



Some steps towards more convenient mixed-model analyses

Roger Payne

VSN International, 5 The Waterhouse,
Waterhouse Street, Hemel Hempstead, UK

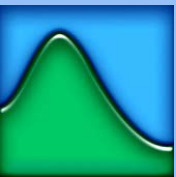
Palm Cove, 14th July 2011

GenStat 14th Edition

- highlights

- new start page for quick links to files and help
- ability to link to *Computer-Assisted Statistics Textbook (CAST)* e-books
- graphics – "child" viewer, new object oriented wizard and tabbed menus, automatic definition of colour bands, 2-d graphs with marginal distribution plots, keys in trellis plots, names for symbols
- regression – within-dose error estimation, ability to specify d.f. for fixed dispersions, paired-preference comparison tests, logistic ridge regression
- analysis-of-variance – procedure to decide best algorithm automatically
- design – spreadsheet representation of field plans
- REML – comparison contrasts, easier assessment of random models, easier meta-analysis specification
- random permutation tests for two-dimensional contingency tables
- statistical genetics – construction of genetic linkage maps and kinship matrices, association mapping of data from multi-environment trials
- time series – Kalman filter, harmonic and cross-spectrum analyses
- spatial statistics – more anisotropic models in variogram & kriging menus
- data mining – radial basis functions
- survey analysis – CPro data files

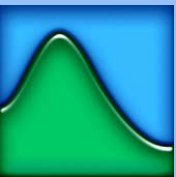
..



REML

- the REML algorithm is very powerful and provides everything you need, but some analyses can be complicated
- they should now be easier using recent new procedures
 - VAIC calculates Akaike and Schwarz information coefficients
 - VRACCUMULATE forms a summary accumulating the results of a sequence of REML random models
 - VRMETA forms the random model for a REML meta analysis
 - VLSD prints approximate least significant differences of means
 - VMCOMPARISON performs pairwise comparisons between means
 - VTCOMPARISON calculates comparison contrasts within a multi-way table of REML predicted means
- see *Release 14: new features* in the on-line help for full info

..

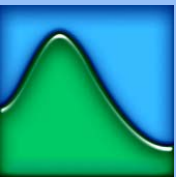


Example – plasticity of *Parus major*



- study by John Quinn and colleagues at University of Oxford of the individual responsiveness of Great Tits to the risks of starvation and predation
 - studied how individual birds balance risks of predation (feeders in exposed or sheltered positions) and starvation (feeding more vital in morning than afternoon)
 - and related that to their “personality” (behavioural plasticity)
- full fixed model
 - risk*ampm*age*sex
 - risk – position of feeder
 - ampm – time of day
- full random model
 - bird*location + bird*risk*ampm
 - location – 8 independent sites in Wytham Wood
- y-variate
 - $\text{Logit}(100 \times (\text{time at low starvation-risk feeder}) / (\text{total time present}))$

..



Analysis

Determine random model

Fit full fixed model of `mintemp + risk*ampm*age*sex`, and simplify random model from full model of `bird*location + bird.risk + bird.ampm + bird.risk.ampm`.

```
14 VCOMPONENTS [FIXED= mintemp + risk*ampm*age*sex; FACTORIAL=9]\
15     RANDOM=bird*location + bird.risk + bird.ampm + bird.risk.ampm;\
16     CONSTRAINTS=none
17 REML [PRINT=*] logit
18 VRACCUMULATE [PRINT=*] 'Full random model'
19 VCOMPONENTS [FIXED= mintemp + risk*ampm*age*sex; FACTORIAL=9]\
20     RANDOM=bird*location + bird.risk + bird.ampm; CONSTRAINTS=none,none
21 REML [PRINT=*] logit
22 VRACCUMULATE [PRINT=deviance,aic,dfrandom,change]
```

← save info from full random model

← save and print from model without `bird.risk.ampm`

Accumulated summary of REML random models

	Deviance	AIC	Random d.f.	Change in deviance	Change in random d.f.	Change chi-prob
Full random model	3183.32	3197.32	7	*	*	*
- bird.risk.ampm	3185.19	3197.19	6	1.87	1	0.172

Note: omits constant, $-\log(\det(\mathbf{X}\mathbf{X}))$, that depends only on the fixed model.

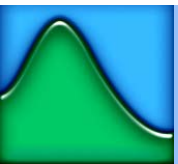
```
23 CAPTION 'Interim conclusion',\
24     !t('Output shows no evidence of a bird.risk.ampm random term,',\
25     'drop this and continue.');
```

Interim conclusion

Output shows no evidence of a `bird.risk.ampm` random term, drop this and continue.

Analysis

```
26
27 VRACCUMULATE [PRINT=*; METHOD=restart] 'Full random model (but without bird.risk.ampm) .'
28 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\
29 RANDOM=bird + location + bird.risk + bird.ampm; CONSTRAINTS=none
30 REML [PRINT=*] logit
31 VRACCUMULATE [PRINT=*] 'Drop bird.location from full random model' ← assess
32 bird.location
33 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\
34 RANDOM=bird*location + bird.risk + bird.ampm; CONSTRAINTS=none
35 REML [PRINT=*] logit ← assess
36 VRACCUMULATE [PRINT=*] 'Return to full random model' bird.risk
37 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\
38 RANDOM=bird*location + bird.ampm; CONSTRAINTS=none,none
39 REML [PRINT=*] logit
40 VRACCUMULATE [PRINT=*] 'Drop bird.risk from full random model'
41
42 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\
43 RANDOM=bird*location + bird.risk + bird.ampm; CONSTRAINTS=none
44 REML [PRINT=*] logit
45 VRACCUMULATE [PRINT=*] 'Return to full random model' ← assess
46 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\ bird.ampm
47 RANDOM=bird*location + bird.risk; CONSTRAINTS=none
48 REML [PRINT=*] logit
49 VRACCUMULATE [PRINT=*] 'Drop bird.ampm from full random model'
50
51 VRACCUMULATE [PRINT=deviance,aic,dfrandom,change; METHOD=printonly]
```



Analysis

Accumulated summary of REML random models

	Deviance	AIC	Random d.f.
Full random model (but without bird.risk.ampm).	3185.19	3197.19	6
Drop bird.location from full random model	3188.85	3198.85	5
Return to full random model	3185.19	3197.19	6
Drop bird.risk from full random model	3217.17	3227.17	5
Return to full random model	3185.19	3197.19	6
Drop bird.ampm from full random model	3185.88	3195.88	5
	Change in deviance	Change in random d.f.	
Full random model (but without bird.risk.ampm).	*	*	
Drop bird.location from full random model	3.66	1	
Return to full random model	-3.66	-1	
Drop bird.risk from full random model	31.98	1	
Return to full random model	-31.98	-1	
Drop bird.ampm from full random model	0.70	1	
	Change chi-prob		
Full random model (but without bird.risk.ampm).	*		
Drop bird.location from full random model	0.056		
Return to full random model	0.056		
Drop bird.risk from full random model	0.000		
Return to full random model	0.000		
Drop bird.ampm from full random model	0.404		

← conclusion drop
bird.ampm

Note: omits constant, $-\log(\det(\mathbf{X}\mathbf{X}))$, that depends only on the fixed model.

Analysis

```

57 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\
58     RANDOM=bird*location + bird.risk; CONSTRAINTS=none
59 REML        [PRINT=*] logit
60 VRACCUMULATE [PRINT=*; METHOD=restart] 'Full random model but without bird.ampm'
61 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\
62     RANDOM=bird*location; CONSTRAINTS=none
63 REML        [PRINT=*] logit
64 VRACCUMULATE [PRINT=*] 'Drop bird.risk from model'
65
66 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\
67     RANDOM=bird*location + bird.risk; CONSTRAINTS=none,none
68 REML        [PRINT=*] logit
69 VRACCUMULATE [PRINT=*] 'Return to full random model without bird.ampm'
70 VCOMPONENTS [FIXED= mintemp + risk * ampm * age * sex; FACTORIAL=9]\
71     RANDOM=bird + location + bird.risk; CONSTRAINTS=none,none
72 REML        [PRINT=*] logit
73 VRACCUMULATE [PRINT=*] 'Drop bird.location from model'
74
75 VRACCUMULATE [PRINT=deviance,aic,dfrandom,change; METHOD=printonly]

```

← recheck
random
model

Accumulated summary of REML random models

	Deviance	AIC	Random d.f.
Full random model but without bird.ampm	3185.88	3195.88	5
Drop bird.risk from model	3233.36	3241.36	4
Return to full random model without bird.ampm	3185.88	3195.88	5
Drop bird.location from model	3189.82	3197.82	4
	Change in deviance	Change in random d.f.	Change chi-prob
Full random model but without bird.ampm	*	*	*
Drop bird.risk from model	47.47	1	0.000
Return to full random model without bird.ampm	-47.47	-1	0.000
Drop bird.location from model	3.94	1	0.047

← keep other
random
terms

Analysis

REML variance components analysis

Response variate: logit
Fixed model: Constant + risk + ampm + age + risk.ampm + risk.age + ampm.age
Random model: bird + location + bird.location + bird.risk
Number of units: 636
Residual term has been added to model
Sparse algorithm with AI optimisation

← fixed model
after dropping
unnecessary terms
(using F statistics)

Tests for fixed effects

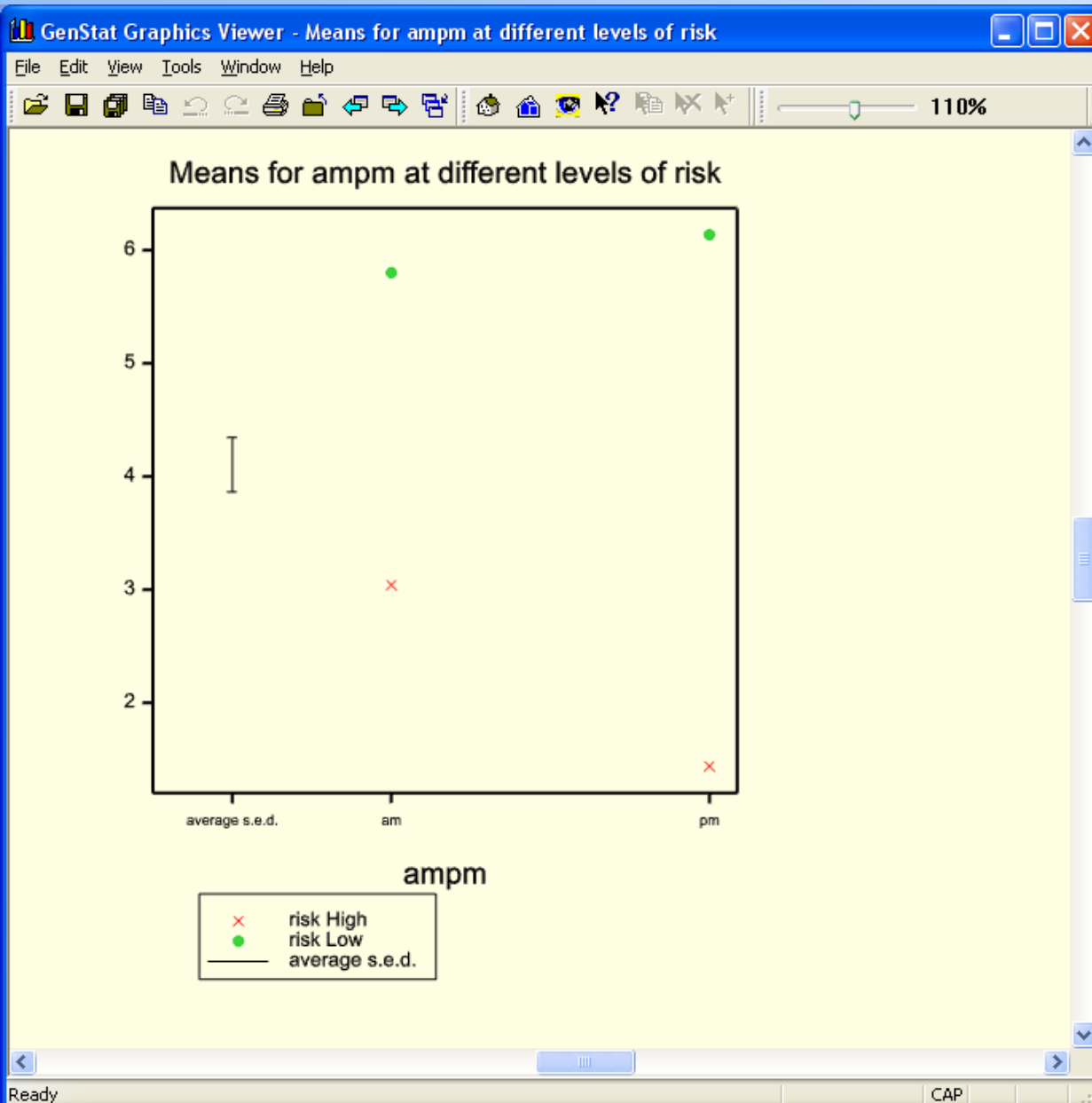
Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
risk	84.85	1	84.85	167.6	<0.001
ampm	7.88	1	7.88	318.9	0.005
age	0.48	1	0.48	115.0	0.490
risk.ampm	12.26	1	12.26	320.9	<0.001
risk.age	4.34	1	4.34	164.1	0.039
ampm.age	4.06	1	4.06	315.3	0.045

Dropping individual terms from full fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
risk.ampm	12.83	1	12.83	320.9	<0.001
risk.age	4.30	1	4.30	164.1	0.040
ampm.age	4.06	1	4.06	315.3	0.045

Analysis

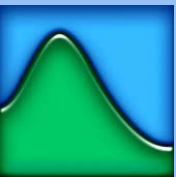


← overall, birds are less willing to feed on high-risk feeders in the afternoon when the starvation risk is lower

Further analyses included a covariate EB (exploratory behaviour) which allowed the reaction to risk to be assessed against the birds' personalities..

Example – meta analysis

- meta analysis of variety trial somewhere in North America
- 1324 test genotypes and 2 standards assessed at 17 sites
- initial program analysed the sites individually to determine the most appropriate random model to use at each one
 - random rows and/or columns and/or spatial correlation
- program then written to give the combined analysis
- ..



Analysis

```
4  CALCULATE kg_ha = MVINSERT(kg_ha; kg_ha== -9)
5  DUMMY      Location,Row,Column,Genotype,Kg_ha; VALUE=loc,prow,pcol,NAME,kg_ha
6
7  VRMETA     [RANDOM=rmodel; EXPERIMENTS=Location] \
8             'Altona','Banff','Drumheller','Edmonton','Guelph',\
9             'Kentville','Lethbridge','Regina','St. Paul','Thunder Bay';\
10 LOCAL=2(!f(Row+Column)),!f(Row),2(!f(Row+Column)),\
11          !f(Row),!f(Row+Column),!f(Row),2(!f(Row+Column))
12
13 VCOMPONENTS [EXPERIMENT=Location] Genotype + Location + #rmodel
14 VRESIDUAL [EXPERIMENT=4; TERM=Row.Column] MODEL=AR,AR; FACTOR=Row,Column; ORDER=1
15 VRESIDUAL [EXPERIMENT=7; TERM=Row.Column] MODEL=AR; FACTOR=Column; ORDER=1
16 VRESIDUAL [EXPERIMENT=9; TERM=Row.Column] MODEL=AR,AR; FACTOR=Row,Column; ORDER=1
17 VRESIDUAL [EXPERIMENT=13; TERM=Row.Column] MODEL=AR,AR; FACTOR=Row,Column; ORDER=1
18 REML       [PRINT=model,components,waldTests; MAXCYCLE=100;\
19            FMETHOD=none; PSE=differences; MVINCLUDE=explanatory,yvariate;\
20            METHOD=AI; MAXCYCLE=100] Kg_ha
```

← define random terms to be fitted only at specific sites

REML variance components analysis

Response variate: kg_ha
Fixed model: Constant
Random model: NAME + loc + prow@Altona + pcol@Altona + prow@Banff + pcol@Banff + prow@Drumheller + prow@Edmonton + pcol@Edmonton + prow@Guelph + pcol@Guelph + prow@Kentville + prow@Lethbridge + pcol@Lethbridge + prow@Regina + prow@St. Paul + pcol@St. Paul + prow@Thunder Bay + pcol@Thunder Bay
Number of units: 13032



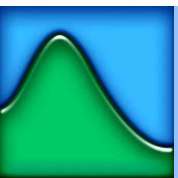
Analysis

Residual models for multi-experiment analysis

← different residual models, defined by VRESIDUAL

Experiment factor: loc

Experiment	Term	Factor	Model	Order	Nrows
Aberdeen	Residual	Whole term	Identity	0	144
Altona	Residual	Whole term	Identity	0	1152
Banff	Residual	Whole term	Identity	0	648
Cape Breton	prow.pcol	prow	Auto-regressive	1	36
		pcol	Auto-regressive	1	26
Drumheller	Residual	Whole term	Identity	0	864
Edmonton	Residual	Whole term	Identity	0	864
Elk Point	prow.pcol	prow	Identity	0	36
		pcol	Auto-regressive	1	10
Guelph	Residual	Whole term	Identity	0	1296
Humboldt	prow.pcol	prow	Auto-regressive	1	36
		pcol	Auto-regressive	1	18
LLoydminster	Residual	Whole term	Identity	0	576
Kentville	Residual	Whole term	Identity	0	648
Lethbridge	Residual	Whole term	Identity	0	1008
North Battleford	prow.pcol	prow	Auto-regressive	1	36
		pcol	Auto-regressive	1	18
Regina	Residual	Whole term	Identity	0	432
St. Paul	Residual	Whole term	Identity	0	1440
Thunder Bay	Residual	Whole term	Identity	0	1008
Waterloo	Residual	Whole term	Identity	0	360

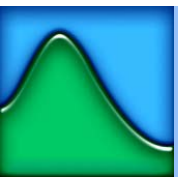


Analysis

Estimated variance components

Random term NAME	component	s.e.
loc	768626.	272579.
prow@Altona	5859.	1503.
pcol@Altona	8378.	2968.
prow@Banff	7890.	7437.
pcol@Banff	30066.	18533.
prow@Drumheller	13722.	3280.
prow@Edmonton	89205.	20749.
pcol@Edmonton	152105.	54362.
prow@Guelph	11998.	2829.
pcol@Guelph	36754.	10149.
prow@Kentville	160426.	28481.
prow@Lethbridge	28428.	5773.
pcol@Lethbridge	27753.	9325.
prow@Regina	8707.	3938.
prow@St. Paul	48988.	9656.
pcol@St. Paul	18078.	5844.
prow@Thunder Bay	59302.	12367.
pcol@Thunder Bay	31443.	11983.

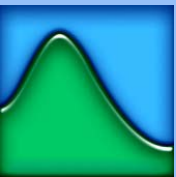
← different random terms,
defined by VRMETA



Regression and ANOVA

- FITINDIVIDUALLY can now estimate lack of fit of a model
 - MODEL & RDISPLAY now allow d.f. to be specified for fixed dispersions
 - FACCOMBINATIONS forms a factor to indicate observations with identical combinations of values of a set of variates, texts or factors
 - available from *Generalized Linear Models* menu
- ADSPREADSHEET puts data and plan of a design into spreadsheets
- AOVANYHOW performs analysis of variance using either ANOVA, regression or REML as appropriate
- AOVDISPLAY provides further output from an AOVANYHOW analysis
- AEFICIENCY calculates efficiency factors for experimental designs

..



Regression lack of fit

The screenshot shows the GenStat software interface. The main window is titled 'Generalized Linear Models' and contains the following settings:

- Available Data: Age, Pressure
- Analysis: General Model
- Response Variate: Pressure
- Maximal Model: (empty)
- Model to be Fitted: Age
- Distribution: Normal
- Link Function: Identity

The 'Generalized Linear Model Options' dialog is open, showing the following options:

- Display: Model, Summary, E-probability, Correlations, Fitted values
- Estimates: Estimates, E-probability, Confidence intervals, Accumulated, Wald tests
- Dispersion Parameter: Fix, Estimate (Value: 0)
- Confidence limit for estimates (%): 95
- Other options: Estimate constant term, Fit model terms individually, Estimate lack of fit

Below the dialog, the 'Accumulated analysis of variance' table is shown:

Change	d.f.	s.s.
+ Age	1	2647.69
Lack of fit	24	338.98
Residual	12	222.60
Total	37	3209.28

A message below the table reads: "Message: ratios are based on dispersion parameter with va...".

- implemented in FITINDIVIDUALLY
 - so currently available only from the Generalized Linear Models menu

Analysis

```
22 "General Model."  
23 MODEL [DISTRIBUTION=normal; LINK=identity; DISPERSION=*] Pressure  
24 FITINDIVIDUALLY [PRINT=model,summary,accumulated; CONSTANT=estimate; FPROB=yes; TPROB=yes;\  
25 FACT=9; LACKOFFIT=estimate] Age
```

Regression analysis

Response variate: Pressure
Fitted terms: Constant + Age

← option set to estimate lack of fit

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	1	2647.7	2647.69	142.73	<.001
Residual	36	561.6	15.60		
Total	37	3209.3	86.74		
Change	24	339.0	14.12	0.76	0.726

Percentage variance accounted for 82.0
Standard error of observations is fixed at 4.31.

Message: deviance ratios are based on dispersion parameter with value 18.6, and F probabilities on its specified d.f. of 12.

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Age	1	2647.69	2647.69	142.73	<.001
Lack of fit	24	338.98	14.12	0.76	0.726
Residual	12	222.60	18.55		
Total	37	3209.28	86.74		

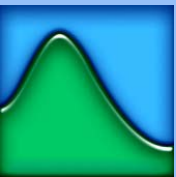
← dispersion parameter & d.f. set to values from lack of fit

Message: ratios are based on dispersion parameter with value 18.6, and F probabilities on its specified d.f. of 12.

AOVANYHOW – automatic aov

- is design balanced (EXIT parameter of ANOVA)?
 - yes – *use ANOVA* e.g. *Oats.gsh Canola.gsh*
 - no – is there a single error term?
 - yes – are there less than 2 treatment factors?
 - yes – *use A2WAY* e.g. *Foster.gsh*
 - no – *use AUNBALANCED*
 - no – useful information in higher error terms (AEFFICIENCY)?
 - no – are here less than 2 treatment factors?
 - yes – *use A2WAY*
 - no – *use AUNBALANCED* e.g. *Product.gsh*
 - no – *use REML* e.g. *Wheat72.gsh*
- output from AOVANYHOW (or AOVDISPLAY) only covers aspects available from all methods
- others obtainable from appropriate display command with SAVE[1]

..



Wish list / on-going

- 15th or 16th Edition..?
 - new client interface
 - sparse data structures
 - large data sets
 - further statistical genetics
 - additional maths and stats functions
 - wavelets
 - Bayesian methods
 - additional links to NAG Library
 - further data mining using NAG Data Mining Components
- + please send your requests/comments to *roger@vsni.co.uk*

..

