

## Pneumonia Survival Analysis

Ioannis Matzakos Chorianopoulos

### Generate survival estimates

To perform a survival analysis and to generate the survival estimates and plots of the Kaplan-Meier product-limit estimator regarding the different number of siblings of the pneumon data, are needed the "survival" and "KMsurv" packages. Firstly, the time variable for the survival object is chldage, which is when the babies had pneumonia in their first 12 months. The status variable had to be created independently using the chldage variable. Secondly, the status is represented as a binary variable, 1 being all the babies that had pneumonia before their 12th month (chldage<12) and 0 being every other case (chldage>12). Moreover, the nsibs variable, which is the number of siblings of each child, needed to be re-coded by assigning value 3 to all the instances that have values greater than 3. Finally, the model is fitted using the survfit() function and then the results can be shown in the summary of the model and at the plots as well. The R code and the output of the above task is the following.

```
#creating the status variable and including it in the dataset
pneumon$status = (pneumon$chldage<12) * 1 + 0
#recoding nsibs variable so that any number greater than 3 will get the value
of 3
numsibs = pneumon$nsibs
pneumon$nsibs[pneumon$nsibs>=3] = 3
#recoding wmonth, ages 4-6 months into value 4 and ages 7+ months into value
5
wage = pneumon$wmonth
pneumon$wmonth[pneumon$wmonth>=4 & pneumon$wmonth<=6] = 4
pneumon$wmonth[wage>=7] = 5
#the original values of nsibs & wmonth are assigned to new variables

#fitting the model
model = survfit(Surv(chldage, status) ~ nsibs, data=pneumon, type = "kaplan")
summary(model)

## Call: survfit(formula = Surv(chldage, status) ~ nsibs, data = pneumon,
##      type = "kaplan")
##
##                nsibs=0
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    0.5  1801     35   0.981 0.00325   0.974   0.987
##    1.0  1766     43   0.957 0.00480   0.947   0.966
##    2.0  1723     45   0.932 0.00594   0.920   0.943
##    3.0  1678     48   0.905 0.00691   0.892   0.919
##    4.0  1630     42   0.882 0.00761   0.867   0.897
```

```

## 5.0 1588 54 0.852 0.00837 0.835 0.868
## 6.0 1534 48 0.825 0.00895 0.808 0.843
## 7.0 1486 50 0.797 0.00947 0.779 0.816
## 8.0 1436 47 0.771 0.00990 0.752 0.791
## 9.0 1389 33 0.753 0.01016 0.733 0.773
## 10.0 1356 35 0.733 0.01042 0.713 0.754
## 11.0 1321 30 0.717 0.01062 0.696 0.738
##
## nsibs=1
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.5 1156 33 0.971 0.00490 0.962 0.981
## 1.0 1123 46 0.932 0.00742 0.917 0.946
## 2.0 1077 26 0.909 0.00845 0.893 0.926
## 3.0 1051 27 0.886 0.00935 0.868 0.904
## 4.0 1024 40 0.851 0.01047 0.831 0.872
## 5.0 984 39 0.817 0.01136 0.796 0.840
## 6.0 945 34 0.788 0.01202 0.765 0.812
## 7.0 911 31 0.761 0.01254 0.737 0.786
## 8.0 880 35 0.731 0.01304 0.706 0.757
## 9.0 845 37 0.699 0.01349 0.673 0.726
## 10.0 808 31 0.672 0.01381 0.646 0.700
## 11.0 777 31 0.645 0.01407 0.618 0.674
##
## nsibs=2
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.5 383 9 0.977 0.00774 0.961 0.992
## 1.0 374 12 0.945 0.01163 0.923 0.968
## 2.0 362 24 0.883 0.01645 0.851 0.915
## 3.0 338 15 0.843 0.01857 0.808 0.881
## 4.0 323 11 0.815 0.01986 0.777 0.854
## 5.0 312 13 0.781 0.02114 0.740 0.823
## 6.0 299 14 0.744 0.02230 0.702 0.789
## 7.0 285 14 0.708 0.02324 0.663 0.755
## 8.0 271 12 0.676 0.02391 0.631 0.725
## 9.0 259 15 0.637 0.02457 0.591 0.687
## 10.0 244 7 0.619 0.02482 0.572 0.669
## 11.0 237 7 0.601 0.02503 0.553 0.652
##
## nsibs=3
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.5 130 7 0.946 0.0198 0.908 0.986
## 1.0 123 3 0.923 0.0234 0.878 0.970
## 2.0 120 3 0.900 0.0263 0.850 0.953
## 3.0 117 5 0.862 0.0303 0.804 0.923
## 4.0 112 3 0.838 0.0323 0.778 0.904
## 5.0 109 7 0.785 0.0361 0.717 0.859
## 6.0 102 5 0.746 0.0382 0.675 0.825
## 7.0 97 2 0.731 0.0389 0.658 0.811
## 8.0 95 3 0.708 0.0399 0.634 0.790
## 9.0 92 4 0.677 0.0410 0.601 0.762

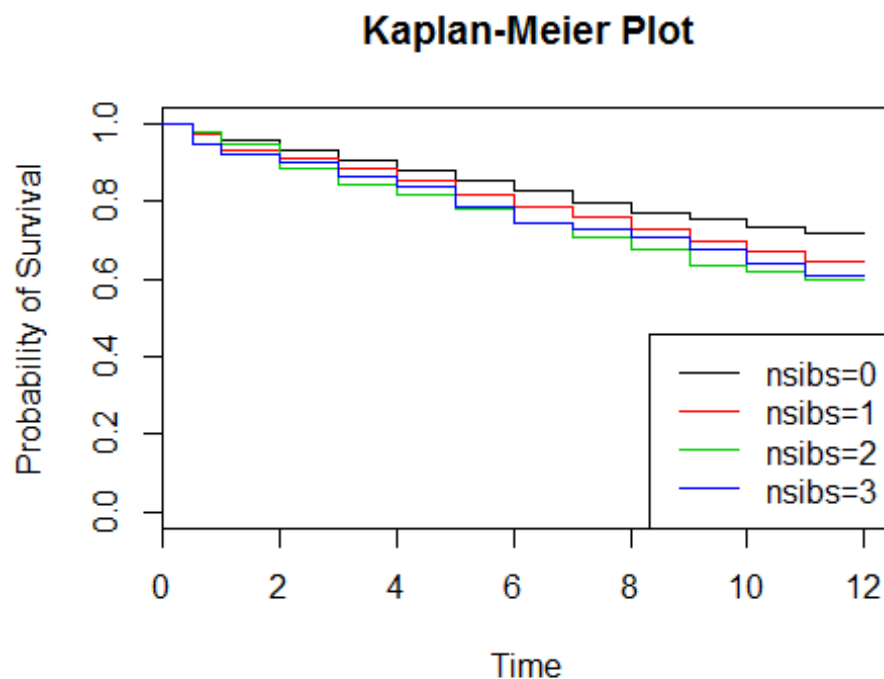
```

##	10.0	88	5	0.638	0.0421	0.561	0.727
##	11.0	83	4	0.608	0.0428	0.529	0.698

In the summary of the fitted model is described for each level of number of siblings the number of the observed instances at risk at a specific moment, the number of instances that the event occurred, the probability of survival, the standard error, the 95% confidence intervals and how all these aspects are distributed over time. As it can be seen from the summary of the model above, for each number of siblings the probability of a baby not getting pneumonia is reducing over time gradually. Although, the number of events that occurred varies for each moment in time without having any specific patterns. Moreover, the 95% confidence intervals are also reducing over time as the probability of survival. Finally, overall the probabilities of survival remain in quite high levels.

*#plotting the fitted model*

```
par(mfrow=c(1,1))
plot(model, main="Kaplan-Meier Plot", xlab="Time", ylab="Probability of Survival", col=c(1,2,3,4), lwd=1)
legend('bottomright', c("nsibs=0", "nsibs=1", "nsibs=2", "nsibs=3"), lty=1, col=c(1,2,3,4))
```



Regarding the Kaplan-Meier plot, all the different levels of number of siblings have similar survival functions. Firstly, they begin and end at the same points in time. Secondly, the differences that can be observed concern the course of these survival functions over time. The probabilities of survival for each function differ for different number of siblings but for all functions these probabilities change at the same points in time. However, the

probability of survival decreases steadily over time but overall it remains in fairly high levels.

## Statistical significance using log-rank test

In this task, firstly the null and the alternative hypothesis need to be defined as they are stated below:

H0: There is no difference in hazard rates of pneumonia between groups.

H1: There is a difference in hazard rates of pneumonia between groups.

Essentially, the null hypothesis suggests that different number of siblings does not affect the hazard rates of babies in getting pneumonia, while the alternative hypothesis suggests the opposite. Secondly, the `survdifff()` function is used in order to determine if there is or not a significant difference in the hazard rates of pneumonia between groups, the long-rank test was performed considering the different number of siblings for each instance of the data. The R code and the output of the above task is the following:

```
#log-rank test using survdifff
survdifff(Surv(chldage, status) ~ nsibs, rho = 0, data=pneumon)

## Call:
## survdifff(formula = Surv(chldage, status) ~ nsibs, data = pneumon,
##           rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## nsibs=0 1801         510     597.1      12.71      28.00
## nsibs=1 1156         410     369.2       4.51       6.93
## nsibs=2  383         153     117.3      10.86      12.53
## nsibs=3  130          51      40.4       2.79       2.99
##
##  Chisq= 31.9  on 3 degrees of freedom, p= 5.51e-07
```

Given the p-value of the log-rank test that the `survdifff()` function provided ( $p=5.51e-07$ ) the difference of hazard rates of pneumonia between groups is highly significant. Therefore, multiple comparisons are needed be performed in order to find out which groups differ at 0.05 significance level.

The comparisons that need to be made are 6 because there are 4 levels for the number of siblings in order to decide which levels are different. Moreover, because there are two hypotheses H0 and H1 the significance level will be  $\alpha=0.05/2$ , but in this case because the Bonferroni correction method is involved the significance level will have to be divided by the number of comparisons in order to compute the Bonferroni critical value. So  $\alpha=(0.05/2)/6$ , which is  $\alpha=0.004166667$ . The R code and the output of the above task is the following:

```
alpha = 0.05/2/6
alpha
## [1] 0.004166667

#Multiple Comparisons
#0-1
```

```

survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[pneumon$nsibs<2,])

## Call:
## survdif(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[pneumon$nsibs < 2, ], rho = 0)
##
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=0 1801      510      568      6.02      16.3
## as.factor(nsibs)=1 1156      410      352      9.73      16.3
##
## Chisq= 16.3 on 1 degrees of freedom, p= 5.55e-05

#0-2
survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==0|pneumon$nsibs==2),])

## Call:
## survdif(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[(pneumon$nsibs == 0 | pneumon$nsibs == 2),
## ], rho = 0)
##
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=0 1801      510      554      3.5      21.9
## as.factor(nsibs)=2  383      153      109     17.8      21.9
##
## Chisq= 21.9 on 1 degrees of freedom, p= 2.83e-06

#0-3
survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==0|pneumon$nsibs==3),])

## Call:
## survdif(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[(pneumon$nsibs == 0 | pneumon$nsibs == 3),
## ], rho = 0)
##
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=0 1801      510     525.4      0.454      7.37
## as.factor(nsibs)=3  130       51      35.6      6.711      7.37
##
## Chisq= 7.4 on 1 degrees of freedom, p= 0.00662

#1-2
survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==1|pneumon$nsibs==2),])

## Call:
## survdif(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[(pneumon$nsibs == 1 | pneumon$nsibs == 2),
## ], rho = 0)

```

```
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=1 1156      410      427      0.7      3.01
## as.factor(nsibs)=2  383      153      136      2.2      3.01
##
##  Chisq= 3   on 1 degrees of freedom, p= 0.0826

#1-3
survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==1|pneumon$nsibs==3),])

## Call:
## survdiff(formula = Surv(chldage, status) ~ as.factor(nsibs),
##          data = pneumon[(pneumon$nsibs == 1 | pneumon$nsibs == 3),
##                          ], rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=1 1156      410      415.6      0.0745      0.784
## as.factor(nsibs)=3  130       51      45.4      0.6811      0.784
##
##  Chisq= 0.8   on 1 degrees of freedom, p= 0.376

#2-3
survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==2|pneumon$nsibs==3),])

## Call:
## survdiff(formula = Surv(chldage, status) ~ as.factor(nsibs),
##          data = pneumon[(pneumon$nsibs == 2 | pneumon$nsibs == 3),
##                          ], rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=2  383      153      151.8      0.00977      0.0398
## as.factor(nsibs)=3 130       51      52.2      0.02839      0.0398
##
##  Chisq= 0   on 1 degrees of freedom, p= 0.842
```

Between levels 0 and 1 gives p-value  $p = 5.55e-05$ , for levels 0 and 2 the p-value is  $p = 2.83e-06$  and for levels 0 and 3 the p-value is  $p = 0.00662$ . All of these p-values that log-rank test provided are less than alpha. In these cases, the null hypothesis is rejected. For levels 1 and 2 the p-value is  $p = 0.0826$ , for levels 1 and 3 the p-value is  $p = 0.376$  and for levels 2 and 3 the p-value is  $p = 0.842$ . All these p-values are greater than alpha. In these cases, the null hypothesis is accepted. Therefore, regarding the results provided by performing the long-rank test for each comparison of different levels, the groups 0-1, 0-2 and 0-3 differ from the groups 1-2, 1-3 and 2-3.

## Cox regression and backward elimination

In this part, the data are going to be analyzed using Cox regression. The variables that must be explored along with their 2-way interactions are the number of siblings

(nsibs), weaning age (wmonth), maternal age (mthage), race, poverty, birthweight (bweight) and maternal smoking (smoke). For this analysis, the method used in order to construct the optimum model is backward elimination, using 0.05 significance level.

Backward Elimination method in principle has the following steps:

1. Build the maximum model.
2. Remove the attribute with the higher p-value.
3. Repeat step 2 until all the attributes in the model have p-values less than or equal to the 0.05 significance level.

However, in this case some additional rules apply, because the maximum Cox regression model, apart from the variables themselves, includes all the possible interactions. All the interactions are treated first until the significant interactions remain in the model. Furthermore, after dealing with the interactions, the p-values of all attributes are being investigated considering the 0.05 significance level. Additionally, it is important to point out that all the variables that are involved in a significant interaction with another variable cannot be removed from the model, because otherwise the interaction will be removed too and all the significant aspects are needed in the model in order to build the best model possible. So, the initial and the final models are shown below (see in the appendix the R output of the whole process):

➤ Initial Model:

```
model1 = coxph(Surv(chldage,status) ~
nsibs*(wmonth+mthage+factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
mthage*(factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
factor(race)*(factor(poverty)+factor(bweight)+factor(smoke))+
factor(poverty)*(factor(bweight)+factor(smoke))+
factor(bweight)*factor(smoke), data=pneumon)
anova(model1)
```

```
## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
```

	loglik	Chisq	Df	Pr(> Chi )	
## NULL	-8957.3				
## nsibs	-8943.5	27.6205	1	1.476e-07	***
## wmonth	-8942.8	1.2350	1	0.266430	
## mthage	-8843.5	198.6766	1	< 2.2e-16	***
## factor(race)	-8842.5	2.0356	2	0.361386	
## factor(poverty)	-8841.4	2.1914	1	0.138781	
## factor(bweight)	-8840.1	2.5743	1	0.108615	
## factor(smoke)	-8839.3	1.5650	2	0.457256	
## nsibs:wmonth	-8838.3	2.0451	1	0.152700	
## nsibs:mthage	-8836.0	4.6102	1	0.031783	*

```
## nsibs:factor(race) -8835.6 0.7539 2 0.685935
## nsibs:factor(poverty) -8834.9 1.3595 1 0.243619
## nsibs:factor(bweight) -8834.8 0.3565 1 0.550440
## nsibs:factor(smoke) -8834.6 0.3611 2 0.834793
## wmonth:mthage -8831.6 6.0126 1 0.014204 *
## wmonth:factor(race) -8830.6 2.0194 2 0.364328
## wmonth:factor(poverty) -8829.2 2.7117 1 0.099618 .
## wmonth:factor(bweight) -8825.5 7.4019 1 0.006516 **
## wmonth:factor(smoke) -8825.4 0.1932 2 0.907920
## mthage:factor(race) -8824.8 1.1767 2 0.555243
## mthage:factor(poverty) -8824.8 0.0008 1 0.977833
## mthage:factor(bweight) -8824.7 0.1920 1 0.661257
## mthage:factor(smoke) -8823.5 2.3529 2 0.308367
## factor(race):factor(poverty) -8823.1 0.9024 2 0.636856
## factor(race):factor(bweight) -8819.4 7.3028 2 0.025955 *
## factor(race):factor(smoke) -8816.8 5.2414 4 0.263416
## factor(poverty):factor(bweight) -8816.7 0.2016 1 0.653410
## factor(poverty):factor(smoke) -8815.7 1.9566 2 0.375952
## factor(bweight):factor(smoke) -8809.5 12.5177 2 0.001913 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#### ➤ Final Model:

```
model23 = coxph(Surv(chldage,status) ~
nsibs+wmonth+mthage+factor(bweight)*factor(smoke), data=pneumon)
anova(model23)
```

#### ## Analysis of Deviance Table

```
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
```

```
##
##          loglik      Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.2664303
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(bweight) -8842.3 2.4252 1 0.1193975
## factor(smoke) -8841.4 1.7362 2 0.4197501
## factor(bweight):factor(smoke) -8833.2 16.5059 2 0.0002605 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model23)
```

#### ## Call:

```
## coxph(formula = Surv(chldage, status) ~ nsibs + wmonth + mthage +
##       factor(bweight) * factor(smoke), data = pneumon)
```

```
##
```

```
## n= 3470, number of events= 1124
```

```
##
```

```
##          coef exp(coef) se(coef)      z
```



```

## nsibs                0.01179    1.01186    0.03832    0.308
## wmonth               -0.06581    0.93631    0.01779   -3.699
## mthage               0.16861    1.18366    0.01208   13.964
## factor(bweight)1     -0.11319    0.89298    0.08780   -1.289
## factor(smoke)1       -0.11799    0.88870    0.09367   -1.260
## factor(smoke)2       -0.14469    0.86529    0.13815   -1.047
## factor(bweight)1:factor(smoke)1  0.48601    1.62581    0.14649    3.318
## factor(bweight)1:factor(smoke)2  0.60248    1.82665    0.19794    3.044
##
## Pr(>|z|)
## nsibs                0.758410
## wmonth               0.000217 ***
## mthage               < 2e-16 ***
## factor(bweight)1     0.197341
## factor(smoke)1       0.207791
## factor(smoke)2       0.294911
## factor(bweight)1:factor(smoke)1  0.000908 ***
## factor(bweight)1:factor(smoke)2  0.002336 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## exp(coef) exp(-coef) lower .95 upper .95
## nsibs      1.0119    0.9883    0.9386    1.0908
## wmonth      0.9363    1.0680    0.9042    0.9695
## mthage      1.1837    0.8448    1.1560    1.2120
## factor(bweight)1  0.8930    1.1198    0.7518    1.0607
## factor(smoke)1    0.8887    1.1252    0.7396    1.0678
## factor(smoke)2    0.8653    1.1557    0.6600    1.1344
## factor(bweight)1:factor(smoke)1  1.6258    0.6151    1.2200    2.1665
## factor(bweight)1:factor(smoke)2  1.8266    0.5475    1.2393    2.6924
##
## Concordance= 0.631 (se = 0.009 )
## Rsquare= 0.069 (max possible= 0.994 )
## Likelihood ratio test= 248.2 on 8 df, p=0
## Wald test = 248.7 on 8 df, p=0
## Score (logrank) test = 252.7 on 8 df, p=0

```

## Validating the probability assumptions

In this task, after building a Cox regression model, it is important to check if the assumptions for this kind of models are violated. The First assumption is the proportionality assumption, which assumes that given two observations with different values for the independent variables, the ratio of the hazard functions for those two observations does not depend on time, considering a multiplicative relationship between the hazard function and the log-linear function of the covariates. The second assumption suggests that there is a log-linear relationship between the independent variables and the hazard function. In the following code a Cox test is run in order to determine if the proportional hazards assumption of Cox regression is violated or not.

```
#checking the proportionality assumption
cox.zph(model23)
```

```
##              rho      chisq      p
## nsibs          -0.00622  0.0454 0.8313
## wmonth          0.07196  5.6500 0.0175
## mthage          0.02498  0.7119 0.3988
## factor(bweight)1 0.00482  0.0260 0.8719
## factor(smoke)1    0.01366  0.2106 0.6463
## factor(smoke)2    0.02652  0.7957 0.3724
## factor(bweight)1:factor(smoke)1 0.02863  0.9214 0.3371
## factor(bweight)1:factor(smoke)2 -0.02650  0.7933 0.3731
## GLOBAL              NA 10.9643 0.2037
```

Given the results of the test performed above, all the p-values of the variables included in the final model are not statistically significant, which means greater than 0.05. Moreover, considering the global p-value, we are led to the same conclusion. Therefore, the hazards are proportional and the assumption is not violated.

## Summary

By looking at the summary of the final Cox regression model produced using the backward elimination method, the hazard ratios are provided under the column labeled as `exp(coef)`. Essentially, these hazard ratios describe what effect each variable included in the model has on hazard. For the number of siblings the hazard ratio is 1.01186, which means that the number of siblings of a baby does not increase the hazard of getting pneumonia dramatically but it can increase it by 1.01186. In the case of weaning month, the hazard ratio is 0.93631, which means that weaning month slightly decreases the hazard of getting pneumonia by 0.93631. However, in the case if mothers age, the hazard ratio shows that this attribute can increase hazard by 1.18366. Furthermore, regarding the weight of the baby and if the mother was smoking or not, the hazard ratios show that these variables do not increase hazard by 0.89298 and 0.88870 respectively.

The hazard ratio of `nsibs` is 1.0119 and confidence interval (0.9386,1.0908). This suggests that is better to have less siblings. For `wmonth` the hazard ratio is 0.9363 and the confidence interval (0.9042,0.9695) which means that is preferred that the child to be weaned older. For `mthage` the hazard ratio is 1.1837 and the confidence interval (1.1560,1.2120), which suggests that a younger mother considered to be a good factor that lowers the risks. Additionally, for `bweight` and the two levels of smoke the hazard ratios are 0.8930, 0.8887 and 0.8653. Nevertheless, there is an interaction between them, which gives hazard ratios 1.6258 and 1.8266. Therefore, the likelihood of the event happening can be determined by looking the hazard rates provided in the model summary. In general, if the hazard ratios are greater than 1 then the more likely it is the event will happen, unless the hazard ratios are less than 1 then the less likely it is that the event will happen.

Finally, by considering the output that the test of the proportional hazards assumption of Cox regression provided, it is not statistically significant for every covariate in the final model because all the p-values are greater than 0.05. The same conclusion applies to the global result of the test, which was also not significant. Therefore, considering all the above the proportional hazards assumption is not violated, which suggests that all are proportionally distributed over time.

## Appendix

### R source code including the output:

```
#Pneumonia Survival Analysis

#Author: Ioannis Matzakos Chorianopoulos

require(survival)
require(KMsurv)
data(pneumon)

#creating the status variable and including it in the dataset
pneumon$status = (pneumon$chldage<12) * 1 + 0

#recoding nsibs variable so that any number greater than 3 will get the value of 3
numsibs = pneumon$nsibs

pneumon$nsibs[pneumon$nsibs>=3] = 3

#recoding wmonth, ages 4-6 months into value 4 and ages 7+ months into value 5
wage = pneumon$wmonth

pneumon$wmonth[pneumon$wmonth>=4 & pneumon$wmonth<=6] = 4
pneumon$wmonth[wage>=7] = 5

#the original values of pneumon$nsibs and pneumon$wmonth are assigned to new attributes outside the dataset, numsibs and wage respectively, just in case they are needed later for any reason.

#fitting the model
model = survfit(Surv(chldage, status) ~ nsibs, data=pneumon, type = "kaplan")
summary(model)

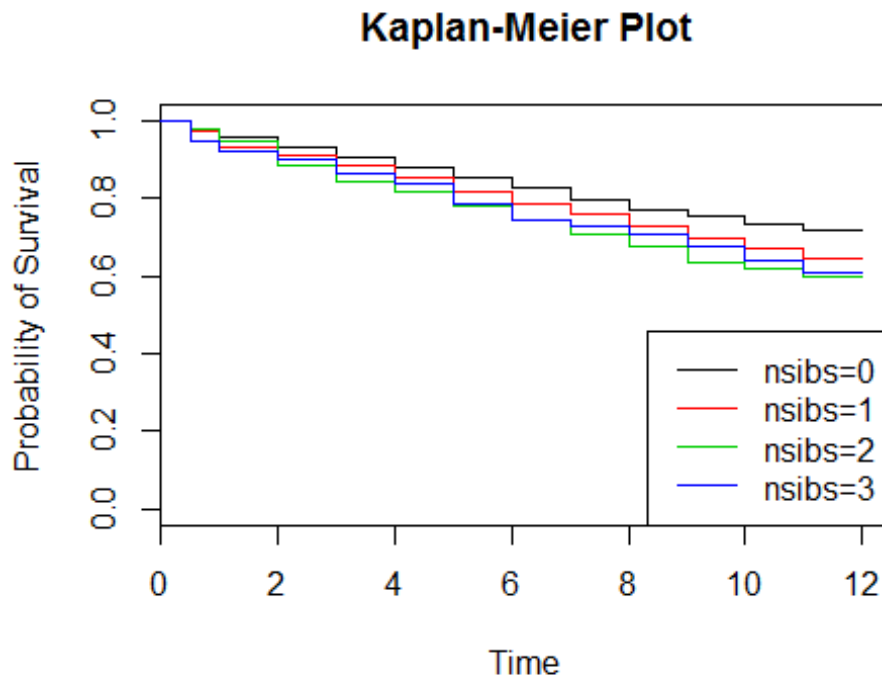
## Call: survfit(formula = Surv(chldage, status) ~ nsibs, data = pneumon,
##      type = "kaplan")
##
##               nsibs=0
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   0.5   1801     35   0.981 0.00325   0.974   0.987
##   1.0   1766     43   0.957 0.00480   0.947   0.966
##   2.0   1723     45   0.932 0.00594   0.920   0.943
##   3.0   1678     48   0.905 0.00691   0.892   0.919
##   4.0   1630     42   0.882 0.00761   0.867   0.897
##   5.0   1588     54   0.852 0.00837   0.835   0.868
##   6.0   1534     48   0.825 0.00895   0.808   0.843
##   7.0   1486     50   0.797 0.00947   0.779   0.816
##   8.0   1436     47   0.771 0.00990   0.752   0.791
```

```

## 9.0 1389 33 0.753 0.01016 0.733 0.773
## 10.0 1356 35 0.733 0.01042 0.713 0.754
## 11.0 1321 30 0.717 0.01062 0.696 0.738
##
## nsibs=1
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.5 1156 33 0.971 0.00490 0.962 0.981
## 1.0 1123 46 0.932 0.00742 0.917 0.946
## 2.0 1077 26 0.909 0.00845 0.893 0.926
## 3.0 1051 27 0.886 0.00935 0.868 0.904
## 4.0 1024 40 0.851 0.01047 0.831 0.872
## 5.0 984 39 0.817 0.01136 0.796 0.840
## 6.0 945 34 0.788 0.01202 0.765 0.812
## 7.0 911 31 0.761 0.01254 0.737 0.786
## 8.0 880 35 0.731 0.01304 0.706 0.757
## 9.0 845 37 0.699 0.01349 0.673 0.726
## 10.0 808 31 0.672 0.01381 0.646 0.700
## 11.0 777 31 0.645 0.01407 0.618 0.674
##
## nsibs=2
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.5 383 9 0.977 0.00774 0.961 0.992
## 1.0 374 12 0.945 0.01163 0.923 0.968
## 2.0 362 24 0.883 0.01645 0.851 0.915
## 3.0 338 15 0.843 0.01857 0.808 0.881
## 4.0 323 11 0.815 0.01986 0.777 0.854
## 5.0 312 13 0.781 0.02114 0.740 0.823
## 6.0 299 14 0.744 0.02230 0.702 0.789
## 7.0 285 14 0.708 0.02324 0.663 0.755
## 8.0 271 12 0.676 0.02391 0.631 0.725
## 9.0 259 15 0.637 0.02457 0.591 0.687
## 10.0 244 7 0.619 0.02482 0.572 0.669
## 11.0 237 7 0.601 0.02503 0.553 0.652
##
## nsibs=3
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 0.5 130 7 0.946 0.0198 0.908 0.986
## 1.0 123 3 0.923 0.0234 0.878 0.970
## 2.0 120 3 0.900 0.0263 0.850 0.953
## 3.0 117 5 0.862 0.0303 0.804 0.923
## 4.0 112 3 0.838 0.0323 0.778 0.904
## 5.0 109 7 0.785 0.0361 0.717 0.859
## 6.0 102 5 0.746 0.0382 0.675 0.825
## 7.0 97 2 0.731 0.0389 0.658 0.811
## 8.0 95 3 0.708 0.0399 0.634 0.790
## 9.0 92 4 0.677 0.0410 0.601 0.762
## 10.0 88 5 0.638 0.0421 0.561 0.727
## 11.0 83 4 0.608 0.0428 0.529 0.698

```

```
#plotting the fitted model
par(mfrow=c(1,1))
plot(model, main="Kaplan-Meier Plot", xlab="Time", ylab="Probability of Survival", col=c(1,2,3,4), lwd=1)
legend('bottomright', c("nsibs=0","nsibs=1","nsibs=2","nsibs=3"), lty=1, col=c(1,2,3,4))
```



```
#log-rank test using survdiff
survdiff(Surv(chldage, status) ~ nsibs, rho = 0, data=pneumon)

## Call:
## survdiff(formula = Surv(chldage, status) ~ nsibs, data = pneumon,
##          rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## nsibs=0 1801      510     597.1    12.71    28.00
## nsibs=1 1156      410     369.2     4.51     6.93
## nsibs=2  383      153     117.3    10.86    12.53
## nsibs=3  130       51      40.4     2.79     2.99
##
##  Chisq= 31.9  on 3 degrees of freedom, p= 5.51e-07

#log-rank test using the summary of a cox regression model
summary(coxph(Surv(chldage, status) ~ as.factor(nsibs), data=pneumon))

## Call:
## coxph(formula = Surv(chldage, status) ~ as.factor(nsibs), data = pneumon)
```

```
##
## n= 3470, number of events= 1124
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## as.factor(nsibs)1 0.26663   1.30555  0.06633 4.019 5.83e-05 ***
## as.factor(nsibs)2 0.43084   1.53855  0.09219 4.673 2.96e-06 ***
## as.factor(nsibs)3 0.39756   1.48819  0.14687 2.707 0.00679 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## as.factor(nsibs)1     1.306      0.766     1.146     1.487
## as.factor(nsibs)2     1.539      0.650     1.284     1.843
## as.factor(nsibs)3     1.488      0.672     1.116     1.985
##
## Concordance= 0.544 (se = 0.008 )
## Rsquare= 0.009 (max possible= 0.994 )
## Likelihood ratio test= 31.21 on 3 df, p=7.679e-07
## Wald test = 31.59 on 3 df, p=6.386e-07
## Score (logrank) test = 31.92 on 3 df, p=5.438e-07

alpha = 0.05/2/6
alpha

## [1] 0.004166667

#Multiple Comparisons
#0-1
survdifff(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[pneumon$nsibs<2,])

## Call:
## survdifff(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[pneumon$nsibs < 2, ], rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=0 1801      510      568      6.02      16.3
## as.factor(nsibs)=1 1156      410      352      9.73      16.3
##
## Chisq= 16.3 on 1 degrees of freedom, p= 5.55e-05

#0-2
survdifff(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==0|pneumon$nsibs==2),])

## Call:
## survdifff(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[(pneumon$nsibs == 0 | pneumon$nsibs == 2),
## ], rho = 0)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## as.factor(nsibs)=0 1801      510      554      3.5      21.9
## as.factor(nsibs)=2  383      153      109     17.8      21.9
##
##  Chisq= 21.9  on 1 degrees of freedom, p= 2.83e-06

#0-3
survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==0|pneumon$nsibs==3),])

## Call:
## survdiff(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[(pneumon$nsibs == 0 | pneumon$nsibs == 3),
## ], rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=0 1801      510      525.4      0.454      7.37
## as.factor(nsibs)=3  130       51      35.6      6.711      7.37
##
##  Chisq= 7.4  on 1 degrees of freedom, p= 0.00662

#1-2
survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==1|pneumon$nsibs==2),])

## Call:
## survdiff(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[(pneumon$nsibs == 1 | pneumon$nsibs == 2),
## ], rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=1 1156      410      427      0.7      3.01
## as.factor(nsibs)=2  383      153      136      2.2      3.01
##
##  Chisq= 3  on 1 degrees of freedom, p= 0.0826

#1-3
survdif(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,
data=pneumon[(pneumon$nsibs==1|pneumon$nsibs==3),])

## Call:
## survdiff(formula = Surv(chldage, status) ~ as.factor(nsibs),
## data = pneumon[(pneumon$nsibs == 1 | pneumon$nsibs == 3),
## ], rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## as.factor(nsibs)=1 1156      410      415.6      0.0745      0.784
## as.factor(nsibs)=3  130       51      45.4      0.6811      0.784
##
##  Chisq= 0.8  on 1 degrees of freedom, p= 0.376
```

#2-3

```
survdifff(Surv(chldage, status) ~ as.factor(nsibs), rho = 0,  
data=pneumon[(pneumon$nsibs==2|pneumon$nsibs==3),])
```

```
## Call:
```

```
## survdifff(formula = Surv(chldage, status) ~ as.factor(nsibs),  
##      data = pneumon[(pneumon$nsibs == 2 | pneumon$nsibs == 3),  
##      ], rho = 0)
```

```
##  
##              N Observed Expected (O-E)^2/E (O-E)^2/V  
## as.factor(nsibs)=2 383      153    151.8   0.00977   0.0398  
## as.factor(nsibs)=3 130       51     52.2   0.02839   0.0398  
##
```

```
##  Chisq= 0  on 1 degrees of freedom, p= 0.842
```

*#Survival analysis using Cox regression*

*#Backward Elimination using 0.05 significance level*

```
modell1 = coxph(Surv(chldage,status) ~  
nsibs*(wmonth+mthage+factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
```

```
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
```

```
mthage*(factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
```

```
factor(race)*(factor(poverty)+factor(bweight)+factor(smoke))+  
              factor(poverty)*(factor(bweight)+factor(smoke))+  
              factor(bweight)*factor(smoke), data=pneumon)
```

```
anova(modell1)
```

```
## Analysis of Deviance Table
```

```
## Cox model: response is Surv(chldage, status)
```

```
## Terms added sequentially (first to last)
```

```
##
```

```
##              loglik      Chisq Df Pr(>|Chi|)  
## NULL              -8957.3  
## nsibs              -8943.5    27.6205  1  1.476e-07 ***  
## wmonth             -8942.8     1.2350  1   0.266430  
## mthage             -8843.5   198.6766  1  < 2.2e-16 ***  
## factor(race)       -8842.5     2.0356  2   0.361386  
## factor(poverty)    -8841.4     2.1914  1   0.138781  
## factor(bweight)    -8840.1     2.5743  1   0.108615  
## factor(smoke)      -8839.3     1.5650  2   0.457256  
## nsibs:wmonth       -8838.3     2.0451  1   0.152700  
## nsibs:mthage       -8836.0     4.6102  1   0.031783 *  
## nsibs:factor(race) -8835.6     0.7539  2   0.685935  
## nsibs:factor(poverty) -8834.9     1.3595  1   0.243619  
## nsibs:factor(bweight) -8834.8     0.3565  1   0.550440
```



```

## nsibs:factor(smoke) -8834.6 0.3611 2 0.834793
## wmonth:mthage -8831.6 6.0126 1 0.014204 *
## wmonth:factor(race) -8830.6 2.0194 2 0.364328
## wmonth:factor(poverty) -8829.2 2.7117 1 0.099618 .
## wmonth:factor(bweight) -8825.5 7.4019 1 0.006516 **
## wmonth:factor(smoke) -8825.4 0.1932 2 0.907920
## mthage:factor(race) -8824.8 1.1767 2 0.555243
## mthage:factor(poverty) -8824.8 0.0008 1 0.977833
## mthage:factor(bweight) -8824.7 0.1920 1 0.661257
## mthage:factor(smoke) -8823.5 2.3529 2 0.308367
## factor(race):factor(poverty) -8823.1 0.9024 2 0.636856
## factor(race):factor(bweight) -8819.4 7.3028 2 0.025955 *
## factor(race):factor(smoke) -8816.8 5.2414 4 0.263416
## factor(poverty):factor(bweight) -8816.7 0.2016 1 0.653410
## factor(poverty):factor(smoke) -8815.7 1.9566 2 0.375952
## factor(bweight):factor(smoke) -8809.5 12.5177 2 0.001913 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove mthage:factor(poverty) 0.977833
model2 = coxph(Surv(chldage,status) ~
nsibs*(wmonth+mthage+factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
mthage*(factor(race)+factor(bweight)+factor(smoke))+
factor(race)*(factor(poverty)+factor(bweight)+factor(smoke))+
factor(poverty)*(factor(bweight)+factor(smoke))+
factor(bweight)*factor(smoke), data=pneumon)
anova(model2)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
## loglik Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.266430
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(race) -8842.5 2.0356 2 0.361386
## factor(poverty) -8841.4 2.1914 1 0.138781
## factor(bweight) -8840.1 2.5743 1 0.108615
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:wmonth -8838.3 2.0451 1 0.152700
## nsibs:mthage -8836.0 4.6102 1 0.031783 *
## nsibs:factor(race) -8835.6 0.7539 2 0.685935
## nsibs:factor(poverty) -8834.9 1.3595 1 0.243619
## nsibs:factor(bweight) -8834.8 0.3565 1 0.550440

```

```

## nsibs:factor(smoke) -8834.6 0.3611 2 0.834793
## wmonth:mthage -8831.6 6.0126 1 0.014204 *
## wmonth:factor(race) -8830.6 2.0194 2 0.364328
## wmonth:factor(poverty) -8829.2 2.7117 1 0.099618 .
## wmonth:factor(bweight) -8825.5 7.4019 1 0.006516 **
## wmonth:factor(smoke) -8825.4 0.1932 2 0.907920
## mthage:factor(race) -8824.8 1.1767 2 0.555243
## mthage:factor(bweight) -8824.7 0.1923 1 0.661045
## mthage:factor(smoke) -8823.5 2.3532 2 0.308323
## factor(race):factor(poverty) -8823.1 0.8600 2 0.650500
## factor(race):factor(bweight) -8819.5 7.2908 2 0.026111 *
## factor(race):factor(smoke) -8816.8 5.2674 4 0.260947
## factor(poverty):factor(bweight) -8816.7 0.1780 1 0.673074
## factor(poverty):factor(smoke) -8815.8 1.9043 2 0.385909
## factor(bweight):factor(smoke) -8809.5 12.4987 2 0.001932 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove wmonth:factor(smoke) 0.907920
model3 = coxph(Surv(chldage,status) ~
nsibs*(wmonth+mthage+factor(race)+factor(poverty)+factor(bweight)+factor(smoke))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
mthage*(factor(race)+factor(bweight)+factor(smoke))+
factor(race)*(factor(poverty)+factor(bweight)+factor(smoke))+
factor(poverty)*(factor(bweight)+factor(smoke))+
factor(bweight)*factor(smoke), data=pneumon)
anova(model3)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
```

	loglik	Chisq	Df	Pr(> Chi )	
## NULL	-8957.3				
## nsibs	-8943.5	27.6205	1	1.476e-07	***
## wmonth	-8942.8	1.2350	1	0.266430	
## mthage	-8843.5	198.6766	1	< 2.2e-16	***
## factor(race)	-8842.5	2.0356	2	0.361386	
## factor(poverty)	-8841.4	2.1914	1	0.138781	
## factor(bweight)	-8840.1	2.5743	1	0.108615	
## factor(smoke)	-8839.3	1.5650	2	0.457256	
## nsibs:wmonth	-8838.3	2.0451	1	0.152700	
## nsibs:mthage	-8836.0	4.6102	1	0.031783	*
## nsibs:factor(race)	-8835.6	0.7539	2	0.685935	
## nsibs:factor(poverty)	-8834.9	1.3595	1	0.243619	
## nsibs:factor(bweight)	-8834.8	0.3565	1	0.550440	
## nsibs:factor(smoke)	-8834.6	0.3611	2	0.834793	

```

## wmonth:mthage -8831.6 6.0126 1 0.014204 *
## wmonth:factor(race) -8830.6 2.0194 2 0.364328
## wmonth:factor(poverty) -8829.2 2.7117 1 0.099618 .
## wmonth:factor(bweight) -8825.5 7.4019 1 0.006516 **
## mthage:factor(race) -8824.9 1.1768 2 0.555202
## mthage:factor(bweight) -8824.8 0.1957 1 0.658200
## mthage:factor(smoke) -8823.6 2.4264 2 0.297239
## factor(race):factor(poverty) -8823.2 0.8401 2 0.657013
## factor(race):factor(bweight) -8819.5 7.2935 2 0.026076 *
## factor(race):factor(smoke) -8816.9 5.3624 4 0.252091
## factor(poverty):factor(bweight) -8816.8 0.1770 1 0.673955
## factor(poverty):factor(smoke) -8815.8 1.9104 2 0.384743
## factor(bweight):factor(smoke) -8809.7 12.2887 2 0.002146 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove nsibs:factor(smoke) 0.834793
model4 = coxph(Surv(chldage,status) ~
nsibs*(wmonth+mthage+factor(race)+factor(poverty)+factor(bweight))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
mthage*(factor(race)+factor(bweight)+factor(smoke))+
factor(race)*(factor(poverty)+factor(bweight)+factor(smoke))+
factor(poverty)*(factor(bweight)+factor(smoke))+
factor(bweight)*factor(smoke), data=pneumon)
anova(model4)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
## loglik Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.266430
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(race) -8842.5 2.0356 2 0.361386
## factor(poverty) -8841.4 2.1914 1 0.138781
## factor(bweight) -8840.1 2.5743 1 0.108615
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:wmonth -8838.3 2.0451 1 0.152700
## nsibs:mthage -8836.0 4.6102 1 0.031783 *
## nsibs:factor(race) -8835.6 0.7539 2 0.685935
## nsibs:factor(poverty) -8834.9 1.3595 1 0.243619
## nsibs:factor(bweight) -8834.8 0.3565 1 0.550440
## wmonth:mthage -8831.7 6.0753 1 0.013708 *
## wmonth:factor(race) -8830.7 2.0123 2 0.365633
## wmonth:factor(poverty) -8829.3 2.7654 1 0.096321 .
## wmonth:factor(bweight) -8825.6 7.3743 1 0.006616 **

```

```

## mthage:factor(race) -8825.1 1.1713 2 0.556756
## mthage:factor(bweight) -8825.0 0.2072 1 0.648932
## mthage:factor(smoke) -8823.9 2.0217 2 0.363914
## factor(race):factor(poverty) -8823.5 0.8023 2 0.669555
## factor(race):factor(bweight) -8819.9 7.2297 2 0.026921 *
## factor(race):factor(smoke) -8817.2 5.5477 4 0.235565
## factor(poverty):factor(bweight) -8817.1 0.1778 1 0.673276
## factor(poverty):factor(smoke) -8816.1 1.9138 2 0.384085
## factor(bweight):factor(smoke) -8809.9 12.4392 2 0.001990 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove nsibs:factor(race) 0.685935
model5 = coxph(Surv(chldage,status) ~
nsibs*(wmonth+mthage+factor(poverty)+factor(bweight))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
mthage*(factor(race)+factor(bweight)+factor(smoke))+
factor(race)*(factor(poverty)+factor(bweight)+factor(smoke))+
factor(poverty)*(factor(bweight)+factor(smoke))+
factor(bweight)*factor(smoke), data=pneumon)
anova(model5)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
## loglik Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.266430
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(poverty) -8842.4 2.2201 1 0.136220
## factor(bweight) -8841.3 2.1222 1 0.145181
## factor(race) -8840.1 2.4590 2 0.292441
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:wmonth -8838.3 2.0451 1 0.152700
## nsibs:mthage -8836.0 4.6102 1 0.031783 *
## nsibs:factor(poverty) -8835.2 1.5599 1 0.211686
## nsibs:factor(bweight) -8834.9 0.5764 1 0.447708
## wmonth:mthage -8831.9 6.0343 1 0.014030 *
## wmonth:factor(race) -8830.9 2.0793 2 0.353587
## wmonth:factor(poverty) -8829.5 2.7670 1 0.096224 .
## wmonth:factor(bweight) -8825.8 7.4394 1 0.006381 **
## mthage:factor(race) -8825.1 1.2909 2 0.524433
## mthage:factor(bweight) -8825.0 0.1679 1 0.682012
## mthage:factor(smoke) -8824.0 2.0338 2 0.361720
## factor(poverty):factor(race) -8823.6 0.7849 2 0.675408
## factor(bweight):factor(race) -8820.2 6.9403 2 0.031113 *

```

```

## factor(race):factor(smoke)      -8817.4    5.5216  4    0.237841
## factor(poverty):factor(bweight) -8817.3    0.1906  1    0.662414
## factor(poverty):factor(smoke)   -8816.3    1.9438  2    0.378366
## factor(bweight):factor(smoke)   -8810.1   12.4423  2    0.001987 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove mthage:factor(bweight) 0.682012
model6 = coxph(Surv(chldage, status) ~
nsibs*(wmonth+mthage+factor(poverty)+factor(bweight))+

wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
      mthage*(factor(race)+factor(smoke))+

factor(race)*(factor(poverty)+factor(bweight)+factor(smoke))+
      factor(poverty)*(factor(bweight)+factor(smoke))+
      factor(bweight)*factor(smoke), data=pneumon)
anova(model6)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
##              loglik      Chisq Df Pr(>|Chi|)
## NULL                      -8957.3
## nsibs                      -8943.5   27.6205  1  1.476e-07 ***
## wmonth                     -8942.8    1.2350  1   0.266430
## mthage                     -8843.5  198.6766  1  < 2.2e-16 ***
## factor(poverty)            -8842.4    2.2201  1   0.136220
## factor(bweight)            -8841.3    2.1222  1   0.145181
## factor(race)                -8840.1    2.4590  2   0.292441
## factor(smoke)               -8839.3    1.5650  2   0.457256
## nsibs:wmonth                -8838.3    2.0451  1   0.152700
## nsibs:mthage                -8836.0    4.6102  1   0.031783 *
## nsibs:factor(poverty)       -8835.2    1.5599  1   0.211686
## nsibs:factor(bweight)       -8834.9    0.5764  1   0.447708
## wmonth:mthage               -8831.9    6.0343  1   0.014030 *
## wmonth:factor(race)          -8830.9    2.0793  2   0.353587
## wmonth:factor(poverty)       -8829.5    2.7670  1   0.096224 .
## wmonth:factor(bweight)       -8825.8    7.4394  1   0.006381 **
## mthage:factor(race)          -8825.1    1.2909  2   0.524433
## mthage:factor(smoke)         -8824.1    2.1159  2   0.347166
## factor(poverty):factor(race) -8823.7    0.7702  2   0.680383
## factor(bweight):factor(race) -8820.3    6.7682  2   0.033909 *
## factor(race):factor(smoke)   -8817.5    5.5360  4   0.236581
## factor(poverty):factor(bweight) -8817.4    0.1866  1   0.665756
## factor(poverty):factor(smoke) -8816.4    1.9915  2   0.369453
## factor(bweight):factor(smoke) -8810.2   12.3784  2   0.002052 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

#remove factor(poverty):factor(race) 0.680383
model7 = coxph(Surv(chldage,status) ~
nsibs*(wmonth+mthage+factor(poverty)+factor(bweight))+

wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
      mthage*(factor(race)+factor(smoke))+
      factor(race)*(factor(bweight)+factor(smoke))+
      factor(poverty)*(factor(bweight)+factor(smoke))+
      factor(bweight)*factor(smoke), data=pneumon)

anova(model7)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
##          loglik      Chisq Df Pr(>|Chi|)
## NULL                -8957.3
## nsibs                -8943.5   27.6205   1  1.476e-07 ***
## wmonth               -8942.8    1.2350   1   0.266430
## mthage               -8843.5  198.6766   1 < 2.2e-16 ***
## factor(poverty)      -8842.4    2.2201   1   0.136220
## factor(bweight)      -8841.3    2.1222   1   0.145181
## factor(race)         -8840.1    2.4590   2   0.292441
## factor(smoke)        -8839.3    1.5650   2   0.457256
## nsibs:wmonth         -8838.3    2.0451   1   0.152700
## nsibs:mthage         -8836.0    4.6102   1   0.031783 *
## nsibs:factor(poverty) -8835.2    1.5599   1   0.211686
## nsibs:factor(bweight) -8834.9    0.5764   1   0.447708
## wmonth:mthage        -8831.9    6.0343   1   0.014030 *
## wmonth:factor(race)   -8830.9    2.0793   2   0.353587
## wmonth:factor(poverty) -8829.5    2.7670   1   0.096224 .
## wmonth:factor(bweight) -8825.8    7.4394   1   0.006381 **
## mthage:factor(race)   -8825.1    1.2909   2   0.524433
## mthage:factor(smoke)  -8824.1    2.1159   2   0.347166
## factor(bweight):factor(race) -8820.5    7.1414   2   0.028136 *
## factor(race):factor(smoke) -8817.7    5.6297   4   0.228560
## factor(poverty):factor(bweight) -8817.5    0.3127   1   0.576018
## factor(poverty):factor(smoke) -8816.8    1.4702   2   0.479466
## factor(bweight):factor(smoke) -8810.5   12.6359   2   0.001804 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove factor(poverty):factor(bweight) 0.576018
model8 = coxph(Surv(chldage,status) ~
nsibs*(wmonth+mthage+factor(poverty)+factor(bweight))+

wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
      mthage*(factor(race)+factor(smoke))+
      factor(race)*(factor(bweight)+factor(smoke))+
      factor(poverty)*factor(smoke)+

```



```

                                factor(bweight)*factor(smoke), data=pneumon)
anova(model8)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
##           loglik      Chisq Df Pr(>|Chi|)
## NULL                -8957.3
## nsibs                -8943.5   27.6205   1  1.476e-07 ***
## wmonth              -8942.8    1.2350   1   0.266430
## mthage              -8843.5  198.6766   1  < 2.2e-16 ***
## factor(poverty)     -8842.4    2.2201   1   0.136220
## factor(bweight)     -8841.3    2.1222   1   0.145181
## factor(race)        -8840.1    2.4590   2   0.292441
## factor(smoke)       -8839.3    1.5650   2   0.457256
## nsibs:wmonth        -8838.3    2.0451   1   0.152700
## nsibs:mthage        -8836.0    4.6102   1   0.031783 *
## nsibs:factor(poverty) -8835.2    1.5599   1   0.211686
## nsibs:factor(bweight) -8834.9    0.5764   1   0.447708
## wmonth:mthage       -8831.9    6.0343   1   0.014030 *
## wmonth:factor(race)  -8830.9    2.0793   2   0.353587
## wmonth:factor(poverty) -8829.5    2.7670   1   0.096224 .
## wmonth:factor(bweight) -8825.8    7.4394   1   0.006381 **
## mthage:factor(race)  -8825.1    1.2909   2   0.524433
## mthage:factor(smoke) -8824.1    2.1159   2   0.347166
## factor(bweight):factor(race) -8820.5    7.1414   2   0.028136 *
## factor(race):factor(smoke) -8817.7    5.6297   4   0.228560
## factor(poverty):factor(smoke) -8816.9    1.6024   2   0.448800
## factor(bweight):factor(smoke) -8810.6   12.5067   2   0.001924 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove mthage:factor(race) 0.524433
model9 = coxph(Surv(chldage, status) ~
nsibs*(wmonth+mthage+factor(poverty)+factor(bweight))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
          mthage*factor(smoke)+
          factor(race)*(factor(bweight)+factor(smoke))+
          factor(poverty)*factor(smoke)+
          factor(bweight)*factor(smoke), data=pneumon)
anova(model9)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
##           loglik      Chisq Df Pr(>|Chi|)
## NULL                -8957.3

```

```
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.266430
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(poverty) -8842.4 2.2201 1 0.136220
## factor(bweight) -8841.3 2.1222 1 0.145181
## factor(race) -8840.1 2.4590 2 0.292441
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:wmonth -8838.3 2.0451 1 0.152700
## nsibs:mthage -8836.0 4.6102 1 0.031783 *
## nsibs:factor(poverty) -8835.2 1.5599 1 0.211686
## nsibs:factor(bweight) -8834.9 0.5764 1 0.447708
## wmonth:mthage -8831.9 6.0343 1 0.014030 *
## wmonth:factor(race) -8830.9 2.0793 2 0.353587
## wmonth:factor(poverty) -8829.5 2.7670 1 0.096224 .
## wmonth:factor(bweight) -8825.8 7.4394 1 0.006381 **
## mthage:factor(smoke) -8825.0 1.5347 2 0.464248
## factor(bweight):factor(race) -8822.0 6.0738 2 0.047983 *
## factor(race):factor(smoke) -8819.2 5.4571 4 0.243528
## factor(poverty):factor(smoke) -8818.5 1.4927 2 0.474097
## factor(bweight):factor(smoke) -8811.9 13.1237 2 0.001413 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#remove factor(poverty):factor(smoke) 0.474097
```

```
model10 = coxph(Surv(chldage, status) ~
nsibs*(wmonth+mthage+factor(poverty)+factor(bweight))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
mthage*factor(smoke)+
factor(race)*(factor(bweight)+factor(smoke))+
factor(poverty)+
factor(bweight)*factor(smoke), data=pneumon)
anova(model10)
```

```
## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##          loglik      Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.266430
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(poverty) -8842.4 2.2201 1 0.136220
## factor(bweight) -8841.3 2.1222 1 0.145181
## factor(race) -8840.1 2.4590 2 0.292441
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:wmonth -8838.3 2.0451 1 0.152700
## nsibs:mthage -8836.0 4.6102 1 0.031783 *
## nsibs:factor(poverty) -8835.2 1.5599 1 0.211686
```



```

## nsibs:factor(bweight)          -8834.9    0.5764    1    0.447708
## wmonth:mthage                 -8831.9    6.0343    1    0.014030 *
## wmonth:factor(race)           -8830.9    2.0793    2    0.353587
## wmonth:factor(poverty)        -8829.5    2.7670    1    0.096224 .
## wmonth:factor(bweight)        -8825.8    7.4394    1    0.006381 **
## mthage:factor(smoke)          -8825.0    1.5347    2    0.464248
## factor(bweight):factor(race)  -8822.0    6.0738    2    0.047983 *
## factor(race):factor(smoke)    -8819.2    5.4571    4    0.243528
## factor(bweight):factor(smoke) -8813.0   12.5082    2    0.001923 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove mthage:factor(smoke) 0.464248
model11 = coxph(Surv(chldage, status) ~
nsibs*(wmonth+mthage+factor(poverty)+factor(bweight))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
      mthage+
      factor(race)*(factor(bweight)+factor(smoke))+
      factor(poverty)+
      factor(bweight)*factor(smoke), data=pneumon)
anova(model11)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##              loglik      Chisq Df Pr(>|Chi|)
## NULL                  -8957.3
## nsibs                 -8943.5   27.6205    1 1.476e-07 ***
## wmonth                -8942.8    1.2350    1  0.266430
## mthage                -8843.5  198.6766    1 < 2.2e-16 ***
## factor(poverty)       -8842.4    2.2201    1  0.136220
## factor(bweight)       -8841.3    2.1222    1  0.145181
## factor(race)          -8840.1    2.4590    2  0.292441
## factor(smoke)         -8839.3    1.5650    2  0.457256
## nsibs:wmonth          -8838.3    2.0451    1  0.152700
## nsibs:mthage          -8836.0    4.6102    1  0.031783 *
## nsibs:factor(poverty) -8835.2    1.5599    1  0.211686
## nsibs:factor(bweight) -8834.9    0.5764    1  0.447708
## wmonth:mthage         -8831.9    6.0343    1  0.014030 *
## wmonth:factor(race)    -8830.9    2.0793    2  0.353587
## wmonth:factor(poverty) -8829.5    2.7670    1  0.096224 .
## wmonth:factor(bweight) -8825.8    7.4394    1  0.006381 **
## factor(bweight):factor(race) -8822.6    6.3451    2  0.041896 *
## factor(race):factor(smoke) -8819.9    5.4154    4  0.247270
## factor(bweight):factor(smoke) -8813.6   12.5743    2  0.001860 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

#remove nsibs:factor(bweight) 0.447708
model12 = coxph(Surv(chldage,status) ~ nsibs*(wmonth+mthage+factor(poverty))+
wmonth*(mthage+factor(race)+factor(poverty)+factor(bweight))+
mthage+
factor(race)*(factor(bweight)+factor(smoke))+
factor(poverty)+
factor(bweight)*factor(smoke), data=pneumon)
anova(model12)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
##          loglik      Chisq Df Pr(>|Chi|)
## NULL          -8957.3
## nsibs          -8943.5    27.6205  1  1.476e-07 ***
## wmonth         -8942.8     1.2350  1   0.266430
## mthage         -8843.5   198.6766  1  < 2.2e-16 ***
## factor(poverty) -8842.4     2.2201  1   0.136220
## factor(race)    -8841.4     2.0069  2   0.366613
## factor(bweight) -8840.1     2.5743  1   0.108615
## factor(smoke)   -8839.3     1.5650  2   0.457256
## nsibs:wmonth    -8838.3     2.0451  1   0.152700
## nsibs:mthage    -8836.0     4.6102  1   0.031783 *
## nsibs:factor(poverty) -8835.2     1.5599  1   0.211686
## wmonth:mthage   -8832.2     5.9816  1   0.014456 *
## wmonth:factor(race) -8831.1     2.1575  2   0.340012
## wmonth:factor(poverty) -8829.7     2.8015  1   0.094174 .
## wmonth:factor(bweight) -8826.0     7.5088  1   0.006140 **
## factor(race):factor(bweight) -8822.8     6.4569  2   0.039618 *
## factor(race):factor(smoke) -8820.1     5.4121  4   0.247566
## factor(bweight):factor(smoke) -8813.8    12.4441  2   0.001985 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove wmonth:factor(race) 0.340012
model13 = coxph(Surv(chldage,status) ~ nsibs*(wmonth+mthage+factor(poverty))+
wmonth*(mthage+factor(poverty)+factor(bweight))+
mthage+
factor(race)*(factor(bweight)+factor(smoke))+
factor(poverty)+
factor(bweight)*factor(smoke), data=pneumon)
anova(model13)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
##          loglik      Chisq Df Pr(>|Chi|)

```

```

## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.266430
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(poverty) -8842.4 2.2201 1 0.136220
## factor(bweight) -8841.3 2.1222 1 0.145181
## factor(race) -8840.1 2.4590 2 0.292441
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:wmonth -8838.3 2.0451 1 0.152700
## nsibs:mthage -8836.0 4.6102 1 0.031783 *
## nsibs:factor(poverty) -8835.2 1.5599 1 0.211686
## wmonth:mthage -8832.2 5.9816 1 0.014456 *
## wmonth:factor(poverty) -8830.9 2.6060 1 0.106458
## wmonth:factor(bweight) -8827.5 6.8924 1 0.008656 **
## factor(bweight):factor(race) -8824.3 6.2875 2 0.043120 *
## factor(race):factor(smoke) -8821.6 5.4812 4 0.241391
## factor(bweight):factor(smoke) -8815.1 12.8811 2 0.001596 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove factor(race):factor(smoke) 0.241391
model14 = coxph(Surv(chldage,status) ~ nsibs*(wmonth+mthage+factor(poverty))+
               wmonth*(mthage+factor(poverty)+factor(bweight))+
               mthage+
               factor(race)*factor(bweight)+
               factor(poverty)+
               factor(bweight)*factor(smoke), data=pneumon)

anova(model14)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##          loglik      Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.266430
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(poverty) -8842.4 2.2201 1 0.136220
## factor(bweight) -8841.3 2.1222 1 0.145181
## factor(race) -8840.1 2.4590 2 0.292441
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:wmonth -8838.3 2.0451 1 0.152700
## nsibs:mthage -8836.0 4.6102 1 0.031783 *
## nsibs:factor(poverty) -8835.2 1.5599 1 0.211686
## wmonth:mthage -8832.2 5.9816 1 0.014456 *
## wmonth:factor(poverty) -8830.9 2.6060 1 0.106458
## wmonth:factor(bweight) -8827.5 6.8924 1 0.008656 **
## factor(bweight):factor(race) -8824.3 6.2875 2 0.043120 *
## factor(bweight):factor(smoke) -8818.9 10.8332 2 0.004442 **

```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove nsibs:factor(poverty) 0.211686
model15 = coxph(Surv(chldage,status) ~ nsibs*(wmonth+mthage)+
               wmonth*(mthage+factor(poverty)+factor(bweight))+
               mthage+
               factor(race)*factor(bweight)+
               factor(poverty)+
               factor(bweight)*factor(smoke), data=pneumon)
anova(model15)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
##          loglik      Chisq Df Pr(>|Chi|)
## NULL                    -8957.3
## nsibs                    -8943.5   27.6205  1  1.476e-07 ***
## wmonth                   -8942.8    1.2350  1   0.266430
## mthage                   -8843.5  198.6766  1  < 2.2e-16 ***
## factor(poverty)          -8842.4    2.2201  1   0.136220
## factor(bweight)          -8841.3    2.1222  1   0.145181
## factor(race)             -8840.1    2.4590  2   0.292441
## factor(smoke)            -8839.3    1.5650  2   0.457256
## nsibs:wmonth             -8838.3    2.0451  1   0.152700
## nsibs:mthage             -8836.0    4.6102  1   0.031783 *
## wmonth:mthage            -8832.9    6.1226  1   0.013347 *
## wmonth:factor(poverty)   -8831.7    2.5400  1   0.110991
## wmonth:factor(bweight)   -8828.2    6.8251  1   0.008989 **
## factor(bweight):factor(race) -8825.0    6.4834  2   0.039097 *
## factor(bweight):factor(smoke) -8819.7   10.6025  2   0.004985 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove nsibs:wmonth 0.152700
model16 = coxph(Surv(chldage,status) ~ nsibs*mthage+
               wmonth*(mthage+factor(poverty)+factor(bweight))+
               mthage+
               factor(race)*factor(bweight)+
               factor(poverty)+
               factor(bweight)*factor(smoke), data=pneumon)
anova(model16)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
##          loglik      Chisq Df Pr(>|Chi|)
## NULL                    -8957.3
## nsibs                    -8943.5   27.6205  1  1.476e-07 ***
```

```

## mthage -8851.7 183.4255 1 < 2.2e-16 ***
## wmonth -8843.5 16.4862 1 4.901e-05 ***
## factor(poverty) -8842.4 2.2201 1 0.136220
## factor(bweight) -8841.3 2.1222 1 0.145181
## factor(race) -8840.1 2.4590 2 0.292441
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:mthage -8837.6 3.4961 1 0.061513 .
## mthage:wmonth -8835.4 4.2468 1 0.039326 *
## wmonth:factor(poverty) -8834.1 2.7141 1 0.099467 .
## wmonth:factor(bweight) -8831.9 4.3562 1 0.036876 *
## factor(bweight):factor(race) -8828.8 6.1892 2 0.045293 *
## factor(bweight):factor(smoke) -8823.5 10.6719 2 0.004815 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove wmonth:factor(poverty) 0.099467
model17 = coxph(Surv(chldage,status) ~ nsibs*mthage+
                wmonth*(mthage+factor(bweight))+
                mthage+
                factor(race)*factor(bweight)+
                factor(poverty)+
                factor(bweight)*factor(smoke), data=pneumon)
anova(model17)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
## loglik Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## mthage -8851.7 183.4255 1 < 2.2e-16 ***
## wmonth -8843.5 16.4862 1 4.901e-05 ***
## factor(bweight) -8842.3 2.4252 1 0.119397
## factor(race) -8841.0 2.4963 2 0.287031
## factor(poverty) -8840.1 1.8798 1 0.170361
## factor(smoke) -8839.3 1.5650 2 0.457256
## nsibs:mthage -8837.6 3.4961 1 0.061513 .
## mthage:wmonth -8835.4 4.2468 1 0.039326 *
## wmonth:factor(bweight) -8833.6 3.6904 1 0.054728 .
## factor(bweight):factor(race) -8830.5 6.1800 2 0.045502 *
## factor(bweight):factor(smoke) -8825.1 10.8268 2 0.004457 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove nsibs:mthage 0.061513
model18 = coxph(Surv(chldage,status) ~ nsibs+
                wmonth*(mthage+factor(bweight))+
                mthage+
                factor(race)*factor(bweight)+

```

```

        factor(poverty)+
        factor(bweight)*factor(smoke), data=pneumon)
anova(model18)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##               loglik      Chisq Df Pr(>|Chi|)
## NULL                -8957.3
## nsibs                -8943.5   27.6205  1  1.476e-07 ***
## wmonth                -8942.8    1.2350  1   0.266430
## mthage                -8843.5  198.6766  1  < 2.2e-16 ***
## factor(bweight)       -8842.3    2.4252  1   0.119397
## factor(race)          -8841.0    2.4963  2   0.287031
## factor(poverty)       -8840.1    1.8798  1   0.170361
## factor(smoke)         -8839.3    1.5650  2   0.457256
## wmonth:mthage         -8837.4    3.7427  1   0.053040 .
## wmonth:factor(bweight) -8835.6    3.7609  1   0.052464 .
## factor(bweight):factor(race) -8832.5    6.1442  2   0.046325 *
## factor(bweight):factor(smoke) -8826.8   11.4791  2   0.003216 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove wmonth:mthage 0.053040
model19 = coxph(Surv(chldage,status) ~ nsibs+
                wmonth*factor(bweight)+
                mthage+
                factor(race)*factor(bweight)+
                factor(poverty)+
                factor(bweight)*factor(smoke), data=pneumon)
anova(model19)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##               loglik      Chisq Df Pr(>|Chi|)
## NULL                -8957.3
## nsibs                -8943.5   27.6205  1  1.476e-07 ***
## wmonth                -8942.8    1.2350  1   0.266430
## factor(bweight)       -8940.9    3.7861  1   0.051681 .
## mthage                -8842.3  197.3157  1  < 2.2e-16 ***
## factor(race)          -8841.0    2.4963  2   0.287031
## factor(poverty)       -8840.1    1.8798  1   0.170361
## factor(smoke)         -8839.3    1.5650  2   0.457256
## wmonth:factor(bweight) -8838.1    2.3633  1   0.124220
## factor(bweight):factor(race) -8835.1    6.0113  2   0.049506 *
## factor(bweight):factor(smoke) -8829.2   11.8249  2   0.002706 **

```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove wmonth:factor(bweight) 0.124220
model20 = coxph(Surv(chldage,status) ~ nsibs+
               wmonth+
               mthage+
               factor(race)*factor(bweight)+
               factor(poverty)+
               factor(bweight)*factor(smoke), data=pneumon)
anova(model20)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
```

	loglik	Chisq	Df	Pr(> Chi )	
## NULL	-8957.3				
## nsibs	-8943.5	27.6205	1	1.476e-07	***
## wmonth	-8942.8	1.2350	1	0.266430	
## mthage	-8843.5	198.6766	1	< 2.2e-16	***
## factor(race)	-8842.5	2.0356	2	0.361386	
## factor(bweight)	-8841.0	2.8859	1	0.089357	.
## factor(poverty)	-8840.1	1.8798	1	0.170361	
## factor(smoke)	-8839.3	1.5650	2	0.457256	
## factor(race):factor(bweight)	-8836.7	5.2744	2	0.071561	.
## factor(bweight):factor(smoke)	-8830.1	13.0695	2	0.001452	**

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove factor(race):factor(bweight) 0.071561
model21 = coxph(Surv(chldage,status) ~ nsibs+
               wmonth+
               mthage+
               factor(race)+
               factor(poverty)+
               factor(bweight)*factor(smoke), data=pneumon)
anova(model21)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##
```

	loglik	Chisq	Df	Pr(> Chi )	
## NULL	-8957.3				
## nsibs	-8943.5	27.6205	1	1.476e-07	***
## wmonth	-8942.8	1.2350	1	0.2664303	
## mthage	-8843.5	198.6766	1	< 2.2e-16	***
## factor(race)	-8842.5	2.0356	2	0.3613858	
## factor(poverty)	-8841.4	2.1914	1	0.1387807	
## factor(bweight)	-8840.1	2.5743	1	0.1086147	



```

## factor(smoke) -8839.3 1.5650 2 0.4572559
## factor(bweight):factor(smoke) -8831.1 16.5049 2 0.0002606 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove factor(race) 0.3613858
model22 = coxph(Surv(chldage,status) ~ nsibs+
                wmonth+
                mthage+
                factor(poverty)+
                factor(bweight)*factor(smoke), data=pneumon)
anova(model22)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##              loglik      Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.2664303
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(poverty) -8842.4 2.2201 1 0.1362200
## factor(bweight) -8841.3 2.1222 1 0.1451805
## factor(smoke) -8840.6 1.5546 2 0.4596526
## factor(bweight):factor(smoke) -8832.2 16.6211 2 0.0002459 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#remove factor(poverty) 0.1362200
model23 = coxph(Surv(chldage,status) ~
                nsibs+wmonth+mthage+factor(bweight)*factor(smoke), data=pneumon)
anova(model23)

## Analysis of Deviance Table
## Cox model: response is Surv(chldage, status)
## Terms added sequentially (first to last)
##
##              loglik      Chisq Df Pr(>|Chi|)
## NULL -8957.3
## nsibs -8943.5 27.6205 1 1.476e-07 ***
## wmonth -8942.8 1.2350 1 0.2664303
## mthage -8843.5 198.6766 1 < 2.2e-16 ***
## factor(bweight) -8842.3 2.4252 1 0.1193975
## factor(smoke) -8841.4 1.7362 2 0.4197501
## factor(bweight):factor(smoke) -8833.2 16.5059 2 0.0002605 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model23)

```



```
## Call:
## coxph(formula = Surv(chldage, status) ~ nsibs + wmonth + mthage +
##       factor(bweight) * factor(smoke), data = pneumon)
##
##      n= 3470, number of events= 1124
##
##               coef exp(coef) se(coef)      z
## nsibs           0.01179   1.01186  0.03832  0.308
## wmonth          -0.06581   0.93631  0.01779 -3.699
## mthage           0.16861   1.18366  0.01208 13.964
## factor(bweight)1 -0.11319   0.89298  0.08780 -1.289
## factor(smoke)1    -0.11799   0.88870  0.09367 -1.260
## factor(smoke)2    -0.14469   0.86529  0.13815 -1.047
## factor(bweight)1:factor(smoke)1  0.48601   1.62581  0.14649  3.318
## factor(bweight)1:factor(smoke)2  0.60248   1.82665  0.19794  3.044
##
##               Pr(>|z|)
## nsibs           0.758410
## wmonth          0.000217 ***
## mthage           < 2e-16 ***
## factor(bweight)1 0.197341
## factor(smoke)1    0.207791
## factor(smoke)2    0.294911
## factor(bweight)1:factor(smoke)1 0.000908 ***
## factor(bweight)1:factor(smoke)2 0.002336 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## nsibs           1.0119   0.9883   0.9386   1.0908
## wmonth           0.9363   1.0680   0.9042   0.9695
## mthage           1.1837   0.8448   1.1560   1.2120
## factor(bweight)1 0.8930   1.1198   0.7518   1.0607
## factor(smoke)1    0.8887   1.1252   0.7396   1.0678
## factor(smoke)2    0.8653   1.1557   0.6600   1.1344
## factor(bweight)1:factor(smoke)1 1.6258   0.6151   1.2200   2.1665
## factor(bweight)1:factor(smoke)2 1.8266   0.5475   1.2393   2.6924
##
## Concordance= 0.631 (se = 0.009 )
## Rsquare= 0.069 (max possible= 0.994 )
## Likelihood ratio test= 248.2 on 8 df, p=0
## Wald test = 248.7 on 8 df, p=0
## Score (logrank) test = 252.7 on 8 df, p=0

#All the variables in the model are significant so the backward elimination process stops.
#Additionally, the variables involved in the factor(bweight):factor(smoke) interaction can not be removed,
#because their interaction is significant. So the final model is model23
```

*#checking the proportionality assumption*

`cox.zph(model23)`

```
##               rho      chisq      p
## nsibs          -0.00622  0.0454 0.8313
## wmonth          0.07196  5.6500 0.0175
## mthage          0.02498  0.7119 0.3988
## factor(bweight)1 0.00482  0.0260 0.8719
## factor(smoke)1    0.01366  0.2106 0.6463
## factor(smoke)2    0.02652  0.7957 0.3724
## factor(bweight)1:factor(smoke)1 0.02863  0.9214 0.3371
## factor(bweight)1:factor(smoke)2 -0.02650  0.7933 0.3731
## GLOBAL              NA 10.9643 0.2037
```

`plot(cox.zph(model23), col=c(2,4))`

