8.68e-01
## age=44  5    2.071   2.7720  1.77e-01  2.83e-01
## age=45  1    0.000   0.5170  5.17e-01  7.01e-01
## age=46  1    0.495   0.5110  5.00e-04  6.75e-04
## age=47  1    0.446   0.5660  2.53e-02  3.55e-02
## age=48  4    1.419   2.3439  3.65e-01  5.45e-01
## age=49  2    1.551   0.4628  2.56e+00  3.02e+00
## age=50  6    2.634   2.7372  3.86e-03  5.89e-03
## age=51  2    0.142   1.3302  1.06e+00  1.68e+00
## age=52  2    0.899   0.5343  2.49e-01  3.06e-01
## age=53  9    4.412   3.6751  1.48e-01  1.99e-01
## age=54  4    1.380   2.3924  4.29e-01  6.61e-01
## age=55  6    2.728   2.3754  5.24e-02  7.19e-02
## age=56  9    4.064   4.5337  4.86e-02  7.34e-02
## age=57  9    3.975   3.4100  9.36e-02  1.29e-01
## age=58  8    2.045   4.4291  1.28e+00  1.93e+00
## age=59  8    2.349   4.4139  9.66e-01  1.49e+00
## age=60 11    5.071   4.5276  6.51e-02  9.62e-02
## age=61  5    3.237   1.2689  3.05e+00  3.71e+00
## age=62  7    3.356   2.7847  1.17e-01  1.67e-01
## age=63 11    3.705   5.4489  5.58e-01  8.71e-01
## age=64 11    3.575   4.9357  3.75e-01  5.83e-01
## age=65  8    4.476   2.3603  1.90e+00  2.76e+00
## age=66  7    3.351   2.8839  7.57e-02  1.09e-01
## age=67  8    4.217   2.4942  1.19e+00  1.51e+00
## age=68 10    3.877   5.7985  6.37e-01  1.05e+00
## age=69 11    3.872   5.1121  3.01e-01  4.47e-01
## age=70 10    3.914   4.8619  1.85e-01  2.74e-01
## age=71  7    2.920   3.4621  8.49e-02  1.19e-01
## age=72  7    4.318   2.7357  9.16e-01  1.34e+00
## age=73  6    4.870   1.1617  1.18e+01  1.38e+01
## age=74 10    4.080   4.6617  7.26e-02  1.14e-01
## age=75  5    1.825   2.3451  1.16e-01  1.64e-01
## age=76  5    4.047   0.9842  9.53e+00  1.12e+01
## age=77  2    0.000   1.0536  1.05e+00  1.54e+00
## age=80  2    1.003   1.0148  1.33e-04  1.82e-04
## age=81  1    0.996   0.0175  5.45e+01  5.53e+01
## age=82  1    0.956   0.0482  1.71e+01  1.76e+01
##
##  Chisq= 130  on 41 degrees of freedom, p= 4e-11
```


Strata
age_cat=a
age_cat=b
age_cat=c
Survival probability
1.00
0.75
0.50
0.25
0.00
0
250
500
750
1000
Time


Strata
age_cat=a
age_cat=b
age_cat=c
Survival probability
1.00
0.75
0.50
0.25
0.00
0
250
500
750
1000
Time



Czas przeżycia w każdej z metod jest najniższy dla trzeciej (najstarszej) kategorii wiekowej. Widzimy również dla pierwszej grupy wiekowej (w małym stopniu) najwwyższy czas przeżycia. Czas przeżycia zależy od grupy wiekowej

2 Lista 2

2.1 Zadanie 1

Parametryczne dopasowanie modelu przyspieszonego czasu przeżycia na podstawie rozkładu Weibulla za pomocą funkcji survreg. Przyjmujemy za zmienną zależną $time$**, a za charakterystki zmienne** $age, sex, ph.ecog, ph.karno$**.**

```
#zad1
lung %<>% as.data.table()


lung[ph.ecog == 3, ph.ecog.cat := 2]

x <- survreg(Surv(time, status)~as.factor(ph.ecog) + age + as.factor(sex) + as.factor(ph
summary(x)

##
## Call:
## survreg(formula = Surv(time, status) ~ as.factor(ph.ecog) + age +
##     as.factor(sex) + as.factor(ph.karno), data = lung, dist = "weibull")
##                          Value Std. Error     z       p
## (Intercept)            7.33464    0.59689 12.29 < 2e-16
## as.factor(ph.ecog)1   -0.31592    0.19501 -1.62 0.10523
```

```
## as.factor(ph.ecog)2    -0.78827    0.27866 -2.83 0.00467
## as.factor(ph.ecog)3    -1.50386    0.78468 -1.92 0.05530
## age                    -0.00868    0.00674 -1.29 0.19746
## as.factor(sex)2         0.42202    0.12230  3.45 0.00056
## as.factor(ph.karno)60  -0.45232    0.37928 -1.19 0.23303
## as.factor(ph.karno)70  -0.52667    0.35112 -1.50 0.13362
## as.factor(ph.karno)80  -0.64729    0.35046 -1.85 0.06475
## as.factor(ph.karno)90  -0.61515    0.36524 -1.68 0.09213
## as.factor(ph.karno)100 -0.62742    0.42367 -1.48 0.13863
## Log(scale)             -0.33709    0.06258 -5.39 7.2e-08
##
## Scale= 0.714
##
## Weibull distribution
## Loglik(model)= -1124.3   Loglik(intercept only)= -1141.1
##  Chisq= 33.42 on 10 degrees of freedom, p= 0.00023
## Number of Newton-Raphson Iterations: 6
## n=226 (2 observations deleted due to missingness)
```

2.2 Zadanie nr 2

Interpretacja współczynników modelu dopasowanego w zadaniu nr 1.

$$\hat{\ln} X = \hat{\mu} + \hat{\gamma}_1 age + \hat{\lambda}_i^{sex} + \hat{\lambda}_j^{ph.ecog} + \hat{\lambda}_k^{ph.karno} + \sigma * W$$

Znaczenia poszczególnych symboli:

- $\hat{\mu}$ - intercept, współczynnik zerowy
- $z = (wiek, sex, ph.ecog, ph.karno)$ - wektor charakterystyk zmiennych modelu
- $\hat{\gamma}_1$ - współczynnik charakterystyki zmiennej (typu number) *age*
- $\hat{\lambda}_{i,j,k}$ - wektory charakterystyk zmiennych (typu factor) z odpowiadającymi im wartościami dla poszczególnych podgrup.
- W - zmienna losowa rozkładu Weibulla

3 Lista 3.

3.1 Zadanie 1.

`Wyznaczymy szacowany rozkład czasu przeżycia dla kobiety w wieku 70 lat o charakterystyc`
`ph.ecog=1 i ph.karno=90.`

Na podstawie wyznaczonego rozkładu czasu przeżycia obliczamy prawdopodobieństwo, że 70-letnia kobieta o danej charakterystyce z zadania nr 1 przeżyje więcej niż 300 dni.

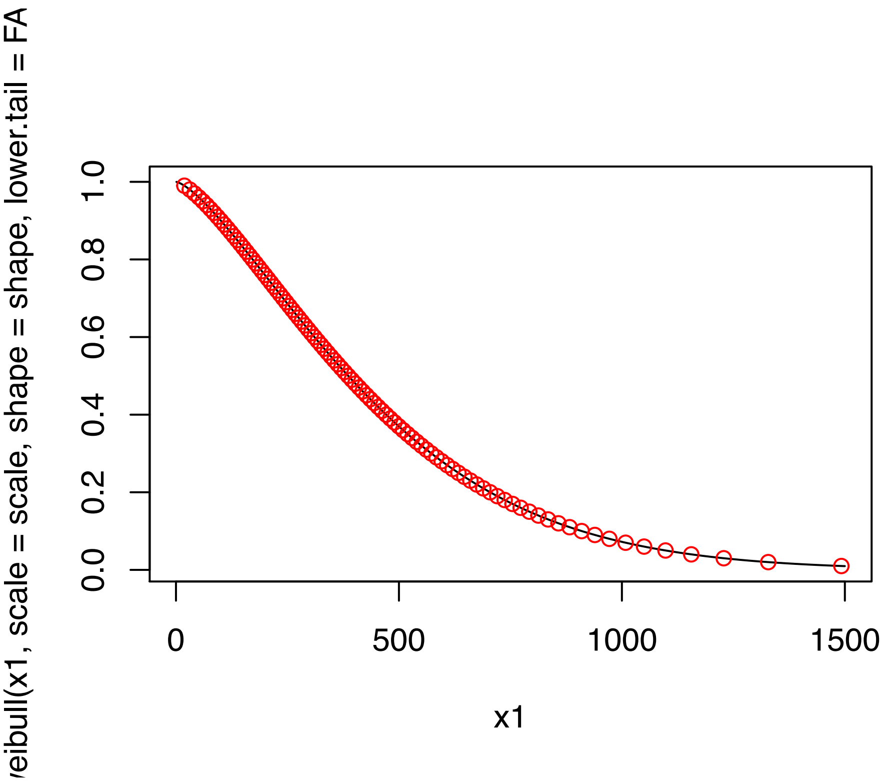
```r
model <- survreg(Surv(lung$time, status)~as.factor(ph.ecog) + age + as.factor(sex) + as.
summary(model)

##
## Call:
## survreg(formula = Surv(lung$time, status) ~ as.factor(ph.ecog) +
##     age + as.factor(sex) + as.factor(ph.karno), data = lung,
##     dist = "weibull")
##                            Value Std. Error     z       p
## (Intercept)              7.33464    0.59689 12.29 < 2e-16
## as.factor(ph.ecog)1     -0.31592    0.19501 -1.62 0.10523
## as.factor(ph.ecog)2     -0.78827    0.27866 -2.83 0.00467
## as.factor(ph.ecog)3     -1.50386    0.78468 -1.92 0.05530
## age                     -0.00868    0.00674 -1.29 0.19746
## as.factor(sex)2          0.42202    0.12230  3.45 0.00056
## as.factor(ph.karno)60   -0.45232    0.37928 -1.19 0.23303
## as.factor(ph.karno)70   -0.52667    0.35112 -1.50 0.13362
## as.factor(ph.karno)80   -0.64729    0.35046 -1.85 0.06475
## as.factor(ph.karno)90   -0.61515    0.36524 -1.68 0.09213
## as.factor(ph.karno)100  -0.62742    0.42367 -1.48 0.13863
## Log(scale)              -0.33709    0.06258 -5.39 7.2e-08
##
## Scale= 0.714
##
## Weibull distribution
## Loglik(model)= -1124.3   Loglik(intercept only)= -1141.1
##  Chisq= 33.42 on 10 degrees of freedom, p= 0.00023
## Number of Newton-Raphson Iterations: 6
## n=226 (2 observations deleted due to missingness)

pred <- predict(model, list(sex=2, age=70, ph.ecog=1, ph.karno=90), type="quantile", p =

shape <- 1/0.714
scale <- exp(7.33464 + (-0.31592) + (-0.61515) + (-0.00868)*70 + 0.42202)

pweibull(300, scale = scale, shape = shape, lower.tail = FALSE) #prawdo przezycia

## [1] 0.6146727
```

Prawdopodobieństwo wynosi około 61,5 procenta.

3.2 Zadanie 2.

Wizualizacja dystrybuanty i funkcji przeżycia z zadania nr 1.

```r
x1 <- seq(1,1500)
plot(x1, pweibull(x1, scale = scale, shape = shape, lower.tail = FALSE), type = "l")
lines(pred, 1 - seq(0.01,0.99,by=0.01), col = "red", type = "p")
```



3.3 Zadanie 3.

Zweryfikujmy hipotezę, czy zmienna wiek jest istotna (na poziomie istotności alpha =0.05) w modelu przyjętym powyżej. Zrobimy to za pomocą określenia p-value dla charakterystyki age.

```
summary(x)

##
## Call:
## survreg(formula = Surv(time, status) ~ as.factor(ph.ecog) + age +
##     as.factor(sex) + as.factor(ph.karno), data = lung, dist = "weibull")
##                             Value Std. Error     z       p
## (Intercept)               7.33464    0.59689 12.29 < 2e-16
## as.factor(ph.ecog)1      -0.31592    0.19501 -1.62 0.10523
## as.factor(ph.ecog)2      -0.78827    0.27866 -2.83 0.00467
## as.factor(ph.ecog)3      -1.50386    0.78468 -1.92 0.05530
## age                      -0.00868    0.00674 -1.29 0.19746
## as.factor(sex)2           0.42202    0.12230  3.45 0.00056
## as.factor(ph.karno)60    -0.45232    0.37928 -1.19 0.23303
## as.factor(ph.karno)70    -0.52667    0.35112 -1.50 0.13362
## as.factor(ph.karno)80    -0.64729    0.35046 -1.85 0.06475
## as.factor(ph.karno)90    -0.61515    0.36524 -1.68 0.09213
## as.factor(ph.karno)100   -0.62742    0.42367 -1.48 0.13863
## Log(scale)               -0.33709    0.06258 -5.39 7.2e-08
##
## Scale= 0.714
##
```

```
## Weibull distribution
## Loglik(model)= -1124.3   Loglik(intercept only)= -1141.1
##  Chisq= 33.42 on 10 degrees of freedom, p= 0.00023
## Number of Newton-Raphson Iterations: 6
## n=226 (2 observations deleted due to missingness)
```

Zauważmy, że p-value jest większe od przyjętego alpha, zatem możemy uznać, że zmienna age nie jest statystycznie istotna.

4 lista 4.

4.1 Zadanie 1.

```
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno) + as.factor(pat.karno) +
##     meal.cal + wt.loss, data = df, dist = "weibull")
##                              Value Std. Error     z       p
## (Intercept)               7.16e+00   1.04e+00  6.91 4.8e-12
## age                      -2.76e-03   8.10e-03 -0.34  0.7329
## as.factor(sex)2           4.26e-01   1.47e-01  2.91  0.0037
## as.factor(ph.ecog)1      -4.09e-01   2.36e-01 -1.73  0.0839
## as.factor(ph.ecog)2      -9.05e-01   3.64e-01 -2.48  0.0130
## as.factor(ph.ecog)3      -1.72e+00   7.89e-01 -2.19  0.0288
## as.factor(ph.karno)60    -7.27e-01   4.70e-01 -1.55  0.1220
## as.factor(ph.karno)70    -7.08e-01   4.41e-01 -1.61  0.1081
## as.factor(ph.karno)80    -8.37e-01   4.41e-01 -1.90  0.0578
## as.factor(ph.karno)90    -9.40e-01   4.53e-01 -2.07  0.0382
## as.factor(ph.karno)100   -1.02e+00   5.09e-01 -2.00  0.0454
## as.factor(pat.karno)40    1.80e-01   1.04e+00  0.17  0.8623
## as.factor(pat.karno)50   -6.73e-01   8.35e-01 -0.81  0.4201
## as.factor(pat.karno)60   -1.96e-01   7.09e-01 -0.28  0.7827
## as.factor(pat.karno)70    1.95e-02   7.31e-01  0.03  0.9787
## as.factor(pat.karno)80    9.09e-02   7.29e-01  0.12  0.9008
## as.factor(pat.karno)90   -4.44e-02   7.32e-01 -0.06  0.9516
## as.factor(pat.karno)100   2.75e-01   7.45e-01  0.37  0.7116
## meal.cal                  4.88e-05   1.94e-04  0.25  0.8016
## wt.loss                   9.27e-03   5.64e-03  1.64  0.1002
## Log(scale)               -3.84e-01   7.24e-02 -5.30 1.2e-07
##
## Scale= 0.681
##
## Weibull distribution
## Loglik(model)= -829.8   Loglik(intercept only)= -847.6
##  Chisq= 35.65 on 19 degrees of freedom, p= 0.012
## Number of Newton-Raphson Iterations: 7
```

```
## n=168 (60 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno) + as.factor(pat.karno) +
##     meal.cal + wt.loss, data = df, dist = "weibull")
##                             Value Std. Error     z       p
## (Intercept)              6.97e+00   8.85e-01  7.88 3.4e-15
## as.factor(sex)2          4.28e-01   1.46e-01  2.92  0.0035
## as.factor(ph.ecog)1     -4.04e-01   2.36e-01 -1.71  0.0865
## as.factor(ph.ecog)2     -9.13e-01   3.65e-01 -2.50  0.0125
## as.factor(ph.ecog)3     -1.73e+00   7.88e-01 -2.20  0.0278
## as.factor(ph.karno)60   -7.21e-01   4.69e-01 -1.54  0.1240
## as.factor(ph.karno)70   -6.97e-01   4.39e-01 -1.59  0.1124
## as.factor(ph.karno)80   -8.27e-01   4.41e-01 -1.88  0.0607
## as.factor(ph.karno)90   -9.28e-01   4.53e-01 -2.05  0.0406
## as.factor(ph.karno)100  -9.93e-01   5.05e-01 -1.97  0.0492
## as.factor(pat.karno)40   1.53e-01   1.03e+00  0.15  0.8820
## as.factor(pat.karno)50  -7.05e-01   8.30e-01 -0.85  0.3955
## as.factor(pat.karno)60  -2.09e-01   7.08e-01 -0.30  0.7673
## as.factor(pat.karno)70   5.46e-04   7.28e-01  0.00  0.9994
## as.factor(pat.karno)80   8.53e-02   7.29e-01  0.12  0.9068
## as.factor(pat.karno)90  -6.24e-02   7.30e-01 -0.09  0.9319
## as.factor(pat.karno)100  2.62e-01   7.44e-01  0.35  0.7244
## meal.cal                 5.88e-05   1.92e-04  0.31  0.7594
## wt.loss                  9.39e-03   5.61e-03  1.67  0.0940
## Log(scale)              -3.84e-01   7.23e-02 -5.31 1.1e-07
##
## Scale= 0.681
##
## Weibull distribution
## Loglik(model)= -829.9   Loglik(intercept only)= -847.6
##  Chisq= 35.53 on 18 degrees of freedom, p= 0.0081
## Number of Newton-Raphson Iterations: 6
## n=168 (60 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(ph.ecog) +
##     as.factor(ph.karno) + as.factor(pat.karno) + meal.cal + wt.loss,
##     data = df, dist = "weibull")
##                             Value Std. Error     z     p
## (Intercept)              7.31e+00   1.05e+00  6.98 3e-12
## age                     -4.03e-03   8.31e-03 -0.49 0.628
## as.factor(ph.ecog)1     -4.24e-01   2.41e-01 -1.76 0.078
## as.factor(ph.ecog)2     -8.18e-01   3.66e-01 -2.24 0.025
## as.factor(ph.ecog)3     -1.79e+00   8.03e-01 -2.23 0.026
## as.factor(ph.karno)60   -5.31e-01   4.73e-01 -1.12 0.261
## as.factor(ph.karno)70   -5.51e-01   4.40e-01 -1.25 0.210
```

```
## as.factor(ph.karno)80   -6.21e-01   4.44e-01 -1.40 0.162
## as.factor(ph.karno)90   -7.82e-01   4.57e-01 -1.71 0.087
## as.factor(ph.karno)100  -9.61e-01   5.19e-01 -1.85 0.064
## as.factor(pat.karno)40   3.70e-01   1.05e+00  0.35 0.725
## as.factor(pat.karno)50  -5.28e-01   8.53e-01 -0.62 0.536
## as.factor(pat.karno)60  -2.61e-01   7.20e-01 -0.36 0.717
## as.factor(pat.karno)70  -2.30e-02   7.42e-01 -0.03 0.975
## as.factor(pat.karno)80   1.58e-01   7.42e-01  0.21 0.831
## as.factor(pat.karno)90   1.41e-02   7.42e-01  0.02 0.985
## as.factor(pat.karno)100  3.42e-01   7.57e-01  0.45 0.652
## meal.cal                -3.75e-05   1.88e-04 -0.20 0.842
## wt.loss                  6.61e-03   5.47e-03  1.21 0.227
## Log(scale)              -3.68e-01   7.33e-02 -5.03 5e-07
##
## Scale= 0.692
##
## Weibull distribution
## Loglik(model)= -834.3   Loglik(intercept only)= -847.6
##  Chisq= 26.62 on 18 degrees of freedom, p= 0.086
## Number of Newton-Raphson Iterations: 7
## n=168 (60 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(sex) +
##     as.factor(ph.karno) + as.factor(pat.karno) + meal.cal + wt.loss,
##     data = df, dist = "weibull")
##                             Value Std. Error     z       p
## (Intercept)              6.39e+00   1.03e+00  6.17 6.8e-10
## age                     -3.06e-03   8.44e-03 -0.36  0.7170
## as.factor(sex)2          4.24e-01   1.52e-01  2.79  0.0052
## as.factor(ph.karno)60   -9.59e-01   4.84e-01 -1.98  0.0473
## as.factor(ph.karno)70   -7.92e-01   4.61e-01 -1.72  0.0860
## as.factor(ph.karno)80   -6.43e-01   4.45e-01 -1.44  0.1485
## as.factor(ph.karno)90   -5.89e-01   4.48e-01 -1.31  0.1886
## as.factor(ph.karno)100  -4.80e-01   4.66e-01 -1.03  0.3033
## as.factor(pat.karno)40   4.23e-01   1.07e+00  0.40  0.6926
## as.factor(pat.karno)50  -5.63e-01   8.63e-01 -0.65  0.5145
## as.factor(pat.karno)60  -1.52e-01   7.32e-01 -0.21  0.8360
## as.factor(pat.karno)70   2.13e-01   7.49e-01  0.28  0.7760
## as.factor(pat.karno)80   2.73e-01   7.42e-01  0.37  0.7129
## as.factor(pat.karno)90   2.36e-01   7.43e-01  0.32  0.7503
## as.factor(pat.karno)100  4.43e-01   7.59e-01  0.58  0.5594
## meal.cal                 6.53e-05   1.91e-04  0.34  0.7327
## wt.loss                  5.75e-03   5.54e-03  1.04  0.2988
## Log(scale)              -3.50e-01   7.18e-02 -4.87 1.1e-06
##
## Scale= 0.705
##
```

```
## Weibull distribution
## Loglik(model)= -839.7   Loglik(intercept only)= -854.1
##  Chisq= 28.82 on 16 degrees of freedom, p= 0.025
## Number of Newton-Raphson Iterations: 6
## n=169 (59 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(pat.karno) + meal.cal + wt.loss,
##     data = df, dist = "weibull")
##                             Value Std. Error     z       p
## (Intercept)              6.05e+00   9.04e-01  6.70 2.1e-11
## age                     -1.94e-04   8.07e-03 -0.02  0.9808
## as.factor(sex)2          3.90e-01   1.48e-01  2.63  0.0085
## as.factor(ph.ecog)1     -2.60e-01   1.75e-01 -1.49  0.1374
## as.factor(ph.ecog)2     -5.92e-01   2.62e-01 -2.26  0.0239
## as.factor(ph.ecog)3     -1.54e+00   7.43e-01 -2.07  0.0383
## as.factor(pat.karno)40   2.28e-01   1.02e+00  0.22  0.8226
## as.factor(pat.karno)50  -7.10e-01   8.26e-01 -0.86  0.3905
## as.factor(pat.karno)60  -2.21e-01   7.27e-01 -0.30  0.7611
## as.factor(pat.karno)70   1.25e-01   7.40e-01  0.17  0.8662
## as.factor(pat.karno)80   1.77e-01   7.42e-01  0.24  0.8115
## as.factor(pat.karno)90  -8.12e-03   7.49e-01 -0.01  0.9914
## as.factor(pat.karno)100  2.76e-01   7.59e-01  0.36  0.7160
## meal.cal                -1.06e-05   1.86e-04 -0.06  0.9547
## wt.loss                  7.87e-03   5.64e-03  1.39  0.1631
## Log(scale)              -3.53e-01   7.16e-02 -4.93 8.0e-07
##
## Scale= 0.702
##
## Weibull distribution
## Loglik(model)= -832.6   Loglik(intercept only)= -847.6
##  Chisq= 30 on 14 degrees of freedom, p= 0.0076
## Number of Newton-Raphson Iterations: 6
## n=168 (60 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno) + meal.cal + wt.loss,
##     data = df, dist = "weibull")
##                            Value Std. Error     z       p
## (Intercept)             7.36e+00   7.77e-01  9.47 < 2e-16
## age                    -5.99e-03   8.11e-03 -0.74  0.4600
## as.factor(sex)2         4.45e-01   1.44e-01  3.09  0.0020
## as.factor(ph.ecog)1    -2.93e-01   2.29e-01 -1.28  0.2002
## as.factor(ph.ecog)2    -9.25e-01   3.38e-01 -2.74  0.0061
## as.factor(ph.ecog)3    -1.61e+00   7.96e-01 -2.02  0.0433
## as.factor(ph.karno)60  -8.28e-01   4.68e-01 -1.77  0.0768
```

```
## as.factor(ph.karno)70  -8.43e-01   4.39e-01 -1.92  0.0552
## as.factor(ph.karno)80  -9.51e-01   4.40e-01 -2.16  0.0308
## as.factor(ph.karno)90  -1.00e+00   4.53e-01 -2.21  0.0270
## as.factor(ph.karno)100 -9.74e-01   5.10e-01 -1.91  0.0562
## meal.cal                7.36e-05   1.87e-04  0.39  0.6943
## wt.loss                 9.35e-03   5.50e-03  1.70  0.0889
## Log(scale)             -3.56e-01   7.22e-02 -4.93 8.3e-07
##
## Scale= 0.701
##
## Weibull distribution
## Loglik(model)= -845.9   Loglik(intercept only)= -861
##  Chisq= 30.14 on 12 degrees of freedom, p= 0.0027
## Number of Newton-Raphson Iterations: 6
## n=170 (58 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno) + as.factor(pat.karno) +
##     wt.loss, data = df, dist = "weibull")
##                            Value Std. Error     z       p
## (Intercept)              7.39310    0.95507  7.74 9.9e-15
## age                     -0.00812    0.00700 -1.16 0.24608
## as.factor(sex)2          0.44675    0.12468  3.58 0.00034
## as.factor(ph.ecog)1     -0.34608    0.20075 -1.72 0.08472
## as.factor(ph.ecog)2     -0.72571    0.30988 -2.34 0.01919
## as.factor(ph.ecog)3     -1.67653    0.76235 -2.20 0.02787
## as.factor(ph.karno)60   -0.64588    0.45117 -1.43 0.15227
## as.factor(ph.karno)70   -0.77876    0.42375 -1.84 0.06609
## as.factor(ph.karno)80   -0.97830    0.42231 -2.32 0.02053
## as.factor(ph.karno)90   -0.87690    0.43285 -2.03 0.04278
## as.factor(ph.karno)100  -0.94786    0.47721 -1.99 0.04701
## as.factor(pat.karno)40   0.06734    1.00594  0.07 0.94663
## as.factor(pat.karno)50  -0.71656    0.81435 -0.88 0.37890
## as.factor(pat.karno)60  -0.12311    0.69351 -0.18 0.85910
## as.factor(pat.karno)70   0.08103    0.70542  0.11 0.90856
## as.factor(pat.karno)80   0.14374    0.70434  0.20 0.83829
## as.factor(pat.karno)90   0.20208    0.70895  0.29 0.77561
## as.factor(pat.karno)100  0.30040    0.71699  0.42 0.67523
## wt.loss                  0.00939    0.00496  1.89 0.05837
## Log(scale)              -0.40134    0.06514 -6.16 7.2e-10
##
## Scale= 0.669
##
## Weibull distribution
## Loglik(model)= -1020.4   Loglik(intercept only)= -1042.6
##  Chisq= 44.52 on 18 degrees of freedom, p= 0.00049
## Number of Newton-Raphson Iterations: 6
```

```
## n=210 (18 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno) + meal.cal + as.factor(pat.karno),
##     data = df, dist = "weibull")
##                             Value Std. Error     z       p
## (Intercept)              6.96e+00   1.05e+00  6.64 3.2e-11
## age                     -3.65e-03   8.11e-03 -0.45  0.6527
## as.factor(sex)2          4.31e-01   1.48e-01  2.91  0.0036
## as.factor(ph.ecog)1     -3.85e-01   2.32e-01 -1.66  0.0973
## as.factor(ph.ecog)2     -7.57e-01   3.53e-01 -2.14  0.0321
## as.factor(ph.ecog)3     -1.53e+00   8.07e-01 -1.90  0.0574
## as.factor(ph.karno)60   -6.31e-01   4.32e-01 -1.46  0.1438
## as.factor(ph.karno)70   -5.54e-01   4.07e-01 -1.36  0.1733
## as.factor(ph.karno)80   -5.88e-01   4.04e-01 -1.45  0.1458
## as.factor(ph.karno)90   -7.87e-01   4.17e-01 -1.89  0.0588
## as.factor(ph.karno)100  -8.08e-01   4.83e-01 -1.67  0.0941
## meal.cal                 8.97e-05   1.86e-04  0.48  0.6296
## as.factor(pat.karno)40   5.32e-01   1.06e+00  0.50  0.6150
## as.factor(pat.karno)50  -7.81e-01   8.40e-01 -0.93  0.3523
## as.factor(pat.karno)60  -8.27e-02   7.35e-01 -0.11  0.9104
## as.factor(pat.karno)70   1.30e-01   7.56e-01  0.17  0.8638
## as.factor(pat.karno)80   1.12e-01   7.55e-01  0.15  0.8826
## as.factor(pat.karno)90  -9.03e-03   7.59e-01 -0.01  0.9905
## as.factor(pat.karno)100  2.41e-01   7.70e-01  0.31  0.7539
## Log(scale)              -3.43e-01   6.99e-02 -4.90 9.4e-07
##
## Scale= 0.71
##
## Weibull distribution
## Loglik(model)= -896   Loglik(intercept only)= -912.6
##  Chisq= 33.21 on 18 degrees of freedom, p= 0.016
## Number of Newton-Raphson Iterations: 7
## n=178 (50 observations deleted due to missingness)
## [1] 1
## [1] 0.002699796
## [1] 0.006522388
## [1] 0.01796048
## [1] 0.02534732
## [1] 1
## [1] 0.1068637
```

Na podstawie testu wiarygodności i twierdzeniu Wilksa możemy stwierdzić, że zmienna age i zmienna meal.cal nie jest istotna statystycznie

```
##
## Call:
```

```
## survreg(formula = Surv(time, status == 2) ~ as.factor(ph.ecog) +
##     as.factor(ph.karno) + as.factor(pat.karno) + wt.loss, data = df,
##     dist = "weibull")
##                             Value Std. Error     z       p
## (Intercept)               6.84840    0.85633  8.00 1.3e-15
## as.factor(ph.ecog)1      -0.30837    0.20468 -1.51   0.132
## as.factor(ph.ecog)2      -0.63283    0.30948 -2.04   0.041
## as.factor(ph.ecog)3      -1.74848    0.77914 -2.24   0.025
## as.factor(ph.karno)60    -0.48107    0.45681 -1.05   0.292
## as.factor(ph.karno)70    -0.56916    0.42455 -1.34   0.180
## as.factor(ph.karno)80    -0.71188    0.42756 -1.66   0.096
## as.factor(ph.karno)90    -0.62893    0.43910 -1.43   0.152
## as.factor(ph.karno)100   -0.73749    0.48658 -1.52   0.130
## as.factor(pat.karno)40    0.24087    1.02155  0.24   0.814
## as.factor(pat.karno)50   -0.64220    0.83169 -0.77   0.440
## as.factor(pat.karno)60   -0.24145    0.70758 -0.34   0.733
## as.factor(pat.karno)70   -0.00643    0.71839 -0.01   0.993
## as.factor(pat.karno)80    0.12134    0.71989  0.17   0.866
## as.factor(pat.karno)90    0.17900    0.72171  0.25   0.804
## as.factor(pat.karno)100   0.27243    0.73146  0.37   0.710
## wt.loss                   0.00791    0.00491  1.61   0.107
## Log(scale)               -0.38023    0.06568 -5.79 7.1e-09
##
## Scale= 0.684
##
## Weibull distribution
## Loglik(model)= -1028   Loglik(intercept only)= -1042.6
##  Chisq= 29.29 on 16 degrees of freedom, p= 0.022
## Number of Newton-Raphson Iterations: 6
## n=210 (18 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ as.factor(sex) +
##     as.factor(ph.karno) + as.factor(pat.karno) + wt.loss, data = df,
##     dist = "weibull")
##                             Value Std. Error     z       p
## (Intercept)               6.18232    0.81185  7.62 2.6e-14
## as.factor(sex)2           0.42918    0.12704  3.38 0.00073
## as.factor(ph.karno)60    -0.79774    0.44682 -1.79 0.07420
## as.factor(ph.karno)70    -0.76156    0.43195 -1.76 0.07789
## as.factor(ph.karno)80    -0.73058    0.41862 -1.75 0.08095
## as.factor(ph.karno)90    -0.49458    0.42353 -1.17 0.24290
## as.factor(ph.karno)100   -0.40397    0.44056 -0.92 0.35916
## as.factor(pat.karno)40    0.21006    1.02514  0.20 0.83764
## as.factor(pat.karno)50   -0.67601    0.83053 -0.81 0.41567
## as.factor(pat.karno)60   -0.10159    0.71112 -0.14 0.88640
## as.factor(pat.karno)70    0.19983    0.71974  0.28 0.78129
## as.factor(pat.karno)80    0.33420    0.71308  0.47 0.63931
```

```
## as.factor(pat.karno)90   0.38735    0.71766  0.54 0.58937
## as.factor(pat.karno)100  0.42926    0.72829  0.59 0.55559
## wt.loss                  0.00676    0.00484  1.40 0.16263
## Log(scale)              -0.37315    0.06482 -5.76 8.6e-09
##
## Scale= 0.689
##
## Weibull distribution
## Loglik(model)= -1031.2   Loglik(intercept only)= -1049.2
##  Chisq= 35.97 on 14 degrees of freedom, p= 0.0011
## Number of Newton-Raphson Iterations: 5
## n=211 (17 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(pat.karno) + wt.loss, data = df,
##     dist = "weibull")
##                            Value Std. Error     z       p
## (Intercept)              5.94406    0.72426  8.21 2.3e-16
## as.factor(sex)2          0.40738    0.12754  3.19  0.0014
## as.factor(ph.ecog)1     -0.32362    0.15244 -2.12  0.0338
## as.factor(ph.ecog)2     -0.51690    0.21471 -2.41  0.0161
## as.factor(ph.ecog)3     -1.48794    0.71868 -2.07  0.0384
## as.factor(pat.karno)40   0.23429    0.98812  0.24  0.8126
## as.factor(pat.karno)50  -0.70450    0.80188 -0.88  0.3796
## as.factor(pat.karno)60  -0.16550    0.71065 -0.23  0.8158
## as.factor(pat.karno)70   0.16336    0.71533  0.23  0.8194
## as.factor(pat.karno)80   0.28683    0.71734  0.40  0.6893
## as.factor(pat.karno)90   0.24176    0.72288  0.33  0.7380
## as.factor(pat.karno)100  0.33677    0.73030  0.46  0.6447
## wt.loss                  0.00756    0.00489  1.54  0.1224
## Log(scale)              -0.37031    0.06442 -5.75 9.0e-09
##
## Scale= 0.691
##
## Weibull distribution
## Loglik(model)= -1024.3   Loglik(intercept only)= -1042.6
##  Chisq= 36.6 on 12 degrees of freedom, p= 0.00026
## Number of Newton-Raphson Iterations: 5
## n=210 (18 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno) + wt.loss, data = df,
##     dist = "weibull")
##                           Value Std. Error     z       p
## (Intercept)             6.87208    0.41977 16.37 < 2e-16
## as.factor(sex)2         0.45234    0.12448  3.63 0.00028
```

```
## as.factor(ph.ecog)1    -0.29636    0.19909 -1.49 0.13660
## as.factor(ph.ecog)2    -0.93112    0.29847 -3.12 0.00181
## as.factor(ph.ecog)3    -1.78684    0.77579 -2.30 0.02126
## as.factor(ph.karno)60  -0.47246    0.40399 -1.17 0.24221
## as.factor(ph.karno)70  -0.63488    0.37480 -1.69 0.09028
## as.factor(ph.karno)80  -0.86311    0.38203 -2.26 0.02387
## as.factor(ph.karno)90  -0.71081    0.39495 -1.80 0.07190
## as.factor(ph.karno)100 -0.70384    0.44563 -1.58 0.11424
## wt.loss                 0.00790    0.00468  1.69 0.09149
## Log(scale)             -0.36706    0.06454 -5.69 1.3e-08
##
## Scale= 0.693
##
## Weibull distribution
## Loglik(model)= -1044.7   Loglik(intercept only)= -1062.6
##  Chisq= 35.79 on 10 degrees of freedom, p= 9.2e-05
## Number of Newton-Raphson Iterations: 5
## n=213 (15 observations deleted due to missingness)
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno) + as.factor(pat.karno),
##     data = df, dist = "weibull")
##                           Value Std. Error     z       p
## (Intercept)              6.6722     0.8476  7.87 3.5e-15
## as.factor(sex)2          0.4219     0.1241  3.40 0.00068
## as.factor(ph.ecog)1     -0.3359     0.1985 -1.69 0.09055
## as.factor(ph.ecog)2     -0.6298     0.3013 -2.09 0.03660
## as.factor(ph.ecog)3     -1.5437     0.7778 -1.98 0.04717
## as.factor(ph.karno)60   -0.5083     0.4054 -1.25 0.20992
## as.factor(ph.karno)70   -0.5672     0.3847 -1.47 0.14033
## as.factor(ph.karno)80   -0.6652     0.3864 -1.72 0.08514
## as.factor(ph.karno)90   -0.6668     0.3993 -1.67 0.09493
## as.factor(ph.karno)100  -0.6871     0.4487 -1.53 0.12572
## as.factor(pat.karno)40   0.3317     1.0197  0.33 0.74498
## as.factor(pat.karno)50  -0.8788     0.8122 -1.08 0.27927
## as.factor(pat.karno)60  -0.0694     0.7169 -0.10 0.92286
## as.factor(pat.karno)70   0.1504     0.7287  0.21 0.83649
## as.factor(pat.karno)80   0.1753     0.7307  0.24 0.81041
## as.factor(pat.karno)90   0.2054     0.7349  0.28 0.77987
## as.factor(pat.karno)100  0.2662     0.7416  0.36 0.71960
## Log(scale)              -0.3619     0.0627 -5.77 7.9e-09
##
## Scale= 0.696
##
## Weibull distribution
## Loglik(model)= -1101.8   Loglik(intercept only)= -1121.3
##  Chisq= 39.03 on 16 degrees of freedom, p= 0.0011
```

```
## Number of Newton-Raphson Iterations: 6
## n=223 (5 observations deleted due to missingness)
## [1] 0.002699796
## [1] 0.006522388
## [1] 0.02013675
## [1] 0.02534732
## [1] 0.09426431
```

Na podstawie testu ilorazu wiarygodności możemy przyjąć, że przyjęty model1 jest odpowiedni i jest zależny od zmiennych: sex, ph.ecog, ph.karno, pat.karno oraz wt.loss

4.2 Zadanie 3.

Korzystając z kryterium informacyjnego Akaike'a AIC, dokonujemy wyboru najlepszego modelu liniowego logarytmu czasu.

```
WparametersALL <- survreg(Surv(time, status==2)~age+as.factor(sex)+as.factor(ph.ecog)+as
                        dist = "weibull")
step(WparametersALL)

## Start:  AIC=1701.59
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     as.factor(ph.karno) + as.factor(pat.karno) + meal.cal + wt.loss

## Error in drop1.default(fit, scope$drop, scale = scale, trace = trace, : number
of rows in use has changed: remove missing values?
```

Według funkcji step, optymalnym wyborem modelu liniowego logarytmu jest model oparty o charakterystki: sex, ph.ecog i wt.loss

4.3 Zadanie. 4

Korzystając z kryterium informacyjnego bayesowskigo BIC, dokonujemy wyboru najlepszego modelu liniowego logarytmu czasu.

```
n = length(df$status==1)
step(WparametersALL, k = log(n))

## Start:  AIC=1773.61
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     as.factor(ph.karno) + as.factor(pat.karno) + meal.cal + wt.loss

## Error in drop1.default(fit, scope$drop, scale = scale, trace = trace, : number
of rows in use has changed: remove missing values?
```

Według funkcji step, optymalnym wyborem modelu liniowego logarytmu jest model oparty jedynie o charakterystykę sex.