

Stat 574 Biostatistical Methods Final

Erik Erhardt

May 12, 2005

The National Longitudinal Study of Youth included 3,470 personal interviews conducted between 1979 and 1986. A portion of this data can be used to study whether a mother's feeding choice (breast feeding vs. never breast fed) has a protective effect on her child against hospitalization for pneumonia during the first year of life. The data for each mother-child pair consist of time to hospitalization with pneumonia or age at observation if hospitalization for pneumonia never occurred, as well as a collection of other covariates.

The question of primary interest in this assignment is whether there is a difference in survival time (which in this case is time until hospitalization due to pneumonia) for the group of children who were breastfed vs. those who were not. To adequately make this comparison, relevant covariates (from among the group in Table 1) must be included in the model.

1. *Use a model selection procedure to find a model which adjusts for relevant covariates using the Cox PH model.*

First, variables are coded. A `breastfed` indicator based on `weanage>0` is included. Group variables for `alcohol`, `region`, and `race` are added.

Table 1 on page 2 shows the best models selected by best subset. To save room, and because the results lend themselves to this, the table has been organized as if adding or removing one variable at a time to arrive at the best model with a given number of variables. Tables 2 and 3 on page 3 shows the best models arrived at from backward and forward selection. Stepwise selection yields the same model as forward selection. The result of backward selection does not include `breastfed`. The result of forward selection does not include variables that I would think would be important, such as `momage` and `weanage`.

I will use the results from the best subset selection for guidance at selecting a reduced model from the full model. Best subset necessarily tests every possible model and so considers many more possible models than either the forward or backward selection methods.

Table 4 on page 4 shows the reduced model I choose. Categorical variables were grouped in many ways, and `breastfed` was forced in the model.

The model seems reasonable, with variables that are somewhat intuitive but without redundant ones (such as `education` for `momage`). If `cigarettes` are smoked in the house, a person's immune system is compromised, also siblings carry additional germs, both increasing risk for pneumonia. If the mother is very young, simple ignorance can cause many problems.

Number of Variables	Score Chi-Square	Variables added/removed to Model
1	15.0369	breastfed
2	23.4174	cigarette0
3	28.7768	siblings
4	34.3902	momage
5	37.6754	race2
6	40.0508	region2
7	41.4167	region3 region4 (-region2)
8	43.0463	urban
9	44.0328	weanage
10	44.9232	birthweight
11	45.4488	alcohol1
12	45.8874	alcohol2 alcohol3 (-alcohol1)
13	46.2067	alcohol4
14	46.4142	alcohol1
15	46.6315	race3
16	46.7398	education
17	46.7636	region2
18	46.7676	foodage
19	46.7683	poverty

Table 1: Regression Models Selected by Score Criterion

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits
momage	1	-0.12271	0.04954	6.1354	0.0133	0.885	0.803 0.975
cigarette0	1	-0.69087	0.23488	8.6515	0.0033	0.501	0.316 0.794
siblings	1	0.39033	0.12150	10.3210	0.0013	1.477	1.164 1.875
weanage	1	-0.21083	0.08129	6.7265	0.0095	0.810	0.691 0.950

Summary of Backward Elimination

Step	Variable Removed	Number In	Wald Chi-Square	Pr > ChiSq
1	alcohol0	22	.	.
2	cigarette1	21	.	.
3	region1	20	.	.
4	race1	19	.	.
5	poverty	18	0.0003	0.9859
6	foodage	17	0.0009	0.9759
7	region2	16	0.0331	0.8557
8	alcohol1	15	0.2292	0.6321
9	education	14	0.1922	0.6611
10	alcohol4	13	0.3462	0.5563
11	race3	12	0.4337	0.5102
12	alcohol3	11	0.4418	0.5062
13	alcohol2	10	0.5821	0.4455
14	birthweight	9	0.7236	0.3950
15	breastfed	8	0.9172	0.3382
16	region4	7	1.5940	0.2068
17	race2	6	1.3960	0.2374
18	urban	5	2.7521	0.0971
19	region3	4	1.9387	0.1638

Table 2: Summary of Backward Elimination

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits
breastfed	1	-0.94416	0.30051	9.8711	0.0017	0.389	0.216 0.701
cigarette0	1	-0.67989	0.23565	8.3243	0.0039	0.507	0.319 0.804
education	1	-0.13744	0.05676	5.8640	0.0155	0.872	0.780 0.974

Summary of Forward Selection

Step	Variable Entered	Number In	Score Chi-Square	Pr > ChiSq
1	breastfed	1	15.0369	0.0001
2	cigarette0	2	10.3933	0.0013
3	education	3	5.7471	0.0165

Table 3: Summary of Forward and Stepwise Selection

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	1174.115	1136.791
AIC	1174.115	1144.791
SBC	1174.115	1153.953

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	37.3237	4	<.0001
Score	36.5682	4	<.0001
Wald	34.4511	4	<.0001

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits
breastfed	1	-0.87842	0.30235	8.4410	0.0037	0.415	0.230 0.751
cigarette0	1	-0.71427	0.23493	9.2441	0.0024	0.490	0.309 0.776
siblings	1	0.38291	0.12303	9.6868	0.0019	1.467	1.152 1.866
momage	1	-0.12110	0.04986	5.8990	0.0151	0.886	0.803 0.977

Table 4: Final reduced model

2. Once you have determined your model, test the assumption of proportional hazards for each of the covariates in the model. Stratify on any covariates which are found to violate the assumption of proportional hazards.

Table 5 on page 5 gives tests for the PH assumption. For all but `momage`, the PH assumption appears reasonable.

Analysis of Maximum Likelihood Estimates								
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
<code>breastfed</code>	1	-1.40327	0.48818	8.2628	0.0040	0.246	0.094	0.640
<code>cigarette0</code>	1	-0.12609	0.37847	0.1110	0.7390	0.882	0.420	1.851
<code>siblings</code>	1	0.05978	0.19974	0.0896	0.7647	1.062	0.718	1.570
<code>momage</code>	1	0.11492	0.06155	3.4858	0.0619	1.122	0.994	1.266
<code>Zbreastfed</code>	1	0.43958	0.34221	1.6500	0.1990	1.552	0.794	3.035
<code>Zcigarette0</code>	1	-0.43361	0.28012	2.3961	0.1216	0.648	0.374	1.122
<code>Zsiblings</code>	1	0.28020	0.15699	3.1858	0.0743	1.323	0.973	1.800
<code>Zmomage</code>	1	-0.38829	0.04231	84.2200	<.0001	0.678	0.624	0.737

The PHREG Procedure

Linear Hypotheses Testing Results

Label	Chi-Square	DF	Pr > ChiSq
TESTPH	89.2721	4	<.0001
TESTPHZBREASTFED	1.6500	1	0.1990
TESTPHZCIGARETTE0	2.3961	1	0.1216
TESTPHZSIBLINGS	3.1858	1	0.0743
TESTPHZMOMAGE	84.2200	1	<.0001

Table 5: Testing PH assumption

The plots in Figure 1 on page 6 show whether or not the PH assumption seems reasonable for the variables included in the model. Both `breastfed` and `cigarette0` show divergence at time 0. `siblings` indicates one outlier and does not seem to violate PH. There are so many categories for `momage` and it is not clear how to partition the set to realize the violation in PH that the Chi-Square test indicates. If `momage` is partitioned at age 20, `momage1` shows a stratification at time 3.

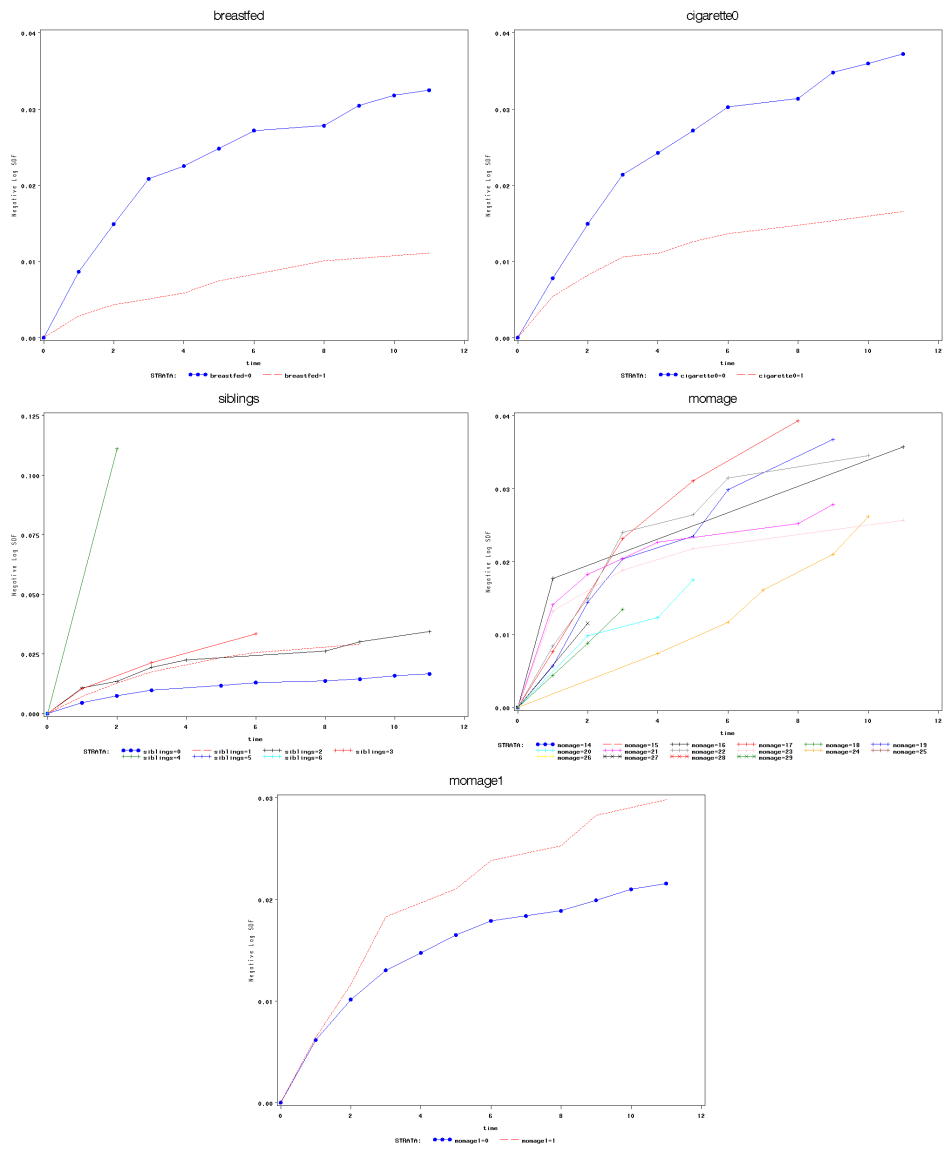


Figure 1: Checking PH assumption for all covariates.

3. *Using only the covariates that you found to be important in part (1), fit the most appropriate parametric model.*

Table 6 on page 7 gives the $\ln L$ and AIC values for five possible parametric models. GAMMA is listed twice, the first is one that was run to convergence, and the second only for 50 iterations. While the GAMMA (conv) model fits significantly better than all the other models, the parameter estimates are also very different and with large standard errors (reference Table 7 on page 8. I discard this model due to inconsistency with the other models and continue the remaining analysis using the Log Normal model as the choice model as it is the best fitting two parameter model.

Name of Distribution	Log Likelihood	AIC
GAMMA (conv)	-416.6692495	849.3385
GAMMA (not conv)	-421.7147999	859.4296
LNORMAL	-424.3139613	862.6279
LLOGISTIC	-427.5807444	869.1615
WEIBULL	-427.8000651	869.6001
EXPONENT	-432.2840295	876.5681

Table 6: Parametric model selection

The plots in Figure 2 on page 9 give the Nelson-Aalen estimator of the cumulative hazard function to compare what's expected under the parametric model, under a variety of models.

Three of the four plots show roughly parallel lines for each group, which is an indication the model fits well for those groups. The exponential distribution does not fit the data well.

The plot in Figure 3 on page 9 gives the log cum baseline hazard rates for each group. They are nearly "parallel" similar, so there does not appear to be strong evidence they are different.

The plot in Figure 4 on page 10 gives difference in log cum baseline hazard rates. For all but the first time point the curve is roughly a horizontal line at 1.15, so there is evidence the hazard rates are different, but proportional.

The plot in Figure 5 on page 10 gives the Anderson plot and this does not follow a 45 degree line, instead it is consistently below 45 degrees, but roughly linear.

The plots in Figure 6 on page 11 gives the score residual plots with horizontal axes `time` and `breastfed`. These are rather unsettling. There is a tight band of residuals around 0, with left skewness into negative values. At the same time there is a band around .75 of 14 "outliers".

```

Name of Distribution      GAMMA (conv)
Log Likelihood          -416.6692495

Standard
Variable  DF  Estimate  Error Chi-Square Pr > ChiSq Label
Intercept 1  -1.61882  1.65386  0.9581  0.3277 Intercept
breastfed 1  0.87534  0.32261  7.3621  0.0067
cigarette0 1  0.22532  0.33223  0.4599  0.4977
siblings 1  -0.60136  0.23034  6.8161  0.0090
momage 1  0.20335  0.07307  7.7449  0.0054
Scale 1  6.77848  1.65331  Gamma scale
Shape 1  -12.47096  5.56675  Gamma shape

Name of Distribution      GAMMA (not conv)
Log Likelihood          -421.7147999
WARNING: Iteration limit exceeded.

Standard
Variable  DF  Estimate  Error Chi-Square Pr > ChiSq Label
Intercept 1  5.19185  1.54486  11.2944  0.0008 Intercept
breastfed 1  1.16097  0.38683  9.0075  0.0027
cigarette0 1  0.85751  0.34731  6.0961  0.0135
siblings 1  -0.64161  0.20249  10.0400  0.0015
momage 1  0.18522  0.07543  6.0294  0.0141
Scale 1  5.87903  0.58632  Gamma scale
Shape 1  -1.26616  0.04822  Gamma shape

Name of Distribution      LNORMAL
Log Likelihood          -424.3139613

Standard
Variable  DF  Estimate  Error Chi-Square Pr > ChiSq Label
Intercept 1  4.95057  1.48360  11.1347  0.0008 Intercept
breastfed 1  1.18855  0.41105  8.3608  0.0038
cigarette0 1  0.94412  0.34621  7.4366  0.0064
siblings 1  -0.59099  0.19395  9.2850  0.0023
momage 1  0.17238  0.07426  5.3892  0.0203
Scale 1  3.34478  0.34915  Normal scale

Name of Distribution      LLOGISTIC
Log Likelihood          -427.5807444

Standard
Variable  DF  Estimate  Error Chi-Square Pr > ChiSq Label
Intercept 1  3.65211  1.39457  6.8581  0.0088 Intercept
breastfed 1  1.20566  0.43557  7.6617  0.0056
cigarette0 1  0.99000  0.34226  8.3666  0.0038
siblings 1  -0.53503  0.18193  8.6480  0.0033
momage 1  0.15846  0.07190  4.8568  0.0275
Scale 1  1.35951  0.15244  Logistic scale

Name of Distribution      WEIBULL
Log Likelihood          -427.8000651

Standard
Variable  DF  Estimate  Error Chi-Square Pr > ChiSq Label
Intercept 1  3.77162  1.38230  7.4447  0.0064 Intercept
breastfed 1  1.20410  0.43709  7.5891  0.0059
cigarette0 1  0.98804  0.34115  8.3878  0.0038
siblings 1  -0.52708  0.17915  8.6557  0.0033
momage 1  0.15651  0.07141  4.8032  0.0284
Scale 1  1.37531  0.15504  Extreme value scale

Name of Distribution      EXPONENT
Log Likelihood          -432.2840295

Standard
Variable  DF  Estimate  Error Chi-Square Pr > ChiSq Label
Intercept 1  3.49413  1.00273  12.1426  0.0005 Intercept
breastfed 1  0.87756  0.30241  8.4208  0.0037
cigarette0 1  0.72250  0.23492  9.4589  0.0021
siblings 1  -0.38444  0.12313  9.7490  0.0018
momage 1  0.10870  0.04989  4.7479  0.0293
Scale 0  1.00000  0  Extreme value scale

Lagrange Multiplier Statistics
Variable  Chi-Square  Pr > ChiSq
Scale 6.3545  0.0117

```

Table 7: Parametric model selection

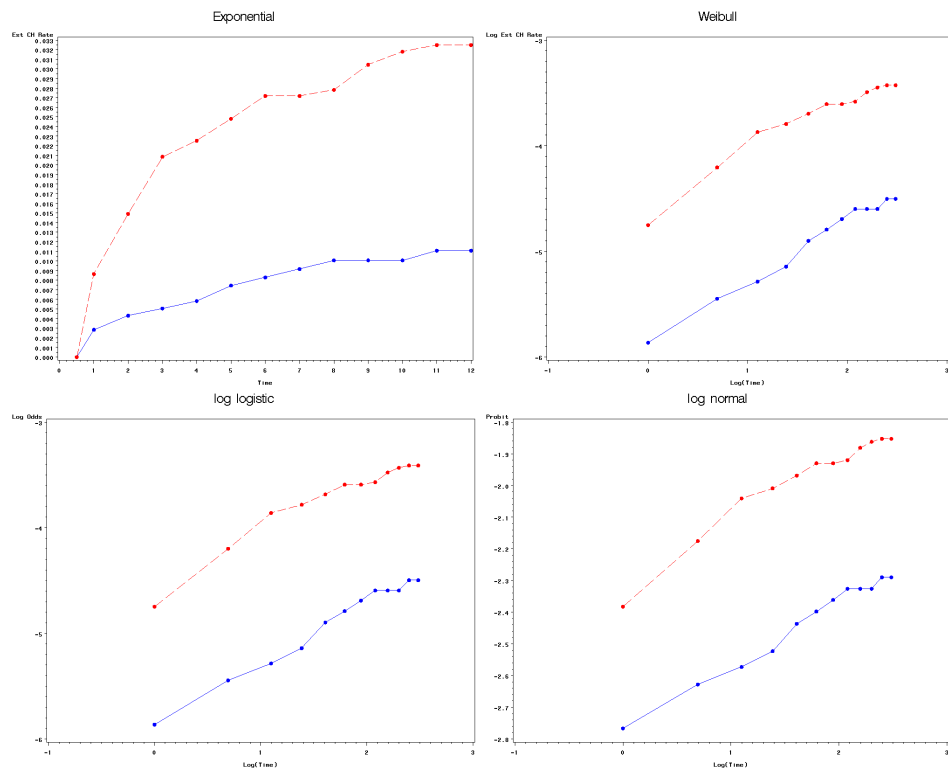


Figure 2: Graphical model fits under four models.

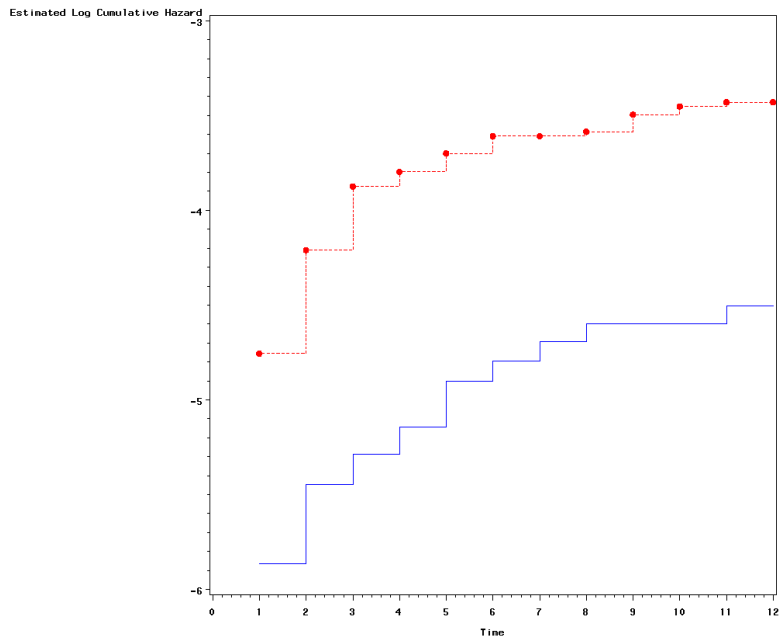


Figure 3: Log cum baseline hazard rates.

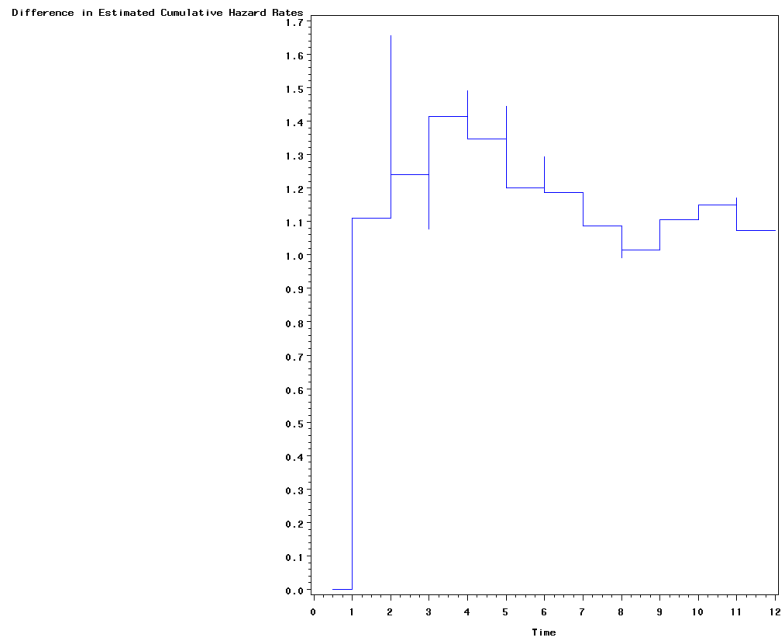


Figure 4: Difference in log cum baseline hazard rates.

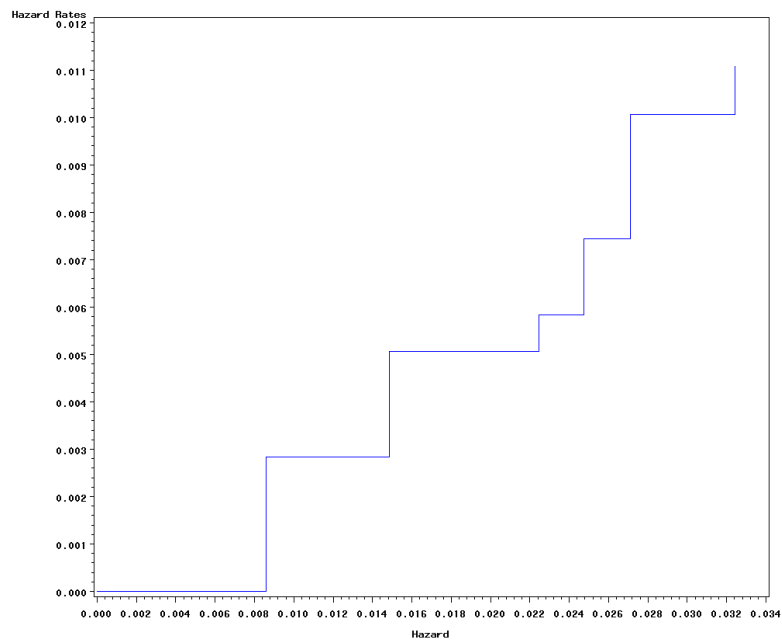


Figure 5: Anderson plot.

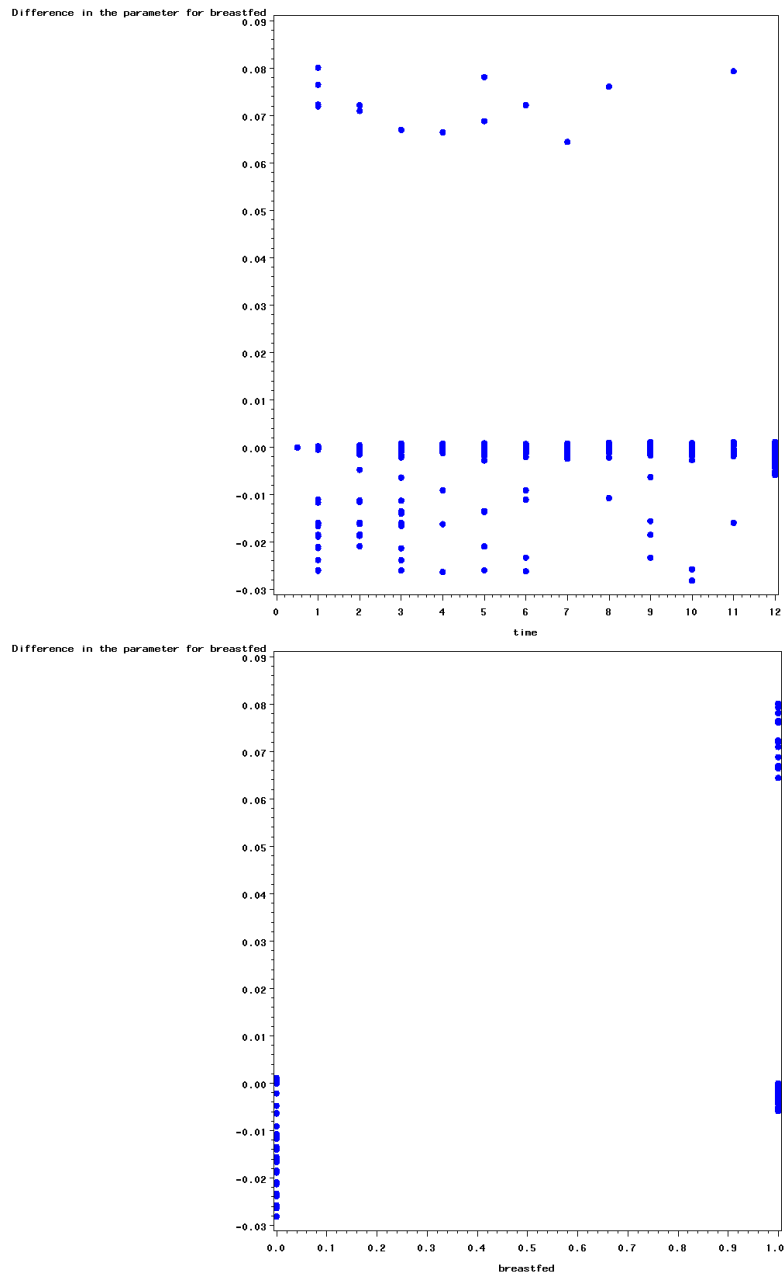


Figure 6: Score residuals.

4. *Using the results of all of your analysis, return to the question of interest: is there a difference in time to hospitalization due to pneumonia for those children who were breastfed vs. those who were not?*

After adjusting for other relevant covariates, based on the parameter estimates in Table 4 on page 4, whether a mother **breastfed** her infant is a significant factor in whether the event of hospitalization from pneumonia occurs. If an infant is **breastfed** from birth, the relative hazard for hospitalization from pneumonia is less than half (0.415 with 95% CI (0.230, 0.751)) the hazard for those not **breastfed**.

Appendix

SAS code used for the above analysis

```

options nocenter linesize=120 pagesize=60 nodate number;
*****;
/* Data */
data pneumonia;
  infile 'F:\USERS\Erik\UNM\biostat_stat574\pneumonia.txt';
  input time cens momage urban alcohol cigarette region poverty
        birthweight race education siblings weanage foodage;
  agecut=20;
  momage1=(momage<agecut);
  momage2=(momage>=agecut);
  nonph_time=2;
  znonph=momage2*(time>nonph_time);
  alcohol0=(alcohol=0);
  alcohol1=(alcohol=1);
  alcohol2=(alcohol=2);
  alcohol3=(alcohol=3);
  alcohol4=(alcohol>=4);
  alcohol01=(alcohol<=1);
  alcohol1234=(alcohol>=2);
  alcohol1234=(alcohol>=1);
  cigarette0=(cigarette=0);
  cigarette1=(cigarette>0);
  region1=(region=1);
  region2=(region=2);
  region3=(region=3);
  region4=(region=4);
  region12=(region<=2);
  region34=(region>=3);
  region124=(region<=2|region=4);
  race1=(race=1);
  race2=(race=2);
  race3=(race=3);
  race13=(race=1|race=3);
  breastfed=(weanage>0);
  * variables to test PH assumption;
  Zbreastfed =breastfed*log(time);
  Zcigarette0 =cigarette0*log(time);
  Zsiblings =siblings*log(time);
  Zmomage =momage*log(time);
  Zregion12 =region12*log(time);
  Zregion34 =region34*log(time);
run;
*proc print data=pneumonia;run;

*****;
/* Model Selection */
/* Best subset results */
title ;
proc phreg data=pneumonia;
  model time*cens(0) =
    breastfed momage urban
    alcohol1 alcohol2 alcohol3 alcohol4 alcohol0
    cigarette0 cigarette1
    region2 region3 region4 region1
    poverty birthweight
    race2 race3 race1
    education siblings weanage foodage
    / ties=efron risklimits SELECTION=SCORE best=1;
run;
/* Backward elimination results */
title ;
proc phreg data=pneumonia;
  model time*cens(0) =
    breastfed momage urban
    alcohol1 alcohol2 alcohol3 alcohol4 alcohol0
    cigarette0 cigarette1
    region2 region3 region4 region1
    poverty birthweight
    race2 race3 race1
    education siblings weanage foodage
    / ties=efron risklimits SELECTION=BACKWARD;
run;
/* Forward elimination results */
title ;

```

```

proc phreg data=pneumonia;
  model time*cens(0) =
    breastfed momage urban
    alcohol1 alcohol2 alcohol3 alcohol4 alcohol0
    cigarette0 cigarette1
    region2 region3 region4 region1
    poverty birthweight
    race2 race3 race1
    education siblings weanage foodage
  / ties=efron risklimits SELECTION=FORWARD;
run;

/* Stepwise elimination results */
title ;
proc phreg data=pneumonia;
  model time*cens(0) =
    breastfed momage urban
    alcohol1 alcohol2 alcohol3 alcohol4 alcohol0
    cigarette0 cigarette1
    region2 region3 region4 region1
    poverty birthweight
    race2 race3 race1
    education siblings weanage foodage
  / ties=efron risklimits SELECTION=STEPWISE;
run;

proc corr data=pneumonia; run;

/* Full model */
proc phreg data=pneumonia;
  model time*cens(0) =
    breastfed momage urban
    alcohol1 alcohol2 alcohol3 alcohol4 alcohol0
    cigarette0 cigarette1
    region2 region3 region4 region1
    poverty birthweight
    race2 race3 race1
    education siblings weanage foodage
  / ties=efron risklimits corrb;
run;

/* Reduced model results (may not include region12 region34) */
title ;
proc phreg data=pneumonia;
  model time*cens(0) =
    breastfed cigarette0 siblings momage
  / ties=efron risklimits;
  output out=pneumonia_diag dfbeta=dfbeta lmax=lmax
  resdev=resdev resmart=resmart ressch=ressch rescco=ressco
  logsurv=logsurv loglogs=loglogs xbeta=xbeta;
  id breastfed;
run;
*****;
/* PH assumption test for each covariate in model */

proc phreg data=pneumonia;
  model time*cens(0) =
    breastfed cigarette0 siblings momage
    Zbreastfed Zcigarette0 Zsiblings Zmomage
  / ties=efron risklimits;
  testph: test Zbreastfed=Zcigarette0=Zsiblings=Zmomage=0;
  testphZbreastfed : test Zbreastfed =0;
  testphZcigarette0: test Zcigarette0=0;
  testphZsiblings  : test Zsiblings  =0;
  testphZmomage    : test Zmomage    =0;
run;

/* Plot hazards -- appears momage changes slope at*/
title "breastfed";
proc lifetest data=pneumonia plots=(ls);
  time time*cens(0);
  strata breastfed;
run;
title "cigarette0";
proc lifetest data=pneumonia plots=(ls);
  time time*cens(0);
  strata cigarette0;
run;
title "siblings";
proc lifetest data=pneumonia plots=(ls);
  time time*cens(0);
  strata siblings;
run;
title "momage";
proc lifetest data=pneumonia plots=(ls);
  time time*cens(0);

```

```

strata momage;
run;
title "momage1";
proc lifetest data=pneumonia plots=(ls);
time time*cens(0);
strata momage1;
run;

title ;
*****;
proc lifereg data=pneumonia;
*model time*cens(0) =breastfed cigarette0 siblings momage / dist=weibull covb;
*model time*cens(0) =breastfed cigarette0 siblings momage / dist=exponential covb;
*model time*cens(0) =breastfed cigarette0 siblings momage / dist=llogistic covb;
*model time*cens(0) =breastfed cigarette0 siblings momage / dist=lnormal covb;
*model time*cens(0) =breastfed cigarette0 siblings momage / dist=gamma covb MAXITER=1000;
run;

aic.x=c(-416.6692495,-421.7147999,-424.3139613,-427.5807444,-427.8000651,-432.2840295);
p=c(3,3,2,2,2,1);k=5;
aic=-2*aic.x+2*(p+k)
t(t(aic))

/* 4 */
/* plots */
data pneumonia;
set pneumonia;
cons=1;
run;

proc phreg data=pneumonia;
model time*cens(0) = cons;
where breastfed=1;
output out=out_pneumonia1 survival=survival1 logsurv=logsurv1 loglogs=loglogs1;
run;

proc phreg data=pneumonia;
model time*cens(0) = cons;
where breastfed=0;
output out=out_pneumonia2 survival=survival2 logsurv=logsurv2 loglogs=loglogs2;
run;

proc print data=out_pneumonia1;run;
proc print data=out_pneumonia2;run;
proc sort data=out_pneumonia1;by time;run;
proc sort data=out_pneumonia2;by time;run;
data out_pneumonia_all;
merge out_pneumonia1 out_pneumonia2;
haz1=-logsurv1; * Exponential;
haz2=-logsurv2;
loghaz1=log(haz1); * Weibull;
loghaz2=log(haz2);
logtime=log(time);
lohaz1=log(exp(haz1)-1); * loglogistic;
lohaz2=log(exp(haz2)-1);
lnhaz1=probit(1-exp(-haz1)); * lognormal;
lnhaz2=probit(1-exp(-haz2));
by time;
run;
proc print data=out_pneumonia_all;run;
symbol1 v=dot i=j;
symbol2 v=dot i=j;
title "Exponential";
proc gplot data = out_pneumonia_all;
symbol i=j;
plot haz1*time = 1 haz2*time = 2 / overlay;
label haz1 = "Est CH Rate";
label time = "Time";
run;

title "Weibull";
proc gplot data = out_pneumonia_all;
symbol i=j;
plot loghaz1*logtime = 1 loghaz2*logtime = 2 / overlay;
label loghaz1 = "Log Est CH Rate";
label logtime = "Log(Time)";
run;

title "log logistic";
proc gplot data = out_pneumonia_all;
symbol i=j;
plot lohaz1*logtime = 1 lohaz2*logtime = 2 / overlay;
label lohaz1 = "Log Odds";

```

```

    label logtime = "Log(Time)";
run;
title "log normal";
proc gplot data = out_pneumonia_all;
    symbol i=j;
    plot ln haz1*logtime = 1 ln haz2*logtime = 2 / overlay;
    label ln haz1 = "Probit";
    label logtime = "Log(Time)";
run;

/* Diagnostitics */
/* Test PH assumption */
/* plot log cum hazard vs time for groups */
data pneumonia1;
    set pneumonia;
    cons=1;
run;
proc phreg data=pneumonia1;
    model time*cens(0) = cons;
    strata breastfed;
    output out=pneumonia_diag6 logsurv = ls / method=ch;
run;
proc print data=pneumonia_diag6;run;
data pneumonia_diag7;
    set pneumonia_diag6;
    if breastfed=1 then cumhaz1=log(-ls);
    if breastfed=0 then cumhaz2=log(-ls);
run;
proc print data=pneumonia_diag7;run;
proc sort data=pneumonia_diag7; by time;run;
title ;
proc gplot data=pneumonia_diag7;
    symbol1 i=stepjl v=none c=blue l=1;
    symbol2 i=stepjl c=red l=2;
    plot cumhaz1*time=1 cumhaz2*time=2 / overlay haxis=axis vaxis=axis2;
    axis label=('Time');
    axis2 label=('Estimated Log Cumulative Hazard' rotate=90);
run;
/* Look at difference in estimated cumulative hazards */
data pneumonia_diag8;
    set pneumonia_diag7;
    retain l1 l2 l3 l4 -5;
    if cumhaz1 ~= . then l1=cumhaz1;
    if cumhaz2 ~= . then l2=cumhaz2;
    diff1 = l2-l1;
run;
proc print data=pneumonia_diag8;
run;
proc gplot data=pneumonia_diag8;
    plot diff1*time=1 / overlay haxis=axis vaxis=axis2;
    axis label=('Time');
    axis2 label=('Difference in Estimated Cumulative Hazard Rates' rotate=90);
run;
/* Andersen plot */
data pneumonia_diag9;
    set pneumonia_diag8;
    retain haz1 haz2;
    if breastfed=1 then haz1=-ls;
    if breastfed=0 then haz2=-ls;
run;
proc gplot data=pneumonia_diag9;
    plot haz1*haz2=1 / overlay haxis=axis vaxis=axis2;
    axis label=('Hazard');
    axis2 label=('Hazard Rates' rotate=90);
run;
/* Plot for score residuals vs. risk scores */
proc gplot data=pneumonia_diag;
    symbol1 v=dot i=none;
    symbol2 v=none i=none;
    plot dfbeta*time;
    plot dfbeta*breastfed;
run;
quit;

```