

NAG Toolbox

nag_correg_linregm_obs_edit (g02dc)

1 Purpose

nag_correg_linregm_obs_edit (g02dc) adds or deletes an observation from a general regression model fitted by nag_correg_linregm_fit (g02da).

2 Syntax

```
[q, rss, ifail] = nag_correg_linregm_obs_edit(update, mean, isx, q, x, ix, y,
      rss, 'm', m, 'ip', ip, 'wt', wt)

[q, rss, ifail] = g02dc(update, mean, isx, q, x, ix, y, rss, 'm', m, 'ip', ip,
      '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

At Mark 22: **ip** was made optional.

3 Description

nag_correg_linregm_fit (g02da) fits a general linear regression model to a dataset. You may wish to change the model by either adding or deleting an observation from the dataset. nag_correg_linregm_obs_edit (g02dc) takes the results from nag_correg_linregm_fit (g02da) and makes the required changes to the vector c and the upper triangular matrix R produced by nag_correg_linregm_fit (g02da). The regression coefficients, standard errors and the variance-covariance matrix of the regression coefficients can be obtained from nag_correg_linregm_update (g02dd) after all required changes to the dataset have been made.

nag_correg_linregm_fit (g02da) performs a QR decomposition on the (weighted) X matrix of independent variables. To add a new observation to a model with p arguments, the upper triangular matrix R and vector c_1 (the first p elements of c) are augmented by the new observation on independent variables in x^T and dependent variable y_{new} . Givens rotations are then used to restore the upper triangular form.

$$\begin{pmatrix} R : c_1 \\ x : y_{\text{new}} \end{pmatrix} \rightarrow \begin{pmatrix} R^* : c_1^* \\ 0 : y_{\text{new}}^* \end{pmatrix}.$$

Note: only R and the upper part of c are updated the remainder of the Q matrix is unchanged.

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

5 Parameters

5.1 Compulsory Input Parameters

1: **update** – CHARACTER(1)

Indicates if an observation is to be added or deleted.

update = 'A'

The observation is added.

update = 'D'

The observation is deleted.

Constraint: **update** = 'A' or 'D'.

2: **mean_p** – CHARACTER(1)

Indicates if a mean has been used in the model.

mean = 'M'

A mean term or intercept will have been included in the model by `nag_correg_linregm_fit` (g02da).

mean = 'Z'

A model with no mean term or intercept will have been fitted by `nag_correg_linregm_fit` (g02da).

Constraint: **mean** = 'M' or 'Z'.

3: **isx(m)** – INTEGER array

If **isx**(j) is greater than 0, the value contained in $\mathbf{x}((j-1) \times \mathbf{ix} + 1)$ is to be included as a value of x^T , for $j = 1, 2, \dots, \mathbf{m}$.

Constraint: if **mean** = 'M', exactly **ip** – 1 elements of **isx** must be > 0 and if **mean** = 'Z', exactly **ip** elements of **isx** must be > 0 .

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

ldq, the first dimension of the array, must satisfy the constraint $ldq \geq \mathbf{ip}$.

Must be array **q** as output by `nag_correg_linregm_fit` (g02da), `nag_correg_linregm_var_add` (g02de), `nag_correg_linregm_var_del` (g02df) or `nag_correg_linregm_fit_onestep` (g02ee), or a previous call to `nag_correg_linregm_obs_edit` (g02dc).

5: **x(:)** – REAL (KIND=nag_wp) array

The dimension of the array **x** must be at least $(\mathbf{m} - 1) \times \mathbf{ix} + 1$

The **ip** values for the dependent variables of the new observation, x^T . The positions will depend on the value of **ix**.

6: **ix** – INTEGER

The increment for elements of **x**. Two situations are common:

ix = 1

The values of x are to be chosen from consecutive locations in **x**, i.e., $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(\mathbf{m})$.

ix = **ldx**

The values of x are to be chosen from a row of a two-dimensional array with first dimension **ldx**, i.e., $\mathbf{x}(1), \mathbf{x}(\mathbf{ldx} + 1), \dots, \mathbf{x}((\mathbf{m} - 1)\mathbf{ldx} + 1)$.

Constraint: **ix** ≥ 1 .

7: **y** – REAL (KIND=nag_wp)

The value of the dependent variable for the new observation, y_{new} .

8: **rss** – REAL (KIND=nag_wp)

The value of the residual sums of squares for the original set of observations.

Constraint: **rss** \geq 0.0.

5.2 Optional Input Parameters

1: **m** – INTEGER

Default: the dimension of the array **ix**.

m , the total number of independent variables in the dataset.

Constraint: **m** \geq 1.

2: **ip** – INTEGER

Default: the first dimension of the array **q**.

The number of linear terms in general linear regression model (including mean if there is one).

Constraint: **ip** \geq 1.

3: **wt** – REAL (KIND=nag_wp)

Default: 0

If provided, **wt** must contain the weight to be used with the new observation.

If **wt** = 0.0, the observation is not included in the model.

Constraint: if **wt** \geq 0.0, *weight* = 'W'.

5.3 Output Parameters

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

The first **ip** elements of the first column of **q** will contain c_1^* the upper triangular part of columns 2 to **ip** + 1 will contain R^* the remainder is unchanged.

2: **rss** – REAL (KIND=nag_wp)

The updated values of the residual sums of squares.

Note: this will only be valid if the model is of full rank.

3: **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 *ldq* < **ip**,
 or **m** < 1,
 or **ix** < 1,
 or **rss** < 0.0,
 or **update** \neq 'A' or 'D',

or **mean** \neq 'M' or 'Z',
 or **weight** \neq 'U' or 'W',
 or **mean** = 'M' and there are not exactly **ip** – 1 nonzero values of **isx**,
 or **mean** = 'Z' and there are not exactly **ip** nonzero values of **isx**,

ifail = 2

On entry, *weight* = 'W' and **wt** < 0.0.

ifail = 3

The *R* matrix could not be updated. This may occur if an attempt is made to delete an observation which was not in the original dataset or to add an observation to a *R* matrix with a zero diagonal element. This error is also possible when removing an observation which reduces the rank of design matrix. In such cases the model should be recomputed using `nag_correg_linregm_fit` (g02da).

ifail = 4

The residual sums of squares cannot be updated. This will occur if the input residual sum of squares is less than the calculated decrease in residual sum of squares when the new observation is deleted.

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

Higher accuracy is achieved by updating the *R* matrix rather than the traditional methods of updating $X'X$.

8 Further Comments

Care should be taken with the use of `nag_correg_linregm_obs_edit` (g02dc).

- (a) It is possible to delete observations which were not included in the original model.
- (b) If several additions/deletions have been performed you are advised to recompute the regression using `nag_correg_linregm_fit` (g02da).
- (c) Adding or deleting observations can alter the rank of the model. Such changes will only be detected when a call to `nag_correg_linregm_update` (g02dd) has been made. `nag_correg_linregm_update` (g02dd) should also be used to compute the new residual sum of squares when the model is not of full rank.

`nag_correg_linregm_obs_edit` (g02dc) may also be used after `nag_correg_linregm_var_add` (g02de), `nag_correg_linregm_var_del` (g02df) and `nag_correg_linregm_fit_onestep` (g02ee).

9 Example

A dataset consisting of 12 observations with four independent variables is read in and a general linear regression model fitted by `nag_correg_linregm_fit` (g02da) and parameter estimates printed. The last observation is then dropped and the parameter estimates recalculated, using `nag_correg_linregm_update`

(g02dd), and printed. Finally a new observation is added and new parameter estimates computed and printed.

9.1 Program Text

```
function g02dc_example

fprintf('g02dc 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;
     1, 1, 1, 1];
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];

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

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

% Display initial results
fprintf('Results from initial model fit using g02da\n\n');
if svd
    fprintf('Model not of full rank, rank = %4d\n\n', irank);
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]');

% Add or delete observations one at a time
action = {'dropping'; 'adding'};
xobs = [1  1  1  1;
        0  1  0  0];
yobs = [37.89;
        37.89];
nobs = numel(yobs);
ix = nag_int(1);

for j = 1:nobs
    % update regression
    [q, rss, ifail] = g02dc( ...
        action{j}, mean_p, isx, q, xobs(j,:), ix, ...
        yobs(j), rss, 'ip', ip);
    % Parameter estimate update

    if strcmp(action{j}, 'adding')
        n = n + 1;
    else
        n = n - 1;
    end
    [rss, idf, b, se, covar, svd, irank, p, ifail] = ...
        g02dd(nag_int(n), ip, q, rss);

    fprintf('\nResults from %s an observation\n',action{j});
    fprintf('Residual sum of squares = %12.4e\n', rss);
end
```

```
fprintf('Degrees of freedom      = %4d\n', idf);
fprintf('\nVariable   Parameter estimate   Standard error\n\n');
fprintf('%6d%20.4e%20.4e\n',[ivar b se]');
```

```
end
```

9.2 Program Results

g02dc example results

Results from initial model fit using g02da

```
Residual sum of squares = 5.2748e+03
Degrees of freedom      = 8
```

| Variable | Parameter estimate | Standard error |
|----------|--------------------|----------------|
| 1 | 2.0724e+01 | 1.3801e+01 |
| 2 | 1.4085e+01 | 1.6240e+01 |
| 3 | 2.6324e+01 | 1.3801e+01 |
| 4 | 2.2597e+01 | 1.3801e+01 |

Results from dropping an observation

```
Residual sum of squares = 2.1705e+01
Degrees of freedom      = 7
```

| Variable | Parameter estimate | Standard error |
|----------|--------------------|----------------|
| 1 | 3.6003e+01 | 1.0166e+00 |
| 2 | 3.7005e+01 | 1.2451e+00 |
| 3 | 4.1603e+01 | 1.0166e+00 |
| 4 | 3.7877e+01 | 1.0166e+00 |

Results from adding an observation

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

| Variable | Parameter estimate | Standard error |
|----------|--------------------|----------------|
| 1 | 3.6003e+01 | 9.6235e-01 |
| 2 | 3.7300e+01 | 9.6235e-01 |
| 3 | 4.1603e+01 | 9.6235e-01 |
| 4 | 3.7877e+01 | 9.6235e-01 |
