QV Calculator

Quasi-variances for group and factor effects in generalized linear models, survival analysis, etc.

David Firth
http://www.stats.ox.ac.uk/~firth/qvcalc/
September 1998

The basic idea

A standard presentational device, both in computer output and in published work, is a list of model parameter estimates and associated standard errors. Such a summary is not sufficient to provide standard errors for estimated *combinations* of the parameters, such as differences or ratios, which may often be of interest: for this, estimated covariances are also needed. Space constraints, however, typically mean that the full covariance matrix cannot routinely be reported.

In the important case of parameters $\beta_1, \ldots, \beta_k$ associated with the k levels of an experimental factor or with a k-category classifying variable, the identifiable linear combinations are *contrasts* of the form $\sum_{j=1}^{k} c_j \beta_j$, with $\sum_{j=1}^{k} c_j = 0$, including the simple between-level or between-group comparisons $\beta_i - \beta_j$. From the usual list of estimates and standard errors, with for example the constraint $\beta_1 = 0$ imposed, only the $k - 1$ comparisons $\{\beta_j = \beta_j - \beta_1, \ j = 2, \ldots, k\}$ can be made; standard errors for other contrasts are not available in the absence of the relevant estimated covariances $\{\widehat{\text{cov}}(\hat{\beta}_i, \hat{\beta}_j) : i, j = 2, \ldots, k\}$. The solution implemented in the *QV Calculator* is as follows: from the full covariance matrix, determine k 'quasi-variances', represented by $t_1, \ldots, t_k$, such that

$$\widehat{\text{var}}\left(\sum_{i=1}^{k} c_i \hat{\beta}_i\right) \approx \sum_{i=1}^{k} c_i^2 t_i,$$

for all contrast vectors $\mathbf{c} = (c_1, \ldots, c_k)$. Then if $t_1, \ldots, t_k$ are reported along with the k values $\hat{\beta}_1, \ldots, \hat{\beta}_k$, an approximate standard error is available for any contrast $\sum c_i \hat{\beta}_i$ in the intuitively appealing form $(\sum c_i^2 t_i)^{1/2}$; the variance of a simple contrast $\hat{\beta}_i - \hat{\beta}_j$ is approximated simply by $t_i + t_j$. This basic idea appears to have been suggested first by Ridout (1989) and then, independently and under the name 'floating absolute risk', by Easton, Peto & Babiker (1991).

Approximation method

In the *QV Calculator* the quasi-variances $\{t_i\}$ are chosen to minimize the mean square of errors on the log scale, $\log(t_i + t_j) - \log v_{ij}$, over the set of variances v_{ij} computed for simple contrasts $\hat{\beta}_i - \hat{\beta}_j$. Thus *relative* error is controlled: small contrast variances are deliberately approximated more accurately than large ones, but the relative errors $\{(t_i + t_j)/v_{ij} - 1\}$ are of similar magnitude.

Ridout (1989) also suggests controlling relative error, by minimizing the mean of $\{v_{ij}/(t_i + t_j)\}^{1/2} - 1 + \frac{1}{2}\log\{(t_i + t_j)/v_{ij}\}$. In practice Ridout's method gives very similar results to those obtained from the *QV Calculator*. Easton, Peto and Babiker (1991), on the other hand, suggest choosing the quasi-variances to minimize the mean square of $(t_i + t_j) - v_{ij}$. This controls absolute rather than relative error, and can lead to poor approximation of small contrast variances (in terms of relative error) in order to achieve good approximations to larger ones, which seems undesirable.

An example

(McCullagh & Nelder, 1989, pp204-8): The main effects model for the ship damage rates is summarized as follows by GLIM:

```
? $display e$
! displays parameter estimates for the model
! along with "conventional" standard errors
         estimate        s.e.      parameter
    1       -6.406      0.2827     1
    2      -0.5433      0.2309      TYPE(2)
    3      -0.6874      0.4271      TYPE(3)
    4     -0.07596      0.3779      TYPE(4)
    5       0.3256      0.3067      TYPE(5)
    6       0.6971      0.1946      YEAR(2)
    7       0.8184      0.2208      YEAR(3)
    8       0.4534      0.3032      YEAR(4)
    9       0.3845      0.1538      OP(2)
scale parameter 1.691


? $display v$ ! prints the variance-covariance matrix of estimates
  0.0799
 -0.0530  0.0533
 -0.0458  0.0428   0.1824
 -0.0396  0.0390   0.0384   0.1428
 -0.0408  0.0404   0.0412   0.0387   0.0941
 -0.0266  0.0038   0.0030   0.0020 -0.0002   0.0379
 -0.0343  0.0138   0.0043   0.0024 -0.0025   0.0272   0.0487
 -0.0344  0.0160   0.0126 -0.0111   0.0049   0.0281   0.0367   0.0919
 -0.0094  0.0009 -0.0002 -0.0003   0.0013 -0.0036 -0.0089 -0.0147   0.0237
```

Suppose we are interested in the 5-level factor TYPE (the type of ship). The relevant part of the covariance matrix would be input to the *QV Calculator* as

```
0
0   0.0533
0   0.0428   0.1824
0   0.0390   0.0384   0.1428
0   0.0404   0.0412   0.0387   0.0941
```

Notice the column of zeros here, which reflects the fact that GLIM sets the parameter for TYPE(1) to zero and omits the corresponding column from the displayed variance-covariance matrix. Other model-fitting packages have different conventions, and it does not matter which one is used as long as the covariance matrix corresponding to all levels of the factor (i.e, all 5 levels here rather than just 4) is provided.

The *QV Calculator* displays the quasi-variances in this case as follows:

```
Level     Quasi-variance      Quasi-s.e.
-----     --------------      ----------
  1            4.038E-2          0.201
  2            1.269E-2          0.113
  3           0.140              0.374
  4           0.105              0.324
  5            5.393E-2          0.232
```

The 'quasi-s.e.' column simply presents square roots of the quasi-variances $\{t_i\}$, which are then quasi standard errors on the same scale as the estimated effects.

The *QV Calculator* also gives an indication of the accuracy achieved in the approximation of contrast variances. Typically this is very good (*e.g.*, relative errors all less than 1% in the above example). The quality of this type of approximation in general is the subject of current research by R de Menezes, a doctoral student in the Dept of Statistics at Oxford (expected completion date August 1999).

Suggested use

The quasi-variances $\{t_i\}$ (*i.e.*, in the above example, the numbers 0.0404, 0.0127, 0.140, 0.105, 0.0539) can easily be reported alongside the 'conventional' standard errors, or even in their place, to allow the reader to compute a standard error for any contrast of interest.

Technicalities

The computation is done by Xlisp-Stat (Tierney, 1990), through CGI written in AppleScript.

Please cite use

The *QV Calculator* is provided free, but its author would appreciate acknowledgement in the form of citation in any published work that uses it. For the moment, an appropriate form of citation might be

Firth, D. (1998). *The QV Calculator*. At `http://www.stats.ox.ac.uk/~firth/qvcalc/`.

References

Easton, D.F., Peto, J. & Babiker, A.G.A.G. (1991). Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. *Statistics in Medicine* **10**, 1025–1035.

McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models.* London: Chapman & Hall.

Ridout, M.S. (1989). Summarizing the results of fitting generalized linear models to data from designed experiments. In: *Statistical Modelling: Proceedings of GLIM89 and the 4th International Workshop on Statistical Modelling held in Trento, Italy, July 17-21, 1989* (A. Decarli *et al.*, eds.), pp 262–269. New York: Springer.

Tierney, L. (1990). *LISP-STAT: An Object-oriented Environment for Statistical Computing and Dynamic Graphics.* New York: Wiley.