```
options ls=76;

title1 'Hegde & Naik, 1999 SUGI 24 proceedings paper';
title2 'Canonical Correspondence Analysis';
title3 'Hunting Spider Abundance Data';

/* If you have a large data set, you may want to read your data sets using infile st
You may also import your data files (from excel) using File menu.

data c1;
  infile 'C:\Users\lhegde2\Desktop\Mysas\species.txt';
  input y1-y12;
run;

data c2;
  infile 'C:\Users\lhegde2\Desktop\Mysas\env.txt';
  input z1-z6;
run;

*/

/* Otherwise you may directly read your data as we have done here */

data c1;
*Species abundance (y1 to y12) at 28 sites;
input y1-y12;
cards;
0  2  1  0  0  0  5  0  0  0  0  0
0  3  1  1  0  0  4  1  0  0  0  0
0  3  1  0  0  0  4  1  0  0  0  0
0  2  2  1  0  0  5  1  0  0  0  0
0  1  1  0  0  0  4  0  0  0  0  0
0  2  0  0  0  0  5  1  0  0  0  0
0  1  3  3  6  5  8  1  1  0  0  0
0  7  1  1  1  2  5  3  1  0  0  0
0  4  1  0  1  0  4  1  1  0  0  0
1  1  4  9  8  3  9  4  1  1  0  0
2  0  5  5  4  2  7  2  3  0  0  0
1  1  5  3  8  2  9  1  3  0  0  0
1  1  5  5  9  4  9  2  2  1  0  0
3  1  4  9  9  4  9  2  5  1  0  0
1  1  4  7  8  4  9  6  4  1  1  0
1  1  1  4  6  3  8  4  5  3  1  0
```

```
0  0  2  3  6  2  7  3  7  5  0  0
0  0  0  1  1  0  1  1  5  1  0  0
0  0  0  1  2  0  3  3  9  4  0  0
0  1  2  2  0  1  4  1  3  3  3  0
0  0  0  0  1  1  2  1  9  3  1  0
0  0  0  0  0  0  1  0  4  1  1  0
0  0  0  0  0  0  1  0  2  3  3  1
0  1  0  0  0  0  1  0  2  4  3  2
0  0  0  0  0  0  1  0  1  2  4  1
0  0  0  0  0  0  0  0  1  5  3  2
0  0  0  0  0  0  0  0  1  3  4  2
0  0  0  0  0  0  1  0  0  1  2  4
run;

data c2;
*Data on Environmental Variables (z1 to z6) on 28 sites;
input z1-z6;
cards;
9  0  1  1  9  5
7  0  3  0  9  2
8  0  1  0  9  0
8  0  1  0  9  0
9  0  1  2  9  5
8  0  0  2  9  5
8  0  2  3  3  9
6  0  2  1  9  6
7  0  1  0  9  2
8  0  0  5  0  9
9  5  5  1  7  6
8  0  4  2  0  9
6  0  5  6  0  9
8  0  1  5  0  9
9  3  1  7  3  9
6  0  5  8  0  9
5  0  7  8  0  9
5  0  9  7  0  6
6  0  8  8  0  8
3  7  2  5  0  8
4  0  9  8  0  7
4  8  7  8  0  5
0  7  8  8  0  6
0  6  9  9  0  6
1  7  9  8  0  0
```

```
0  5  8  8  0  6
2  7  9  9  0  5
0  9  4  9  0  2
run;

data a;
merge c1;
merge c2;

proc print data=a noobs;
run;

* PROC IML is used to do all the calculations;
*Create Data and Other Matrices;

proc iml;
use a;
read all var{y1,y2,y3,y4,y5,y6,y7,y8,y9,y10,y11,y12} into Y;
read all var{z1,z2,z3,z4,z5,z6} into Z;
close a;


m=ncol(Y); /*Number of species */
q=ncol(Z); /*Number of environmental variables */
n=nrow(Y); /*Number of sites */
GT = Y[+,+]; /* Grand Total of all species */
F = Y/GT; /* We analyze this matrix of relative frequencies instead of original cour
*Create species and site totals;


/* This module just calculates column sums of any given matrix.
Note that the same module will compute the row sums of a matrix provided we transpos
Then we make a diagonal matrix of these sums.
These diagonal matrices are generally called masses or weights.
*/

Start Weights(mat);
size= ncol(mat);
sum=j(size,1,0);
do i=1 to size;
sum[i,1]=mat[+,i];
end;
sum =diag(sum);
```

```
Return (sum);
Finish;


*Standardize (Weighted Mean=0, SD=1) a matrix of quantitative variables;
/* Make sure there are no reduncies in these variables. In other words, there are
no environmental variables that are linear combinations of some other variables incl
*/

Start Scale(mat,w);
n =nrow(mat);
mat =mat -j(n,n,1)*w*mat;
temp1=mat`*w*mat;
temp2=diag(temp1);
temp2=sqrt(temp2);
scalem=inv(temp2);
mat=mat*scalem;
Return (mat);
Finish;

* Computing Square Root of a matrix;
/* These are mathematical requirements for doing our analysis*/

Start SqrtMat(mat);
q = nrow(mat);
call eigen(lambd,ev,mat);
lambda=diag(lambd);
lmda_hf=j(q,q,0);
do i=1 to q;
lmda_hf[i,i]=sqrt(lambda[i,i]);
end;
mat_hf=ev*lmda_hf*ev`;
return (mat_hf);
Finish;

ColMass = Weights(F); /* Species weights */
RowMass  = Weights(F`);  /*Site weights */
w1 = ColMass; /* A researcher may modify these weights */
w2 = RowMass;
Z_s = Scale(Z, w2);  /* Standardize environmental variables */
Psi = Z_s`*w2*Z_s; /* Correlation of observed env variables */
w3 = inv(Psi);
/* This matrix is called metric for environmental variables.
```

```
Lower the weights, higher the variability */


*Create fundamental matrices for the analysis;

F1 = Inv(w1)*F‘; /* Relative Frequency distribution of a given species over sites */
F2 = inv(w2)*F; /* Relative Frequency distribution of a given site over species */
w1_hf = SqrtMat(w1);
w1_nhf = inv(w1_hf);
w2_hf = SqrtMat(w2);
w2_nhf = inv(w2_hf);
w3_hf = SqrtMat(w3);
w3_nhf = inv(w3_hf);
A = F1*Z_s; /* Weighted environmental (standardized) values */
print ’Weighted environmental (standardized) values-An important matrix of CCA’;
print A;

W= w1_hf*A*w3_hf; /* W is a mathematical trick to perform CCA on A to satisfy scalin

*The SVD of the fundamental matrix W ;
call svd(P_mat,sv,Q_mat,W);

*The diagonal elements of D matrix are the singularvalues of A;
D=diag(sv);
Print D;
*The diagonal elements of Lambda matrix are the eigenvalues;
Lambda=D*D;
print ’Eigenvalues’;
print Lambda;
print ’ ’;

*Solutions to Canonical Correspondence Analysis;

umat = w1_nhf*P_mat;

/* This matrix is called left singular vectors of A. */

bmat= w3_nhf*Q_mat;
/* This matrix is called right singular vectors of A.*/

power =1;

/* This number can change from -2 to 2.
```

```
Interpretation of some results in our analysis heavily depends on this number
The researchers may vary this number and get different sets of output.
Most popular choices are 0,0.5,1.
*/


DL= diag(sv ## power);  /* This scales umat, the left singular vector */
DR= diag(sv ## (1-power)); /* This scales bmat, the right singular vector */

U_mat = umat*DL;  /* Called Species Scores. Note this can change for different power
B_mat = w3*bmat*inv(DR); /* Called Canonical Coefficients or Regression Coefficients

X_mat=Z_s*B_mat; /* Called Linear Combination of environmental variables */

print 'Solutions to Canonical Correspondence Analysis';

print 'Solutions to Canonical Coefficients';
print B_mat;

print 'Solutions to Species Scores';
print U_mat;

Print 'Sample Scores: Linear combinations of environmental variables' ;
Print X_mat;

X_mat = Scale(X_mat,w2);
COEVO=Z_s`*w2*X_mat;
print COEVO;


*Sample Scores;
X_s =F2*U_mat; /* These are not generally standardized scores */
print 'Sample Scores';
print X_s;



*Standardize (Weighted Mean=0, SD=1) the sample scores;

X_s = Scale(X_s,w2);

*Correlation of an Environmental variable with an
ordination axis;
```

```
EOCORR=Z_s‘*w2*X_s;
print ’Correlation of an Environmental variable with
an ordination axis’;
print ’OR Inter set Correlations’;
print EOCORR;

*Species-Environment Correlations;
SECORR=X_mat‘*w2*X_s;
SECORR=diag(secorr);
print ’Species-Environment Correlations’;
print SECORR;

reset fw=8 noname;
percent = 100*sv##2 /sv[##];
*Cumulate by multiplying by lower triangular matrix of 1’s;
j = nrow(sv);
tri = (1:j)‘ * repeat(1,1,j) >= repeat(1,j,1)*(1:j);
cum = tri*percent;
Print "Singular values and variance accounted for",,
    sv [colname={’Singular Values’} format=9.4]
    percent [colname={’Percent’} format=8.2]
    cum [colname={’cum % ’} format = 8.2];


*Biplots;
print ’ ’;
print ’Biplot Information’;

*USER NEEDS TO PROVIDE NAMES FOR OBSERVATIONS IN id AND VARIABLES IN vars BELOW;

id={’Arct_lute’, ’Pard_lugu’, ’Zora_spin’, ’Pard_nigr’,
    ’Pard_pull’, ’Aulo_albi’, ’Troc_terr’, ’Alop_cune’,
    ’Pard_mont’, ’Alop_acce’, ’Alop_fabr’,  ’Arct_peri’};

vars={"WATER_CONTENT" "BARE_SAND" "COVER_MOSS"
"LIGHT_REFL" "FALLEN_TWIGS" "COVER_HERBS"};

reset fw=8 noname;

dim=2;
power= 1;
U=umat;
```

```
V=bmat;
factor = sqrt(GT);
U=U[,1:dim];
V=V[,1:dim];
Lambda=sv[1:dim];

DL= diag(Lambda ## power);
DR= diag(Lambda ## (1-power));
A = (1/factor)# U * DL;
 B = factor# V * DR;

OUT=A // B;
Print OUT;
*Create observation labels;
id = id // vars‘;

type = repeat({"OBS "},m,1) // repeat({"VAR "},q,1);
       id  = concat(type,id);
   cvar = concat(shape({"DIM"},1,dim),char(1:dim,1.));

* Create sas data set BIPLOT;
create plot from out[rowname=id colname=cvar];
append from out[rowname=id];
close plot;
quit;

*Split id into _type_ and _Name_;
    data plot;
    set plot;
    drop id;
    length _type_  $3 _name_ $16;
    _type_ = scan(id,1);
    _name_ = scan(id,2);
    run;


DATA PLOT1;SET PLOT; IF _TYPE_='OBS'; RUN;
DATA PLOT2;SET PLOT; IF _TYPE_='VAR';RUN;

PROC PRINT DATA=PLOT1;
RUN;

DATA OBSPLOT;
```

```
 SET PLOT1;
    length text $16;
    xsys='2'; ysys='2';
    text=_name_;
    if _type_='OBS' then do;
    x = dim1;
    y = dim2;
    position='5';
    function='LABEL';
    output;
    end;
RUN;

PROC PRINT DATA=PLOT2;RUN;

DATA VARPLOT;
SET PLOT2;
length text $16;
    xsys='2'; ysys='2';
    text=_name_;
* Draw line from the origin to the variable point;
    if _type_ ='VAR' then do;
    x=0; y=0;
    function ='MOVE ';
    output;
    x=dim1;
    y=dim2;
    function ='DRAW ';
    output;
    if dim1>=0 then position ='6';  /*left justify*/
    else position ='2';             /*right justify*/
    function='LABEL';               /*variable name*/
    output;
    end;
    run;

*PLOT ONLY OBSERVATIONS;
   proc gplot data=plot1;
   plot dim2*dim1/anno=OBSPLOT frame href=0 vref=0 ;
   title1 h=1.2 'Biplot of Hunting Spider Data ';
   title2 f=duplex 'Observations are points';
   run;
```

```
*PLOT ONLY VARIABLES;
proc gplot data=plot2;
plot dim2*dim1/anno=VARPLOT frame href=0 vref=0;
    title1 h=1.2 'Biplot of Hunting Spider Data ';
    title2 f=duplex 'Variables are vectors';
    run;

*PLOT BOTH OBSERVATIONS AND VARIABLES ON THE SAME PLOT;
*PRINT THE DATA SET TO DETERMINE A VALUE FOR SCALE;
proc print DATA=PLOT;
run;

*IN ORDER TO DRAW BOTH PLOTS ON THE SAME GRAPH, ONE NEED TO SELECT A VALUE FOR SCALE
*FOR THE DATA IN HAND, SCALE=0.002 WORKED FINE;

data plot; set plot;
SCALE1=1 ; *USE USER SELECTED VALUE INSTEAD OF 1 HERE;
SCALE2=0.002; *USE USER SELECTED VALUE INSTEAD OF 0.002 HERE;

if _type_='OBS' then do;
dim1=dim1*scale1; dim2=dim2*scale1; output; end;
if _type_='VAR' then do;
dim1=dim1*scale2; dim2=dim2*scale2; output; end;
run;

proc print DATA=PLOT;
run;

*Annotate observation labels and variable vectors;
    data label;
    set plot;
    length text $16;
    xsys='2'; ysys='2';
    text=_name_;
    if _type_='OBS' then do;
    x = dim1;
    y = dim2;
    position='5';
    function='LABEL';
    output;
    end;
* Draw line from the origin to the variable point;
```

```
    if _type_ ='VAR' then do;
    x=0; y=0;
    function ='MOVE';
    output;
    x=dim1;
    y=dim2;
    function ='DRAW';
    output;
    if dim1>=0 then position ='6';  /*left justify*/
    else position ='2';             /*right justify*/
    function='LABEL';               /*variable name*/
    output;
    end;
    run;

* Plot the biplot using proc gplot;

    proc gplot data=plot;
    plot dim2*dim1/anno=label frame href=0 vref=0 ;
   title1 h=1.2 'Biplot of Hunting Spider Data ';
    title2 f=duplex 'Observations are points,
                          Variables are vectors';
    run;

quit;
```