```
options ls=76;

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

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=12; /*Number of species */
q=6; /*Number of environmental variables */
n=28; /*Number of sites */

one=j(n,1,1);

sum1=j(m,1,0);
do i=1 to m;
sum1[i,1]=y[+,i];
end;
S11=diag(sum1);

r1=j(n,1,0);
do i=1 to n;
r1[i,1]=y[i,+];
end;
R=diag(r1);
R_star=R/y[+,+];

*Standardize (Weighted Mean=0, SD=1) environmental
variables;
Z=Z-j(n,n,1)*R_star*Z;
temp1=Z‘*R_star*Z;
temp2=diag(temp1);
temp2=sqrt(temp2);
scalem=inv(temp2);
Z=Z*scalem;

*Create W, the fundamental matrix for the analysis;
S12=Y‘*Z;
S22=Z‘*R*Z;
* Find S11^(1/2) and S11^(-1/2);
s11_hf=j(m,m,0);
s11_nhf=j(m,m,0);
do i=1 to m;
s11_hf[i,i]=sqrt(S11[i,i]);
s11_nhf[i,i]=1/sqrt(S11[i,i]);
end;
```

```
* Find S22^(1/2) and S22^(-1/2);
call eigen(lambd,ev,S22);
lambda=diag(lambd);
lmda_hf=j(q,q,0);
lmda_nhf=j(q,q,0);
do i=1 to q;
lmda_hf[i,i]=sqrt(lambda[i,i]);
lmda_nhf[i,i]=1/sqrt(lambda[i,i]);
end;
s22_hf=ev*lmda_hf*ev‘;
s22_nhf=ev*lmda_nhf*ev‘;
W=s11_nhf*S12*s22_nhf;
print ’The Fundamental Matrix W’;
print W;

*The SVD of the fundamental matrix W ;
call svd(P_mat,Lambda,Q_mat,W);
D=diag(Lambda);
*The diagonal elements of D matrix are the eigenvalues;
D=D*D;
print ’Eigenvalues’;
print D;
print ’Eigenvectors’;
print P_mat Q_mat;
print ’ ’;

*Solutions to Canonical Correspondence Analysis;
u_mat=s11_nhf*P_mat;
beta_mat=s22_nhf*Q_mat;
print ’Solutions to Canonical Correspondence Analysis’;
print u_mat beta_mat;

*Sample Scores: Linear combinations of environmental variables;
X=Z*beta_mat;

*Species scores;
U_hat=inv(s11)*Y‘*X*diag(1/lambda); * Assuming alpha=0;
print ’Species scores’;
print U_hat;

*Standardize (Weighted Mean=0, SD=1) the X matrix;
X=X-j(n,n,1)*R_star*X;
temp1=X‘*R_star*X;
```

```
temp2=diag(temp1);
temp2=sqrt(temp2);
scalem=inv(temp2);
X=X*scalem;
print 'Sample Scores: Linear combinations of environmental
variables';
print X;

*Canonical Coefficients corresponding to eigenvectors;
B_hat=inv(Z`*Z)*Z`*X;
print 'Canonical Coefficients corresponding to eigenvectors';
print B_hat;

* Biplot scores of environmental variables;
*COEVO=Z`*R_star*X;
*print COEVO;

*Species scores;
U_hat=inv(s11)*Y`*X; * Assuming alpha=0;
print 'Species scores';
print U_hat;

*Sample Scores;
X_star=inv(R)*Y*U_hat*inv(D);
print 'Sample Scores';
print X_star;

*Standardize (Weighted Mean=0, SD=1) the X_star matrix;
X_star=X_star-j(n,n,1)*R_star*X_star;
temp1=X_star`*R_star*X_star;
temp2=diag(temp1);
temp2=sqrt(temp2);
scalem=inv(temp2);
X_star=X_star*scalem;

*Correlation of an Environmental variable with an
ordination axis;
EOCORR=Z`*R_star*X_star;
print 'Correlation of an Environmental variable with
an ordination axis';
print 'OR Inter set Correlations';
print EOCORR;
```

```
*Species-Environment Correlations;
SECORR=X‘*R_star*X_star;
SECORR=diag(secorr);
print ’Species-Environment Correlations’;
print SECORR;

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

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

dim=2;
power=0;
scale=0.01;
U=s11_nhf*P_mat;
V=s22_hf*Q_mat;
U=U[,1:dim];
V=V[,1:dim];
Lambda=Lambda[1:dim];
DL= diag(Lambda ## power);
DR= diag(Lambda ## (1-power));
A = U * DL;
B = V * DR # scale;

OUT=A // B;
*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;

proc print;
run;

*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;
*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;
    filename gsasfile "prog710.graph";
    goptions reset=all gaccess=gsasfile autofeed dev=pslmono;
    goptions horigin=1in vorigin=2in;
    goptions hsize=6in vsize=8in;

    proc gplot data=plot;
    plot dim2*dim1/anno=label frame href=0 vref=0
    lvref=3 lhref=3 vaxis=axis2 haxis=axis1 vminor=1 hminor=1;
    axis1 length=5 in order=(-.20 to .20 by .05)  offset=(2)
        label = (h=1.2 'Dimension 1');
    axis2 length=5 in order =(-.15 to .15 by .05) offset=(2)
label=(h=1.2 a=90 r=0  'Dimension 2');
    symbol v=none;
    title1 h=1.2 'Biplot of Hunting Spider Data ';
    title2 j=l 'Output 7.10';
    title3 f=duplex 'Observations are points,
                          Variables are vectors';
    run;
```