Neil Yetz

03/09/2016

ERHS 642: Applied Logistic Regression

Dr. Bachand

**ERHS 642 Logistic Regression Spring 2016**

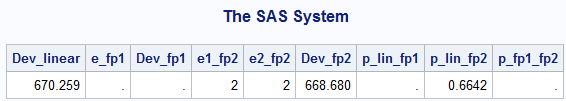
**Homework Assignment 5 – New Version**

Consider the BURN1000 data set, outcome variable DEATH, risk factor AGE.

1. Assess the scale of age using the fp method. Note: When the linear scale is best,

missing values appear in the fp results table for the best one-power model

Table 1.1: fp method output through Dr. Bachand’s fp macro.



1. What is the best one-power transformation? What is its deviance?

The Best one-power transformation is **Linear** and the Deviance= **670.259**

1. What is the best two-power transformation? What is its deviance?

The Best 2 Power Transformation is **AGE2 and AGE2ln(AGE)**  and the Deviance =**668.680**

c. Is the best two-power transformation significantly better than the best one-power

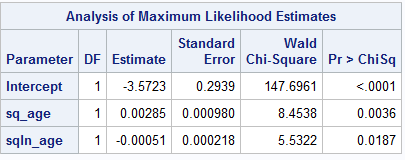
transformation?

**No,** it is not. This can be seen by looking at p\_lin\_fp2 which is a significance test comparing 2 power transformation deviance to the linear deviance. As can be seen, it is **NOT** significant (**p=0.6642)**.

d. Transform age using the best two-power transformation and use it in place of the

linear AGE variable in proc logistic.

Table 1.2: proc logistic analyses of maximum likelihood estimates of best 2 power transformation for age.



e. Obtain ORs and 95% CIs for an increase in AGE of 10 years starting at ages 10, 30,

50 and 70. **Interpret your results.**

g(age+10) – g(age) =

[β0 + β1(age+10)2 + β2(age+10)2ln(age+10)] – [β0 + β1(age)2 + β2(age)2ln(age)]

**β1[(age+10)2 – (age)2]+ β2[(age+10)2ln(age+10) – age2ln(age)]**

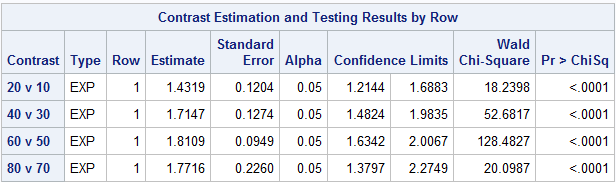
Table 1.3: Calculation for β1

|  |  |  |  |
| --- | --- | --- | --- |
| AGE | (AGE+10)^2 | (AGE)^2 | (AGE+10)^2-(AGE)^2 |
| 10 | 400 | 100 | 300 |
| 30 | 1600 | 900 | 700 |
| 50 | 3600 | 2500 | 1100 |
| 70 | 6400 | 4900 | 1500 |

Table 1.4: Calculation for β2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AGE | (AGE+10)^2 | ln(AGE+10) | (AGE+10)^2\*ln(AGE+10) | AGE^2 | ln(AGE) | (AGE)^2\*ln(AGE) | (AGE+10)^2\*ln(AGE+10)-(AGE)^2\*ln(AGE) |
| 10 | 400 | 2.995732274 | 1198.292909 | 100 | 2.302585 | 230.2585093 | 968.0344001 |
| 30 | 1600 | 3.688879454 | 5902.207127 | 900 | 3.401197 | 3061.077643 | 2841.129483 |
| 50 | 3600 | 4.094344562 | 14739.64042 | 2500 | 3.912023 | 9780.057514 | 4959.58291 |
| 70 | 6400 | 4.382026635 | 28044.97046 | 4900 | 4.248495 | 20817.62669 | 7227.343776 |

Table 1.5: Contrast table for OR’s & 95% CI’s for 10 year increase in AGE at ages 10, 30, 50, and 70



**Interpretation**: Looking at table 1.5, we can see that the effect of age with a 10 year increase on hospital discharge status of “dead” depends on where the individuals starting age was. For example, a 10 year increase of 10 from 10 years of age does not have as much of an effect as a 10 year increase from 50 years of age. It should also be noted, that a 10 year increase from age 50 on hospital discharge status of “dead” is the highest, even when compared to 80 vs 70. Although, please note the wider confidence intervals (especially at older ages) that may lead you to question the results.

f. Obtain ORs and 95% CIs for AGE= 30, 50 and 70 vs. AGE=10. **Interpret your**

**results.**

g(age) – g(10) =

[β0 + β1(age)2 + β2(age)2ln(age)] – [β0(10)2 + β2(10)2ln(10)]

**β1[(age)2 – 102 + β2[(age)2ln(age) – (10)2ln(10)]**

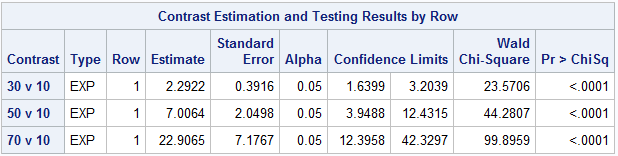
Table 1.6: Calculation for β1

|  |  |  |  |
| --- | --- | --- | --- |
| AGE | (AGE)^2 | 10^2 | Age^2-10^2 |
| 30 | 900 | 100 | 800 |
| 50 | 2500 | 100 | 2400 |
| 70 | 4900 | 100 | 4800 |

Table 1.7: Calculation for β2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AGE | Age^2 | ln(Age) | Age^2\*ln(AGE) | 10^2 | ln(10) | [(10^2)\*ln(10)] | [Age^2ln(Age)]-[(10^2)ln(10)] |
| 30 | 900 | 3.40119738 | 3061.07764 | 100 | 2.30258509 | 230.2585093 | 2830.819134 |
| 50 | 2500 | 3.91202301 | 9780.05751 | 100 | 2.30258509 | 230.2585093 | 9549.799004 |
| 70 | 4900 | 4.24849524 | 20817.6267 | 100 | 2.30258509 | 230.2585093 | 20587.36818 |

Table 1.5: Contrast table for OR’s & 95% CI’s in AGE at ages 30, 50, and 70 vs. Age = 10



**Interpretation:** There is definitely an increase of hospital discharge status of “dead” when ages 30, 50, and 70 are compared to age 10, all appear to be significantly different form age 10 as well. As the disparity gets larger (i.e 70 vs. 10) the 95% confidence interval gets larger. Inferring that the sample size including these two groups is most likely small and the results should be interpreted with caution.

2.

**a. Based on your results in question 1, how would you model age? Explain!**

Based on the results from question 1, I feel as though I would model age in a **linear** fashion for a couple of reasons.

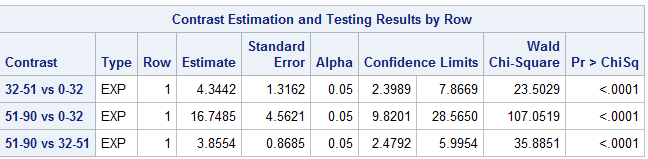
1. Looking at our fp procedure, it seems clear that the best option is linear. Yes, I notice that the deviance is lower for the 2-power transformation, but it is not significant by much. Because there is not much of a difference between the two, I would stick with linear because it is overall *easier to interpret*.
2. Once we ran the fp procedure and looked at the results, they seemed to really indicate a more linear trend as opposed to any real transformation.

a. Assume you create a 3-level age variable (0-32, 32-51, 51-90). Obtain OR(s) and

95% CI(s) and interpret your results. Use the following contrasts: 32-51 vs. 0-32, 51-

90 vs. 0-32 and 51-90 vs. 32-51.

Table 2.1: Contrast estimates of 32-51 vs 0-32, 51-90 vs 0-32 and 51-90 vs 32.51.



3. If you included age in a multivariate model, would you reassess the scale of age?

Explain!

Yes, you would have to reassess the scale of age. The reason is because the other variables that you put into your multivariate model are likely to have an effect of AGE on the outcome of DEATH. It is important to have your model determined before trying to assess the best scale to fit your variables.

SAS CODE

libname sdat 'C:\Users\ndyet\_000\Desktop\Class Folders\Spring 2016\ERHS 642\Data';

/\*data sdat.BURN1000; set BURN1000; run\*/

**data** BURN1000; set sdat.BURN1000;

sq\_age=age\*\***2**;

sqln\_age=sq\_age\*log(age);

if **0**<age<=**32** then a=**1**;

else if **32**< age<=**51** then a=**2**;

else if **51**< age<=**90** then a=**3**;

**run**;

**proc** **print** data=BURN1000;

var sq\_age sqln\_age;

**run**;

**proc** **means** data = burn1000; var AGE; **run**;

\*1a-c;

\*\* Macro for fp assessment \*\*;

**%macro** fp1(dset,y,var,lb,p1);

%do %until(&p1=**7**);

%put \*\*\*\*\* &p1 \*\*\*\*\*;

ODS output FitStatistics = mfs;

data fpdat; set &dset; if &var>&lb; pc=&p1/**2**;

if pc ne **0** then F1=&var\*\*pc; else if pc = **0** then F1=log(&var);

run;

proc logistic descending data=fpdat;

model &y=F1; \*-------------------F1 represents the variable being tested

for scale;

run;

data mfs; set mfs; if criterion='-2 Log L'; drop Criterion InterceptOnly;

run;

proc append data=mfs base=tres; run;

proc datasets; delete fpdat mfs; run;

quit;

%let p1=%eval(&p1+1);

%end;

**%mend** fp1;

%***fp1***(burn1000,death,age,**0**,-**4**); \*-----------Enter data set name, outcome

variable name and name of variable being tested for scale;

**data** pvals; do p1=-**4** to **6**; output; end; **run**;

**data** pvals; set pvals; p1=p1/**2**; **run**;

**data** tres; merge pvals tres; if p1 in (-**1.5**, **1.5**, **2.5**) then delete; **run**;

**proc** **sort** data=tres; by InterceptAndCovariates; **run**;

**data** tres; set tres; if \_N\_=**1** or p1=**1**; **run**;

**%macro** fp2(dset,y,var,lb,p1,p2);

%do %until(&p1=**7**);

%do %until(&p2=**7**);

%put \*\*\*\*\* &p1 &p2 \*\*\*\*\*;

ODS output FitStatistics = mfs;

data fpdat; set &dset; if &var>&lb; pc1=&p1/**2**; pc2=&p2/**2**;

if pc1 ne **0** then F1=&var\*\*pc1; else if pc1 = **0** then F1=log(&var);

if pc1 ne pc2 then do; if pc2 ne **0** then F2=&var\*\*pc2;

else if pc2 = **0** then F2=log(&var); end;

if pc1=pc2 then F2=F1\*log(&var);

run;

proc logistic descending data=fpdat;

model &y=F1 F2; \*------------F1 and F2 represent the variable being tested

for scale;

run;

data mfs; set mfs; if criterion='-2 Log L'; drop Criterion InterceptOnly;

run;

proc append data=mfs base=tres2; run;

proc datasets; delete fpdat mfs; run;

quit;

%let p2=%eval(&p2+1);

%end;

%let p2=%eval(-4);

%let p1=%eval(&p1+1);

%end;

**%mend** fp2;

%***fp2***(burn1000,death,age,**0**,-**4**,-**4**); \*-----------Enter data set name, outcome

variable name and name of variable being tested for scale;

**data** pvals2; do p1=-**4** to **6**; do p2=-**4** to **6**; output;end; end; **run**;

**data** pvals2; set pvals2; p1=p1/**2**; p2=p2/**2**; **run**;

**data** tres2; merge pvals2 tres2;

if p1 in (-**1.5**, **1.5**, **2.5**) or p2 in (-**1.5**, **1.5**, **2.5**) then delete; **run**;

**proc** **sort** data=tres2; by InterceptAndCovariates; **run**;

**data** tres2; set tres2; if \_N\_=**1**; **run**;

**data** comb; set tres tres2; **run**;

**data** c1; set comb; if p1=**1** and p2=**.**; rename

InterceptAndCovariates=Dev\_linear;

drop p1 p2; **run**;

**data** c2; set comb; if p1 ne **1** and p2=**.**; rename

InterceptAndCovariates=Dev\_fp1;

rename p1=e\_fp1; drop p2; **run**;

**data** c3; set comb; if p2 ne **.**; rename InterceptAndCovariates=Dev\_fp2;

rename p1=e1\_fp2; rename p2=e2\_fp2; **run**;

**data** c;

merge c1 c2 c3;

diff\_lin\_fp1=Dev\_linear-Dev\_fp1;

diff\_lin\_fp2=Dev\_linear-Dev\_fp2;

diff\_fp1\_fp2=Dev\_fp1-Dev\_fp2;

p\_lin\_fp1=**1**-probchi(diff\_lin\_fp1,**1**);

p\_lin\_fp2=**1**-probchi(diff\_lin\_fp2,**3**);

p\_fp1\_fp2=**1**-probchi(diff\_fp1\_fp2,**2**);

**run**;

**proc** **print** noobs data=c;

var Dev\_linear e\_fp1 Dev\_fp1 e1\_fp2 e2\_fp2 Dev\_fp2 p\_lin\_fp1 p\_lin\_fp2

p\_fp1\_fp2;

format p\_lin\_fp1 p\_lin\_fp2 p\_fp1\_fp2 **6.4**;

**run**;

**proc** **datasets**; delete tres tres2 pvals pvals2 comb c c1 c2 c3; **run**; **quit**;

\* End macro for fp assessment \*;

\*1d;

**proc** **logistic** descending data=burn1000;

model DEATH=sq\_age sqln\_age;

**RUN**;

**proc** **logistic** descending data=burn1000;

model DEATH=sq\_age sqln\_age;

contrast '20 v 10' sq\_age **300** sqln\_age **968.0344001**/estimate=exp;

contrast '40 v 30' sq\_age **700** sqln\_age **2841.129483**/estimate=exp;

contrast '60 v 50' sq\_age **1100** sqln\_age **4959.58291**/estimate=exp;

contrast '80 v 70' sq\_age **1500** sqln\_age **7227.343776**/estimate=exp;

**RUN**;

**proc** **logistic** descending data=burn1000;

model DEATH=sq\_age sqln\_age;

contrast '30 v 10' sq\_age **800** sqln\_age **2830.819134**/estimate=exp;

contrast '50 v 10' sq\_age **2400** sqln\_age **9549.799004**/estimate=exp;

contrast '70 v 10' sq\_age **4800** sqln\_age **20587.36818**/estimate=exp;

**RUN**;

\*2a;

**proc** **logistic** descending data=burn1000;

class a/param=ref ref=first;

model DEATH=a;

contrast '32-51 vs 0-32' a **1** **0**/estimate=exp;

contrast '51-90 vs 0-32' a **0** **1**/estimate=exp;

contrast '51-90 vs 32-51' a -**1** **1**/estimate=exp;

**run**;