GENERALIZED LINEAR MODELS:Logistic Regression – Model checking

Γιώτα Τουλούμη

Καθηγήτρια Βιοστατιστικής και Επιδημιολογίας
Εργ. Υγιεινής, Επιδημιολογίας και Ιατρικής Στατιστικής
Ιατρική Σχολή Πανεπιστημίου Αθήνας

gtouloum@med.uoa.gr

C. Model checking

The contraceptive use example

Consider the contraceptive use data set:

```
. list
           age   educat   more   cuse     N
  1.       <25      Low     No      0    10
  2.       <25      Low     No      1     4
  3.       <25      Low    Yes      0    53
  4.       <25      Low    Yes      1     6
  5.       <25     High     No      0    50
  6.       <25     High     No      1    10
  7.       <25     High    Yes      0   212
  8.       <25     High    Yes      1    52
  9.     25-29      Low     No      0    19
 10.     25-29      Low     No      1    10
 11.     25-29      Low    Yes      0    60
 12.     25-29      Low    Yes      1    14
 13.     25-29     High     No      0    65
 14.     25-29     High     No      1    27
 15.     25-29     High    Yes      0   155
 16.     25-29     High    Yes      1    54
 17.     30-39      Low     No      0    77
 18.     30-39      Low     No      1    80
 19.     30-39      Low    Yes      0   112
 20.     30-39      Low    Yes      1    33
 21.     30-39     High     No      0    68
 22.     30-39     High     No      1    78
 23.     30-39     High    Yes      0   118
 24.     30-39     High    Yes      1    46
 25.     40-49      Low     No      0    46
 26.     40-49      Low     No      1    48
 27.     40-49      Low    Yes      0    35
 28.     40-49      Low    Yes      1     6
 29.     40-49     High     No      0    12
 30.     40-49     High     No      1    31
 31.     40-49     High    Yes      0     8
 32.     40-49     High    Yes      1     8
```

Measures of goodness of fit

Goodness of fit tests are, by definition, those that compare the observed to the fitted values. In logistic regression (as in any GLM) there are two such statistics: The **Pearson chi-square** and the **deviance**.

The deviance is the likelihood ratio test comparing a model against a *saturated* model as follows:

$$\frac{D(\mathbf{y};\hat{\theta})}{\phi} = -2\left\{l(\hat{\theta};\mathbf{y}) - l(\tilde{\theta};\mathbf{y})\right\}$$

where $l(\tilde{\theta};\mathbf{y})$ is the maximized likelihood of the saturated model and $l(\hat{\theta};\mathbf{y})$ is the maximized likelihood under the model in consideration. In the case of the binomial likelihood (i.e., when data are grouped in k categories of n_i observations each).

Binomial deviance

In the case of the binomial likelihood (grouped in k categories of n_i obs.) the deviance is given by,

$$D = \frac{D(\tilde{\pi};\hat{\pi})}{\phi} = 2\left\{\sum_{i=1}^{k}\left\{y_i \log(\tilde{\pi}_i) + (n_i - y_i)\log(1-\tilde{\pi}_i)\right\} - \left\{y_i \log(\hat{\pi}_i) + (n_i - y_i)\log(1-\hat{\pi}_i)\right\}\right\}$$

$$= \left\{\sum_{i=1}^{k} 2\left\{y_i \log\left[\frac{y_i/n_i}{\hat{\mu}_i/n_i}\right] + (n_i - y_i)\log\left[\frac{(1-y_i/n_i)}{(1-\hat{\mu}_i/n_i)}\right]\right\}\right\}$$

$$= \left\{\sum_{i=1}^{k} 2\left\{y_i \log\left[\frac{y_i}{\hat{\mu}_i}\right] + (n_i - y_i)\log\left[\frac{n_i - y_i}{n_i - \hat{\mu}_i}\right]\right\}\right\} = \left\{\sum_{i=1}^{k} d_i^2\right\}$$

$\hat{\mu}_i = n_i\hat{\pi}_i$ and $\hat{\pi}_i = \hat{\mu}_i/n_i$ and $\tilde{\pi}_i = y_i/n_i$ and $d_i(y_i,\hat{\pi}_i) = \pm\sqrt{2\left[y_i \log\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\log\left(\frac{(n_i - y_i)}{n_i(1-\hat{\pi}_i)}\right)\right]}$, where the sign is determined from the sign of $(y_i - n_i\hat{\pi}_i)$. The deviance has an asymptotic chi-square distribution with k-(p+1) degrees of freedom **IF** the number of categories is small compared to n and does not increase with increasing n. Such would be the case if some of the covariates were continuous and the data could not be grouped in a small number of categories. d_i is the Deviance residual, which we will encounter later in this lecture. **NOTE**: The X^2 approximation is usually quite accurate for differences of deviances even if it is inaccurate for the deviances themselves.

The Pearson chi-square statistic

The Pearson chi-square statistic is given by

$$X^2 = \sum_{i=1}^{k} \left\{ \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)} \right\} = \sum_{i=1}^{k} r_i^2$$

where $r(y_i, \hat{\pi}_i) = \frac{(y_i - n_i \hat{\pi}_i)}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}$. X^2 has an asymptotic chi-square distribution with k-(p+1) degrees of freedom, **IF** the data are grouped in a number of categories that is less than n and does not increase as $n \to \infty$. This means, that the Pearson chi-square statistic does not have a goodness of fit interpretation in cases of individual data (where $k \approx n$). r_i is the Pearson residual for covariate pattern i, which we will encounter later on. It is a good practice not to rely on either deviance or Pearson X^2 when data are sparse. It is much better to look for specific deviations from the model (e.g. test for interactions, non-linear effects).

Contraceptive use example

In the contraceptive data example, if age is not used as a continuous variable, there are 8 covariate categories (=2×4) in each category of contraceptive use. Some data manipulation is in order:

```
. reshape wide N, i(age more educat) j(cuse)
(note:  j = 0 1)

Data                               long   ->   wide
---------------------------------------------------------------------------
Number of obs.                       32   ->      16
Number of variables                   6   ->       6
j variable (2 values)              cuse   ->   (dropped)
xij variables:
                                      N   ->   N0 N1
---------------------------------------------------------------------------
. sort age more educat
. by age more: gen n1=sum(N1)
. by age more: gen n0=sum(N0)
. by age more: drop if _n<_N
. drop educat N0 N1
. rename n1 N1
. rename n0 N0
. generate tot=N0+N1
. label var tot "Total observations (n_i)"

. list
          age       more    contage         N1          N0          tot
  1.      <25         No         20         14          60           74
  2.      <25        Yes         20         58         265          323
  3.    25-29         No       27.5         37          84          121
  4.    25-29        Yes       27.5         68         215          283
  5.    30-39         No         35        158         145          303
  6.    30-39        Yes         35         79         230          309
  7.    40-49         No         45         79          58          137
  8.    40-49        Yes         45         14          43           57
```

Consider the following alternative analysis of contraceptive use by age and desire for more children:

```
. char more[omit] 0

. xi: blogit N1 tot i.age i.more
i.age                Iage_1-4     (naturally coded; Iage_1 omitted)
i.more               Imore_0-1    (naturally coded; Imore_0 omitted)

Logit estimates                                   Number of obs   =       1607
                                                  LR chi2(4)      =     128.88
                                                  Prob > chi2     =     0.0000
Log likelihood = -937.40449                       Pseudo R2       =     0.0643

------------------------------------------------------------------------------
_outcome |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  Iage_2 |   .3678306   .1753673      2.097   0.036        .024117    .7115443
  Iage_3 |   .8077888   .1597533      5.056   0.000        .494678      1.1209
  Iage_4 |   1.022618   .2039337      5.014   0.000       .6229158    1.422321
 Imore_1 |   -.824092   .1171128     -7.037   0.000      -1.053629   -.5945552
   _cons |  -.8698414   .1571298     -5.536   0.000       -1.17781   -.5618727
------------------------------------------------------------------------------
```

Here, `N1` is the number of women using contraceptives in each of the eight `age`×`more` categories and `tot` is the total number of women. `blogit` performs the logistic regression on this binomial sample (i.e., the sample of `N1` out of `tot` women using contraception). Compare these estimates with the output in the previous lecture.

Deviance

We can now derive the deviance manually by following the formula given above. To derive $\hat{\mu}_i$ the expected number of women using contraception in each of the sixteen `age×more` categories we proceed as follows (note that `blogit` produces estimates of *counts* not probabilities):

```
. predict yhat
(option n assumed; predicted no. of cases)
```

Then the deviance is generated as follows:

```
. gen di = 2*(N1*log(N1/yhat) + (tot-N1)*log((tot-N1)/(tot-yhat)) )

. gen D=sum(di)

. display "Deviance = " D[_N]
Deviance = 16.788813

.  display " p = " chiprob(3, D[_N])
 p = .00078105
```

So the p value is p=0.0008, which means that the additive two-factor model does not fit the data adequately. This result is consistent to the analyses shown in the previous lecture.

Note that the square root of `di` is the *deviance* residual. We'll take this up again later on.

Pearson chi-square

The Pearson chi-square statistic is derived similarly:

```
. gen r=(N1-yhat)/sqrt(yhat*(1-yhat/tot))

. gen X2=sum(r^2)

. display "Pearson X2=" X2[_N]
Pearson X2=16.283419

. display " p = " chiprob(3, X2[_N])
 p = .00099191
```

The Pearson chi-square statistic is close to the deviance statistic and is associated with a highly significant p value, which is further evidence for the inadequacy of the two-factor additive model. Notice that `r` is called the *Pearson* residual (we will take this up again momentarily).

The Hosmer and Lemeshow statistic

Consider the models where age was entered as a continuous covariate (dismiss for a second the fact that we assigned a mean age to each group). When individual data are involved, there is a definite need for a goodness of fit statistic. The Hosmer-Lemeshow (HL) statistic fills this need.

The Hosmer and Lemeshow statistic is essentially a Pearson chi-square statistic based on a grouping of the subject group into *g* groups (usually *g* is taken to be ten). Then the Pearson chi-square statistic is derived by considering the $2 \times g$ contingency table.

The grouping can be done by assigning one tenth of subjects to each of the 10 (or *g*) groups, or by assigning one tenth of the estimated probabilities to each group. STATA uses the latter method.

A problem that may arise is "breaking the ties" in a category with a great deal of the observations (i.e., in which group the software will assign the superfluous observations). See Hosmer & Lemeshow for a lucid discussion of this matter.

The HL statistic in the contraceptive-data example

STATA implements the HL statistic as part of the `lfit` command that follows the `logistic` command and the latter can only handle individual-level data. We thus return to the original dataset. The HR statistic is computed as follows:

Step 1. Carry out the logistic regression and generate the predicted probabilities

Step 2. Sort the predicted probabilities

Step 3. Group observations based on the predicted probabilities. Resolve (STATA) ties by assigning all observations with the same predicted value in the same group.

Step 4. Calculate a Pearson chi-square statistic based on the $2 \times g$ contingency table that results from step 3 and the response variable. Based on simulation studies: X^2 degr. of fr.=g-2

Here is the output:

```
. quietly xi: logit cuse i.more contage [freq=N]

. lfit, group(6) table

Logistic model for cuse, goodness-of-fit test
(Table collapsed on quantiles of estimated probabilities)

_Group      _Prob     _Obs_1     _Exp_1     _Obs_0     _Exp_0     _Total
     1     0.1632         58       52.7        265      270.3        323
     2     0.2135         68       60.4        215      222.6        283
     3     0.2743         79       84.8        230      224.2        309
     4     0.3828         65       90.2        187      161.8        252
     5     0.4633        158      140.4        145      162.6        303
     6     0.5730         79       78.5         58       58.5        137

       number of observations =      1607
             number of groups =         6
      Hosmer-Lemeshow chi2(4) =        17.48
                  Prob > chi2 =         0.0016
```

The p value of the Hosmer-Lemeshow chi-square is 17.48, which compared to a chi-square with 4 degrees of freedom results in a p value of 0.0016. This is evidence that the two-factor covariance model with no interaction does not fit the data adequately. Note that we chose *g*=6 as the total number of groups was 8.

Just for clarifying further, let's compute the statistic manually (note that the size of the groups would be close to 1607/6=268 subjects):

```
.  quietly xi: logit cuse i.more contage [freq=N]

. predict phat
(option p assumed; Pr(cuse))

. sort phat
. list age more phat N

           age       more        phat           N
  1.       <25        Yes    .1632108         212
  2.       <25        Yes    .1632108          52
  3.       <25        Yes    .1632108           6
  4.       <25        Yes    .1632108          53
  5.     25-29        Yes    .2135374         155
  6.     25-29        Yes    .2135374          14
  7.     25-29        Yes    .2135374          54
  8.     25-29        Yes    .2135374          60
  9.     30-39        Yes    .2742955         112
 10.     30-39        Yes    .2742955         118
 11.     30-39        Yes    .2742955          33
 12.     30-39        Yes    .2742955          46
 13.       <25         No    .3081821          50
 14.       <25         No    .3081821          10
 15.       <25         No    .3081821          10
 16.       <25         No    .3081821           4
 17.     40-49        Yes    .3700797           8
 18.     40-49        Yes    .3700797          35
 19.     40-49        Yes    .3700797           8
 20.     40-49        Yes    .3700797           6
 21.     25-29         No    .3827633          27
 22.     25-29         No    .3827633          19
 23.     25-29         No    .3827633          65
 24.     25-29         No    .3827633          10
 25.     30-39         No    .4633063          77
 26.     30-39         No    .4633063          78
 27.     30-39         No    .4633063          68
 28.     30-39         No    .4633063          80
 29.     40-49         No    .5729807          46
 30.     40-49         No    .5729807          31
 31.     40-49         No    .5729807          48
 32.     40-49         No    .5729807          12
```

(rows 1–4) = 323 subjects group 1

(rows 5–8) = 283 subjects group 2

(rows 9–12) = 309 subjects group 3

(rows 13–24) = 252 subjects group 4

(rows 25–28) = 303 subjects group 5

(rows 29–32) = 137 subjects group 6

Hand calculation of the HR statistic

The Hosmer-Lemeshow statistic is calculated as a Pearson chi-square statistic based on the 2×6 table

$$X^2 = \sum_{i=1}^{2\times 6} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(58-52.7)^2}{52.7} + \frac{(68-60.4)^2}{60.4} \cdots + \frac{(58-58.5)^2}{58.5}$$

$$\approx 17.48$$

where O_i is the observed count, E_i is the expected count. The expected counts (E_i) are derived by multiplying `_Total` (i.e., the total number of women in this group) in the above output, by `_Prob` (the estimated probability of using contraceptives) and `_Total` by 1-`_Prob`. The associated p value is

```
. di "p = " chiprob(4, 17.48)
p = .00155892
```

which is the same as before and is indicative of the inadequacy of the model.

Model checking

Recall the best model as identified in the previous lecture:

```
. gen contage2=contage*contage

.  xi: logit cuse contage contage2 i.more i.more*contage [freq=N], nolog
i.more               Imore_0-1     (naturally coded; Imore_0 omitted)
i.more*contage       ImXcon_#      (coded as above)
Note: Imore_1 dropped due to collinearity.
Note: contage dropped due to collinearity.

Logit estimates                                   Number of obs   =       1607
                                                  LR chi2(4)      =     143.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -930.18024                       Pseudo R2       =     0.0715

------------------------------------------------------------------------------
    cuse |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 contage |   .2331551   .0651087     3.581   0.000       .1055445    .3607658
contage2 |  -.0024113   .0009398    -2.566   0.010      -.0042532   -.0005693
 Imore_1 |   1.292637   .5810191     2.225   0.026       .1538601    2.431413
ImXcon_1 |  -.0659373   .0176673    -3.732   0.000      -.1005645   -.0313101
   _cons |  -5.216035   1.123734    -4.642   0.000      -7.418513   -3.013557
```

Model checking, is based on residuals and influence measures as was the case in linear regression.

Residuals and influence measures

There are three residuals that we will be focusing on. These are:

1. The Pearson residual for covariate pattern i is $p_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i (n_i - \hat{\mu}_i)/n_i}}$ where $\hat{\mu}_i = n_i \hat{\pi}_i$, and n_i is the number of subjects in the i^{th} covariate pattern. The Pearson residual is produced with the `predict` command in STATA and the option `r`.

2. The *standardized* Pearson residual for covariate pattern i is $s_i = \frac{p_i}{\sqrt{1-h_i}} = \frac{y_i - \hat{\mu}_i}{\sqrt{(1-h_i)\hat{\mu}_i (n_i - \hat{\mu}_i)/n_i}}$. It is produced by the option `rstan` in the `predict` command in STATA. Note that h_i is similar to the "hat" matrix **H** in the general linear model (as extended by Pregibon, 1981 in logistic regression) and is equal to $\mathbf{H} = \mathbf{V}^{1/2}\mathbf{X}(\mathbf{X'VX})^{-1}\mathbf{X'V}^{1/2}$, $\mathbf{V}$ is a diagonal matrix, $v_{ii} = \hat{\mu}_i (n_i - \hat{\mu}_i)/n_i = n_i \hat{\pi}(\mathbf{x}_i)[1 - \hat{\pi}(\mathbf{x}_i)]$.

3. The deviance residual for covariate pattern j is $d_i = \pm\sqrt{2\left[y_i \ln\left(\frac{y_i}{n_i \hat{\pi}_i}\right) + (n_i - y_i)\ln\left(\frac{(n_i - y_i)}{n_i(1-\hat{\pi}_i)}\right)\right]}$. It is produced by the option `deviance` in the `predict` command in STATA.

Residuals and influence measures: Leverage and distance

As an extension of the Cook's distance measure that was introduced in the linear model's discussion, in logistic regression we have its extension in logistic regression (Pregibon, 1981). It is essentially the (standardized) difference between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{(-i)}$ the ML estimate of b excluding all n_i subjects with covariate pattern *i*. The approximate Cook's distance *D* is

$$D_i = \frac{p_i^2 h_i}{(1-h_i)^2} = \frac{s_i^2 h_i}{(1-h_i)}$$

where h_i is the i^th diagonal element of the hat matrix **H**. The Pregibon approximation of the Cook's distance is produced in STATA with option `dbeta` in the STATA command `predict`.

The diagonal elements of the hat matrix can be considered as *leverages* in a similar manner as in the general linear models. These are produced with the option `hat` in the STATA command `predict`.

Leverage

Let h_i denote the i[th] diagonal element of the matrix H defined in page 15. Then, we can show that

$$h_i = \underbrace{n_i\hat{\pi}(\mathbf{x}_i)\left[1-\hat{\pi}(\mathbf{x}_i)\right]}_{v_i}\underbrace{(1,\mathbf{x}_i')(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\begin{pmatrix}1\\ \mathbf{x}_i'\end{pmatrix}}_{b_i} \quad \text{where} \quad b_i = (1,\mathbf{x}_i')(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\begin{pmatrix}1\\ \mathbf{x}_i'\end{pmatrix}$$

A point that must be kept in mind when interpreting the magnitude of h_i is the effect that v_i has on it. Note that, the fit determines the estimated coefficients and since these determine $\hat{\pi}_i$, points with large values of h_i are extreme in the covariate space and thus lie far from the mean. This is if you ignore v_i. Because of v_i at extreme values of $\hat{\pi}_i$ the leverage decreases rapidly and approaches 0. That is, *the points most extreme in the covariate space may have the smallest leverage.*

This is the exact opposite of the situation in linear regression, where the leverage is a monotonic increasing function of the distance of a covariate pattern to the mean. **The practical consequence of this is that to correctly interpret a** particular value of the leverage in logistic regression, we need to know whether or not $\hat{\boldsymbol{\pi}}$ is small (<0.1) or large (>0.9). If 0.1< $\hat{\boldsymbol{\pi}}$ <0.9 then the leverage will give a value that may be thought of as distance. When the estimated probability lies outside (0.1,0.9) then the **value of leverage may not measure distance in the sense that further from the mean implies a larger value.**

Residuals and influence measures: ΔX^2 and ΔD

As a similar idea of the Cook's distance derived above, two more measures of goodness of fit of individual covariate patterns exist. These are ΔX_i^2 and ΔD_i, that is, the difference in the Pearson chi square statistic and the deviance due to removal of the j^{th} covariate pattern. The former measure is

$$\boxed{\Delta X_i^2 = \frac{p_i^2}{(1-h_i)} = s_i^2}$$

ΔX_i^2 is produced in STATA by option `dx2` in the command `predict`. The latter measure is

$\Delta D_i = d_i^2 + \frac{p_i^2 h_i}{(1-h_i)}$ and upon substitution of p_i^2 for d_i^2 this becomes,

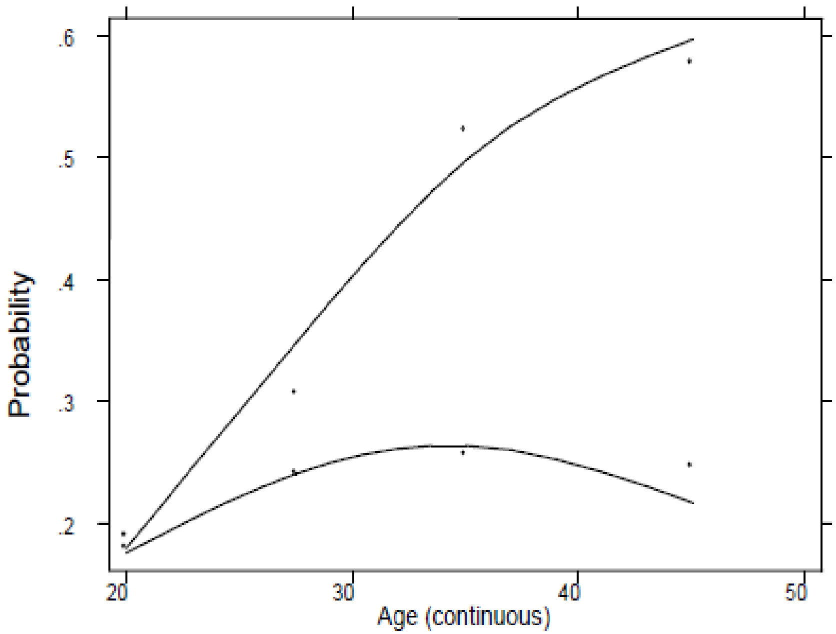
$$\boxed{\Delta D_i = \frac{d_i^2}{(1-h_i)}}$$

ΔD_i is produced by the option `dd` in the STATA command `predict`.

Contraceptive data example

In the example, we produce the fitted values for the probability of contraceptive use as follows:

```
. quietly xi: logit cuse i.more i.age i.more*i.age
> [freq=N]
. predict prob
(option p assumed; Pr(cuse))
. label var prob "Probability"
. quietly xi: logit cuse i.more contage
> contage*contage i.more*contage [freq=N]
. predict phat
. gen phat1=phat if more==1
(16 missing values generated)
. gen phat0=phat if more==0
(16 missing values generated)
. graph phat1 phat0 prob contage, c(ss.) s(iio) xlab
> ylab border
```




```
. sort more

. table contage, contents(mean prob
> mean phat) by(more)

----------+-----------------------
Desires   |
more      |
children? |
and       |
contage   | mean(prob)  mean(phat)
----------+-----------------------
No        |
       20 |   .1891892    .1798393
     27.5 |   .3057851     .348013
       35 |   .5214521    .4976501
       45 |   .5766423     .597039
----------+-----------------------
Yes       |
       20 |   .1795666    .1760204
     27.5 |   .2402827     .240777
       35 |   .2556634    .2641383
       45 |    .245614     .217312
----------+-----------------------
```

Model checking through residuals and influence measures

```
. quietly  xi: logit cuse contage contage2 i.more i.more*contage [freq=N],nolog
. predict p, resid
. predict s, rstand
. predict d, deviance
. predict h,hat
. predict D, dbeta
. predict DX2, dx2
. predict Dd, dd
. predict n, n
```

Notice that n is the number of the covariate pattern. These are

n (Covariate pattern)	more	age	D (~Cook's D)	h (leverage)
1	Yes	<25	0.805561	0.830118
2	No	<25	0.176814	0.610767
3	Yes	25-29	0.000463	0.416496
4	No	25-29	0.965923	0.384639
5	Yes	30-39	0.563625	0.63994
6	No	30-39	4.459163	0.677098
7	Yes	40-49	1.001881	0.599291
8	No	40-49	7.95146	0.841646

Residuals

```
. sum p s d

Variable |       Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
       p |        32   -.0119643    .5481008  -.9751577   .8286497
       s |        32   -.0045499    .9110746  -1.243111   1.458263
       d |        32   -.0143877    .5493285   -.985256   .8287445
```

In situations where the number of subjects per category is fairly large (as is the case here), the central-limit theorem provides a criterion for deciding how large a residual has to be before is considered problematic. A residual larger than 2.0 should be inspected more carefully. We see that no residuals are too large as no residual reaches that threshold. However, the 6th and 8th categories (`more==No` and `age==30-39/40-49`) are associated with a large Cook's distance. Here a criterion similar to the linear-regression situation of a Cook's distance larger than 1.0 being considered large is adopted.

Distance and influence measures

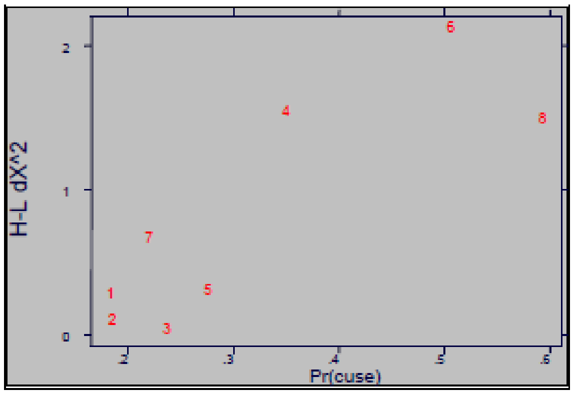
The leverage can be considered in a similar manner as in the linear-regression case. The sum of the diagonal elements of the hat matrix is $(p+1)$ so any leverage twice the average value or higher should be considered further (Pregibon, 1981). The average value ($=(p+1)/k$) here is 5/8=0.625 (the critical value is 2*0.625=1.25), so there are no overly influential categories.

Hosmer and Lemeshow also recommend inspecting graphically the model fit by plotting ΔX^2 and ΔD as well as D against the estimated probability $\hat{\pi}_i = P(Y=1|X=i)$ for covariate pattern i. Then, poorly fit points will be located at the top left and top right corner of the graph, and in general do not conform to the pattern defined by the majority of the points. In the following plots, we identify the points by the covariate pattern n.

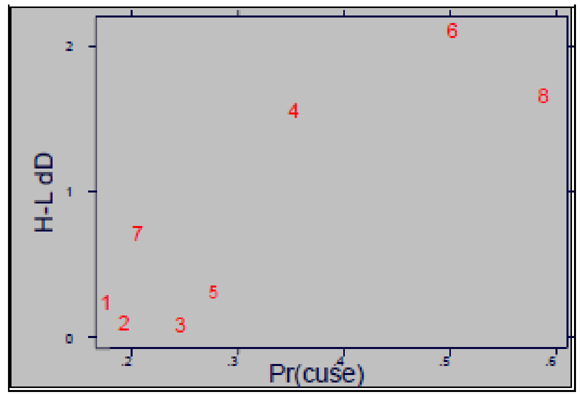
Distance and influence measures

The crude threshold for ΔX^2 and ΔD is 4.0, the approximation of the 95th percentile of the chi-square distribution with one degree of freedom (recall that $\chi^2_{1;0.95} = 3.84$). By extension of the criterion of the Cook's distance, the threshold of D is 1.0.

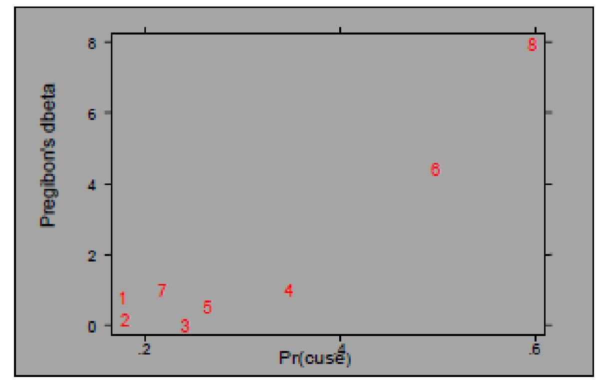
```
.graph DX2 phat, xlab ylab border s([n])
```




```
. graph Dd phat, xlab ylab border s([n])
```




```
.graph D phat, xlab ylab border s([n])
```



We see that no point in the graphs above satisfies any criterion for an unusually poorly fit or influential point. The model fits the data well. At the most, we would like to explore category `n==6` and `n==8` (women ages 30-39 and 40-49 wanting no more children) a bit further.