NAG Toolbox

nag_correg_linregm_fit_newvar (g02dg)

1 Purpose

nag_correg_linregm_fit_newvar (g02dg) calculates the estimates of the arguments of a general linear regression model for a new dependent variable after a call to nag_correg_linregm_fit (g02da).

2 Syntax

```
[rss, cov, q, b, se, res, ifail] = nag_correg_linregm_fit_newvar(rss, ip, irank, cov, q, svd, p, y, wk, 'n', n, 'wt', wt)
[rss, cov, q, b, se, res, ifail] = g02dg(rss, ip, irank, cov, q, svd, p, y, wk, 'n', n, 'wt', wt)
```

Note: the interface to this routine has changed since earlier releases of the toolbox:

At Mark 23: *weight* was removed from the interface; **wt** was made optional.

3 Description

nag_correg_linregm_fit_newvar (g02dg) uses the results given by nag_correg_linregm_fit (g02da) to fit the same set of independent variables to a new dependent variable.

nag_correg_linregm_fit (g02da) computes a QR decomposition of the matrix of p independent variables and also, if the model is not of full rank, a singular value decomposition (SVD). These results can be used to compute estimates of the arguments for a general linear model with a new dependent variable. The QR decomposition leads to the formation of an upper triangular p by p matrix R and an n by n orthogonal matrix Q. In addition the vector $c = Q^{\mathrm{T}}y$ (or $Q^{\mathrm{T}}W^{1/2}y$) is computed. For a new dependent variable, y_{new}, nag_correg_linregm_fit_newvar (g02dg) computes a new value of $c = Q^{\mathrm{T}}y_{\mathrm{new}}$ or $Q^{\mathrm{T}}W^{1/2}y_{\mathrm{new}}$.

If R is of full rank, then the least squares parameter estimates, $\hat{\beta}$, are the solution to

$$R\hat{\beta} = c_1,$$

where c_1 is the first p elements of c.

If R is not of full rank, then nag_correg_linregm_fit (g02da) will have computed an SVD of R,

$$R = Q_* \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} P^{\mathrm{T}},$$

where D is a k by k diagonal matrix with nonzero diagonal elements, k being the rank of R, and Q_* and P are p by p orthogonal matrices. This gives the solution

$$\hat{\beta} = P_1 D^{-1} Q_{*_1}^{\mathrm{T}} c_1,$$

P_1 being the first k columns of P, i.e., $P = (P_1 P_0)$, and Q_{*_1} being the first k columns of Q_*. Details of the SVD are made available by nag_correg_linregm_fit (g02da) in the form of the matrix P^*:

$$P^* = \begin{pmatrix} D^{-1}P_1^{\mathrm{T}} \\ P_0^{\mathrm{T}} \end{pmatrix}.$$

The matrix Q_* is made available through the workspace of nag_correg_linregm_fit (g02da).

In addition to parameter estimates, the new residuals are computed and the variance-covariance matrix of the parameter estimates are found by scaling the variance-covariance matrix for the original regression.

4 References

Golub G H and Van Loan C F (1996) *Matrix Computations* (3rd Edition) Johns Hopkins University Press, Baltimore

Hammarling S (1985) The singular value decomposition in multivariate statistics *SIGNUM Newsl.* **20(3)** 2–25

Searle S R (1971) *Linear Models* Wiley

5 Parameters

5.1 Compulsory Input Parameters

1: **rss** – REAL (KIND=nag_wp)

The residual sum of squares for the original dependent variable.

Constraint: **rss** > 0.0.

2: **ip** – INTEGER

p, the number of independent variables (including the mean if fitted).

Constraint: $1 \leq$ **ip** $\leq$ **n**.

3: **irank** – INTEGER

The rank of the independent variables, as given by nag_correg_linregm_fit (g02da).

Constraint: **irank** > 0, and if **svd** $= \mathit{false}$, then **irank** $=$ **ip**, else **irank** $\leq$ **ip**.

4: **covar**(**ip** $\times$ (**ip** $+$ **1**)/**2**) – REAL (KIND=nag_wp) array

The covariance matrix of the parameter estimates as given by nag_correg_linregm_fit (g02da).

5: **q**(ldq, **ip** $+$ **1**) – REAL (KIND=nag_wp) array

ldq, the first dimension of the array, must satisfy the constraint $ldq \geq$ **n**.

The results of the QR decomposition as returned by nag_correg_linregm_fit (g02da).

6: **svd** – LOGICAL

Indicates if a singular value decomposition was used by nag_correg_linregm_fit (g02da).

svd $= \mathit{true}$
A singular value decomposition was used by nag_correg_linregm_fit (g02da).

svd $= \mathit{false}$
A singular value decomposition was not used by nag_correg_linregm_fit (g02da).

7: **p**(:) – REAL (KIND=nag_wp) array

The dimension of the array **p** must be at least **ip** if **svd** $= \mathit{false}$, and at least **ip** $\times$ **ip** $+ 2 \times$ **ip** otherwise

Details of the QR decomposition and SVD, if used, as returned in array **p** by nag_correg_linregm_fit (g02da).

If **svd** $= \mathit{false}$, only the first **ip** elements of **p** are used; these contain the zeta values for the QR decomposition (see nag_lapack_dgeqrf (f08ae) for details).

If **svd** $= true$, the first **ip** elements of **p** contain the zeta values for the QR decomposition (see nag_lapack_dgeqrf (f08ae) for details) and the next **ip** $\times$ **ip** $+$ **ip** elements of **p** contain details of the singular value decomposition.

8: **y**(**n**) – REAL (KIND=nag_wp) array

The new dependent variable, y_{new}.

9: **wk**(**5** $\times$ (**ip** $-$ **1**) $+$ **ip** $\times$ **ip**) – REAL (KIND=nag_wp) array

If **svd** $= true$, **wk** must be unaltered from the previous call to nag_correg_linregm_fit (g02da) or nag_correg_linregm_fit_newvar (g02dg).

If **svd** $= false$, **wk** is used as workspace.

5.2 Optional Input Parameters

1: **n** – INTEGER

Default: the dimension of the array **y** and the first dimension of the array **q**. (An error is raised if these dimensions are not equal.)

n, the number of observations.

Constraint: **n** $\geq$ **ip**.

2: **wt**(:) – REAL (KIND=nag_wp) array

The dimension of the array **wt** must be at least **n** if $weight =$ 'W', and at least 1 otherwise

If provided>, **wt** must contain the weights to be used in the weighted regression.

If **wt**$(i) = 0.0$, the ith observation is not included in the model, in which case the effective number of observations is the number of observations with nonzero weights.

If **wt** is not provided the effective number of observations is n.

Constraint: if $weight =$ 'W', **wt**$(i) \geq 0.0$, for $i = 1, 2, \ldots, n$.

5.3 Output Parameters

1: **rss** – REAL (KIND=nag_wp)

The residual sum of squares for the new dependent variable.

2: **covar**(**ip** $\times$ (**ip** $+$ **1**)/**2**) – REAL (KIND=nag_wp) array

The upper triangular part of the variance-covariance matrix of the **ip** parameter estimates given in **b**. They are stored packed by column, i.e., the covariance between the parameter estimate given in **b**(i) and the parameter estimate given in **b**(j), $j \geq i$, is stored in **cov**$((j \times (j-1)/2 + i))$.

3: **q**$(ldq,$ **ip** $+$ **1**) – REAL (KIND=nag_wp) array

The first column of **q** contains the new values of c, the remainder of **q** will be unchanged.

4: **b**(**ip**) – REAL (KIND=nag_wp) array

The least squares estimates of the parameters of the regression model, $\hat{\beta}$.

5: **se**(**ip**) – REAL (KIND=nag_wp) array

The standard error of the estimates of the parameters.

6: **res**(**n**) – REAL (KIND=nag_wp) array

The residuals for the new regression model.

7: **ifail** – INTEGER

ifail $= 0$ unless the function detects an error (see Section 5).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail $= 1$

On entry, **ip** < 1,
or **n** $<$ **ip**,
or **irank** ≤ 0,
or **svd** $= \mathit{false}$ and **irank** $\neq$ **ip**,
or **svd** $= \mathit{true}$ and **irank** $>$ **ip**,
or $ldq <$ **n**,
or **rss** ≤ 0.0,
or $\mathit{weight} \neq$ 'U' or 'W'.

ifail $= 2$

On entry, $\mathit{weight} =$ 'W' and a value of **wt** < 0.0.

ifail $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

ifail $= -399$

Your licence key may have expired or may not have been installed correctly.

ifail $= -999$

Dynamic memory allocation failed.

7 Accuracy

The same accuracy as nag_correg_linregm_fit (g02da) is obtained.

8 Further Comments

The values of the leverages, h_i, are unaltered by a change in the dependent variable so a call to nag_correg_linregm_stat_resinf (g02fa) can be made using the value of **h** from nag_correg_linregm_fit (g02da).

9 Example

A dataset consisting of 12 observations with four independent variables and two dependent variables are read in. A model with all four independent variables is fitted to the first dependent variable by nag_correg_linregm_fit (g02da) and the results printed. The model is then fitted to the second dependent variable by nag_correg_linregm_fit_newvar (g02dg) and those results printed.

9.1 Program Text

```
function g02dg_example

fprintf('g02dg example results\n\n');

x = [1, 0, 0, 0;
     0, 0, 0, 1;
     0, 1, 0, 0;
     0, 0, 1, 0;
```

```
     0, 0, 0, 1;
     0, 1, 0, 0;
     0, 0, 0, 1;
     1, 0, 0, 0;
     0, 0, 1, 0;
     1, 0, 0, 0;
     0, 0, 1, 0;
     0, 1, 0, 0];
y = [33.63;     39.62;     38.18;     41.46;     38.02;     35.83;
     35.99;     36.58;     42.92;     37.80;     40.43;     37.89];
ynew = [63; 69; 68; 71; 68; 65; 65; 66; 72; 67; 70; 67];

[n,m]  = size(x);
isx    = ones(m,1,nag_int_name);
mean_p = 'M';
ip     = nag_int(m+1);

% Fit general linear regression model to y
[rss, idf, b, se, covar, res, h, q, svd, irank, p, wk, ifail] = ...
  g02da(mean_p, x, isx, ip, y);

% Display results for y
fprintf('Results for original y-variable using g02da\n\n');
if svd
  fprintf('Model not of full rank\n\n');
end
fprintf('Residual sum of squares = %12.4e\n', rss);
fprintf('Degrees of freedom      = %4d\n', idf);
fprintf('\nVariable   Parameter estimate   Standard error\n\n');
ivar = double([1:ip]');
fprintf('%6d%20.4e%20.4e\n',[ivar b se]');

% Fit same model to ynew
[rss, covar, q, b, se, res, ifail] = ...
    g02dg( ...
           rss, ip, irank, covar, q, svd, p, ynew, wk);

% Display results for ynew
fprintf('\nResults for second y-variable using g02dg\n\n');
fprintf('Residual sum of squares = %12.4e\n', rss);
fprintf('Degrees of freedom      = %4d\n', idf);
fprintf('\nVariable   Parameter estimate   Standard error\n\n');
ivar = double([1:ip]');
fprintf('%6d%20.4e%20.4e\n',[ivar b se]');
```

9.2 Program Results

```
     g02dg example results

Results for original y-variable using g02da

Model not of full rank

Residual sum of squares =   2.2227e+01
Degrees of freedom      =    8

Variable   Parameter estimate   Standard error

     1          3.0557e+01          3.8494e-01
     2          5.4467e+00          8.3896e-01
     3          6.7433e+00          8.3896e-01
     4          1.1047e+01          8.3896e-01
     5          7.3200e+00          8.3896e-01

Results for second y-variable using g02dg

Residual sum of squares =   2.4000e+01
Degrees of freedom      =    8

Variable   Parameter estimate   Standard error
```

1	5.4067e+01	4.0000e-01
2	1.1267e+01	8.7178e-01
3	1.2600e+01	8.7178e-01
4	1.6933e+01	8.7178e-01
5	1.3267e+01	8.7178e-01