

# NAG Library Chapter Introduction

## G22 – Linear Model Specification

**Note:** please be advised that this chapter contains routines classed as ‘experimental’. Please see Section 3.1.1 in How to Use the NAG Library and its Documentation for further information

### Contents

<b>1</b>	<b>Scope of the Chapter</b> .....	2
<b>2</b>	<b>Background to the Problems</b> .....	2
2.1	G22 Handles .....	2
2.2	Specifying a Linear Model .....	2
<b>3</b>	<b>Recommendations on Choice and Use of Available Routines</b> .....	2
3.1	G22 Handles .....	2
3.2	Specifying a Linear Model .....	2
<b>4</b>	<b>Functionality Index</b> .....	3
<b>5</b>	<b>Auxiliary Routines Associated with Library Routine Arguments</b> .....	3
<b>6</b>	<b>Routines Withdrawn or Scheduled for Withdrawal</b> .....	3
<b>7</b>	<b>References</b> .....	3

## 1 Scope of the Chapter

The routines in this chapter provide a mechanism for specifying a linear model using a text based modelling language and are intended to be used in conjunction with the model fitting routines from other chapters, for example Chapter G02.

## 2 Background to the Problems

### 2.1 G22 Handles

Chapter G22 makes heavy use of data structures for passing information between routines. For portability reasons these structures are not referred to directly, but through C pointers. Throughout the documentation these pointers are referred to as G22 handles.

Once the G22 handle is no longer required the associated memory should be released by calling **g22zaf**. It is always safe to release the memory associated with a G22 handle as no G22 handle will ever reference another.

### 2.2 Specifying a Linear Model

Let  $D$  denote a data matrix with  $n$  observations on  $m_d$  independent variables, denoted  $V_1, V_2, \dots, V_{m_d}$ . Let  $y$  denote a vector of  $n$  observations on a dependent variable.

A linear model,  $\mathcal{M}$ , as the term is used in this chapter, expresses a relationship between the independent variables,  $V_j$ , and the dependent variable. This relationship can be expressed as a series of additive terms  $T_1 + T_2 + \dots$ , with each term,  $T_t$ , representing either a single independent variable  $V_j$ , called the main effect of  $V_j$ , or the interaction between two or more independent variables. An interaction term, denoted here using the  $.$  operator, allows the effect of an independent variable on the dependent variable to depend on the value of one or more other independent variables.

## 3 Recommendations on Choice and Use of Available Routines

### 3.1 G22 Handles

Once any G22 handle is no longer required the associated memory should be released by calling **g22zaf**.

All G22 handles have optional arguments associated with them. Some of these optional arguments are specific to a particular type of G22 handle and are described in the documentation of the routine that creates the G22 handle. Other optional arguments are common across all G22 handles. These are described in the documentation for **g22zmf** and **g22znf** which can be used to set and get an optional argument respectively.

### 3.2 Specifying a Linear Model

Prior to specifying a linear model the data matrix,  $D$ , must be described. This is done using **g22ybf**. The linear model,  $\mathcal{M}$ , can then be specified as a text string containing a formula. This allows the model to be specified via variable names and avoids the need to explicitly handle interaction terms. The linear model is specified using **g22yaf** and the documentation for that routine describes the syntax of the formula.

In many of the routines in the NAG Library, for example the regression routines in Chapter G02, a linear model is specified directly via the design matrix,  $X$ . The design matrix is defined by the data matrix,  $D$  and the linear model,  $\mathcal{M}$  and its construction usually requires the use of dummy variables. **g22ycf** constructs the design matrix from the formula supplied to **g22yaf**.

The utility routine **g22ydf** can be used to obtain labels for the parameters of the model as well as a variety of information required by some of the routines in Chapter G02.

## 4 Functionality Index

Linear model,	
construct design matrix .....	g22ycf
data description .....	g22ybf
nested model .....	g22ydf
specification from formula string .....	g22yaf
Service routines,	
destroy a G22 handle .....	g22zaf
general option getting routine .....	g22znf
general option setting routine .....	g22zmf

## 5 Auxiliary Routines Associated with Library Routine Arguments

None.

## 6 Routines Withdrawn or Scheduled for Withdrawal

None.

## 7 References

None.

---