Stat 8053, Fall 2013: Binomial Deviance and Effects Plots (Faraway, Chap. 2)

```
library(alr4)
(g4 <- glm(y ~ (spp + log(d) + s)^2, binomial, Blowdown))
```

```
Call:  glm(formula = y ~ (spp + log(d) + s)^2, family = binomial, data = Blowdown)

Coefficients:
           (Intercept)           sppbalsam fir         sppblack spruce
                -0.439                  -1.626                  -5.959
              sppcedar             sppjackpine          spppaper birch
                -3.855                  -1.548                   3.316
           sppred pine            sppred maple            sppblack ash
                -0.777                  -0.463                   3.761
                log(d)                       s    sppbalsam fir:log(d)
                -0.255                  -6.145                  -0.429
sppblack spruce:log(d)         sppcedar:log(d)      sppjackpine:log(d)
                 2.058                   1.028                  -0.442
 spppaper birch:log(d)      sppred pine:log(d)     sppred maple:log(d)
                -0.708                  -0.545                  -0.388
   sppblack ash:log(d)         sppbalsam fir:s       sppblack spruce:s
                -0.896                   0.566                   2.645
            sppcedar:s           sppjackpine:s        spppaper birch:s
                 3.068                   1.459                  -0.629
         sppred pine:s          sppred maple:s          sppblack ash:s
                 0.775                  -0.623                  -1.038
              log(d):s
                 3.582

Degrees of Freedom: 3665 Total (i.e. Null);  3638 Residual
Null Deviance:            5060
Residual Deviance: 3120         AIC: 3180
```

Log Likelihood and Deviance

The log-likelihood for logistic regression is:

$$\log[L(p)] = \sum_{i=1}^{n} \left[\log \begin{pmatrix} m \\ y \end{pmatrix} + y \log(p) + (m-y) \log(1-p) \right]$$

Suppose $p(\hat{\beta})$ is mle of p based on a logistic regression model with k parameters, and let $\tilde{p} = y/m$ be the mle of p when each observation is fit separately, so it has n parameters. The *deviance* is defined to be

$$
\begin{aligned}
G^2 &= -2\left\{\log[L(\tilde{p})] - \log[L(p(\hat{\beta}))]\right\} \\
&= 2\sum_{i=1}^{n}\left[y\log\left(\frac{y}{mp(\hat{\beta})}\right) + (m-y)\log\left(\frac{m-y}{m-mp(\hat{\beta})}\right)\right]
\end{aligned}
$$

Residual deviance is the deviance of the model you fit, with k parameters. *Null deviance* is the deviance for the model in which all the p are equal, with 1 parameter:

```
g0 <- update(g4, ~ 1)
anova(g0, g4)
```

```
Analysis of Deviance Table

Model 1: y ~ 1
Model 2: y ~ (spp + log(d) + s)^2
  Resid. Df Resid. Dev Df Deviance
1      3665       5058
2      3638       3122 27     1936
```

Deviance provides the basis for performing likelihood ratio tests, using the `Anova` function in the car package

```
Anova(g4)
```

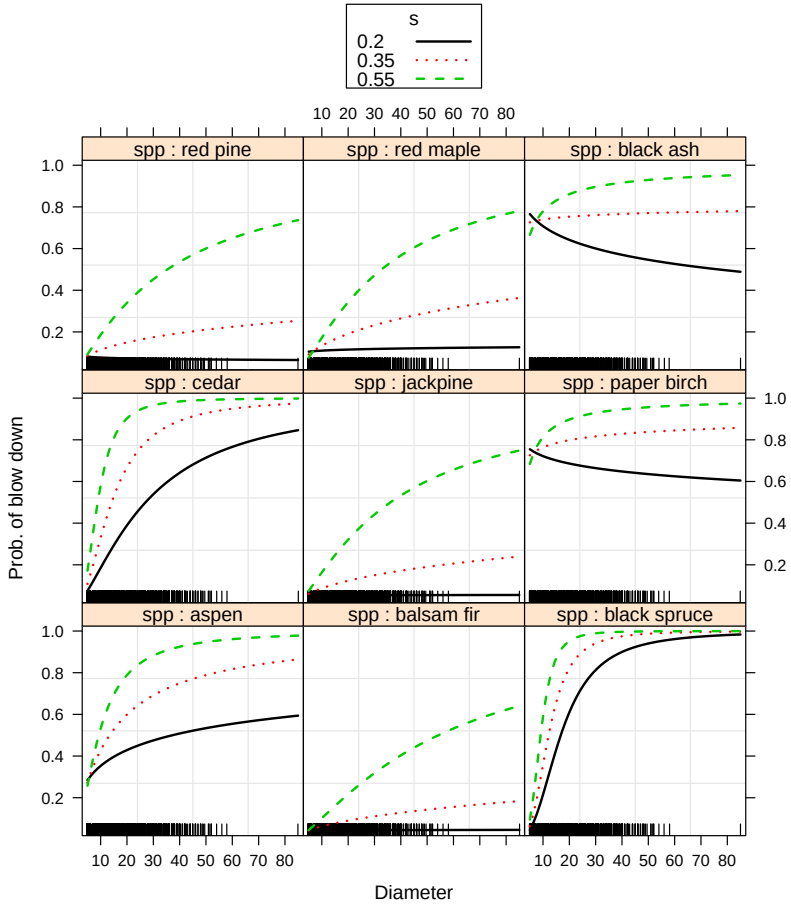
```
Analysis of Deviance Table (Type II tests)

Response: y
           LR Chisq Df Pr(>Chisq)
spp            510  8    < 2e-16
log(d)         228  1    < 2e-16
s              594  1    < 2e-16
spp:log(d)      72  8    2.1e-12
spp:s           36  8    1.4e-05
log(d):s        42  1    1.1e-10
```

Type II tests satisfy the marginality principle suggested by J. A. Nelder (1977), A reformulation of linear models, *Journal of the Royal Statistical Society. Series A (General)*, 140 48-77, `http://www.jstor.org/stable/2344517`:

A lower-order term, such as the *A* main effect, is *never* tested in models that include any of its higher-order relatives like *A:B*, *A:C*, or *A:B:C*. All regressors that are *not* higher-order relatives of the regressor of interest, such as *B*, *C* and *B:C*, are *always* included.

Start at the bottom of the table. Interpret a test for a main effect if and only if all of its higher-order relatives are not "significant".

```
plot(Effect(c("d", "s", "spp"), g4,
     xlevels=list(s=c(.2, .35, .55 ), d=60)), multiline=TRUE, grid=TRUE,
     xlab="Diameter",main="", lines=c(1, 3, 2),
     ylab="Prob. of blow down", rescale.axis=FALSE)
```



AMSsurvey

Counts of Ph.D.s in mathematical sciences categorized by institution `type` (IV is statistics and biostatistics), citizenship and gender for 2008-9 in `count` and 2011-12 in `count11`.

```
head(AMSsurvey)
```

```
   type    sex citizen count count11
1 I(Pu)   Male      US   132     148
2 I(Pu) Female      US    35      40
3 I(Pr)   Male      US    87      63
4 I(Pr) Female      US    20      22
5    II   Male      US    96     161
6    II Female      US    47      53
```

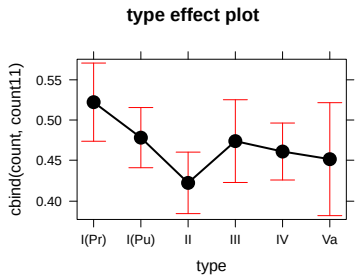
```
xtabs(count ~ type + paste(citizen, sex), AMSsurvey)
```

```
       paste(citizen, sex)
type    Non-US Female Non-US Male US Female US Male
  I(Pr)            25          79        20      87
  I(Pu)            29         130        35     132
  II               50          89        47      96
  III              39          53        32      47
  IV              105         122        54      71
  Va               12          28        14      34
```

```
a1 <- glm(cbind(count, count11) ~ type * sex * citizen, family=binomial, data=AMSsurvey)
Anova(a1)
```

```
Analysis of Deviance Table (Type II tests)

Response: cbind(count, count11)
                LR Chisq Df Pr(>Chisq)
type                11.72  5      0.039
sex                  2.39  1      0.122
citizen              0.42  1      0.518
type:sex             7.58  5      0.181
type:citizen         4.29  5      0.508
sex:citizen          0.17  1      0.684
type:sex:citizen     1.39  5      0.926
```

```
a2 <- update(a1, ~ type)
plot(effect("type", a2))
```



Using `sex` as the response requires reshaping the data file:

```
AMS1 <- reshape(AMSsurvey, varying=c("count", "count11"), v.names="y",
        direction="long", times=c("08-09", "11-12"), timevar="year")
head(AMS1)
```

```
         type    sex citizen  year   y id
1.08-09 I(Pu)   Male      US 08-09 132  1
2.08-09 I(Pu) Female      US 08-09  35  2
3.08-09 I(Pr)   Male      US 08-09  87  3
4.08-09 I(Pr) Female      US 08-09  20  4
5.08-09    II   Male      US 08-09  96  5
6.08-09    II Female      US 08-09  47  6
```

```
AMS2 <- reshape(AMS1[, -6], v.names="y", timevar="sex", idvar=c(1, 3, 4), direction="wide")
head(AMS2)
```

```
          type citizen  year y.Male y.Female
1.08-09  I(Pu)      US 08-09    132       35
3.08-09  I(Pr)      US 08-09     87       20
5.08-09     II      US 08-09     96       47
7.08-09    III      US 08-09     47       32
9.08-09     IV      US 08-09     71       54
11.08-09    Va      US 08-09     34       14
```

```
m1 <- glm(cbind(y.Female, y.Male) ~ type*citizen*year, binomial, AMS2)
summary(m1)
```

```
Call:
glm(formula = cbind(y.Female, y.Male) ~ type * citizen * year,
    family = binomial, data = AMS2)

Deviance Residuals:
 [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -1.15057    0.22947   -5.01  5.3e-07
typeI(Pu)                     -0.34967    0.30795   -1.14  0.25618
typeII                         0.57396    0.28964    1.98  0.04752
typeIII                        0.84384    0.31172    2.71  0.00679
typeIV                         1.00051    0.26529    3.77  0.00016
typeVa                         0.30327    0.41437    0.73  0.46424
citizenUS                     -0.31960    0.33786   -0.95  0.34417
year11-12                      0.00195    0.32143    0.01  0.99516
typeI(Pu):citizenUS            0.49239    0.43872    1.12  0.26172
typeII:citizenUS               0.18202    0.42081    0.43  0.66535
typeIII:citizenUS              0.24192    0.45955    0.53  0.59859
typeIV:citizenUS               0.19597    0.40556    0.48  0.62895
typeVa:citizenUS               0.27960    0.57796    0.48  0.62855
typeI(Pu):year11-12            0.05137    0.42906    0.12  0.90470
typeII:year11-12              -0.15357    0.40128   -0.38  0.70194
typeIII:year11-12             -0.40490    0.44447   -0.91  0.36231
typeIV:year11-12              -0.13739    0.36914   -0.37  0.70975
typeVa:year11-12               0.38273    0.56411    0.68  0.49748
citizenUS:year11-12            0.41613    0.47554    0.88  0.38153
typeI(Pu):citizenUS:year11-12 -0.45033    0.61222   -0.74  0.46199
typeII:citizenUS:year11-12    -0.66142    0.58363   -1.13  0.25709
typeIII:citizenUS:year11-12   -0.55925    0.65015   -0.86  0.38969
typeIV:citizenUS:year11-12    -0.48830    0.56667   -0.86  0.38885
typeVa:citizenUS:year11-12    -0.60665    0.78314   -0.77  0.43855

(Dispersion parameter for binomial family taken to be 1)

    Null deviance:  1.2527e+02  on 23  degrees of freedom
```

```
Residual deviance: -4.0412e-14  on  0  degrees of freedom
AIC: 167.8

Number of Fisher Scoring iterations: 3
```

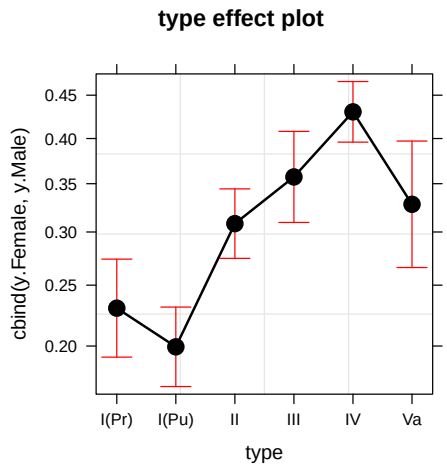
```
Anova(m1)
```

```
Analysis of Deviance Table (Type II tests)

Response: cbind(y.Female, y.Male)
                 LR Chisq Df Pr(>Chisq)
type                104.4  5     <2e-16
citizen               2.3  1       0.13
year                  2.4  1       0.12
type:citizen          2.9  5       0.71
type:year             7.6  5       0.18
citizen:year          0.2  1       0.68
type:citizen:year     1.4  5       0.93
```

```
m2 <- update(m1, ~ type)
Anova(m2)
```

```
Analysis of Deviance Table (Type II tests)

Response: cbind(y.Female, y.Male)
     LR Chisq Df Pr(>Chisq)
type      108  5     <2e-16
```

```
plot(effect("type", m2), grid=TRUE)
```




```
library(lsmeans)
lsmeans(m2, pairwise ~ type)[[2]] # gives pairwise comparisons only
```

```
              estimate     SE df z.ratio p.value
I(Pr) - I(Pu)  0.18277 0.1522 NA  1.2012 0.83649
I(Pr) - II    -0.39950 0.1449 NA -2.7577 0.06455
I(Pr) - III   -0.62027 0.1613 NA -3.8445 0.00169
I(Pr) - IV    -0.92791 0.1389 NA -6.6781 0.00000
I(Pr) - Va    -0.49088 0.1929 NA -2.5442 0.11161
I(Pu) - II    -0.58228 0.1273 NA -4.5743 0.00007
I(Pu) - III   -0.80304 0.1458 NA -5.5092 0.00000
I(Pu) - IV    -1.11068 0.1205 NA -9.2164 0.00000
I(Pu) - Va    -0.67365 0.1801 NA -3.7400 0.00255
II - III      -0.22076 0.1381 NA -1.5982 0.59977
II - IV       -0.52840 0.1112 NA -4.7534 0.00003
II - Va       -0.09137 0.1740 NA -0.5251 0.99519
III - IV      -0.30764 0.1319 NA -2.3321 0.18121
III - Va       0.12939 0.1879 NA  0.6885 0.98328
IV - Va        0.43703 0.1691 NA  2.5843 0.10114
    p values are adjusted using the tukey method for 6 means
```