Analiza Przeżycia Raport 3

Paweł Matławski
album 249732

25 lutego 2021

Spis treści

1 Lista nr 1

```r
library(survival)
library(ggplot2)
library(survminer)

## Loading required package: ggpubr
## Loading required package: magrittr

library(coin)
library(car)

## Loading required package: carData

library(dplyr)

##
## Attaching package: 'dplyr'
## The following object is masked from 'package:car':
##
##     recode
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

df <- lung
df$sex <- as.factor(df$sex)
df$ph.ecog <- as.factor(df$ph.ecog)
df$ph.karno <- as.factor(df$ph.karno)
df$status <- as.factor(df$status)
df$pat.karno <- as.factor(df$pat.karno)
df <- na.omit(df)
attach(df)
```

1.1 Zadanie nr 1

Wykonując odpowiednie testy, chcemy zweryfikować hipotezę o równości rozkładów czasu przeżycia w grupie kobiet i mężczyzn, na poziomie istotności $\alpha = 0.05$.

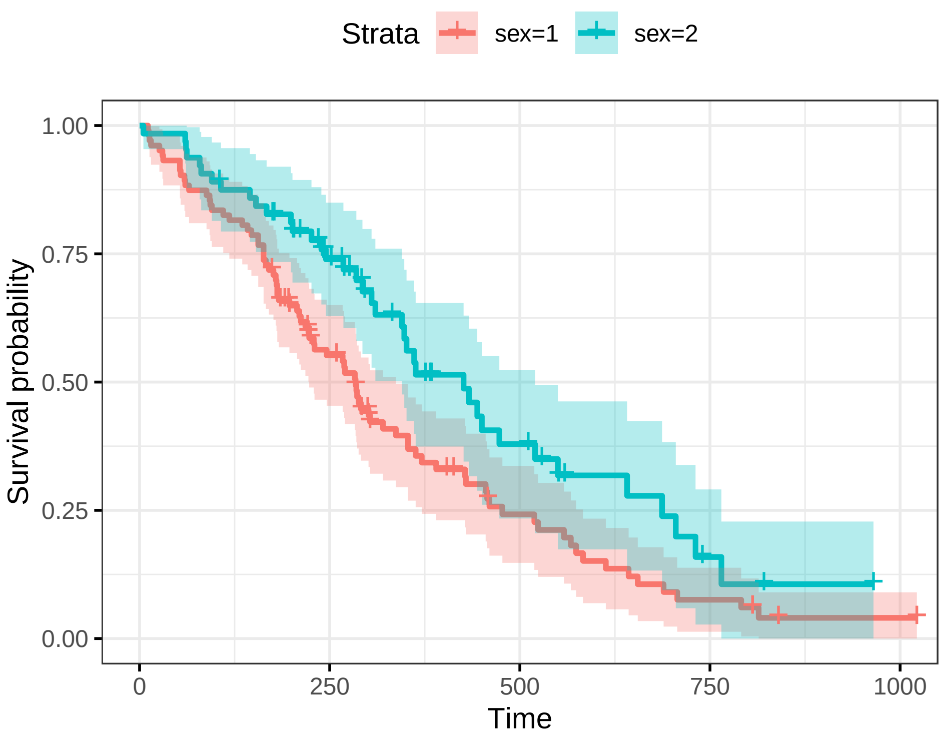
```r
#log-rank
logrank <- survdiff(Surv(time, status==2)~sex, data = df)
#Peto&Peto
petopeto <- survdiff(Surv(time, status==2)~sex, data = df, rho = 1)
#rho=0.5
test05 <- survdiff(Surv(time, status==2)~sex, data = df, rho = 0.5)
```

```
#rho=0.1
test01 <- survdiff(Surv(time, status==2)~sex, data = df, rho = 0.1)
#rho=5
test5 <- survdiff(Surv(time, status==2)~sex, data = df, rho = 5)

logrank

## Call:
## survdiff(formula = Surv(time, status == 2) ~ sex, data = df)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 103       82     68.7      2.57      6.05
## sex=2  64       38     51.3      3.44      6.05
##
##  Chisq= 6  on 1 degrees of freedom, p= 0.01

petopeto

## Call:
## survdiff(formula = Surv(time, status == 2) ~ sex, data = df,
##     rho = 1)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 103     51.0     41.8      2.01      6.76
## sex=2  64     21.1     30.3      2.77      6.76
##
##  Chisq= 6.8  on 1 degrees of freedom, p= 0.009

test05

## Call:
## survdiff(formula = Surv(time, status == 2) ~ sex, data = df,
##     rho = 0.5)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 103     62.8     52.0      2.26      6.72
## sex=2  64     27.4     38.3      3.07      6.72
##
##  Chisq= 6.7  on 1 degrees of freedom, p= 0.01

test01

## Call:
## survdiff(formula = Surv(time, status == 2) ~ sex, data = df,
##     rho = 0.1)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 103     77.3     64.6      2.50      6.26
## sex=2  64     35.4     48.1      3.36      6.26
##
##  Chisq= 6.3  on 1 degrees of freedom, p= 0.01
```

```
test5

## Call:
## survdiff(formula = Surv(time, status == 2) ~ sex, data = df,
##     rho = 5)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=1 103     20.6     16.6     0.952      4.21
## sex=2  64      7.2     11.2     1.413      4.21
##
##  Chisq= 4.2  on 1 degrees of freedom, p= 0.04
```

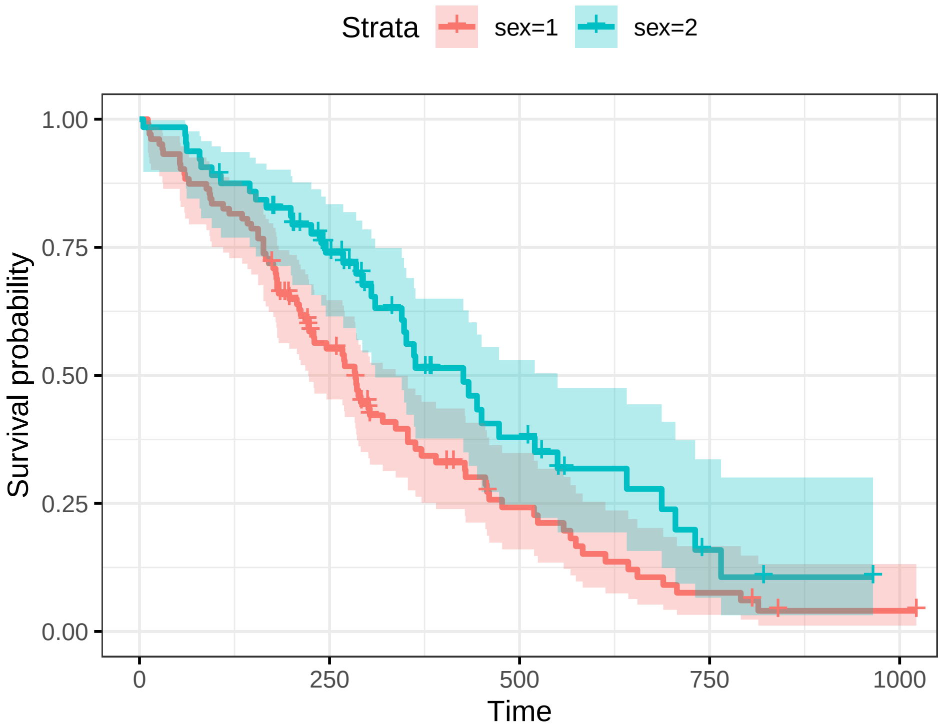
Na podstawie powyższych testów i ich p-value możemy jednoznacznie odrzucić naszą hipotezę o równości czasu przeżycia dla kobiet i mężczyzn. Zwizualizujmy funkcję przeżycia ze względu na płeć, aby dostrzec różnice.

```
fit1 <- survfit(Surv(time, status==2)~sex, data = df, conf.type=c('plain'))
confint1 <- ggsurvplot(fit1, conf.int=TRUE, ggtheme=theme_bw())
confint1
```



Rysunek 1: Przedziały ufności estymacji Kaplana-Meiera typu plain z podziałem ze względu na płeć

```
fit2 <- survfit(Surv(time, status==2)~sex, data = df, conf.type=c('logit'))
confint2 <- ggsurvplot(fit2, conf.int=TRUE, ggtheme=theme_bw())
confint2
```



Rysunek 2: Przedziały ufności estymacji Kaplana-Meiera typu logit z podziałem ze względu na płeć

1.2 Zadanie nr 2

Dokonując kategoryzacji zweryfikujemy hipotezę o równości czasu przeżycia w podgrupach ze względu na wiek. Podzieliłem pacjentów na cztery podgrupy ze względu na wiek: $(30, 50), [50, 70), [70, 90)$.

```
age2 <- cut(age, breaks=c(30, 50, 70, 90), right = FALSE)
#log-rank
logrank1 <- survdiff(Surv(time, status==2)~age2, data = df)
#Peto&Peto
petopeto1 <- survdiff(Surv(time, status==2)~age2, data = df, rho = 1)
#rho=0.5
test051 <- survdiff(Surv(time, status==2)~age2, data = df, rho = 0.5)
#rho=0.1
test011 <- survdiff(Surv(time, status==2)~age2, data = df, rho = 0.1)
#rho=5
test51 <- survdiff(Surv(time, status==2)~age2, data = df, rho = 5)

logrank1

## Call:
## survdiff(formula = Surv(time, status == 2) ~ age2, data = df)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## age2=[30,50)  16        9     11.7     0.618     0.692
## age2=[50,70) 107       76     81.5     0.368     1.156
## age2=[70,90)  44       35     26.8     2.483     3.239
##
##  Chisq= 3.5  on 2 degrees of freedom, p= 0.2
```

```
petopeto1
```

```
## Call:
## survdiff(formula = Surv(time, status == 2) ~ age2, data = df,
##     rho = 1)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## age2=[30,50)  16     5.85     7.12     0.226     0.346
## age2=[50,70) 107    43.52    47.96     0.411     1.711
## age2=[70,90)  44    22.75    17.04     1.913     3.480
##
##  Chisq= 3.5  on 2 degrees of freedom, p= 0.2
```

```
test051
```

```
## Call:
## survdiff(formula = Surv(time, status == 2) ~ age2, data = df,
##     rho = 0.5)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## age2=[30,50)  16     7.22     8.82     0.292     0.404
## age2=[50,70) 107    55.43    60.53     0.430     1.628
## age2=[70,90)  44    27.61    20.91     2.151     3.475
##
##  Chisq= 3.6  on 2 degrees of freedom, p= 0.2
```

```
test011
```

```
## Call:
## survdiff(formula = Surv(time, status == 2) ~ age2, data = df,
##     rho = 0.1)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## age2=[30,50)  16      8.6     11.0     0.512     0.607
## age2=[50,70) 107     70.8     76.3     0.391     1.292
## age2=[70,90)  44     33.2     25.4     2.413     3.325
##
##  Chisq= 3.5  on 2 degrees of freedom, p= 0.2
```

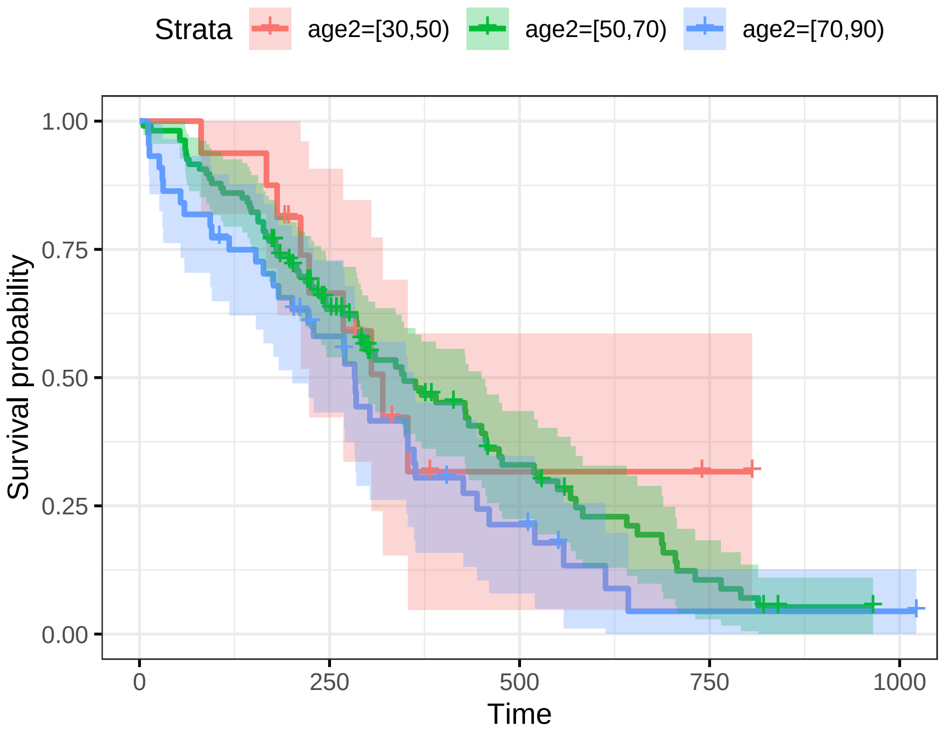
```
test51
```

```
## Call:
## survdiff(formula = Surv(time, status == 2) ~ age2, data = df,
```

```
##     rho = 5)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## age2=[30,50)  16     1.53     2.86     0.618      1.23
## age2=[50,70) 107    15.94    18.16     0.271      1.38
## age2=[70,90)  44    10.29     6.74     1.865      4.32
##
##  Chisq= 4.8  on 2 degrees of freedom, p= 0.09
```

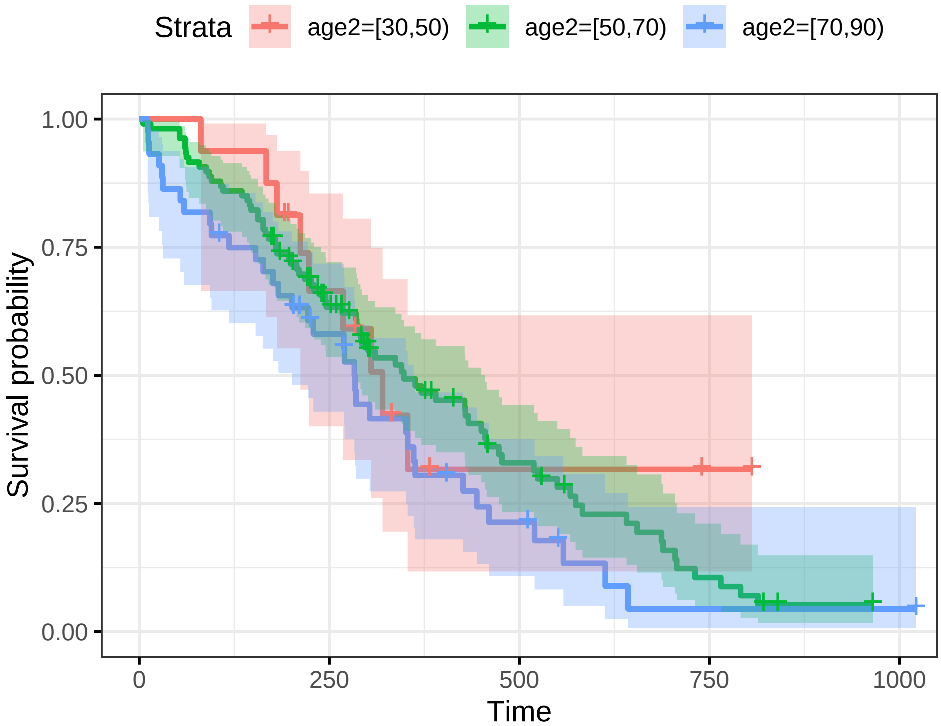
Analizując powyższe testy, należy stwierdzić że nie mamy podstaw do odrzucenia hipotezy o równości czasu przeżycia w podgrupach (na poziomie istotności 0.05). Zwizualizujemy funkcję przeżycia w przyjętych podgrupach, aby to zauważyć.

```
fit3 <- survfit(Surv(time, status==2)~age2, data = df, conf.type=c('plain'))
confint3 <- ggsurvplot(fit3, conf.int=TRUE, ggtheme=theme_bw())
confint3
```



Rysunek 3: Przedziały ufności estymacji Kaplana-Meiera typu plain z podziałem ze względu na wiek

```
fit4 <- survfit(Surv(time, status==2)~age2, data = df, conf.type=c('logit'))
confint4 <- ggsurvplot(fit4, conf.int=TRUE, ggtheme=theme_bw())
confint4
```



Rysunek 4: Przedziały ufności estymacji Kaplana-Meiera typu logit z podziałem ze względu na wiek

2 Lista nr 2

2.1 Zadanie nr 1

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

```
Wparameters <- survreg(Surv(time, status==2)~age+as.factor(sex)+as.factor(ph.ecog)+as.fa
summary(Wparameters)
```

```
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno), data = df, dist = "weibull")
##                          Value Std. Error     z       p
## (Intercept)            7.52926    0.74191 10.15 < 2e-16
## age                   -0.00628    0.00811 -0.77  0.4384
## as.factor(sex)2        0.40246    0.14097  2.85  0.0043
## as.factor(ph.ecog)1   -0.33735    0.23252 -1.45  0.1468
## as.factor(ph.ecog)2   -0.86360    0.33415 -2.58  0.0098
## as.factor(ph.ecog)3   -1.55559    0.79287 -1.96  0.0498
## as.factor(ph.karno)60 -0.76312    0.46078 -1.66  0.0977
```

```
## as.factor(ph.karno)70  -0.83379    0.43870 -1.90  0.0574
## as.factor(ph.karno)80  -0.85848    0.43987 -1.95  0.0510
## as.factor(ph.karno)90  -0.94072    0.45375 -2.07  0.0382
## as.factor(ph.karno)100 -1.02644    0.51353 -2.00  0.0456
## Log(scale)             -0.35680    0.07283 -4.90 9.6e-07
##
## Scale= 0.7
##
## Weibull distribution
## Loglik(model)= -827.4   Loglik(intercept only)= -841.1
##  Chisq= 27.25 on 10 degrees of freedom, p= 0.0024
## Number of Newton-Raphson Iterations: 6
## n= 167
```

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
- $z = (age, sex, ph.ecog, ph.karno)$ - 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 nr 3

3.1 Zadanie nr 1

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

```
Wseq <- seq(0.01, 0.99, by = 0.01)
Wpred <- predict(Wparameters, list(sex = 2, age = 70, ph.ecog = 1, ph.karno = 90), type
Wpred
```

```
##  [1]   19.96737   32.55022   43.38647   53.25566   62.48465   71.25023
##  [7]   79.66184   87.79326   95.69733  103.41354  110.97249  118.39853
## [13]  125.71150  132.92785  140.06149  147.12434  154.12671  161.07770
## [19]  167.98534  174.85686  181.69877  188.51705  195.31715  202.10416
## [25]  208.88280  215.65752  222.43254  229.21185  235.99930  242.79859
```

```
## [31]   249.61329  256.44691  263.30285  270.18448  277.09514  284.03812
## [37]   291.01673  298.03428  305.09408  312.19950  319.35394  326.56088
## [43]   333.82383  341.14643  348.53240  355.98555  363.50985  371.10939
## [49]   378.78843  386.55140  394.40293  402.34784  410.39121  418.53838
## [55]   426.79496  435.16688  443.66042  452.28222  461.03937  469.93939
## [61]   478.99034  488.20082  497.58007  507.13802  516.88538  526.83371
## [67]   536.99557  547.38456  558.01555  568.90477  580.07001  591.53087
## [73]   603.30897  615.42829  627.91553  640.80050  654.11670  667.90197
## [79]   682.19920  697.05740  712.53287  728.69076  745.60709  763.37129
## [85]   782.08963  801.88980  822.92709  845.39315  869.52839  895.64024
## [91]   924.13101  955.54181  990.62564 1030.47628 1076.77374 1132.30262
## [97]  1202.21717 1297.90507 1454.88132
```

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

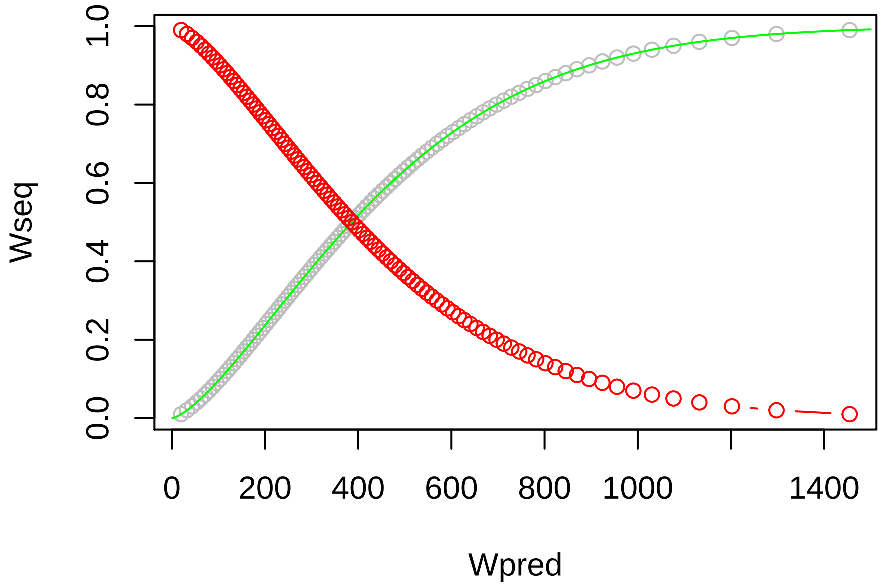
```
Wshape = 1/Wparameters[["scale"]]
Wscale = exp(Wparameters[["coefficients"]][["(Intercept)"]]+
            Wparameters[["coefficients"]][["as.factor(sex)2"]]+
            Wparameters[["coefficients"]][["as.factor(ph.ecog)1"]]+
            Wparameters[["coefficients"]][["as.factor(ph.karno)90"]]+
            70*Wparameters[["coefficients"]][["age"]])
Wprob=length(which(Wpred>300))/length(Wpred)
Wprob

## [1] 0.6161616

pweibull(300, scale = Wscale, shape = Wshape, lower.tail = FALSE)

## [1] 0.6172094
```

Prawdopodobieństwo wynosi około 61,5 procenta.

3.2 Zadanie nr 2

Zwizualizujmy dystrybuantę i funkcję przeżycia z zadania nr 1.

```
x=seq(1, 1500)
plot(Wpred, Wseq, col="gray", type="p")
lines(x, pweibull(x, shape = Wshape, scale = Wscale), col = "green")
lines(Wpred, 1-Wseq, type = "b", col="red")
```



Rysunek 5: Dystrybuanta i funkcja przeżycia rozkładu z zadania nr 1

3.3 Zadanie nr 3

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

```
summary(Wparameters)

##
## Call:
## survreg(formula = Surv(time, status == 2) ~ age + as.factor(sex) +
##     as.factor(ph.ecog) + as.factor(ph.karno), data = df, dist = "weibull")
##                               Value Std. Error     z       p
## (Intercept)                 7.52926    0.74191 10.15 < 2e-16
## age                        -0.00628    0.00811 -0.77  0.4384
## as.factor(sex)2             0.40246    0.14097  2.85  0.0043
## as.factor(ph.ecog)1        -0.33735    0.23252 -1.45  0.1468
## as.factor(ph.ecog)2        -0.86360    0.33415 -2.58  0.0098
## as.factor(ph.ecog)3        -1.55559    0.79287 -1.96  0.0498
## as.factor(ph.karno)60      -0.76312    0.46078 -1.66  0.0977
## as.factor(ph.karno)70      -0.83379    0.43870 -1.90  0.0574
## as.factor(ph.karno)80      -0.85848    0.43987 -1.95  0.0510
## as.factor(ph.karno)90      -0.94072    0.45375 -2.07  0.0382
## as.factor(ph.karno)100     -1.02644    0.51353 -2.00  0.0456
## Log(scale)                 -0.35680    0.07283 -4.90 9.6e-07
##
## Scale= 0.7
##
## Weibull distribution
## Loglik(model)= -827.4   Loglik(intercept only)= -841.1
##  Chisq= 27.25 on 10 degrees of freedom, p= 0.0024
## Number of Newton-Raphson Iterations: 6
## n= 167
```

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

4 Lista nr 4

4.1 Zadanie nr 1

Korzystamy z metody eliminacji i opierając się na teście wiarogodności dokonujemy optymalnego wyboru zmiennych do modelu liniowego logarytmu czasu. Ustalamy poziom istotności jako $\alpha = 0.15$.

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

```
Wsum

##
## 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)              7.19e+00   1.05e+00  6.85 7.4e-12
## age                     -2.92e-03   8.16e-03 -0.36  0.7208
## as.factor(sex)2          4.32e-01   1.49e-01  2.90  0.0037
## as.factor(ph.ecog)1     -4.07e-01   2.38e-01 -1.71  0.0868
## as.factor(ph.ecog)2     -9.16e-01   3.69e-01 -2.48  0.0131
## as.factor(ph.ecog)3     -1.73e+00   7.94e-01 -2.18  0.0293
## as.factor(ph.karno)60   -7.31e-01   4.73e-01 -1.55  0.1223
## as.factor(ph.karno)70   -7.23e-01   4.48e-01 -1.62  0.1063
## as.factor(ph.karno)80   -8.46e-01   4.46e-01 -1.90  0.0577
## as.factor(ph.karno)90   -9.47e-01   4.57e-01 -2.07  0.0382
## as.factor(ph.karno)100  -1.02e+00   5.12e-01 -2.00  0.0458
## as.factor(pat.karno)40   1.67e-01   1.04e+00  0.16  0.8730
## as.factor(pat.karno)50  -6.89e-01   8.41e-01 -0.82  0.4128
## as.factor(pat.karno)60  -2.04e-01   7.13e-01 -0.29  0.7746
## as.factor(pat.karno)70   4.78e-03   7.36e-01  0.01  0.9948
## as.factor(pat.karno)80   6.35e-02   7.42e-01  0.09  0.9317
## as.factor(pat.karno)90  -6.45e-02   7.40e-01 -0.09  0.9305
## as.factor(pat.karno)100  2.55e-01   7.54e-01  0.34  0.7356
## meal.cal                 4.95e-05   1.95e-04  0.25  0.7995
## wt.loss                  9.26e-03   5.67e-03  1.63  0.1023
## Log(scale)              -3.78e-01   7.27e-02 -5.20 2.0e-07
##
## Scale= 0.685
##
## Weibull distribution
## Loglik(model)= -823.4   Loglik(intercept only)= -841.1
##  Chisq= 35.42 on 19 degrees of freedom, p= 0.012
## Number of Newton-Raphson Iterations: 7
## n= 167

#hipoteza bez age
Wparameters_age <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.ecog)+as.fa
                        data = df, dist = 'weibull')
Wsum_age <- summary(Wparameters_age)
Wsum_age

##
## 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)               6.996060   0.895475  7.81 5.6e-15
## as.factor(sex)2           0.433103   0.148673  2.91  0.0036
## as.factor(ph.ecog)1      -0.403045   0.237338 -1.70  0.0895
## as.factor(ph.ecog)2      -0.922961   0.370249 -2.49  0.0127
## as.factor(ph.ecog)3      -1.739677   0.793606 -2.19  0.0284
## as.factor(ph.karno)60    -0.724099   0.471389 -1.54  0.1245
## as.factor(ph.karno)70    -0.709731   0.445610 -1.59  0.1112
## as.factor(ph.karno)80    -0.834768   0.445305 -1.87  0.0608
## as.factor(ph.karno)90    -0.933899   0.456698 -2.04  0.0409
## as.factor(ph.karno)100   -0.995685   0.507989 -1.96  0.0500
## as.factor(pat.karno)40    0.140241   1.041265  0.13  0.8929
## as.factor(pat.karno)50   -0.720630   0.835550 -0.86  0.3884
## as.factor(pat.karno)60   -0.218308   0.711986 -0.31  0.7591
## as.factor(pat.karno)70   -0.013912   0.734511 -0.02  0.9849
## as.factor(pat.karno)80    0.060473   0.741194  0.08  0.9350
## as.factor(pat.karno)90   -0.081490   0.738294 -0.11  0.9121
## as.factor(pat.karno)100   0.242734   0.752656  0.32  0.7471
## meal.cal                  0.000060   0.000193  0.31  0.7558
## wt.loss                   0.009390   0.005633  1.67  0.0956
## Log(scale)               -0.378429   0.072632 -5.21 1.9e-07
##
## Scale= 0.685
##
## Weibull distribution
## Loglik(model)= -823.4   Loglik(intercept only)= -841.1
##  Chisq= 35.29 on 18 degrees of freedom, p= 0.0087
## Number of Newton-Raphson Iterations: 7
## n= 167
```

```r
lambda_age = exp( -823.4-( -823.4))
p_age= 1 - pchisq(-2*log(lambda_age), df = 1)
#hipoteza bez sex
Wparameters_sex <- survreg(Surv(time, status==2)~age+as.factor(ph.ecog)+as.factor(ph.kar
                        data = df, dist = 'weibull')
Wsum_sex <- summary(Wparameters_sex)
Wsum_sex
```

```
##
## 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)               7.29e+00   1.06e+00  6.89 5.5e-12
## age                      -4.00e-03   8.36e-03 -0.48   0.633
## as.factor(ph.ecog)1      -4.27e-01   2.42e-01 -1.76   0.078
```

```
## as.factor(ph.ecog)2     -8.12e-01   3.69e-01 -2.20   0.028
## as.factor(ph.ecog)3     -1.79e+00   8.07e-01 -2.21   0.027
## as.factor(ph.karno)60   -5.29e-01   4.75e-01 -1.11   0.265
## as.factor(ph.karno)70   -5.43e-01   4.44e-01 -1.22   0.221
## as.factor(ph.karno)80   -6.16e-01   4.47e-01 -1.38   0.169
## as.factor(ph.karno)90   -7.79e-01   4.60e-01 -1.69   0.091
## as.factor(ph.karno)100  -9.59e-01   5.22e-01 -1.84   0.066
## as.factor(pat.karno)40   3.73e-01   1.05e+00  0.35   0.723
## as.factor(pat.karno)50  -5.28e-01   8.58e-01 -0.61   0.539
## as.factor(pat.karno)60  -2.62e-01   7.24e-01 -0.36   0.718
## as.factor(pat.karno)70  -1.73e-02   7.47e-01 -0.02   0.981
## as.factor(pat.karno)80   1.72e-01   7.51e-01  0.23   0.819
## as.factor(pat.karno)90   2.24e-02   7.48e-01  0.03   0.976
## as.factor(pat.karno)100  3.52e-01   7.64e-01  0.46   0.645
## meal.cal                -3.76e-05   1.89e-04 -0.20   0.842
## wt.loss                  6.64e-03   5.50e-03  1.21   0.227
## Log(scale)              -3.64e-01   7.36e-02 -4.94 7.9e-07
##
## Scale= 0.695
##
## Weibull distribution
## Loglik(model)= -827.9   Loglik(intercept only)= -841.1
##  Chisq= 26.41 on 18 degrees of freedom, p= 0.091
## Number of Newton-Raphson Iterations: 7
## n= 167
```

```
lambda_sex = exp( -827.9-( -823.4))
p_sex= 1 - pchisq(-2*log(lambda_sex), df = 1)
#hipoteza bez ph.ecog
Wparameters_phecog <- survreg(Surv(time, status==2)~age+as.factor(sex)+as.factor(ph.karn
                        data = df, dist = 'weibull')
Wsum_phecog <- summary(Wparameters_phecog)
Wsum_phecog
```

```
##
## 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)              6.42e+00   1.04e+00  6.16 7.1e-10
## age                     -3.68e-03   8.50e-03 -0.43  0.6649
## as.factor(sex)2          4.18e-01   1.53e-01  2.74  0.0062
## as.factor(ph.karno)60   -9.06e-01   4.90e-01 -1.85  0.0641
## as.factor(ph.karno)70   -7.93e-01   4.67e-01 -1.70  0.0895
## as.factor(ph.karno)80   -6.52e-01   4.46e-01 -1.46  0.1445
## as.factor(ph.karno)90   -5.96e-01   4.49e-01 -1.33  0.1839
## as.factor(ph.karno)100  -4.85e-01   4.66e-01 -1.04  0.2987
```

```
## as.factor(pat.karno)40   3.60e-01   1.07e+00  0.34  0.7364
## as.factor(pat.karno)50  -6.03e-01   8.65e-01 -0.70  0.4856
## as.factor(pat.karno)60  -1.59e-01   7.32e-01 -0.22  0.8281
## as.factor(pat.karno)70   2.22e-01   7.49e-01  0.30  0.7667
## as.factor(pat.karno)80   2.72e-01   7.46e-01  0.36  0.7151
## as.factor(pat.karno)90   2.42e-01   7.44e-01  0.33  0.7444
## as.factor(pat.karno)100  4.48e-01   7.60e-01  0.59  0.5555
## meal.cal                 6.48e-05   1.94e-04  0.33  0.7383
## wt.loss                  6.42e-03   5.61e-03  1.14  0.2524
## Log(scale)              -3.51e-01   7.25e-02 -4.83 1.3e-06
##
## Scale= 0.704
##
## Weibull distribution
## Loglik(model)= -827.1   Loglik(intercept only)= -841.1
##  Chisq= 28 on 16 degrees of freedom, p= 0.032
## Number of Newton-Raphson Iterations: 7
## n= 167

lambda_phecog = exp( -827.1-( -823.4))
p_phecog=1 - pchisq(-2*log(lambda_phecog), df = 1)
#hipoteza bez ph.karno
Wparameters_phkarno <- survreg(Surv(time, status==2)~age+as.factor(sex)+as.factor(ph.eco
                       data = df, dist = 'weibull')
Wsum_phkarno <- summary(Wparameters_phkarno)
Wsum_phkarno

##
## 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)              6.06e+00   9.12e-01  6.65 3.0e-11
## age                     -2.50e-04   8.12e-03 -0.03  0.9754
## as.factor(sex)2          3.92e-01   1.50e-01  2.62  0.0088
## as.factor(ph.ecog)1     -2.61e-01   1.76e-01 -1.48  0.1382
## as.factor(ph.ecog)2     -5.96e-01   2.69e-01 -2.21  0.0269
## as.factor(ph.ecog)3     -1.54e+00   7.47e-01 -2.06  0.0392
## as.factor(pat.karno)40   2.27e-01   1.02e+00  0.22  0.8241
## as.factor(pat.karno)50  -7.15e-01   8.31e-01 -0.86  0.3894
## as.factor(pat.karno)60  -2.26e-01   7.31e-01 -0.31  0.7576
## as.factor(pat.karno)70   1.20e-01   7.44e-01  0.16  0.8714
## as.factor(pat.karno)80   1.70e-01   7.50e-01  0.23  0.8209
## as.factor(pat.karno)90  -1.41e-02   7.55e-01 -0.02  0.9851
## as.factor(pat.karno)100  2.71e-01   7.65e-01  0.35  0.7232
## meal.cal                -9.87e-06   1.87e-04 -0.05  0.9579
## wt.loss                  7.86e-03   5.67e-03  1.39  0.1655
```

```
## Log(scale)              -3.48e-01   7.19e-02 -4.83 1.3e-06
##
## Scale= 0.706
##
## Weibull distribution
## Loglik(model)= -826.2   Loglik(intercept only)= -841.1
##  Chisq= 29.79 on 14 degrees of freedom, p= 0.0082
## Number of Newton-Raphson Iterations: 6
## n= 167
```

```
lambda_phkarno = exp(-826.2-( -823.4))
p_phkarno=1 - pchisq(-2*log(lambda_phkarno), df = 1)
#hipoteza bez pat.karno
Wparameters_patkarno <- survreg(Surv(time, status==2)~age+as.factor(sex)+as.factor(ph.ec
                        data = df, dist = 'weibull')
Wsum_patkarno <- summary(Wparameters_patkarno)
Wsum_patkarno
```

```
##
## 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)             7.325536   0.766556  9.56 < 2e-16
## age                    -0.004628   0.007984 -0.58  0.5622
## as.factor(sex)2         0.446625   0.143764  3.11  0.0019
## as.factor(ph.ecog)1    -0.333386   0.229271 -1.45  0.1459
## as.factor(ph.ecog)2    -0.977054   0.337214 -2.90  0.0038
## as.factor(ph.ecog)3    -1.660770   0.786327 -2.11  0.0347
## as.factor(ph.karno)60  -0.821317   0.461071 -1.78  0.0749
## as.factor(ph.karno)70  -0.846346   0.434840 -1.95  0.0516
## as.factor(ph.karno)80  -0.949250   0.435059 -2.18  0.0291
## as.factor(ph.karno)90  -0.991114   0.447269 -2.22  0.0267
## as.factor(ph.karno)100 -1.003812   0.504726 -1.99  0.0467
## meal.cal                0.000059   0.000185  0.32  0.7492
## wt.loss                 0.009387   0.005436  1.73  0.0842
## Log(scale)             -0.369802   0.072668 -5.09 3.6e-07
##
## Scale= 0.691
##
## Weibull distribution
## Loglik(model)= -825.9   Loglik(intercept only)= -841.1
##  Chisq= 30.42 on 12 degrees of freedom, p= 0.0024
## Number of Newton-Raphson Iterations: 6
## n= 167
```

```
lambda_patkarno = exp( -825.9-( -823.4))
```

```r
p_patkarno=1 - pchisq(-2*log(lambda_patkarno), df = 1)
#hipoteza bez meal.cal
Wparameters_mealcal <- survreg(Surv(time, status==2)~age+as.factor(sex)+as.factor(ph.ec
                        data = df, dist = 'weibull')
Wsum_mealcal <- summary(Wparameters_mealcal)
Wsum_mealcal
```

```
##
## 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)             7.24478    1.02966  7.04 2.0e-12
## age                    -0.00323    0.00808 -0.40  0.6891
## as.factor(sex)2         0.42699    0.14750  2.89  0.0038
## as.factor(ph.ecog)1    -0.40846    0.23789 -1.72  0.0860
## as.factor(ph.ecog)2    -0.91770    0.36882 -2.49  0.0128
## as.factor(ph.ecog)3    -1.73820    0.79322 -2.19  0.0284
## as.factor(ph.karno)60  -0.71105    0.46603 -1.53  0.1271
## as.factor(ph.karno)70  -0.72261    0.44737 -1.62  0.1063
## as.factor(ph.karno)80  -0.83802    0.44512 -1.88  0.0597
## as.factor(ph.karno)90  -0.93429    0.45463 -2.06  0.0399
## as.factor(ph.karno)100 -1.01757    0.51200 -1.99  0.0469
## as.factor(pat.karno)40  0.13470    1.03632  0.13  0.8966
## as.factor(pat.karno)50 -0.68365    0.84068 -0.81  0.4161
## as.factor(pat.karno)60 -0.20144    0.71336 -0.28  0.7776
## as.factor(pat.karno)70  0.01687    0.73493  0.02  0.9817
## as.factor(pat.karno)80  0.07293    0.74057  0.10  0.9215
## as.factor(pat.karno)90 -0.05372    0.73885 -0.07  0.9420
## as.factor(pat.karno)100  0.26543   0.75245  0.35  0.7243
## wt.loss                 0.00922    0.00564  1.64  0.1017
## Log(scale)             -0.37776    0.07274 -5.19 2.1e-07
##
## Scale= 0.685
##
## Weibull distribution
## Loglik(model)= -823.4   Loglik(intercept only)= -841.1
##  Chisq= 35.36 on 18 degrees of freedom, p= 0.0085
## Number of Newton-Raphson Iterations: 7
## n= 167
```

```r
lambda_mealcal = exp(-823.4-( -823.4))
p_mealcal=1 - pchisq(-2*log(lambda_mealcal), df = 1)
#hipoteza bez wt.loss
Wparameters_wtloss <- survreg(Surv(time, status==2)~age+as.factor(sex)+as.factor(ph.ecog
                        data = df, dist = 'weibull')
Wsum_wtloss <- summary(Wparameters_wtloss)
Wsum_wtloss
```

```
##
## 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)                7.20e+00   1.07e+00  6.71 1.9e-11
## age                       -4.23e-03   8.36e-03 -0.51   0.613
## as.factor(sex)2            4.05e-01   1.50e-01  2.69   0.007
## as.factor(ph.ecog)1       -4.04e-01   2.42e-01 -1.67   0.095
## as.factor(ph.ecog)2       -8.07e-01   3.71e-01 -2.18   0.030
## as.factor(ph.ecog)3       -1.61e+00   8.00e-01 -2.01   0.044
## as.factor(ph.karno)60     -7.23e-01   4.80e-01 -1.51   0.132
## as.factor(ph.karno)70     -6.73e-01   4.54e-01 -1.48   0.138
## as.factor(ph.karno)80     -7.46e-01   4.50e-01 -1.66   0.097
## as.factor(ph.karno)90     -8.67e-01   4.62e-01 -1.88   0.060
## as.factor(ph.karno)100    -1.00e+00   5.20e-01 -1.92   0.054
## meal.cal                   4.84e-05   1.85e-04  0.26   0.793
## as.factor(pat.karno)40     5.09e-01   1.04e+00  0.49   0.624
## as.factor(pat.karno)50    -5.99e-01   8.52e-01 -0.70   0.482
## as.factor(pat.karno)60    -9.82e-02   7.19e-01 -0.14   0.891
## as.factor(pat.karno)70     1.51e-01   7.43e-01  0.20   0.838
## as.factor(pat.karno)80     1.58e-01   7.49e-01  0.21   0.833
## as.factor(pat.karno)90     2.50e-03   7.48e-01  0.00   0.997
## as.factor(pat.karno)100    3.10e-01   7.62e-01  0.41   0.684
## Log(scale)                -3.66e-01   7.29e-02 -5.02 5.1e-07
##
## Scale= 0.693
##
## Weibull distribution
## Loglik(model)= -824.7   Loglik(intercept only)= -841.1
##  Chisq= 32.67 on 18 degrees of freedom, p= 0.018
## Number of Newton-Raphson Iterations: 7
## n= 167
```

```r
lambda_wtloss = exp( -824.7-( -823.4))
p_wtloss=1 - pchisq(-2*log(lambda_wtloss), df = 1)
#pvalues
p_age
```

```
## [1] 1
```

```r
p_sex
```

```
## [1] 0.002699796
```

```r
p_phecog
```

```
## [1] 0.006522388
```

```
p_phkarno

## [1] 0.01796048

p_patkarno

## [1] 0.02534732

p_mealcal

## [1] 1

p_wtloss

## [1] 0.1068637
```

Badając 7 hipotez i analizując test wiarogodności przy pomocy twierdzenia Wilksa wykonujemy pierwszy krok eliminacji odrzucając zmienną age, która nie jest statystycznie istotna w naszym modelu. Wykonujemy kolejny krok badając 6 hipotez teraz w opraciu o model zależny od zmiennych: sex,ph.ecog, ph.karno, pat.karno, meal.cal, wt.loss.

```
#hipoteza bez sex
Wparameters_sex2 <- survreg(Surv(time, status==2)~as.factor(ph.ecog)+as.factor(ph.karno)
                        data = df, dist = 'weibull')
Wsum_sex2 <- summary(Wparameters_sex2)
Wsum_sex2

##
## Call:
## survreg(formula = Surv(time, status == 2) ~ 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)               7.03e+00   9.09e-01  7.74 1.0e-14
## as.factor(ph.ecog)1      -4.23e-01   2.41e-01 -1.75   0.080
## as.factor(ph.ecog)2      -8.29e-01   3.70e-01 -2.24   0.025
## as.factor(ph.ecog)3      -1.81e+00   8.05e-01 -2.25   0.025
## as.factor(ph.karno)60    -5.20e-01   4.73e-01 -1.10   0.271
## as.factor(ph.karno)70    -5.33e-01   4.43e-01 -1.20   0.229
## as.factor(ph.karno)80    -6.08e-01   4.48e-01 -1.36   0.175
## as.factor(ph.karno)90    -7.67e-01   4.61e-01 -1.66   0.096
## as.factor(ph.karno)100   -9.29e-01   5.20e-01 -1.79   0.074
## as.factor(pat.karno)40    3.32e-01   1.05e+00  0.32   0.752
## as.factor(pat.karno)50   -5.72e-01   8.52e-01 -0.67   0.502
## as.factor(pat.karno)60   -2.78e-01   7.22e-01 -0.39   0.700
## as.factor(pat.karno)70   -4.29e-02   7.45e-01 -0.06   0.954
## as.factor(pat.karno)80    1.66e-01   7.50e-01  0.22   0.825
## as.factor(pat.karno)90   -1.23e-04   7.46e-01  0.00   1.000
## as.factor(pat.karno)100   3.41e-01   7.62e-01  0.45   0.655
```

```
## meal.cal                   -2.61e-05   1.87e-04 -0.14   0.889
## wt.loss                     6.91e-03   5.43e-03  1.27   0.203
## Log(scale)                 -3.65e-01   7.35e-02 -4.97 6.8e-07
##
## Scale= 0.694
##
## Weibull distribution
## Loglik(model)= -828   Loglik(intercept only)= -841.1
##  Chisq= 26.18 on 17 degrees of freedom, p= 0.071
## Number of Newton-Raphson Iterations: 7
## n= 167
```

```
lambda_sex2 = exp(-828-(-823.4))
p_sex2= 1 - pchisq(-2*log(lambda_sex2), df = 1)
#hipoteza bez ph.ecog
Wparameters_phecog2 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.karno)+
                        data = df, dist = 'weibull')
Wsum_phecog2 <- summary(Wparameters_phecog2)
Wsum_phecog2
```

```
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ 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)                 6.160313   0.844429  7.30 3.0e-13
## as.factor(sex)2             0.421530   0.152286  2.77  0.0056
## as.factor(ph.karno)60      -0.892295   0.486772 -1.83  0.0668
## as.factor(ph.karno)70      -0.770147   0.462880 -1.66  0.0961
## as.factor(ph.karno)80      -0.622351   0.440024 -1.41  0.1573
## as.factor(ph.karno)90      -0.567722   0.443053 -1.28  0.2001
## as.factor(ph.karno)100     -0.443079   0.456315 -0.97  0.3316
## as.factor(pat.karno)40      0.327882   1.067332  0.31  0.7587
## as.factor(pat.karno)50     -0.643386   0.858938 -0.75  0.4538
## as.factor(pat.karno)60     -0.173423   0.730496 -0.24  0.8123
## as.factor(pat.karno)70      0.204037   0.747676  0.27  0.7849
## as.factor(pat.karno)80      0.274917   0.745110  0.37  0.7122
## as.factor(pat.karno)90      0.226668   0.742293  0.31  0.7601
## as.factor(pat.karno)100     0.440956   0.759767  0.58  0.5617
## meal.cal                    0.000078   0.000192  0.41  0.6840
## wt.loss                     0.006558   0.005580  1.18  0.2399
## Log(scale)                 -0.351349   0.072411 -4.85 1.2e-06
##
## Scale= 0.704
##
## Weibull distribution
## Loglik(model)= -827.2   Loglik(intercept only)= -841.1
```

```
##  Chisq= 27.81 on 15 degrees of freedom, p= 0.023
## Number of Newton-Raphson Iterations: 6
## n= 167
```

```
lambda_phecog2 = exp(-827.2-(-823.4))
p_phecog2=1 - pchisq(-2*log(lambda_phecog2), df = 1)
#hipoteza bez ph.karno
Wparameters_phkarno2 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.ecog)+
                        data = df, dist = 'weibull')
Wsum_phkarno2 <- summary(Wparameters_phkarno2)
Wsum_phkarno2
```

```
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ 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   7.65e-01  7.90 2.8e-15
## as.factor(sex)2           3.92e-01   1.50e-01  2.62  0.0087
## as.factor(ph.ecog)1      -2.61e-01   1.75e-01 -1.49  0.1357
## as.factor(ph.ecog)2      -5.97e-01   2.66e-01 -2.24  0.0250
## as.factor(ph.ecog)3      -1.54e+00   7.44e-01 -2.07  0.0383
## as.factor(pat.karno)40    2.24e-01   1.02e+00  0.22  0.8256
## as.factor(pat.karno)50   -7.18e-01   8.24e-01 -0.87  0.3833
## as.factor(pat.karno)60   -2.27e-01   7.30e-01 -0.31  0.7560
## as.factor(pat.karno)70    1.19e-01   7.42e-01  0.16  0.8729
## as.factor(pat.karno)80    1.69e-01   7.50e-01  0.23  0.8213
## as.factor(pat.karno)90   -1.56e-02   7.53e-01 -0.02  0.9835
## as.factor(pat.karno)100   2.70e-01   7.64e-01  0.35  0.7238
## meal.cal                 -8.81e-06   1.84e-04 -0.05  0.9618
## wt.loss                   7.88e-03   5.65e-03  1.40  0.1629
## Log(scale)               -3.48e-01   7.18e-02 -4.85 1.2e-06
##
## Scale= 0.706
##
## Weibull distribution
## Loglik(model)= -826.2   Loglik(intercept only)= -841.1
##  Chisq= 29.79 on 13 degrees of freedom, p= 0.0051
## Number of Newton-Raphson Iterations: 6
## n= 167
```

```
lambda_phkarno2 = exp(-826.2-(-823.4))
p_phkarno2=1 - pchisq(-2*log(lambda_phkarno2), df = 1)
#hipoteza bez pat.karno
Wparameters_patkarno2 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.ecog)
                        data = df, dist = 'weibull')
Wsum_patkarno2 <- summary(Wparameters_patkarno2)
Wsum_patkarno2
```

```
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ 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)              6.99e+00   5.05e-01 13.84 < 2e-16
## as.factor(sex)2          4.51e-01   1.43e-01  3.14  0.0017
## as.factor(ph.ecog)1     -3.20e-01   2.28e-01 -1.40  0.1608
## as.factor(ph.ecog)2     -9.85e-01   3.40e-01 -2.90  0.0038
## as.factor(ph.ecog)3     -1.67e+00   7.88e-01 -2.12  0.0337
## as.factor(ph.karno)60   -8.21e-01   4.60e-01 -1.78  0.0744
## as.factor(ph.karno)70   -8.31e-01   4.34e-01 -1.92  0.0553
## as.factor(ph.karno)80   -9.35e-01   4.36e-01 -2.14  0.0320
## as.factor(ph.karno)90   -9.74e-01   4.48e-01 -2.17  0.0298
## as.factor(ph.karno)100 -9.58e-01   5.00e-01 -1.92  0.0554
## meal.cal                 7.71e-05   1.83e-04  0.42  0.6734
## wt.loss                  9.57e-03   5.40e-03  1.77  0.0764
## Log(scale)              -3.70e-01   7.25e-02 -5.09 3.5e-07
##
## Scale= 0.691
##
## Weibull distribution
## Loglik(model)= -826   Loglik(intercept only)= -841.1
##  Chisq= 30.08 on 11 degrees of freedom, p= 0.0015
## Number of Newton-Raphson Iterations: 5
## n= 167
```

```
lambda_patkarno2 = exp(-826-(-823.4))
p_patkarno2=1 - pchisq(-2*log(lambda_patkarno2), df = 1)
#hipoteza bez meal.cal
Wparameters_mealcal2 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.ecog)+
                        data = df, dist = 'weibull')
Wsum_mealcal2 <- summary(Wparameters_mealcal2)
Wsum_mealcal2
```

```
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ 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)               7.03607    0.88760  7.93 2.2e-15
## as.factor(sex)2           0.42665    0.14712  2.90  0.0037
## as.factor(ph.ecog)1      -0.40410    0.23745 -1.70  0.0888
## as.factor(ph.ecog)2      -0.92587    0.37006 -2.50  0.0124
## as.factor(ph.ecog)3      -1.75171    0.79290 -2.21  0.0272
## as.factor(ph.karno)60    -0.69868    0.46361 -1.51  0.1318
```

```
## as.factor(ph.karno)70   -0.70809    0.44527 -1.59  0.1118
## as.factor(ph.karno)80   -0.82337    0.44460 -1.85  0.0640
## as.factor(ph.karno)90   -0.91713    0.45444 -2.02  0.0436
## as.factor(ph.karno)100  -0.98722    0.50829 -1.94  0.0521
## as.factor(pat.karno)40   0.09606    1.03125  0.09  0.9258
## as.factor(pat.karno)50  -0.71953    0.83545 -0.86  0.3891
## as.factor(pat.karno)60  -0.21764    0.71177 -0.31  0.7598
## as.factor(pat.karno)70  -0.00224    0.73341  0.00  0.9976
## as.factor(pat.karno)80   0.07095    0.74033  0.10  0.9236
## as.factor(pat.karno)90  -0.07113    0.73757 -0.10  0.9232
## as.factor(pat.karno)100  0.25398    0.75182  0.34  0.7355
## wt.loss                  0.00936    0.00559  1.67  0.0941
## Log(scale)              -0.37835    0.07263 -5.21 1.9e-07
##
## Scale= 0.685
##
## Weibull distribution
## Loglik(model)= -823.5   Loglik(intercept only)= -841.1
##  Chisq= 35.2 on 17 degrees of freedom, p= 0.0059
## Number of Newton-Raphson Iterations: 6
## n= 167
```

```
lambda_mealcal2 = exp(-823.5-(-823.4))
p_mealcal2=1 - pchisq(-2*log(lambda_mealcal2), df = 1)
#hipoteza bez wt.loss
Wparameters_wtloss2 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.ecog)+a
                      data = df, dist = 'weibull')
Wsum_wtloss2 <- summary(Wparameters_wtloss2)
Wsum_wtloss2
```

```
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ 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)              6.91e+00   9.08e-01  7.61 2.8e-14
## as.factor(sex)2          4.07e-01   1.50e-01  2.72  0.0066
## as.factor(ph.ecog)1     -3.96e-01   2.42e-01 -1.64  0.1010
## as.factor(ph.ecog)2     -8.13e-01   3.73e-01 -2.18  0.0294
## as.factor(ph.ecog)3     -1.62e+00   8.01e-01 -2.02  0.0433
## as.factor(ph.karno)60   -7.15e-01   4.78e-01 -1.50  0.1342
## as.factor(ph.karno)70   -6.53e-01   4.51e-01 -1.45  0.1482
## as.factor(ph.karno)80   -7.30e-01   4.50e-01 -1.62  0.1046
## as.factor(ph.karno)90   -8.46e-01   4.62e-01 -1.83  0.0670
## as.factor(ph.karno)100  -9.61e-01   5.17e-01 -1.86  0.0630
## meal.cal                 6.31e-05   1.82e-04  0.35  0.7286
## as.factor(pat.karno)40   4.78e-01   1.04e+00  0.46  0.6447
```

```
## as.factor(pat.karno)50  -6.43e-01   8.47e-01 -0.76  0.4480
## as.factor(pat.karno)60  -1.14e-01   7.17e-01 -0.16  0.8736
## as.factor(pat.karno)70   1.29e-01   7.41e-01  0.17  0.8623
## as.factor(pat.karno)80   1.57e-01   7.49e-01  0.21  0.8336
## as.factor(pat.karno)90  -2.00e-02   7.46e-01 -0.03  0.9787
## as.factor(pat.karno)100  2.96e-01   7.61e-01  0.39  0.6968
## Log(scale)              -3.67e-01   7.28e-02 -5.04 4.6e-07
##
## Scale= 0.693
##
## Weibull distribution
## Loglik(model)= -824.9   Loglik(intercept only)= -841.1
##  Chisq= 32.41 on 17 degrees of freedom, p= 0.013
## Number of Newton-Raphson Iterations: 7
## n= 167
```

```
lambda_wtloss2 = exp(-824.9-(-823.4))
p_wtloss2=1 - pchisq(-2*log(lambda_wtloss2), df = 1)
#pvalues
p_sex2
```

```
## [1] 0.002420151
```

```
p_phecog2
```

```
## [1] 0.00583683
```

```
p_phkarno2
```

```
## [1] 0.01796048
```

```
p_patkarno2
```

```
## [1] 0.02258689
```

```
p_mealcal2
```

```
## [1] 0.6547208
```

```
p_wtloss2
```

```
## [1] 0.08326452
```

Na podstawie testu wiarogodności przyjmujemy, że odrzucamy zmienną meal.cal i przyjmujemy model zależny od zmiennych: sex, ph.ecog, ph.karno, pat.karno oraz wt.loss.

```
#hipoteza bez sex
Wparameters_sex3 <- survreg(Surv(time, status==2)~as.factor(ph.ecog)+as.factor(ph.karno)
                        data = df, dist = 'weibull')
```

```
Wsum_sex3 <- summary(Wparameters_sex3)
Wsum_sex3

##
## 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)                7.01773    0.90141  7.79 7.0e-15
## as.factor(ph.ecog)1       -0.42287    0.24136 -1.75   0.080
## as.factor(ph.ecog)2       -0.82658    0.36919 -2.24   0.025
## as.factor(ph.ecog)3       -1.80480    0.80457 -2.24   0.025
## as.factor(ph.karno)60     -0.52929    0.46821 -1.13   0.258
## as.factor(ph.karno)70     -0.53222    0.44302 -1.20   0.230
## as.factor(ph.karno)80     -0.61137    0.44681 -1.37   0.171
## as.factor(ph.karno)90     -0.77199    0.45910 -1.68   0.093
## as.factor(ph.karno)100    -0.93177    0.51923 -1.79   0.073
## as.factor(pat.karno)40     0.34968    1.04218  0.34   0.737
## as.factor(pat.karno)50    -0.57464    0.85163 -0.67   0.500
## as.factor(pat.karno)60    -0.28189    0.72147 -0.39   0.696
## as.factor(pat.karno)70    -0.05122    0.74220 -0.07   0.945
## as.factor(pat.karno)80     0.15929    0.74866  0.21   0.832
## as.factor(pat.karno)90    -0.00624    0.74480 -0.01   0.993
## as.factor(pat.karno)100    0.33279    0.76039  0.44   0.662
## wt.loss                    0.00691    0.00545  1.27   0.205
## Log(scale)                -0.36512    0.07349 -4.97 6.8e-07
##
## Scale= 0.694
##
## Weibull distribution
## Loglik(model)= -828   Loglik(intercept only)= -841.1
##  Chisq= 26.16 on 16 degrees of freedom, p= 0.052
## Number of Newton-Raphson Iterations: 7
## n= 167

lambda_sex3 = exp(-828-(-823.5))
p_sex3= 1 - pchisq(-2*log(lambda_sex3), df = 1)
#hipoteza bez ph.ecog
Wparameters_phecog3 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.karno)+
                        data = df, dist = 'weibull')
Wsum_phecog3 <- summary(Wparameters_phecog3)
Wsum_phecog3

##
## 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)                6.20091    0.83877  7.39 1.4e-13
## as.factor(sex)2            0.41255    0.15062  2.74  0.0062
## as.factor(ph.karno)60     -0.85298    0.47628 -1.79  0.0733
## as.factor(ph.karno)70     -0.75995    0.46161 -1.65  0.0997
## as.factor(ph.karno)80     -0.59910    0.43619 -1.37  0.1696
## as.factor(ph.karno)90     -0.53873    0.43730 -1.23  0.2180
## as.factor(ph.karno)100    -0.42422    0.45432 -0.93  0.3504
## as.factor(pat.karno)40     0.27211    1.05840  0.26  0.7971
## as.factor(pat.karno)50    -0.64036    0.85896 -0.75  0.4560
## as.factor(pat.karno)60    -0.17027    0.73019 -0.23  0.8156
## as.factor(pat.karno)70     0.21912    0.74669  0.29  0.7692
## as.factor(pat.karno)80     0.29213    0.74370  0.39  0.6945
## as.factor(pat.karno)90     0.24437    0.74091  0.33  0.7415
## as.factor(pat.karno)100    0.45802    0.75859  0.60  0.5460
## wt.loss                    0.00650    0.00552  1.18  0.2388
## Log(scale)                -0.35116    0.07240 -4.85 1.2e-06
##
## Scale= 0.704
##
## Weibull distribution
## Loglik(model)= -827.2   Loglik(intercept only)= -841.1
##  Chisq= 27.64 on 14 degrees of freedom, p= 0.016
## Number of Newton-Raphson Iterations: 6
## n= 167
```

```
lambda_phecog3 = exp(-827.2-(-823.5))
p_phecog3=1 - pchisq(-2*log(lambda_phecog3), df = 1)
#hipoteza bez ph.karno
Wparameters_phkarno3 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.ecog)+
                        data = df, dist = 'weibull')
Wsum_phkarno3 <- summary(Wparameters_phkarno3)
Wsum_phkarno3
```

```
##
## 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)                6.03904    0.75340  8.02 1.1e-15
## as.factor(sex)2            0.39318    0.14882  2.64  0.0082
## as.factor(ph.ecog)1       -0.26078    0.17486 -1.49  0.1359
## as.factor(ph.ecog)2       -0.59623    0.26559 -2.24  0.0248
## as.factor(ph.ecog)3       -1.54112    0.74408 -2.07  0.0383
## as.factor(pat.karno)40     0.22622    1.01422  0.22  0.8235
## as.factor(pat.karno)50    -0.72149    0.82127 -0.88  0.3797
```

```
## as.factor(pat.karno)60  -0.22777    0.72967 -0.31  0.7549
## as.factor(pat.karno)70   0.11520    0.73825  0.16  0.8760
## as.factor(pat.karno)80   0.16705    0.74848  0.22  0.8234
## as.factor(pat.karno)90  -0.01835    0.75147 -0.02  0.9805
## as.factor(pat.karno)100  0.26709    0.76181  0.35  0.7259
## wt.loss                  0.00788    0.00565  1.39  0.1635
## Log(scale)              -0.34781    0.07168 -4.85 1.2e-06
##
## Scale= 0.706
##
## Weibull distribution
## Loglik(model)= -826.2   Loglik(intercept only)= -841.1
##  Chisq= 29.79 on 12 degrees of freedom, p= 0.003
## Number of Newton-Raphson Iterations: 5
## n= 167
```

```
lambda_phkarno3 = exp(-826.2-(-823.5))
p_phkarno3=1 - pchisq(-2*log(lambda_phkarno3), df = 1)
#hipoteza bez pat.karno
Wparameters_patkarno3 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.ecog)
                        data = df, dist = 'weibull')
Wsum_patkarno3 <- summary(Wparameters_patkarno2)
Wsum_patkarno3
```

```
##
## Call:
## survreg(formula = Surv(time, status == 2) ~ 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)             6.99e+00   5.05e-01 13.84 < 2e-16
## as.factor(sex)2         4.51e-01   1.43e-01  3.14  0.0017
## as.factor(ph.ecog)1    -3.20e-01   2.28e-01 -1.40  0.1608
## as.factor(ph.ecog)2    -9.85e-01   3.40e-01 -2.90  0.0038
## as.factor(ph.ecog)3    -1.67e+00   7.88e-01 -2.12  0.0337
## as.factor(ph.karno)60  -8.21e-01   4.60e-01 -1.78  0.0744
## as.factor(ph.karno)70  -8.31e-01   4.34e-01 -1.92  0.0553
## as.factor(ph.karno)80  -9.35e-01   4.36e-01 -2.14  0.0320
## as.factor(ph.karno)90  -9.74e-01   4.48e-01 -2.17  0.0298
## as.factor(ph.karno)100 -9.58e-01   5.00e-01 -1.92  0.0554
## meal.cal                7.71e-05   1.83e-04  0.42  0.6734
## wt.loss                 9.57e-03   5.40e-03  1.77  0.0764
## Log(scale)             -3.70e-01   7.25e-02 -5.09 3.5e-07
##
## Scale= 0.691
##
## Weibull distribution
## Loglik(model)= -826   Loglik(intercept only)= -841.1
```

```
##  Chisq= 30.08 on 11 degrees of freedom, p= 0.0015
## Number of Newton-Raphson Iterations: 5
## n= 167
```

```
lambda_patkarno3 = exp(-826-(-823.5))
p_patkarno3=1 - pchisq(-2*log(lambda_patkarno3), df = 1)
#hipoteza bez wt.loss
Wparameters_wtloss3 <- survreg(Surv(time, status==2)~as.factor(sex)+as.factor(ph.ecog)+a
                        data = df, dist = 'weibull')
Wsum_wtloss3 <- summary(Wparameters_wtloss3)
Wsum_wtloss3
```

```
##
## 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)               6.9496     0.9019  7.71 1.3e-14
## as.factor(sex)2           0.3992     0.1482  2.69   0.007
## as.factor(ph.ecog)1      -0.3960     0.2418 -1.64   0.101
## as.factor(ph.ecog)2      -0.8154     0.3734 -2.18   0.029
## as.factor(ph.ecog)3      -1.6341     0.8002 -2.04   0.041
## as.factor(ph.karno)60    -0.6859     0.4697 -1.46   0.144
## as.factor(ph.karno)70    -0.6508     0.4514 -1.44   0.149
## as.factor(ph.karno)80    -0.7163     0.4492 -1.59   0.111
## as.factor(ph.karno)90    -0.8274     0.4599 -1.80   0.072
## as.factor(ph.karno)100   -0.9494     0.5171 -1.84   0.066
## as.factor(pat.karno)40    0.4270     1.0265  0.42   0.677
## as.factor(pat.karno)50   -0.6447     0.8475 -0.76   0.447
## as.factor(pat.karno)60   -0.1170     0.7177 -0.16   0.870
## as.factor(pat.karno)70    0.1410     0.7406  0.19   0.849
## as.factor(pat.karno)80    0.1651     0.7487  0.22   0.825
## as.factor(pat.karno)90   -0.0120     0.7463 -0.02   0.987
## as.factor(pat.karno)100  0.3052     0.7609  0.40   0.688
## Log(scale)               -0.3663     0.0727 -5.04 4.8e-07
##
## Scale= 0.693
##
## Weibull distribution
## Loglik(model)= -824.9   Loglik(intercept only)= -841.1
##  Chisq= 32.29 on 16 degrees of freedom, p= 0.0092
## Number of Newton-Raphson Iterations: 7
## n= 167
```

```
lambda_wtloss3 = exp(-824.9-(-823.5))
p_wtloss3=1 - pchisq(-2*log(lambda_wtloss3), df = 1)
#pvalues
p_sex3
```

```
## [1] 0.002699796
```

```
p_phecog3
```

```
## [1] 0.006522388
```

```
p_phkarno3
```

```
## [1] 0.02013675
```

```
p_patkarno3
```

```
## [1] 0.02534732
```

```
p_wtloss3
```

```
## [1] 0.09426431
```

Ostatecznie optymalny model na podstawie testu ilorazu wiarogodności jest zależny od zmiennych: sex, ph.ecog, ph.karno, pat.karno oraz wt.loss.

4.2 Zadanie nr 2

Korzystając z kryterium informacyjnego Akaike'a (AIC), dokonujemy wyboru najlepszego modelu liniowego logarytmu czasu. Korzystamy w tym wypadku z dostępnej funkcji *step* z pakietu *stats*.

```
step(WparametersALL)
```

```
## Start:  AIC=1688.72
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     as.factor(ph.karno) + as.factor(pat.karno) + meal.cal + wt.loss
##
##                         Df    AIC
## - as.factor(pat.karno)  7 1679.7
## - as.factor(ph.karno)   5 1684.3
## - meal.cal              1 1686.8
## - age                   1 1686.8
## <none>                    1688.7
## - wt.loss               1 1689.5
## - as.factor(ph.ecog)    3 1690.1
## - as.factor(sex)        1 1695.7
##
## Step:  AIC=1679.72
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     as.factor(ph.karno) + meal.cal + wt.loss
##
##                        Df    AIC
## - as.factor(ph.karno)  5 1676.9
```

```
## - meal.cal            1 1677.8
## - age                 1 1678.1
## <none>                  1679.7
## - wt.loss             1 1680.8
## - as.factor(ph.ecog)  3 1683.5
## - as.factor(sex)      1 1688.0
##
## Step:  AIC=1676.87
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     meal.cal + wt.loss
##
##                      Df    AIC
## - meal.cal            1 1674.9
## - age                 1 1675.0
## <none>                  1676.9
## - wt.loss             1 1677.2
## - as.factor(sex)      1 1682.6
## - as.factor(ph.ecog)  3 1685.5
##
## Step:  AIC=1674.88
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     wt.loss
##
##                      Df    AIC
## - age                 1 1673.0
## <none>                  1674.9
## - wt.loss             1 1675.2
## - as.factor(sex)      1 1680.7
## - as.factor(ph.ecog)  3 1683.8
##
## Step:  AIC=1672.99
## Surv(time, status == 2) ~ as.factor(sex) + as.factor(ph.ecog) +
##     wt.loss
##
##                      Df    AIC
## <none>                  1673.0
## - wt.loss             1 1673.4
## - as.factor(sex)      1 1678.9
## - as.factor(ph.ecog)  3 1683.9
## Call:
## survreg(formula = Surv(time, status == 2) ~ as.factor(sex) +
##     as.factor(ph.ecog) + wt.loss, data = df, dist = "weibull")
##
## Coefficients:
##         (Intercept)     as.factor(sex)2 as.factor(ph.ecog)1 as.factor(ph.ecog)2
##         6.128882960         0.390879246        -0.249685286        -0.786653889
## as.factor(ph.ecog)3             wt.loss
##        -1.523148004         0.008247483
```

```
##
## Scale= 0.7174609
##
## Loglik(model)= -829.5   Loglik(intercept only)= -841.1
##  Chisq= 23.15 on 5 degrees of freedom, p= 0.000316
## n= 167
```

Według funkcji step, optymalnym wyborem modelu liniowego logarytmu jest model oparty charakterystki: sex, ph.ecog i wt.loss. Teraz wykonamy tą samą analizę korzystając z bayesowskigo kryterium informacyjnego (BIC).

4.3 Zadanie nr 3

```
n = length(df$status==2)
step(WparametersALL, k = log(n))

## Start:  AIC=1754.19
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     as.factor(ph.karno) + as.factor(pat.karno) + meal.cal + wt.loss
##
##                         Df    AIC
## - as.factor(pat.karno)  7 1723.4
## - as.factor(ph.karno)   5 1734.2
## - as.factor(ph.ecog)    3 1746.3
## - meal.cal              1 1749.1
## - age                   1 1749.2
## - wt.loss               1 1751.8
## <none>                    1754.2
## - as.factor(sex)        1 1758.1
##
## Step:  AIC=1723.37
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     as.factor(ph.karno) + meal.cal + wt.loss
##
##                        Df    AIC
## - as.factor(ph.karno)  5 1704.9
## - as.factor(ph.ecog)   3 1717.8
## - meal.cal             1 1718.3
## - age                  1 1718.6
## - wt.loss              1 1721.4
## <none>                   1723.4
## - as.factor(sex)       1 1728.6
##
## Step:  AIC=1704.93
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     meal.cal + wt.loss
##
```

```
##                       Df    AIC
## - meal.cal             1 1699.8
## - age                  1 1699.9
## - wt.loss              1 1702.1
## - as.factor(ph.ecog)  3 1704.2
## <none>                   1704.9
## - as.factor(sex)       1 1707.5
##
## Step:  AIC=1699.82
## Surv(time, status == 2) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##     wt.loss
##
##                       Df    AIC
## - age                  1 1694.8
## - wt.loss              1 1697.0
## - as.factor(ph.ecog)  3 1699.3
## <none>                   1699.8
## - as.factor(sex)       1 1702.5
##
## Step:  AIC=1694.81
## Surv(time, status == 2) ~ as.factor(sex) + as.factor(ph.ecog) +
##     wt.loss
##
##                       Df    AIC
## - wt.loss              1 1692.1
## <none>                   1694.8
## - as.factor(ph.ecog)  3 1696.4
## - as.factor(sex)       1 1697.6
##
## Step:  AIC=1692.09
## Surv(time, status == 2) ~ as.factor(sex) + as.factor(ph.ecog)
##
##                       Df    AIC
## - as.factor(ph.ecog)  3 1691.2
## <none>                   1692.1
## - as.factor(sex)       1 1693.9
##
## Step:  AIC=1691.25
## Surv(time, status == 2) ~ as.factor(sex)
##
##                   Df    AIC
## <none>               1691.2
## - as.factor(sex)  1 1692.4
## Call:
## survreg(formula = Surv(time, status == 2) ~ as.factor(sex), data = df,
##     dist = "weibull")
##
## Coefficients:
```

```
##     (Intercept) as.factor(sex)2
##       5.9222791       0.3539958
##
## Scale= 0.7400554
##
## Loglik(model)= -837.9   Loglik(intercept only)= -841.1
##  Chisq= 6.24 on 1 degrees of freedom, p= 0.0125
## n= 167
```

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