MATH 523: EXAMPLE - LOW BIRTHWEIGHT DATA

EXAMPLE: Low Birthweight data:- The `birthwt` data frame in `R` has 189 rows and 10 columns. The data were collected at Baystate Medical Center, Springfield, Mass during 1986. This data frame contains the following columns:

- `low` indicator of birth weight less than 2.5 kg (binary response).
- `age` mother's age in years (covariate).
- `lwt` mother's weight in pounds at last menstrual period (covariate).
- `race` mother's race (1 = white, 2 = black, 3 = other) (factor with three levels).
- `smoke` smoking status during pregnancy (factor with two levels).
- `ptl` number of previous premature labours.
- `ht` history of hypertension (factor with two levels).
- `ui` presence of uterine irritability (factor with two levels).
- `ftv` number of physician visits during the first trimester.
- `bwt` birth weight in grams (continuous version of response).

These data are available in the `MASS` library in `R`. Following the recoding and omission of variables devised by Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S, Fourth ed.*, the following code sets up and performs the analysis:

```
1  attach(birthwt)
2  race <- factor(race, labels = c("white", "black", "other"))
3  ptd <- factor(ptl > 0)
4  ftv <- factor(ftv)
5  levels(ftv)[-(1:2)] <- "2+"
6  bwt <- data.frame(low = factor(low), age, lwt, race,
7      smoke = (smoke > 0), ptd, ht = (ht > 0), ui = (ui > 0), ftv)
8  detach("birthwt")
9
10 lbw<-glm(low ~ ., family=binomial, bwt)
```

The model fitted is an additive model; the results are as follows:

```
> summary(lbw)
Call: glm(formula = low ~ ., family = binomial, data = bwt)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7038  -0.8068  -0.5008   0.8835   2.2152

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.82302    1.24471   0.661  0.50848
age         -0.03723    0.03870  -0.962  0.33602
lwt         -0.01565    0.00708  -2.211  0.02705 *
raceblack    1.19241    0.53597   2.225  0.02609 *
raceother    0.74069    0.46174   1.604  0.10869
smokeTRUE    0.75553    0.42502   1.778  0.07546 .
ptdTRUE      1.34376    0.48062   2.796  0.00518 **
htTRUE       1.91317    0.72074   2.654  0.00794 **
uiTRUE       0.68019    0.46434   1.465  0.14296
ftv1        -0.43638    0.47939  -0.910  0.36268
ftv2+        0.17901    0.45638   0.392  0.69488
---
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: 195.48  on 178  degrees of freedom
AIC: 217.48
```

We can attempt to omit individual terms by using the `drop1` command

```
> drop1(lbw,test="Chisq")
Single term deletions

Model:
low ~ age + lwt + race + smoke + ptd + ht + ui + ftv
       Df Deviance    AIC    LRT  Pr(Chi)
<none>      195.48 217.48
age     1   196.42 216.42 0.9419 0.331796
lwt     1   200.95 220.95 5.4739 0.019302 *
race    2   201.23 219.23 5.7513 0.056380 .
smoke   1   198.67 218.67 3.1982 0.073717 .
ptd     1   203.58 223.58 8.1085 0.004406 **
ht      1   202.93 222.93 7.4584 0.006314 **
ui      1   197.59 217.59 2.1100 0.146342
ftv     2   196.83 214.83 1.3582 0.507077
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
```

On this basis, we can attempt to omit several terms, specifically `ftv`, `age` and `ui`.

```
>lbw.1<-update(lbw,~. - ftv - age - ui)
>anova(lbw.1,lbw)

Analysis of Deviance Table

Model 1: low ~ lwt + race + smoke + ptd + ht
Model 2: low ~ age + lwt + race + smoke + ptd + ht + ui + ftv
  Resid. Df Resid. Dev Df Deviance
1       182     200.48
2       178     195.48  4   5.0067

> summary(lbw.1)

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
```

The deviance change of 5.0067 on 4 degrees of freedom is not a significant result, thus the simpler model with the variables omitted is an adequate simplification. It is possible to investigate further by including interactions between factor predictors, and polynomial terms in the covariate `lwt`.