NAG Toolbox

nag_surviv_coxmodel_risksets (g12za)

1 Purpose

nag_surviv_coxmodel_risksets (g12za) creates the risk sets associated with the Cox proportional hazards model for fixed covariates.

2 Syntax

```
[num, ixs, nxs, x, id, nd, tp, irs, ifail] = nag_surviv_coxmodel_risksets(ns, z, isz, ip, t, ic, isi, mxn, 'n', n, 'm', m)
[num, ixs, nxs, x, id, nd, tp, irs, ifail] = g12za(ns, z, isz, ip, t, ic, isi, mxn, 'n', n, 'm', m)
```

3 Description

The Cox proportional hazards model (see Cox (1972)) relates the time to an event, usually death or failure, to a number of explanatory variables known as covariates. Some of the observations may be right-censored, that is, the exact time to failure is not known, only that it is greater than a known time.

Let t_i, for $i = 1, 2, \ldots, n$, be the failure time or censored time for the ith observation with the vector of p covariates z_i. It is assumed that censoring and failure mechanisms are independent. The hazard function, $\lambda(t, z)$, is the probability that an individual with covariates z fails at time t given that the individual survived up to time t. In the Cox proportional hazards model, $\lambda(t, z)$ is of the form

$$\lambda(t, z) = \lambda_0(t) \exp\left(z^{\mathrm{T}}\beta\right),$$

where λ_0 is the base-line hazard function, an unspecified function of time, and β is a vector of unknown arguments. As λ_0 is unknown, the arguments β are estimated using the conditional or marginal likelihood. This involves considering the covariate values of all subjects that are at risk at the time when a failure occurs. The probability that the subject that failed had their observed set of covariate values is computed.

The risk set at a failure time consists of those subjects that fail or are censored at that time and those who survive beyond that time. As risk sets are computed for every distinct failure time, it should be noted that the combined risk sets may be considerably larger than the original data. If the data can be considered as coming from different strata such that λ_0 varies from strata to strata but β remains constant, then nag_surviv_coxmodel_risksets (g12za) will return a factor that indicates to which risk set/strata each member of the risk sets belongs rather than just to which risk set.

Given the risk sets the Cox proportional hazards model can then be fitted using a Poisson generalized linear model (nag_correg_glm_poisson (g02gc) with nag_anova_dummyvars (g04ea) to compute dummy variables) using Breslow's approximation for ties (see Breslow (1974)). This will give the same fit as nag_surviv_coxmodel (g12ba). If the exact treatment of ties in discrete time is required, as given by Cox (1972), then the model is fitted as a conditional logistic model using nag_contab_condl_logistic (g11ca).

4 References

Breslow N E (1974) Covariate analysis of censored survival data *Biometrics* **30** 89–99

Cox D R (1972) Regression models in life tables (with discussion) *J. Roy. Statist. Soc. Ser. B* **34** 187–220

Gross A J and Clark V A (1975) *Survival Distributions: Reliability Applications in the Biomedical Sciences* Wiley

5 Parameters

5.1 Compulsory Input Parameters

1: **ns** – INTEGER

The number of strata. If **ns** > 0 then the stratum for each observation must be supplied in **isi**.

Constraint: **ns** ≥ 0.

2: **z**$(ldz, \mathbf{m})$ – REAL (KIND=nag_wp) array

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

The ith row must contain the covariates which are associated with the ith failure time given in **t**.

3: **isz**$(\mathbf{m})$ – INTEGER array

Indicates which subset of covariates are to be included in the model.

isz$(j) \geq 1$
The jth covariate is included in the model.

isz$(j) = 0$
The jth covariate is excluded from the model and not referenced.

Constraint: **isz**$(j) \geq 0$ and at least one value must be nonzero.

4: **ip** – INTEGER

p, the number of covariates included in the model as indicated by **isz**.

Constraint: **ip** $=$ the number of nonzero values of **isz**.

5: **t**$(\mathbf{n})$ – REAL (KIND=nag_wp) array

The vector of n failure censoring times.

6: **ic**$(\mathbf{n})$ – INTEGER array

The status of the individual at time t given in **t**.

ic$(i) = 0$
Indicates that the ith individual has failed at time **t**(i).

ic$(i) = 1$
Indicates that the ith individual has been censored at time **t**(i).

Constraint: **ic**$(i) = 0$ or 1, for $i = 1, 2, \ldots,$ **n**.

7: **isi**$(:)$ – INTEGER array

The dimension of the array **isi** must be at least **n** if **ns** > 0, and at least 1 otherwise

If **ns** > 0, the stratum indicators which also allow data points to be excluded from the analysis.

If **ns** $= 0$, **isi** is not referenced.

isi$(i) = k$
Indicates that the ith data point is in the kth stratum, where $k = 1, 2, \ldots,$ **ns**.

isi$(i) = 0$
Indicates that the ith data point is omitted from the analysis.

Constraint: if **ns** > 0, $0 \leq$ **isi**$(i) \leq$ **ns**, for $i = 1, 2, \ldots,$ **n**.

8: **mxn** – INTEGER

The first dimension of the array **x** and the dimension of the arrays **ixs** and **id**.

Constraint: **mxn** must be sufficiently large for the arrays to contain the expanded risk sets. The size will depend on the pattern of failures times and censored times. The minimum value will be returned in **num** unless the function exits with **ifail** $= 1$ or 2.

5.2 Optional Input Parameters

1: **n** – INTEGER

Default: the dimension of the arrays **t**, **ic** and the first dimension of the array **z**. (An error is raised if these dimensions are not equal.)

n, the number of data points.

Constraint: **n** ≥ 2.

2: **m** – INTEGER

Default: the dimension of the array **isz** and the second dimension of the array **z**. (An error is raised if these dimensions are not equal.)

The number of covariates in array **z**.

Constraint: **m** ≥ 1.

5.3 Output Parameters

1: **num** – INTEGER

The number of values in the combined risk sets.

2: **ixs**(**mxn**) – INTEGER array

The factor giving the risk sets/strata for the data in **x** and **id**.

If **ns** $= 0$ or 1, **ixs**$(i) = l$ for members of the lth risk set.

If **ns** > 1, **ixs**$(i) = (j-1) \times$ **nd** $+ l$ for the observations in the lth risk set for the jth strata.

3: **nxs** – INTEGER

The number of levels for the risk sets/strata factor given in **ixs**.

4: **x**(**mxn**, **ip**) – REAL (KIND=nag_wp) array

The dimension of the array **x** will be **num**

The first **num** rows contain the values of the covariates for the members of the risk sets.

5: **id**(**mxn**) – INTEGER array

Indicates if the member of the risk set given in **x** failed.

id$(i) = 1$ if the member of the risk set failed at the time defining the risk set and **id**$(i) = 0$ otherwise.

6: **nd** – INTEGER

The number of distinct failure times, i.e., the number of risk sets.

7: **tp**(**n**) – REAL (KIND=nag_wp) array

tp(i) contains the ith distinct failure time, for $i = 1, 2, \ldots,$ **nd**.

8: **irs**(**n**) – INTEGER array

Indicates rows in **x** and elements in **ixs** and **id** corresponding to the risk sets. The first risk set corresponding to failure time **tp**(1) is given by rows 1 to **irs**(1). The lth risk set is given by rows **irs**$(l-1)+1$ to **irs**(l), for $l = 1, 2, \ldots,$ **nd**.

9: **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, **m** < 1,
or **n** < 2,
or **ns** < 0,
or $ldz <$ **n**.

ifail $= 2$

On entry, **isz**$(i) < 0$ for some i,
or the value of **ip** is incompatible with **isz**,
or **ic**$(i) \neq 1$ or 0.
or **ns** > 0 and **isi**$(i) < 0$,
or **ns** > 1 and **isi**$(i) >$ **ns**.

ifail $= 3$

mxn is too small, the minimum value is returned in **num**.

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

Not applicable.

8 Further Comments

When there are strata present, i.e., **ns** > 1, not all the **nxs** groups may be present.

9 Example

The data are the remission times for two groups of leukemia patients (see page 242 of Gross and Clark (1975)). A dummy variable indicates which group they come from. The risk sets are computed using nag_surviv_coxmodel_risksets (g12za) and the Cox's proportional hazard model is fitted using nag_contab_condl_logistic (g11ca).

9.1 Program Text

```
     function g12za_example

fprintf('g12za example results\n\n');

% Data
n = 42;
m = 1;
z          = zeros(n,m);
ic         = zeros(n,1,nag_int_name);
ic(31:end) = 1;
z(22:end)  = 1;
isz        = [nag_int(1)];

t = [ 1;      1;      2;      2;      3;      4;      4;
      5;      5;      8;      8;      8;      8;     11;
     11;     12;     12;     15;     17;     22;     23;
      6;      6;      6;      7;     10;     13;     16;
     22;     23;      6;      9;     10;     11;     17;
     19;     20;     25;     32;     32;     34;     35];

% Parameters
ns  = nag_int(0);
ip  = nag_int(m);
isi = [nag_int(0)];
mxn = nag_int(1000);

% Create risk set
[num, ixs, nxs, x, id, nd, tp, irs, ifail] = ...
  g12za(...
        ns, z, isz, ip, t, ic, isi, mxn);

% Parameters for conditional logistic analysis
tol = 1e-05;
maxit = nag_int(10);

% initial estimate
b   = zeros(m,1);
% Calculate parameter estimates
[dev, b, se, sc, covar, nca, nct, ifail] = ...
g11ca( ...
       nxs, x, isz, id, ixs, b, tol, maxit, 'n', num);

% Display results
ns = max(ns,1);
fprintf(' Parameter      Estimate       Standard Error\n\n');
ivar = [1:m]';
fprintf('%6d%18.4f%18.4f\n',[ivar b se]');
```

9.2 Program Results

```
     g12za example results

 Parameter      Estimate       Standard Error

     1            1.6282            0.4331
```