

Low Birth Weight Study

bweight.data

Col 1 = Identification Code
Col 2 = Low Birth Weight Baby (1=Yes under 2500g, 0=No)
Col 3 = Mother's age in years
Col 4 = Weight at Last Period
Col 5 = Race (1=White, 2=Black, 3=Other)
Col 6 = Smoke during Pregnancy (1=Yes, 0=No)
Col 7 = History of Premature Labour (# of times)
Col 8 = History of Hypertension (1=Yes, 0=No)
Col 9 = Presence of Uterine Irritability (1=Yes, 0=No)
Col 10 = Visits to Doctor During 1st trimester
Col 11 = Baby's birth Weight in Grams

```
> bweight = read.table("http://www.utstat.toronto.edu/~brunner/312f10/code_n_data/
bweight.data")
> bweight[1:5,]
  low age lwt race smoke ptl ht ui ftv bwt
85  0  19 182   2     0   0  0  1  0 2523
86  0  33 155   3     0   0  0  0  3 2551
87  0  20 105   1     1   0  0  0  1 2557
88  0  21 108   1     1   0  0  1  2 2594
89  0  18 107   1     1   0  0  1  0 2600
> # The following is just to save some typing
> low <- bweight$low ; age <- bweight$age ; lwt <- bweight$lwt
> race <- bweight$race ; smoke <- bweight$smoke; ptl <- bweight$ptl
> ht <- bweight$ht; ui <- bweight$ui; ftv <- bweight$ftv
> # Crude descriptive stats
> table(low)
low
 0  1
130 59
> summary(age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 14.00  19.00   23.00   23.24  26.00   45.00
> summary(lwt)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  80.0  110.0   121.0   129.8  140.0   250.0
> table(race) # 1=White, 2=Black, 3=Other
race
 1  2  3
96 26 67
> table(smoke)
smoke
 0  1
115 74
> table(ptl)
ptl
 0  1  2  3
159 24  5  1
> ptl[ptl>1]=1 # Collapsing categories
> table(ptl)
ptl
 0  1
159 30
```

```

> table(ht)
ht
 0  1
177 12
> table(ui)
ui
 0  1
161 28
> table(ftv)
ftv
 0  1  2  3  4  6
100 47 30 7 4 1
> # Don't collapse ftv for now

> # First, some simple examples to illustrate the methods
> # Two continuous independent variables
> modell <- glm(low ~ age + lwt, family=binomial)
> summary(modell)

```

Call:
 glm(formula = low ~ age + lwt, family = binomial)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.1352	-0.9088	-0.7480	1.3392	2.0595

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.748773	0.997097	1.754	0.0795 .
age	-0.039788	0.032287	-1.232	0.2178
lwt	-0.012775	0.006211	-2.057	0.0397 *

 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: 227.12 on 186 degrees of freedom
 AIC: 233.12

Number of Fisher Scoring iterations: 4

$$\text{Deviance} = \sum_{i=1}^n (-2 \log P\{Y_i = y_i | x_i, \hat{\beta}\}) = \sum_{i=1}^n d_i$$

$$\text{Deviance Residual: } r_i^D = \text{sign} \left(y_i - P\{Y_i = y_i | x_i, \hat{\beta}\} \right) \sqrt{d_i}$$

Null deviance is the deviance of a model with just the intercept.

```

> summary(modell)

Call:
glm(formula = low ~ age + lwt, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1352  -0.9088  -0.7480   1.3392   2.0595

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.748773   0.997097   1.754   0.0795 .
age          -0.039788   0.032287  -1.232   0.2178
lwt          -0.012775   0.006211  -2.057   0.0397 *
---
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: 227.12  on 186  degrees of freedom
AIC: 233.12

Number of Fisher Scoring iterations: 4

> modell$coefficients
(Intercept)          age          lwt
 1.74877349 -0.03978793 -0.01277541
> modell$deviance
[1] 227.1234
> modell$null.deviance
[1] 234.672
> # G-squared = Deviance(Reduced)-Deviance(Full)
> # df = difference in number of betas
> G2 = modell$null.deviance-modell$deviance; G2
[1] 7.548608
> 1-pchisq(G2,df=1)
[1] 0.006005646
> anova(modell)
Analysis of Deviance Table

Model: binomial, link: logit

Response: low

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev
NULL    188    234.672
age      1      2.760    187    231.912
lwt      1      4.789    186    227.123

>
> 1-pchisq(4.789,1) # LR test of weight controlling for age
[1] 0.02864205
> 1-pchisq(2.057^2,1) # Wald test of weight controlling for age
[1] 0.03968623
>
> # Estimate probability of low birth weight for a 19 year old
> # mother weighing 120 pounds
> x = c(1,19,120); xb = sum(x*modell$coefficients)
> phat = exp(xb)/(1+exp(xb)); phat
[1] 0.3681301

```

```

> # For constant age, increase of weight by one pound multiplies
> # odds of low birth weight baby by ...
> exp(model1$coefficients[3])
      lwt
0.9873058

> # Represent race with 2 indicator dummy variables. First the hard way:
> n = length(race); n
[1] 189
> r1=numeric(n); r2 = numeric(n)
> r1[race==2]=1; r2[race==3]=1
> table(r1,race)
      race
r1    1  2  3
0  96  0  67
1   0 26   0
> table(r2,race)
      race
r2    1  2  3
0  96 26   0
1   0  0  67
>
> model2a = glm(low ~ r1 + r2, family=binomial); summary(model2a)

Call:
glm(formula = low ~ r1 + r2, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0489  -0.9665  -0.7401   1.4041   1.6905

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.1550    0.2391  -4.830 1.36e-06 ***
r1             0.8448    0.4634   1.823  0.0683 .
r2             0.6362    0.3478   1.829  0.0674 .
---
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: 229.66  on 186  degrees of freedom
AIC: 235.66

Number of Fisher Scoring iterations: 4

>
> G2a = model2a$null.deviance-model2a$deviance; G2a
[1] 5.010366
> 1-pchisq(G2a,2)
[1] 0.08166065
> racelow = table(race,low); racelow
      low
race  0  1
  1  73 23
  2  15 11
  3  42 25
> loglin(racelow,margin=list(1,2))$lrt
2 iterations: deviation 0
[1] 5.010366

```

```

>
> racefac <- factor(bweight$race,label=c("White","Black","Other"))
> contrasts(racefac)
      Black Other
White    0     0
Black    1     0
Other    0     1
> # So the default is indicator dummy variable coding
> model2b = glm(low ~ racefac, family=binomial)
> # summary(model2b) is 100% identical to summary(model2a)

> # Estimated odds of low birth weight baby are ___ times as
> # great for Blacks as Whites:

> exp(model2b$coefficients[2])
racefacBlack
  2.327536

>
> # Control for a continuous variable
> model3 = glm(low ~ lwt + racefac, family=binomial); summary(model3)

Call:
glm(formula = low ~ lwt + racefac, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3491  -0.8919  -0.7196   1.2526   2.0993

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.805753   0.845167   0.953   0.3404
lwt         -0.015223   0.006439  -2.364   0.0181 *
racefacBlack  1.081066   0.488052   2.215   0.0268 *
racefacOther  0.480603   0.356674   1.347   0.1778
---
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: 223.26  on 185  degrees of freedom
AIC: 231.26

Number of Fisher Scoring iterations: 4

```

```

> G2change = model2b$deviance-model3$deviance; G2change
[1] 6.40254
> # What is H0?
> 1-pchisq(G2change,1)
[1] 0.01139572

> # Another way, using anova to compare 2 models
> anova(model2b,model3)
Analysis of Deviance Table

Model 1: low ~ racefac
Model 2: low ~ lwt + racefac
  Resid. Df Resid. Dev  Df Deviance
1         186      229.662
2          185      223.259   1    6.403
>

> # What about race controlling for weight?
> # Could fit a reduced model with just weight, but ...
> anova(model3)
Analysis of Deviance Table

Model: binomial, link: logit

Response: low

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev
NULL                188      234.672
lwt                 1    5.981      187      228.691
racefac             2    5.432      185      223.259
> 1-pchisq(5.432,2)
[1] 0.06613878

> # Still not statistically significant. It's time to get serious
> # about model building.
> bweight[1:5,]
  low age lwt race smoke ptl ht ui ftv bwt
85  0  19 182   2    0  0  0  1  0 2523
86  0  33 155   3    0  0  0  0  3 2551
87  0  20 105   1    1  0  0  0  1 2557
88  0  21 108   1    1  0  0  1  2 2594
89  0  18 107   1    1  0  0  1  0 2600
> fullmod = glm(low ~ age+lwt+racefac+smoke+ptl+ht+ui+ftv,family=binomial)

```

```
> summary(fullmod)
```

```
Call:
```

```
glm(formula = low ~ age + lwt + racefac + smoke + ptl + ht +  
    ui + ftv, family = binomial)
```

```
Deviance Residuals:
```

```
    Min      1Q   Median      3Q      Max  
-1.6305 -0.7894 -0.5094  0.9119  2.2257
```

```
Coefficients:
```

```
            Estimate Std. Error z value Pr(>|z|)  
(Intercept)  0.644476   1.223889   0.527  0.59849  
age          -0.039548   0.038305  -1.032  0.30186  
lwt          -0.015078   0.007034  -2.143  0.03207 *  
racefacBlack  1.218791   0.533168   2.286  0.02226 *  
racefacOther  0.819439   0.450466   1.819  0.06890 .  
smoke        0.859459   0.409836   2.097  0.03599 *  
ptl          1.218512   0.463015   2.632  0.00850 **  
ht           1.860429   0.708161   2.627  0.00861 **  
ui           0.719299   0.463419   1.552  0.12062  
ftv          0.050900   0.175456   0.290  0.77174
```

```
---  
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: 196.75 on 179 degrees of freedom  
AIC: 216.75
```

```
Number of Fisher Scoring iterations: 4
```

```
> # Try dropping age, ui, ftv: Test simultaneously  
> # What is H0?  
> redmod1 = glm(low ~ lwt+racefac+smoke+ptl+ht,family=binomial)  
> G2change1 = redmod1$deviance-fullmod$deviance; G2change1  
[1] 3.732170  
> 1-pchisq(G2change1,3)  
[1] 0.2918750  
> # No problem discarding these.  
> # Controlling for the other vars, they do nothing.  
> summary(redmod1)
```

```
Call:
```

```
glm(formula = low ~ lwt + racefac + smoke + ptl + ht, family = binomial)
```

```
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 *  
racefacBlack  1.26372    0.52933   2.387  0.01697 *  
racefacOther  0.86418    0.43509   1.986  0.04701 *  
smoke        0.87611    0.40071   2.186  0.02879 *  
ptl          1.23144    0.44625   2.760  0.00579 **  
ht           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

```
> # Test all the variables at once.
> G2 = redmod1$null.deviance-redmod1$deviance; G2
[1] 34.18974
> 1-pchisq(G2,6)
[1] 6.182967e-06
> # What about race controlling for the other variables?
> redmod2 = glm(low ~ lwt+smoke+ptl+ht,family=binomial)
> G2race = redmod2$deviance-redmod1$deviance; G2race
[1] 7.47308
> 1-pchisq(G2race,2)
[1] 0.02383643
>
> # Controlling for other variables, the estimated odds
> # of a low birth weight baby are ___ times as great
> # for a Black mother as compared to a White mother.
> redmod1$coefficients
(Intercept)          lwt racefacBlack racefacOther          smoke
0.09461948 -0.01672867  1.26372441  0.86417633  0.87610630
          ptl          ht
1.23143674  1.76744247
> exp(redmod1$coefficients[3])
racefacBlack
3.538576
>
> # Controlling for other variables, the estimated odds
> # of a low birth weight baby are ___ times as great
> # for an Other mother as compared to a White mother.
> exp(redmod1$coefficients[4])
racefacOther
2.373051
>
> # Controlling for other variables, are the odds of
> # a low birth weight baby different for Other and Black mothers?
```

$$\begin{aligned} \log \text{ odds} &= \beta_0 + \beta_1 \text{lwt} + \beta_2 \text{r1} + \beta_2 \text{r2} + \beta_4 \text{smoke} + \beta_5 \text{ptl} + \beta_6 \text{ht} \\ &= \beta_0 + \beta_1 \text{lwt} + \beta_2 (\text{r1} + \text{r2}) + \beta_4 \text{smoke} + \beta_5 \text{ptl} + \beta_6 \text{ht} \end{aligned}$$

```
> r = r1+r2
> redmod3 = glm(low ~ lwt+r+smoke+ptl+ht,family=binomial)
> G2change = redmod3$deviance-redmod1$deviance; G2change
[1] 0.5313281
> 1-pchisq(G2change,1)
[1] 0.4660491
> # Consistent with no difference.
```