

Birth weight data for Assignment 8 (Question 3)

```
> rm(list=ls()); options(scipen=999)
> library(MASS); # help(birthwt)
> head(birthwt)
  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
91   0  21 124   3     0   0  0  0   0 2622
>
> # Explore a bit first
> # with(birthwt,table(ftv))
> # with(birthwt,table(race)) # 1 = white, 2 = black, 3 = other
> # summary(birthwt) # All numeric
>
> bweight = within(birthwt,{
+ race = factor(race, labels = c('White', 'Black', 'Other')) # Order
+   maintained
+ anyftv = ftv; anyftv[anyftv>1] = 1
+ })
> with(bweight,table(race)) # Now what does it look like?
race
White Black Other
  96    26    67
> contrasts(bweight$race) # White still reference category.
      Black Other
White    0     0
Black    1     0
Other    0     1
>
>
> # Full model
> modell = glm(low ~ age + lwt + race + smoke + anyftv, data=bweight,
family=binomial)
> summary(modell)
```

```
Call:
glm(formula = low ~ age + lwt + race + smoke + anyftv, family =
binomial,
    data = bweight)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5289  -0.9075  -0.5797   1.3007   2.0550
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.351747   1.110641   0.317   0.75147
age          -0.019288   0.034977  -0.551   0.58133
lwt          -0.012507   0.006387  -1.958   0.05021 .
raceBlack    1.224274   0.517471   2.366   0.01799 *
raceOther    0.908265   0.425003   2.137   0.03259 *
```

```

smoke          1.024100    0.387048    2.646    0.00815 **
anyftv         -0.149082    0.352620   -0.423    0.67245
---
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: 214.40 on 182 degrees of freedom
AIC: 228.4

```

Number of Fisher Scoring iterations: 4

```

>
> # 12a
> se = sqrt(diag(vcov(modell))); se
(Intercept)      age      lwt    raceBlack    raceOther
smoke          anyftv
1.110641332 0.034976846 0.006386949 0.517471090 0.425003136
0.387048253 0.352619776
>
> # For (b) and (c), see summary
>
> betahat = coefficients(modell); betahat
(Intercept)      age      lwt    raceBlack    raceOther
smoke          anyftv
0.35174720 -0.01928760 -0.01250698 1.22427400 0.90826519
1.02409982 -0.14908197
>
> # 12d Correcting for all the other variables, the odds of a low
birth weight baby are an estimated \underline{\hspace{10mm}} times as
great for a mother who smokes during pregnancy.
> exp(betahat[6]) # 2.78
smoke
2.784588
>
> # 12e (i) LR test for race
> norace = glm(low ~ age + lwt + smoke + anyftv, data=bweight,
family=binomial)
> anova(norace,modell, test="Chisq")
Analysis of Deviance Table

Model 1: low ~ age + lwt + smoke + anyftv
Model 2: low ~ age + lwt + race + smoke + anyftv
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         184        222.19
2         182        214.40  2    7.7919  0.02032 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> # 12e (ii) Wald test for race
> Vhat = vcov(modell)

```

```

> # For Wald tests: Wtest = function(L,Tn,Vn,h=0) # H0: L theta = h
>
source("http://www.utstat.utoronto.ca/brunner/Rfunctions/Wtest.txt")
> LL = rbind(c(0,0,0,1,0,0,0),
+           c(0,0,0,0,1,0,0))
> Wtest(LL,betahat,Vhat)
              W              df      p-value
7.40790875 2.00000000 0.02462595
>
> # (12f) See summary
>
> # 12g Correcting for all the other variables, the odds of a low
birth weight baby are an estimated            times as
great for a Black mother, compared to an Other mother.
> exp(betahat[4]) # 3.4
raceBlack
3.401696
>
> # 12h Give a 95% confidence interval for that last number.
> # Indicator for Black is 4th element
> a = betahat[4] - 1.96*se[4]; b = betahat[4] + 1.96*se[4]
> c(exp(a),exp(b))
raceBlack raceBlack
1.233716 9.379414
>
> # 12i) For a 26-year-old, 130 pound, Black non-smoking mother with
no first trimester visits to the doctor, estimate the probability of
a low birth weight baby and give a 95% confidence interval for the
probability.
>
> # Estimate first without using predict.
> betahat
(Intercept)          age          lwt    raceBlack    raceOther
smoke          anyftv
0.35174720 -0.01928760 -0.01250698 1.22427400 0.90826519
1.02409982 -0.14908197
> x = c(1,26,130,1,0,0,0); xb = sum(x*betahat)
> pihat = exp(xb)/(1+exp(xb)); pihat # 0.3655479
[1] 0.3655479
> # Now try predict
> head(bweight)
  low age lwt  race smoke ptl ht ui ftv  bwt anyftv
85   0  19 182 Black    0   0  0  1   0 2523     0
86   0  33 155 Other    0   0  0  0   3 2551     1
87   0  20 105 White    1   0  0  0   1 2557     1
88   0  21 108 White    1   0  0  1   2 2594     1
89   0  18 107 White    1   0  0  1   0 2600     0
91   0  21 124 Other    0   0  0  0   0 2622     0
> onemom = data.frame(age=26, lwt=130, race='Black', smoke=0,
anyftv=0)

```

```

> pred1 = predict(modell, newdata=onemom, type = "response", se.fit =
TRUE); pred1
$fit
      1
0.3655479

$se.fit
      1
0.1240606

$residual.scale
[1] 1

> # CI using the delta method
> lower95 = pred1$fit - 1.96*pred1$se
> upper95 = pred1$fit + 1.96*pred1$se
> c(lower95,upper95)
      1      1
0.1223890 0.6087068
> # CI starting from difference in log odds. Quoting,
> # The second way is to get an interval for the log odds, and
transform it.
> # It works because the probability is an increasing function of the
log odds.
> # The default for predict is x'betahat
> pred2 = predict(modell,newdata=onemom, se.fit = TRUE); pred2
$fit
      1
-0.5513644

$se.fit
[1] 0.5349226

$residual.scale
[1] 1

> a = pred2$fit - 1.96*pred2$se; b = pred2$fit + 1.96*pred2$se
> Lower95 = exp(a)/(1+exp(a)); Upper95 = exp(b)/(1+exp(b))
> c(Lower95,Upper95)
      1      1
0.1680078 0.6217738
>
> # 12j Comparing Black and Other mothers
> # (i) Wald test
> LblackVSother = rbind(c(0,0,0,1,-1,0,0))
> Wtest(LblackVSother,betahat,Vhat) # W = 0.353, p = 0.5523908
      W      df    p-value
0.3530515 1.0000000 0.5523908
> # (ii) Odds ratio
> orBO = exp(betahat[4]-betahat[5]); orBO # 1.37 (?)
raceBlack
1.371642

```

```

> # (iii) CI for odds ratio
> seBO = sqrt(as.numeric(LblackVSother %*% Vhat %*%
t(LblackVSother))); seBO
[1] 0.5318389
> a = betahat[4]-betahat[5] - 1.96*seBO; b = betahat[4]-betahat[5] +
1.96*seBO
> c(exp(a),exp(b)) # 0.4836492 3.8900154
raceBlack raceBlack
0.4836492 3.8900154
>
> # Do it with a different reference category as a cross check
> # Remember the original coding of race is 123
> bw2 = within(bweight,
+ {contrasts(race) = contr.treatment(3,base=3) # Reference is Other
+ })
> with(bw2,contrasts(race))
      1 2
White 1 0
Black 0 1
Other 0 0
> model2 = glm(low ~ age + lwt + race + smoke + anyftv, family =
binomial, data = bw2)
> summary(model2)

```

```

Call:
glm(formula = low ~ age + lwt + race + smoke + anyftv, family =
binomial,
     data = bw2)

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5289	-0.9075	-0.5797	1.3007	2.0550

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.260012	1.017375	1.238	0.21553
age	-0.019288	0.034977	-0.551	0.58133
lwt	-0.012507	0.006387	-1.958	0.05021 .
race1	-0.908265	0.425003	-2.137	0.03259 *
race2	0.316009	0.531839	0.594	0.55239
smoke	1.024100	0.387048	2.646	0.00815 **
anyftv	-0.149082	0.352620	-0.423	0.67245

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: 214.40 on 182 degrees of freedom
AIC: 228.4

Number of Fisher Scoring iterations: 4

```

> # Notice how the p-value for race2 = 0.55239, same as for the
earlier Wald test

```

```

> # Now odds ratio Black/Other = 1.371642
> betahat2 = coefficients(model2); betahat2
(Intercept)      age      lwt      race1      race2
smoke      anyftv
 1.26001240 -0.01928760 -0.01250698 -0.90826519  0.31600880
1.02409982 -0.14908197
> orBO
raceBlack
 1.371642
> exp(betahat2[5])
  race2
1.371642
> # Maybe I should always name the columns after I set the contrasts,
like this
> with(bw2, contrasts(race))
      1 2
White 1 0
Black 0 1
Other 0 0
> bw2 = within(bw2, {
+ colnames(contrasts(race)) = c('White', 'Black')
+ })
> with(bw2, contrasts(race))
      White Black
White    1     0
Black    0     1
Other    0     0
> model2 = glm(low ~ age + lwt + race + smoke + anyftv, family =
binomial, data = bw2)
> summary(model2)

```

```

Call:
glm(formula = low ~ age + lwt + race + smoke + anyftv, family =
binomial,
    data = bw2)

```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.5289	-0.9075	-0.5797	1.3007	2.0550

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.260012	1.017375	1.238	0.21553
age	-0.019288	0.034977	-0.551	0.58133
lwt	-0.012507	0.006387	-1.958	0.05021 .
raceWhite	-0.908265	0.425003	-2.137	0.03259 *
raceBlack	0.316009	0.531839	0.594	0.55239
smoke	1.024100	0.387048	2.646	0.00815 **
anyftv	-0.149082	0.352620	-0.423	0.67245

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: 214.40 on 182 degrees of freedom