Stepwise Logistic Regression

Veerasak Kritsanapraphan

8/3/2017

## Stepwise Logistic Regression

Import packages necessary first.

library(MASS)  
library(plyr)  
library(ggplot2)  
library(knitr)

Prepare data

# Assign better variable names  
colnames(birthwt) <- c("birthwt.below.2500", "mother.age", "mother.weight",   
 "race", "mother.smokes", "previous.prem.labor", "hypertension", "uterine.irr",   
 "physician.visits", "birthwt.grams")  
  
# Assign better labels to categorical variables  
birthwt <- transform(birthwt,   
 race = as.factor(mapvalues(race, c(1, 2, 3),   
 c("white","black", "other"))),  
 mother.smokes = as.factor(mapvalues(mother.smokes,   
 c(0,1), c("no", "yes"))),  
 hypertension = as.factor(mapvalues(hypertension,   
 c(0,1), c("no", "yes"))),  
 uterine.irr = as.factor(mapvalues(uterine.irr,   
 c(0,1), c("no", "yes"))),  
 birthwt.below.2500 = as.factor(mapvalues(birthwt.below.2500,  
 c(0,1), c("no", "yes")))  
 )

Run logistic regression

formula = birthwt.below.2500 ~ mother.age + mother.weight + physician.visits + mother.smokes + uterine.irr + previous.prem.labor + hypertension  
fullmod = glm(formula, family = binomial, data = birthwt)  
summary(fullmod)

##   
## Call:  
## glm(formula = formula, family = binomial, data = birthwt)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0714 -0.8105 -0.6220 1.0356 2.0334   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.390719 1.090079 1.276 0.2020   
## mother.age -0.043249 0.035404 -1.222 0.2219   
## mother.weight -0.014367 0.006655 -2.159 0.0308 \*   
## physician.visits 0.023433 0.173127 0.135 0.8923   
## mother.smokesyes 0.553932 0.344437 1.608 0.1078   
## uterine.irryes 0.739301 0.456663 1.619 0.1055   
## previous.prem.labor 0.594336 0.348260 1.707 0.0879 .   
## hypertensionyes 1.873160 0.690840 2.711 0.0067 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 234.67 on 188 degrees of freedom  
## Residual deviance: 208.75 on 181 degrees of freedom  
## AIC: 224.75  
##   
## Number of Fisher Scoring iterations: 4

No independent variables

nothing <- glm(birthwt.below.2500 ~ 1,family=binomial, data = birthwt)  
summary(nothing)

##   
## Call:  
## glm(formula = birthwt.below.2500 ~ 1, family = binomial, data = birthwt)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.8651 -0.8651 -0.8651 1.5259 1.5259   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.790 0.157 -5.033 4.84e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 234.67 on 188 degrees of freedom  
## Residual deviance: 234.67 on 188 degrees of freedom  
## AIC: 236.67  
##   
## Number of Fisher Scoring iterations: 4

Stepwise Logistic Regression (“Backward”)

backwards = step(fullmod) # Backwards selection is the default

## Start: AIC=224.75  
## birthwt.below.2500 ~ mother.age + mother.weight + physician.visits +   
## mother.smokes + uterine.irr + previous.prem.labor + hypertension  
##   
## Df Deviance AIC  
## - physician.visits 1 208.77 222.77  
## - mother.age 1 210.29 224.29  
## <none> 208.75 224.75  
## - uterine.irr 1 211.32 225.32  
## - mother.smokes 1 211.33 225.33  
## - previous.prem.labor 1 211.77 225.77  
## - mother.weight 1 213.97 227.97  
## - hypertension 1 216.53 230.53  
##   
## Step: AIC=222.77  
## birthwt.below.2500 ~ mother.age + mother.weight + mother.smokes +   
## uterine.irr + previous.prem.labor + hypertension  
##   
## Df Deviance AIC  
## - mother.age 1 210.31 222.31  
## <none> 208.77 222.77  
## - uterine.irr 1 211.33 223.33  
## - mother.smokes 1 211.33 223.33  
## - previous.prem.labor 1 211.78 223.78  
## - mother.weight 1 213.97 225.97  
## - hypertension 1 216.54 228.54  
##   
## Step: AIC=222.31  
## birthwt.below.2500 ~ mother.weight + mother.smokes + uterine.irr +   
## previous.prem.labor + hypertension  
##   
## Df Deviance AIC  
## <none> 210.31 222.31  
## - previous.prem.labor 1 212.83 222.83  
## - mother.smokes 1 213.01 223.01  
## - uterine.irr 1 213.15 223.15  
## - mother.weight 1 216.63 226.63  
## - hypertension 1 218.45 228.45

formula(backwards)

## birthwt.below.2500 ~ mother.weight + mother.smokes + uterine.irr +   
## previous.prem.labor + hypertension

summary(backwards)

##   
## Call:  
## glm(formula = birthwt.below.2500 ~ mother.weight + mother.smokes +   
## uterine.irr + previous.prem.labor + hypertension, family = binomial,   
## data = birthwt)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0738 -0.7877 -0.6416 1.0657 1.9836   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.562843 0.859815 0.655 0.51272   
## mother.weight -0.015493 0.006597 -2.348 0.01886 \*   
## mother.smokesyes 0.563972 0.342368 1.647 0.09950 .   
## uterine.irryes 0.769617 0.452910 1.699 0.08927 .   
## previous.prem.labor 0.533933 0.341417 1.564 0.11785   
## hypertensionyes 1.905592 0.685990 2.778 0.00547 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 234.67 on 188 degrees of freedom  
## Residual deviance: 210.31 on 183 degrees of freedom  
## AIC: 222.31  
##   
## Number of Fisher Scoring iterations: 4

Stepwise Logistic Regression (“Forward”)

forwards <- step(fullmod, direction="forward")

## Start: AIC=224.75  
## birthwt.below.2500 ~ mother.age + mother.weight + physician.visits +   
## mother.smokes + uterine.irr + previous.prem.labor + hypertension

formula(forwards)

## birthwt.below.2500 ~ mother.age + mother.weight + physician.visits +   
## mother.smokes + uterine.irr + previous.prem.labor + hypertension

summary(forwards)

##   
## Call:  
## glm(formula = birthwt.below.2500 ~ mother.age + mother.weight +   
## physician.visits + mother.smokes + uterine.irr + previous.prem.labor +   
## hypertension, family = binomial, data = birthwt)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0714 -0.8105 -0.6220 1.0356 2.0334   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.390719 1.090079 1.276 0.2020   
## mother.age -0.043249 0.035404 -1.222 0.2219   
## mother.weight -0.014367 0.006655 -2.159 0.0308 \*   
## physician.visits 0.023433 0.173127 0.135 0.8923   
## mother.smokesyes 0.553932 0.344437 1.608 0.1078   
## uterine.irryes 0.739301 0.456663 1.619 0.1055   
## previous.prem.labor 0.594336 0.348260 1.707 0.0879 .   
## hypertensionyes 1.873160 0.690840 2.711 0.0067 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 234.67 on 188 degrees of freedom  
## Residual deviance: 208.75 on 181 degrees of freedom  
## AIC: 224.75  
##   
## Number of Fisher Scoring iterations: 4