Consider data is from the Heart and Estrogen/Progestin Study (HERS), a clinical trial of hormone therapy for prevention of recurrent heart attacks and deaths among 2,763 post-menopausal women with existing coronary heart disease (Hulley et al., 1998). The HERS data is available at: `http://www.epibiostat.ucsf.edu/biostat/vgsm/data/hersdata.txt`, and it is described in **Regression Methods in Biostatistics**, page 30; variable descriptions are also given on the book website `http://www.epibiostat.ucsf.edu/biostat/vgsm/data/hersdata.codebook.txt`. For now, we will try to predict whether the individuals had a medical condition, `medcond`. We will use the variables `age`, `weight`, `diabetes` and `drinkany`.

```
> HERS <- read.table("HERS.txt", sep="\t", header=T, na.strings=".")
> attach(HERS)
```

Forward

```
> summary(glm(medcond ~ age, family="binomial"))$coef
               Estimate  Std. Error   z value      Pr(>|z|)
(Intercept) -1.60404454 0.400644718 -4.003658 6.237044e-05
age          0.01619155 0.005965348  2.714267 6.642259e-03
    Null deviance: 3647.4  on 2762  degrees of freedom
Residual deviance: 3640.0  on 2761  degrees of freedom
AIC: 3644

> summary(glm(medcond ~ age + weight, family="binomial"))$coef
                Estimate  Std. Error   z value      Pr(>|z|)
(Intercept) -2.169846602 0.496466231 -4.370582 1.239155e-05
age          0.018926204 0.006132171  3.086379 2.026105e-03
weight       0.005279148 0.002742218  1.925138 5.421212e-02
    Null deviance: 3645.5  on 2760  degrees of freedom
Residual deviance: 3634.3  on 2758  degrees of freedom
AIC: 3640.3

> summary(glm(medcond ~ age+ diabetes, family="binomial"))$coef
               Estimate  Std. Error   z value      Pr(>|z|)
(Intercept) -1.89085342 0.407555871 -4.639495 3.492616e-06
age          0.01848156 0.006027953  3.065977 2.169602e-03
diabetes     0.48714064 0.088177630  5.524538 3.303543e-08

> summary(glm(medcond ~ age*diabetes, family="binomial"))$coef
               Estimate  Std. Error   z value      Pr(>|z|)
(Intercept)  -2.51762428 0.478275141 -5.263966 1.409802e-07
age           0.02780399 0.007072117  3.931495 8.441917e-05
diabetes      2.83494074 0.913870625  3.102125 1.921369e-03
age:diabetes -0.03540210 0.013718549 -2.580601 9.862861e-03
    Null deviance: 3647.4  on 2762  degrees of freedom
Residual deviance: 3609.7  on 2760  degrees of freedom
AIC: 3615.7
```

Backward

```
> summary(glm(medcond ~ (age + diabetes + weight + drinkany)^2, family="binomial"))
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.1074581  2.1633263  -0.512   0.6087
age               0.0085067  0.0316546   0.269   0.7881
diabetes          1.8870002  1.1692772   1.614   0.1066
weight           -0.0142558  0.0289592  -0.492   0.6225
drinkany         -0.5868600  1.0756086  -0.546   0.5853
age:diabetes     -0.0304417  0.0147880  -2.059   0.0395 *
age:weight        0.0002082  0.0004289   0.486   0.6273
age:drinkany      0.0073401  0.0131778   0.557   0.5775
diabetes:weight   0.0078719  0.0062412   1.261   0.2072
diabetes:drinkany -0.1361005  0.2052152  -0.663   0.5072
weight:drinkany  -0.0016115  0.0061397  -0.262   0.7930
    Null deviance: 3642.6  on 2758  degrees of freedom
Residual deviance: 3585.7  on 2748  degrees of freedom
AIC: 3607.7


> summary(glm(medcond ~ age + diabetes + weight + drinkany, family="binomial"))
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.874672   0.504526  -3.716 0.000203 ***
age          0.018362   0.006195   2.964 0.003038 **
diabetes     0.432470   0.092422   4.679 2.88e-06 ***
weight       0.001427   0.002854   0.500 0.617138
drinkany    -0.252544   0.083452  -3.026 0.002476 **
    Null deviance: 3642.6  on 2758  degrees of freedom
Residual deviance: 3594.8  on 2754  degrees of freedom
AIC: 3604.8
```

- The big model (with all of the interaction terms) has a deviance of 3585.7; the additive model has a deviance of 3594.8.

$$\chi_6^2 = 3594.8 - 3585.7 = 9.1$$
$$p - value = P(\chi_6^2 \geq 9.1) = 1 - pchisq(9.1, 6) = 0.1680318$$

We cannot reject the null hypothesis, so we know that we don't need the 6 interaction terms. Next we will check whether we need weight.

```
summary(glm(medcond ~ age + diabetes + drinkany, family="binomial"))
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.722505   0.412727  -4.173 3.00e-05 ***
age          0.017573   0.006052   2.904  0.00369 **
diabetes     0.442160   0.089529   4.939 7.86e-07 ***
drinkany    -0.251505   0.083424  -3.015  0.00257 **
    Null deviance: 3644.5  on 2760  degrees of freedom
Residual deviance: 3597.3  on 2757  degrees of freedom
AIC: 3605.3
```

- The additive model has a deviance of 3594.8; the model without weight is 3597.3.

$$\chi_1^2 = 3597.3 - 3594.8 = 2.5$$
$$p - value = P(\chi_1^2 \geq 2.5) = 1 - pchisq(2.5, 1) = 0.1138463$$

We cannot reject the null hypothesis, so we know that we don't need the weight in the model either.