Logistic Regression Using LRM

JuJu and Oliver

## Sample Data

The following code reads the titanic data that we will use in our examples.

 titanic = read.csv("http://facweb1.redlands.edu/fac/jim\_bentley/downloads/math111/titanic.csv")
 # titanic = read.csv("E:/Web16/Downloads/Math111/titanic.csv")
 titanic$AGE=factor(titanic$AGE,labels=c("Child","Adult"))
 titanic$CLASS=factor(titanic$CLASS,labels=c("0","1","2","3"))
 titanic$SEX=factor(titanic$SEX, labels=c("Female","Male"))
 titanic$SURVIVED=factor(titanic$SURVIVED,labels=c("No","Yes"))

We can now check to see if the data frame has been created by entering:

 ls()

## [1] "titanic"

## Loading R Packages

Additional functions necessary for validation and graphical analysis of the quality of logistic models can be found in Frank Harrell’s {} package. Harrell provides some functions to make pretty output in {}.

## load a few packages
 p\_load(Hmisc)
 p\_load(xtable)
 p\_load(lattice)
 p\_load(rms) ### Modern R replacement for Design package

## Fitting Logistic Models

The models fitted here are the equivalent of those fitted in the SAS documentation.

### CLASS

A model to test for the difference in odds of survival as determined by class may be fitted using the **lrm** function.

 dd = datadist(titanic)
 options(datadist="dd")
 titanic.lrm.class=lrm(SURVIVED~CLASS, x=TRUE, y=TRUE, data=titanic)
 titanic.lrm.class

## Logistic Regression Model
##
## lrm(formula = SURVIVED ~ CLASS, data = titanic, x = TRUE, y = TRUE)
##
## Model Likelihood Discrimination Rank Discrim.
## Ratio Test Indexes Indexes
## Obs 2201 LR chi2 180.90 R2 0.110 C 0.642
## No 1490 d.f. 3 g 0.545 Dxy 0.283
## Yes 711 Pr(> chi2) <0.0001 gr 1.725 gamma 0.386
## max |deriv| 6e-12 gp 0.124 tau-a 0.124
## Brier 0.200
##
## Coef S.E. Wald Z Pr(>|Z|)
## Intercept -1.1552 0.0788 -14.67 <0.0001
## CLASS=1 1.6643 0.1390 11.97 <0.0001
## CLASS=2 0.8078 0.1438 5.62 <0.0001
## CLASS=3 0.0678 0.1171 0.58 0.5624
##

 anova(titanic.lrm.class)

## Wald Statistics Response: SURVIVED
##
## Factor Chi-Square d.f. P
## CLASS 173.23 3 <.0001
## TOTAL 173.23 3 <.0001

Note that the (log) odds of survival do not differ for classes 0 (viewed as baseline) and 3. However, classes 1 and 2 differ from 0 (and thus 3) as well as from each other. This can most easily be seen using the odds ratios.

 summary(titanic.lrm.class,CLASS="0")

## Effects Response : SURVIVED
##
## Factor Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
## CLASS - 1:0 1 2 NA 1.664300 0.13902 1.39190 1.93680
## Odds Ratio 1 2 NA 5.282200 NA 4.02240 6.93660
## CLASS - 2:0 1 3 NA 0.807850 0.14375 0.52610 1.08960
## Odds Ratio 1 3 NA 2.243100 NA 1.69230 2.97310
## CLASS - 3:0 1 4 NA 0.067846 0.11711 -0.16169 0.29738
## Odds Ratio 1 4 NA 1.070200 NA 0.85071 1.34630

 summary(titanic.lrm.class,CLASS="3")

## Effects Response : SURVIVED
##
## Factor Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
## CLASS - 0:3 4 1 NA -0.067846 0.11711 -0.29738 0.16169
## Odds Ratio 4 1 NA 0.934400 NA 0.74276 1.17550
## CLASS - 1:3 4 2 NA 1.596500 0.14365 1.31500 1.87800
## Odds Ratio 4 2 NA 4.935700 NA 3.72460 6.54070
## CLASS - 2:3 4 3 NA 0.740000 0.14824 0.44946 1.03050
## Odds Ratio 4 3 NA 2.095900 NA 1.56750 2.80260

 summary(titanic.lrm.class,CLASS="1")

## Effects Response : SURVIVED
##
## Factor Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
## CLASS - 0:1 2 1 NA -1.66430 0.13902 -1.93680 -1.39190
## Odds Ratio 2 1 NA 0.18931 NA 0.14416 0.24861
## CLASS - 2:1 2 3 NA -0.85649 0.16609 -1.18200 -0.53097
## Odds Ratio 2 3 NA 0.42465 NA 0.30666 0.58804
## CLASS - 3:1 2 4 NA -1.59650 0.14365 -1.87800 -1.31500
## Odds Ratio 2 4 NA 0.20260 NA 0.15289 0.26849

 summary(titanic.lrm.class,CLASS="2")

## Effects Response : SURVIVED
##
## Factor Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
## CLASS - 0:2 3 1 NA -0.80785 0.14375 -1.08960 -0.52610
## Odds Ratio 3 1 NA 0.44582 NA 0.33635 0.59091
## CLASS - 1:2 3 2 NA 0.85649 0.16609 0.53097 1.18200
## Odds Ratio 3 2 NA 2.35490 NA 1.70060 3.26100
## CLASS - 3:2 3 4 NA -0.74000 0.14824 -1.03050 -0.44946
## Odds Ratio 3 4 NA 0.47711 NA 0.35681 0.63797

While the odds for class 3 relative to class 0 are essentially 1:1, class 1 has a 5.28:1 odds of survival and class 2 has a 2.24:1 odds of survival relative to class 0.

The probability of survival for the different classes may be plotted.

 print(plot(Predict(titanic.lrm.class, fun=plogis), ylab="Probability of Survival"))



A nomogram may be helpful at this point.

 nom = nomogram(titanic.lrm.class, fun=plogis)
 print(plot(nom))



## NULL

### AGE and SEX

A model to test for the difference in odds of survival as determined by age and sex may be fitted using the **lmr** function.

 dd = datadist(titanic)
 options(datadist="dd")
 titanic.lrm.agesex=lrm(SURVIVED~AGE\*SEX, x=TRUE, y=TRUE, data=titanic)
 titanic.lrm.agesex

## Logistic Regression Model
##
## lrm(formula = SURVIVED ~ AGE \* SEX, data = titanic, x = TRUE,
## y = TRUE)
##
## Model Likelihood Discrimination Rank Discrim.
## Ratio Test Indexes Indexes
## Obs 2201 LR chi2 456.68 R2 0.262 C 0.713
## No 1490 d.f. 3 g 0.841 Dxy 0.427
## Yes 711 Pr(> chi2) <0.0001 gr 2.320 gamma 0.787
## max |deriv| 1e-10 gp 0.187 tau-a 0.187
## Brier 0.171
##
## Coef S.E. Wald Z Pr(>|Z|)
## Intercept 0.4990 0.3075 1.62 0.1046
## AGE=Adult 0.5654 0.3269 1.73 0.0837
## SEX=Male -0.6870 0.3970 -1.73 0.0835
## AGE=Adult \* SEX=Male -1.7465 0.4167 -4.19 <0.0001
##

 anova(titanic.lrm.agesex)

## Wald Statistics Response: SURVIVED
##
## Factor Chi-Square d.f. P
## AGE (Factor+Higher Order Factors) 23.88 2 <.0001
## All Interactions 17.57 1 <.0001
## SEX (Factor+Higher Order Factors) 371.97 2 <.0001
## All Interactions 17.57 1 <.0001
## AGE \* SEX (Factor+Higher Order Factors) 17.57 1 <.0001
## TOTAL 391.59 3 <.0001

The odds associated with the model are:

 summary(titanic.lrm.agesex, AGE="Adult", SEX="Male")

## Effects Response : SURVIVED
##
## Factor Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
## AGE - Child:Adult 2 1 NA 1.1811 0.25839 0.67465 1.6875
## Odds Ratio 2 1 NA 3.2579 NA 1.96330 5.4060
## SEX - Female:Male 2 1 NA 2.4335 0.12669 2.18520 2.6818
## Odds Ratio 2 1 NA 11.3990 NA 8.89270 14.6120
##
## Adjusted to: AGE=Adult SEX=Male

 summary(titanic.lrm.agesex, AGE="Child", SEX="Female")

## Effects Response : SURVIVED
##
## Factor Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
## AGE - Adult:Child 1 2 NA 0.56540 0.32692 -0.075348 1.20620
## Odds Ratio 1 2 NA 1.76020 NA 0.927420 3.34060
## SEX - Male:Female 1 2 NA -0.68704 0.39698 -1.465100 0.09102
## Odds Ratio 1 2 NA 0.50306 NA 0.231050 1.09530
##
## Adjusted to: AGE=Child SEX=Female

The probability of survival for the different combinations of sex and age group may be plotted.

 Predict(titanic.lrm.agesex, fun=plogis, AGE=c("Child","Adult"), SEX=c("Female","Male"))

## AGE SEX yhat lower upper
## 1 Child Female 0.6222222 0.4741134 0.7505638
## 2 Adult Female 0.7435294 0.6998704 0.7828084
## 3 Child Male 0.4531250 0.3362143 0.5754460
## 4 Adult Male 0.2027594 0.1841419 0.2227454
##
## Response variable (y):
##
## Limits are 0.95 confidence limits

 print(plot(Predict(titanic.lrm.agesex, fun=plogis, AGE=c("Child","Adult"), SEX=c("Female","Male"))))



A nomogram may be helpful at this point.

 nom = nomogram(titanic.lrm.agesex, fun=plogis)
 print(plot(nom))



## NULL

### CLASS, AGE and SEX

A model to test for the difference in odds of survival as determined by class, age and sex may be fitted using the **lmr** function.

 dd = datadist(titanic)
 options(datadist="dd")
 titanic.lrm.classagesex=lrm(SURVIVED~CLASS\*SEX+AGE\*SEX, data=titanic, x=TRUE, y=TRUE)
 titanic.lrm.classagesex

## Logistic Regression Model
##
## lrm(formula = SURVIVED ~ CLASS \* SEX + AGE \* SEX, data = titanic,
## x = TRUE, y = TRUE)
##
## Model Likelihood Discrimination Rank Discrim.
## Ratio Test Indexes Indexes
## Obs 2201 LR chi2 634.70 R2 0.350 C 0.766
## No 1490 d.f. 9 g 1.341 Dxy 0.532
## Yes 711 Pr(> chi2) <0.0001 gr 3.823 gamma 0.638
## max |deriv| 5e-10 gp 0.233 tau-a 0.233
## Brier 0.157
##
## Coef S.E. Wald Z Pr(>|Z|)
## Intercept 2.0775 0.7171 2.90 0.0038
## CLASS=1 1.6642 0.8003 2.08 0.0376
## CLASS=2 0.0497 0.6874 0.07 0.9424
## CLASS=3 -2.0894 0.6381 -3.27 0.0011
## SEX=Male -1.7888 0.7728 -2.31 0.0206
## AGE=Adult -0.1803 0.3618 -0.50 0.6182
## CLASS=1 \* SEX=Male -1.1033 0.8199 -1.35 0.1784
## CLASS=2 \* SEX=Male -0.7647 0.7271 -1.05 0.2929
## CLASS=3 \* SEX=Male 1.5623 0.6562 2.38 0.0173
## SEX=Male \* AGE=Adult -1.3581 0.4551 -2.98 0.0028
##

 anova(titanic.lrm.classagesex)

## Wald Statistics Response: SURVIVED
##
## Factor Chi-Square d.f. P
## CLASS (Factor+Higher Order Factors) 124.28 6 <.0001
## All Interactions 48.25 3 <.0001
## SEX (Factor+Higher Order Factors) 254.38 5 <.0001
## All Interactions 63.47 4 <.0001
## AGE (Factor+Higher Order Factors) 31.30 2 <.0001
## All Interactions 8.91 1 0.0028
## CLASS \* SEX (Factor+Higher Order Factors) 48.25 3 <.0001
## SEX \* AGE (Factor+Higher Order Factors) 8.91 1 0.0028
## TOTAL INTERACTION 63.47 4 <.0001
## TOTAL 311.38 9 <.0001

The odds associated with the model are

 summary(titanic.lrm.classagesex, CLASS="0", AGE="Adult", SEX="Male")

## Effects Response : SURVIVED
##
## Factor Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
## CLASS - 1:0 1 2 NA 0.56086 0.17821 0.21158 0.91014
## Odds Ratio 1 2 NA 1.75220 NA 1.23560 2.48470
## CLASS - 2:0 1 3 NA -0.71504 0.23692 -1.17940 -0.25068
## Odds Ratio 1 3 NA 0.48917 NA 0.30747 0.77827
## CLASS - 3:0 1 4 NA -0.52708 0.15308 -0.82711 -0.22705
## Odds Ratio 1 4 NA 0.59033 NA 0.43731 0.79688
## SEX - Female:Male 2 1 NA 3.14690 0.62453 1.92290 4.37100
## Odds Ratio 2 1 NA 23.26400 NA 6.84040 79.11900
## AGE - Child:Adult 2 1 NA 1.53850 0.27608 0.99735 2.07960
## Odds Ratio 2 1 NA 4.65740 NA 2.71110 8.00100
##
## Adjusted to: CLASS=0 SEX=Male AGE=Adult

 summary(titanic.lrm.classagesex, CLASS="3", AGE="Child", SEX="Female")

## Effects Response : SURVIVED
##
## Factor Low High Diff. Effect S.E. Lower 0.95 Upper 0.95
## CLASS - 0:3 4 1 NA 2.08940 0.63814 0.83867 3.34010
## Odds Ratio 4 1 NA 8.08010 NA 2.31330 28.22300
## CLASS - 1:3 4 2 NA 3.75360 0.52986 2.71510 4.79210
## Odds Ratio 4 2 NA 42.67500 NA 15.10600 120.56000
## CLASS - 2:3 4 3 NA 2.13910 0.32956 1.49320 2.78500
## Odds Ratio 4 3 NA 8.49190 NA 4.45120 16.20100
## SEX - Male:Female 1 2 NA -0.22646 0.42412 -1.05770 0.60480
## Odds Ratio 1 2 NA 0.79735 NA 0.34725 1.83090
## AGE - Adult:Child 1 2 NA -0.18035 0.36179 -0.88945 0.52876
## Odds Ratio 1 2 NA 0.83498 NA 0.41088 1.69680
##
## Adjusted to: CLASS=3 SEX=Female AGE=Child

The probability of survival for the different combinations of sex and age group may be plotted.

 Predict(titanic.lrm.classagesex, fun=plogis, CLASS=c("0","1","2","3"), AGE=c("Child","Adult"), SEX=c("Female","Male"))

## CLASS AGE SEX yhat lower upper
## 1 0 Child Female 0.8886935 0.66194638 0.9701987
## 2 1 Child Female 0.9768348 0.92575376 0.9930367
## 3 2 Child Female 0.8935160 0.78056016 0.9519104
## 4 3 Child Female 0.4970148 0.33827964 0.6563539
## 5 0 Adult Female 0.8695652 0.66454828 0.9573283
## 6 1 Adult Female 0.9723831 0.92874210 0.9895961
## 7 2 Adult Female 0.8750999 0.79597143 0.9263784
## 8 3 Adult Female 0.4520760 0.37865906 0.5276396
## 9 0 Child Male 0.5716729 0.43150500 0.7012113
## 10 1 Child Male 0.7004700 0.55927805 0.8116614
## 11 2 Child Male 0.3949969 0.25626440 0.5529915
## 12 3 Child Male 0.4406809 0.32190559 0.5666583
## 13 0 Adult Male 0.2227378 0.19619893 0.2517422
## 14 1 Adult Male 0.3342723 0.26910345 0.4064468
## 15 2 Adult Male 0.1229466 0.08312897 0.1781310
## 16 3 Adult Male 0.1446912 0.11604927 0.1789709
##
## Response variable (y):
##
## Limits are 0.95 confidence limits

 print(plot(Predict(titanic.lrm.classagesex, fun=plogis, CLASS=c("0","1","2","3"),SEX=c("Female","Male"), AGE=c("Child","Adult")), pch=c(2,1),col=c(1,2),layout=c(1,2)))



A nomogram may be helpful at this point.

 nom = nomogram(titanic.lrm.classagesex, fun=plogis)
 print(plot(nom))



## NULL