Dispersion Models in Arc

Sanford Weisberg

Department of Applied Statistics, University of Minnesota, St. Paul, MN 55108-6042.

Supported by National Science Foundation Grant DUE 97-52887.

January 16, 2001, minor revisiion, November 19, 2002

Abstract

This paper describes an **Arc** add-in for fitting *normal dispersion models*. The lisp code also provides a template for adding other fitting models to **Arc**.

1 Introduction

The *normal dispersion model* is a version of the 2D model described in Section 18.3.3 of *Applied Regression Including Computing and Graphics*. Suppose we have a set of p predictors $\mathbf{x}$ from which we derive a set of k terms $\mathbf{u}$. We assume that

$$\mathrm{E}(y|\mathbf{x}) = \mathsf{m}(\boldsymbol{\eta}_1^T\mathbf{u}) \text{ and } \mathrm{Var}(y|\mathbf{x}) = \exp(\boldsymbol{\eta}_2^T\mathbf{u}) \tag{1}$$

The parameter vectors $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are estimated separately. m is called the kernel mean function. If $\mathbf{u} = \mathbf{x}$, this is a 2D model; in general the dimension may be higher than two.

This document explains how to use **Arc** to fit normal dispersion models. Section 2 describes the procedure for obtaining the computer code that is needed. Section 3 provides an example on its use. Section 4 provides a brief description of how the code works; further information is available in the comments in the file dispmod.lsp.

2 Getting the Code

The user must use the following procedure:

1. Download the file dispmod.lsp, which is accessible from from www.stat.umn.edu/arc/addons.html. If you are using Windows or Macintosh, place it in the Extras directory in your Arc directory. If you are using Unix, the file should either be put in the director /usr/local/lib/Arc/Extras, or in your local Extras directory. The file will load automatically whenever you start **Arc**.

2. To remove the file, either move it to another directory, or change its name so it no longer ends in ".lsp."

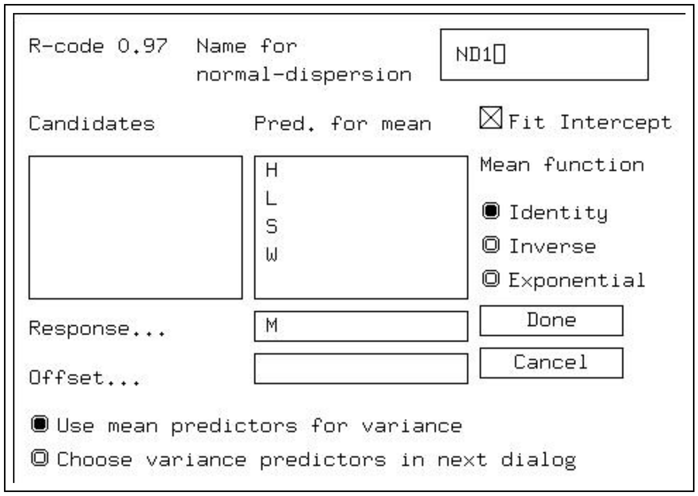

Figure 1: The normal dispersion dialog.

3 Using dispmod

After loading the file `dispmod.lsp`, load the file `mussels.lsp`, described in Section 19.4 of *Applied Regression Including Computing and Graphics*. Fit the regression of M on the four predictors without transformation. Using model checking plots, Chapter 17, it will be clear that the mean function fit by this model matches the data, the variance function is clearly wrong. To fix this problem, we will fit a normal dispersion model.

Select the item "Fit normal dispersion" from the Graph&Fit menu. This will give the dialog shown in Figure 1. Use this dialog to select the predictors for the kernel mean function and the response. There are three choices for `m`: the default identity kernel mean function, as well as the inverse and the exponential kernel mean functions. We will use the identity here. The user can choose to use the same predictors for the variance function or different predictors (using a second dialog) using radio buttons. These models do not now permit the use of weights (although extension to do so would not be hard). Clicking the Done button will fit the model:

```
Normal Dispersion Model:  Model for the mean
Iteration 1: deviance = 570.095
Iteration 2: deviance = 419.407
Iteration 3: deviance = 419.328
Iteration 4: deviance = 419.328    Four iterations were required

Data set = Mussels, Name of Fit = ND1
Normdisp Regression
Kernel mean function = Identity
Response      = M
Terms         = (H L S W)
Coefficient Estimates   Estimates for the mean function
```

```
Label      Estimate        Std. Error     Est/SE
Constant  -6.65794         2.50566        -2.657
H          0.132320        0.0599194       2.208
L         -0.0114649       0.0308125      -0.372
S          0.107331        0.0138865       7.729
W          0.0447194       0.109240        0.409

Scale factor:                   1.
Number of cases:               82
Degrees of freedom:            77
Pearson X2:                82.000
Deviance:                  82.000

Normal Dispersion Model:  Model for the variance
Coefficient Estimates    Estimates for the variance function
Label      Estimate        Std. Error     Est/SE
Constant  -0.943177        1.62283        -0.581
H         -0.00583039      0.0261662      -0.223
L          0.00526995      0.0129822       0.406
S          0.00513276      0.00542125      0.947
W          0.0472197       0.0559652       0.844

Scale factor:              1.41421
Number of cases:               82
Degrees of freedom:            77
Pearson X2:               307.968
Deviance:                 233.766
-2*L(max), const var:     456.664  deviance, assuming variance constant
-2*L(max), model:         419.328  deviance at maximum
```

In addition to the output, the usual graphics are available in **Arc** to study this model further.

4 How it works

This program uses the algorithm proposed by M. Aitkin (1987), Modeling variance heterogeneity in normal regression models using GLIM, *Applied Statistics*, 36, 332–339. Since all the details are carefully laid out in that paper, we give the general outline of the algorithm:

1. Select as starting values $\hat{\boldsymbol{\eta}}_2 = \mathbf{0}$. This is equivalent to assuming the variance function is constant.
2. For the current value of $\hat{\boldsymbol{\eta}}_2$, estimate $\boldsymbol{\eta}_1$ using weighted least squares, with weights given by $1/\exp(\hat{\boldsymbol{\eta}}_2^T\mathbf{u})$.
3. For the current value of $\hat{\boldsymbol{\eta}}_1$, compute the residuals $\hat{e}^2$, and reëstimate $\hat{\boldsymbol{\eta}}_2$ by fitting a gamma regression with response $\hat{e}^2$, scale factor equal to 2, and predictors specified in the dialogs.
4. Repeat steps 2 and 3 until a convergence criterion is met.

The program then reports the results of the two separate regression, including estimates, estimated standard errors of coefficients, and the value of the deviance at the maximum. Also included is the deviance assuming that the variance function is constant. The difference between these two deviances is a test that $\boldsymbol{\eta}_2 = \mathbf{0}$. In the example, this difference is 37.336 with 4 d.f, suggesting strongly that the variance is not constant.

5 Revision

This add-on did not correctly handle missing values. This bug was fixed in November, 2002.