Logistic Regression on BirthWeight

Greg Johnson

We will run a logistic regression model on the birthwt dataset from the MASS library.

```
bwt <- with(birthwt, {
  race <- factor(race, labels = c("white", "black", "other"))
  ptd <- factor(ptl > 0)
  ftv <- factor(ftv)
  levels(ftv)[-(1:2)] <- "2+"
  data.frame(low = factor(low), age, lwt, race, smoke = (smoke > 0),
             ptd, ht = (ht > 0), ui = (ui > 0), ftv)
})

logist2.low<-glm(low ~ lwt+race+smoke+ptd+ht, binomial, bwt)
summary(logist2.low)
```

```
##
## Call:
## glm(formula = low ~ lwt + race + smoke + ptd + ht, family = binomial,
##     data = bwt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8188  -0.8035  -0.5457   0.9667   2.1530
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.09462    0.95704   0.099  0.92124
## lwt         -0.01673    0.00695  -2.407  0.01608 *
## raceblack    1.26372    0.52933   2.387  0.01697 *
## raceother    0.86418    0.43509   1.986  0.04701 *
## smokeTRUE    0.87611    0.40071   2.186  0.02879 *
## ptdTRUE      1.23144    0.44625   2.760  0.00579 **
## htTRUE       1.76744    0.70841   2.495  0.01260 *
## ---
## 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: 200.48  on 182  degrees of freedom
## AIC: 214.48
##
## Number of Fisher Scoring iterations: 4
```

The estimated parameters of the logistic regression are the expected change in log odds of being low weight vs. regular weight when one of the predictors is changed by one unit and the others are held constant. Thus if we exponentiate the coefficients, we can get just the change in odds.

```
exp(logist2.low[["coefficients"]][-1])
```

```
##       lwt raceblack raceother smokeTRUE   ptdTRUE    htTRUE
## 0.9834105 3.5385761 2.3730507 2.4015306 3.4261485 5.8558577
```

If we look explicitly at race (whether the mother is black), we get a change in odds of:

```
exp(logist2.low[["coefficients"]][3])
```

```
## raceblack
##  3.538576
```

Thus if a woman is black (holding all other variables constant), her odds of having a low weight baby increase by 253%.

What is the odds ratio estimate for the mother being black? Better yet, let's give a 95% confidence interval.

```
coef<-coef(summary(logist2.low))[3,1]
se<-coef(summary(logist2.low))[3,2]
CI<-c(exp(coef-1.96*se),exp(coef+1.96*se))
CI
```

```
## [1] 1.253884 9.986187
```

What is the odds ratio of having a low birth baby versus not for a woman who is 20 pounds heavier than another woman with all other factors being equal?

```
delta<-20
coef<-coef(summary(logist2.low))[2,1]
1/exp(delta*coef)
```

```
## [1] 1.397344
```

So everything else being equal, a woman who is 20 pounds heavier than another is 39% more likely to not have a low birth baby.

What is the probability of having a low birth child for a 150 pound white woman who smokes, has had no previous premature labors, but has had a history of hypertension?

```
x<-c(1,150,0,0,1,0,1)
coef<-coef(summary(logist2.low))[,1]
phat<-exp(t(x)%*%coef)/(1+exp(t(x)%*%coef))
phat
```

```
##           [,1]
## [1,] 0.5569684
```

Thus this particular lady is expected to have a probability of 55.7% of having a low birth child.

Next let's estimate and find a confidence interval for the odds ratio of having a low weight child for a black mother versus a (race=other) mother.

We can take advantage of the fact that a ratio of odds ratios (with the same denominator in each odds ratio) is itself an odds ratio where the common factor has been cancelled out. If we want an odds ratio of black mother vs (race=other) mother, we just take the odds ratio of black mother vs white and divide it by the odds ratio of (race=other) mother vs. white and the odds related to being white will cancel!

```
coef_black<-coef(summary(logist2.low))[3,1]
coef_other<-coef(summary(logist2.low))[4,1]
exp(coef_black)/exp(coef_other)
```

```
## [1] 1.491151
```

This is a nice solution in that I don't have to rerun a model where the race variable is split up so that black mothers are the group to be compared to. However our current model doesn't have standard errors for a black vs. race=other parameters. So we'll rerun the logistic model with a reordered factor (with black as the first level) so that we can get the standard error we need.

```r
bwt$race<-relevel(bwt$race,"black")
logist3.low<-glm(low ~ lwt+race+smoke+ptd+ht, binomial, bwt)

coef<-coef(summary(logist3.low))[4,1]
se<-coef(summary(logist3.low))[4,2]
CI<-c(exp(coef-1.96*se),exp(coef+1.96*se))
CI
```

```
## [1] 0.2293846 1.9606165
```

The confidence interval contains 1 so we can't reject the null hypothesis that having race=other vs. black has no effect on birth weight.