CHAPTER 2 4TH GENERATION LANGUAGES FOR STATISTICS

2.1 GENERAL CONCEPTS

4th generation language

- specialised for specific tasks
- as close from natural language as possible
- described in broad terms or even using graphical objects (« object oriented »)
- Examples: programming queries to data bases (e.g. Microsoft Visual Basic, Borland Delphi).

Examples of 4th generation languages in statistics:

- S (then S-Plus and R),
- Gauss (more oriented towards econometrics),
- Matlab (more oriented towards engineering);
- SAS is also considered as a 4th generation language but with stratistical production in mind.

Object oriented programming:

Each objet has a declarative part and processing methods
Examples:

rectangle = object which can be defined by its length and width;
a processing method consists in computing the area;
several objects of the rectangle type can be defined;
other objects such as the square as special case of the rectangle;
the method for computing the area is automatically inherited by the square
but other methods can be defined.

•

2.2 EXAMPLE: MATLAB
Comparison between algorithms-Fortran-MATLAB

	Algorithm	**Fortran**	**MATLAB**
numeric variable declaration	Variable *var*: numeric	INTEGER :: *var* REAL :: *var* REAL(8) :: *var*	N/A
logical variable declaration	Variable *var*: logical	LOGICAL :: *var*	N/A
character variable declaration	Variable *var*: character	CHARACTER(LEN=*n*) :: *var*	*N/A*
relative accuracy	*N/A*	*N/A*	eps
π	*N/A*	COS(-1.0D0)	pi
imaginary unit	*N/A*	COMPLEX(0.0D0,1.0D0)	i *or* j
infinite	*N/A*	*N/A*	inf
character string	'string'	'string'	'string'
matrix	*N/A*	*N/A*	1:n *or* 1:2:n [1 3 2 ; 6 4 5]
scalar operators	+ - * /	+ - * /	+ - .* ./ (element-wise)
matrix operators	*N/A*	*N/A*	* / right division \...left division
exponentiation	^	**	.^

	Algorithm	**Fortran**	**MATLAB**
Kronecker product	*N/A*	*N/A*	kron
transposition	*N/A*	*N/A*	'
transposition-conjugated	*N/A*	*N/A*	.'
equality	=	==	==
<	<	<	<
>	>	>	>
≤	≤	<=	<=
≥	≥	>=	>=
≠	≠	/=	~=
logical or	ou	.OR.	|
logical and	et	.AND.	&
logical not	non	.NOT.	~
Neperian logarithm	*N/A*	LOG(...)	log
decimal logarithm	*N/A*	LOG10(...)	log10
square root	√	SQRT(...)	sqrt
integer part	⌊ ⌋	INT(...)	floor(...)
integer rounding	*N/A*	NINT(...)	round(...)
modulo	*N/A*	MOD(...,...)	rem(...,...)
random number	*N/A*	*N/A*	rand(*n*)

	Algorithm	**Fortran**	**MATLAB**
assignation	*var* ← *expr*	*var* = *expr*	*var* = *expr* ;
keying in a variable	Input *var*	READ(*,*) *var*	*var* = input('var?') *text* = input('text?','s')
displaying an expression	Display *expr*	WRITE(*,*) *expr*	*expr*
open a file (for input)	*N/A*	OPEN(*n*,FILE='*name*')	*fid* = fopen('name','r')
open a file (for output)	*N/A*	OPEN(*n*,FILE='name')	*fid* = fopen('name','w')
reading a variable (binary file)	*N/A*	READ(*n*) *var*	*var* = fread(*fid*)
writing an expression (binary file)	*N/A*	WRITE(*n*) *expr*	*count* = fwrite(fid,expr, real*8)
reading a variable (text file)	Read *var*	READ(*n*,*) *var*	*nat* = fscanf(fid,'%d') *var* = fscanf(fid,'%f') text = fscanf(fid,'%s')
writing an expression (text file)	Write *expr*	WRITE(*n*,*) *expr*	fprintf(fid,'Texe',text) fprintf(fid,'%6.2f %7.3', tabl)
comment		C ... *ou* *	% ...

	Algorithm	**Fortran**	**MATLAB**
sequence structure	*instr1* *instr2*	*instr1* *instr2*	*instr1*, *instr2* *or* *instr1* *instr2*

conditional	If *condition*
	then *block*
	endif
		IF(*condition*) THEN
		block
		ENDIF
			if *condition*
			block
			end

conditional	If *condition*
	then *block*
	else *block*
	endif
		IF(*condition*) THEN
		block
		ELSE
		block
		END IF
			if *condition*
			block
			else
			block
			end

loop with counter	For *var* = 1 to *n* do *block* endfor	DO *var* = 1, *n* *block* END DO	for *i* = 1 : *n* *block* end

iterative	Iterate
	block
	exitif *condition*
	block
	enditerate
		DO *var* = 1, 9999
		block
		IF(*condition*) EXIT
		block
		END DO
			while 1
			block
			if *condition*, break, end
			block
			end

procedure call	Call *proc*(...)	CALL *proc*(...)	*proc*(...)
end of a procedure	endalgorithm	RETURN	return

Example

```
% VARCAL.M
% Algorithm for computing the variance
% I N I T I A  L I S A T I O N
N = input('Input the number of data') ;
X = zeros(1, N) ;
X = input('Input the data') ;
% T R E A T M E N T
[ XBAR, VARI ] = MEANVAR(X, N) ;
% C L O S I N G
fprintf('MEAN     = %12.5g \n', XBAR )
fprintf('VARIANCE = %12.5g \n', VARI )
```

```
function [ XB, VARI ] = MEANVAR( X, N )
% MEANVAR returns mean and variance
% based on the "calculator" procedure
%
%-------------------------------------
XB = mean(X) ;
VARI = mean(X.^2) - XB^2 ;
return
```

REFERENCES

A. BIRAN, A. et M. BREINER, "MATLAB for engineers", Addison-Wesley, Wokingham, 1995.

G. H. GOLUB et C. F. VAN LOAN, «Matrix Computations» ,2nd edition, John Hopkins University Press, 1989.

The MATH WORKS Inc., "MATLAB User's Guide", August 1992 (reprints: November 1994).

The MATH WORKS Inc., "Reference Guide", October 1992 (reprints: April 1995).

N.B.

The MATH WORKS Inc., 24 Prime Park Way, Natick, Mass. 01760-1500 (Phone: (508) 653-1415, Fax: (508) 653-2997, Anonymous FTP server: ftp.mathworks.com, WWW Home Page: http://www.mathworks.com).

2. MATLAB Primer (3rd) edition
by Kermit Sigmon

MATLPRIM.PDF

(available on the Internet)

2.3 Contents of the Statistics Toolbox

(from MATLAB help files; programs are ©The Math Works Inc.)

```
Statistics Toolbox.
 Version 1.0, 15-Sept.-1993
 Copyright (c) 1993 by The MathWorks, Inc.
 $Revision: 1.6 $  $Date: 1993/09/09 21:49:18 $
```

Distributions.

```
 Probability density functions.
   betapdf     - Beta density.
   binopdf     - Binomial density.
   chi2pdf     - Chi square density.
   exppdf      - Exponential density.
   fpdf        - F density.
   gampdf      - Gamma density.
   geopdf      - Geometric density.
   hygepdf     - Hypergeometric density.
   normpdf     - Normal (Gaussian) density.
   poisspdf    - Poisson density.
   tpdf        - T density.
   unidpdf     - Discrete uniform density.
   unifpdf     - Uniform density.
   weibpdf     - Weibull density.
```

Cumulative Distribution functions.

```
   betacdf     - Beta cdf.
   binocdf     - Binomial cdf.
   chi2cdf     - Chi square cdf.
   expcdf      - Exponential cdf.
   fcdf        - F cdf.
   gamcdf      - Gamma cdf.
   geocdf      - Geometric cdf.
   hygecdf     - Hypergeometric cdf.
   normcdf     - Normal (Gaussian) cdf.
   poisscdf    - Poisson cdf.
   tcdf        - T cdf.
   unidcdf     - Discrete uniform cdf.
   unifcdf     - Uniform cdf.
   weibcdf     - Weibull cdf.
```

Inverse Cumulative Distribution functions.

```
   betainv     - Beta critical values.
   binoinv     - Binomial critical values.
   chi2inv     - Chi square critical values.
   expinv      - Exponential critical values.
   finv        - F critical values.
   gaminv      - Gamma critical values.
   geoinv      - Geometric critical values.
   hygeinv     - Hypergeometric critical values.
   norminv     - Normal (Gaussian) critical values.
   poissinv    - Poisson critical values.
```

```
tinv        - T critical values.
unidinv     - Discrete uniform critical values.
unifinv     - Uniform critical values.
weibinv     - Weibull critical values.
```

Random Number Generators.

```
betarnd     - Beta random numbers.
binornd     - Binomial random numbers.
chi2rnd     - Chi square random numbers.
exprnd      - Exponential random numbers.
frnd        - F random numbers.
gamrnd      - Gamma random numbers.
geornd      - Geometric random numbers.
hygernd     - Hypergeometric random numbers.
normrnd     - Normal (Gaussian) random numbers.
poissrnd    - Poisson random numbers.
trnd        - T random numbers.
unidrnd     - Discrete uniform random numbers.
unifrnd     - Uniform random numbers.
weibrnd     - Weibull random numbers.
```

Statistics.

```
betastat    - Beta mean and variance.
binostat    - Binomial mean and variance.
chi2stat    - Chi square mean and variance.
expstat     - Exponential mean and variance.
fstat       - F mean and variance.
gamstat     - Gamma mean and variance.
geostat     - Geometric mean and variance.
hygestat    - Hypergeometric mean and variance.
normstat    - Normal (Gaussian) mean and variance.
poisstat    - Poisson mean and variance.
tstat       - T mean and variance.
unidstat    - Discrete uniform mean and variance.
unifstat    - Uniform mean and variance.
weibstat    - Weibull mean and variance.
```

Descriptive Statistics.

```
geomean     - Geometric mean.
harmmean    - Harmonic mean.
iqr         - Interquartile range.
mad         - Median Absolute Deviation.
prctile     - Percentiles.
range       - Range.
trimmean    - Trimmed mean.
var         - Variance.
```

Linear Models.

```
regress     - Multivariate linear regression.
anova1      - One-Way Analysis of Variance.
anova2      - Two-Way Analysis of Variance.
```

Hypothesis Tests.

```
ztest       - Z test of one mean.
ttest       - T test of one mean.
ttest2      - T test of two means.
```

Statistical Plots.

```
boxplot     - One boxplot per column.
fsurfht     - Interactive contour plot of a function.
normplot    - Normal probability plot.
qqplot      - Quantile-Quantile plot.
surfht      - Interactive contour plot of a data grid.
```

Statistics Demos.

```
disttool    - GUI tool for exploring probability distribution functions.
polytool    - Interactive graph for prediction of fitted polynomials.
randtool    - GUI tool for generating random numbers.
```

2.4 Contents of the GKSLIB library by Gordon K. Smith

(full programs available from StatLib, address: http://lib.stat.cmu.edu/general/gkslib)

```
# Wrapped by Gordon Smyth <gks@umnstat> on Tue Oct  3 13:27:46 1995
#
# This archive contains:
#     readme      betacf.m    betai.m     binp.m
#     chisqaux.m  chisqp.m    chisqq.m    contents.m
#     derivs.m    diagsum.m   digamma.m   digammas.m
#     dmult.m     excise.m    expovec.m   fdistaux.m
#     fdistp.m    fdistq.m    gammad.m    gquad.m
#     gquad6.m    grule.m     indicato.m  int0infy.m
#     invits.m    laguerr2.m  laguerre.m  levels.m
#     levenb.m    lik.m       logist.m    ma.m
#     mprony.m    normp.m     normq.m     normqq.m
#     orthpoly.m  poisp.m     poisson.m   randbin.m
#     randgamm.m  randpois.m  regr.m      runmed.m
#     startup.m   tally.m     trigamma.m  trigamms.m
#

GKSLIB

This is a library of MATLAB statistical routines which I have written
or collected.  See the file contents.m for a table of contents.
Acknowledgements to

   Peter R Shaw, Woods Hole Oceanographic Institution
   Howard Wilson, University of Alabama
   Roger B Sidje, University of Queensland
   Mohamad A Akra, Massachussets Institute of Technology

To install the library you must

1.  Copy the files to a convenient directory
2.  Add the directory to the matlabpath

Under Microsoft Windows 3.1 you can achieve this effect by
1.  Coping all files to a directory, c:\gkslib say
2.  Edit the file c:\gkslib\startup.m so that it contains the line
    "matlabpath(['c:\gkslib;' matlabpath]);"
3.  Edit the properties of the MATLAB icon to make c:\gkslib the
    working directory.

Please direct any problems to me:

Gordon K Smyth
gks@maths.uq.oz.au
Department of Mathematics, Unversity of Queensland, Q 4072, Australia
http://www.maths.uq.oz.au/

2 Oct 95
% Gordon Smyth's library of MATLAB statistical routines.
%
```

% Regression

```
%   regr       - least squares regression
%   logist     - ordinal logistic regression
```



```
%   poisson    - poisson regression
%   mprony     - fit sum of exponential functions by nonlinear l.s.
%
```

% Other functions associated with regression and designed experiments

```
%   excise     - remove missing values
%   levels     - generate factor levels
%   indicato   - generate indicator variables for factor
%   tally      - univariate frequency table
%   orthpoly   - generate orthogonal polynomials
%   levenb     - Levenberg modified Newton-Raphson
%
```

% Probability distributions

```
%   betai      - incomplete Beta function
%   binp       - binomial distribution function
%   chisqp     - chi-squared distribution function
%   chisqq     - chi-squared inverse distribution function
%   fdistp     - Fisher's F distribution function
%   fdispq     - Fisher's F inverse distribution function
%   gammad     - gamma distribution density function
%   normp      - normal distribution function
%   normq      - normal inverse distribution function
%   normqq     - normal probability plot
%   poisp      - Poisson distribution function
%   randbin    - random deviates from binomial distribution
%   randgamm   - random deviates from gamma distribution
%   randpois   - random deviates from Poisson distribution
%
```

% Times series

```
%   ma         - moving average smoother
%   runmed     - running median of three
%
```

% Special functions

```
%   digamma    - digamma function
%   trigamma   - trigamma function
%
```

% Gaussian quadrature

```
%   gquad      - Legendre quadrature
%   gquad6     - six point Legendre quadrature
%   grule      - compute Legendre nodes and weights
%   int0infy   - Laguerre quadrature
%   laguerre   - compute Laguerre nodes and weights
%   laguerr2   - Laguerre nodes and weights for gamma expectation
%
```

% Other matrix functions

```
%   diagsum    - sums of diagonals of matrix
%   expovec    - matrix exponential by Krylov subspace projection
%   invits     - eigenvalue nearest zero by inverse iteration
```