Article

Calculation and statistic test of partial correlation of general correlation measures

WenJun Zhang
School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; International Academy of Ecology and Environmental Sciences, Hong Kong
E-mail: zhwj@mail.sysu.edu.cn, wjzhang@iaees.org

Received 13 July 2015; *Accepted 10 August 2015*; *Published online 1 December 2015*

Abstract
It is well known that Pearson linear correlations between more than two attributes (nodes, taxa, variables, etc) can be adjusted to partial linear correlations for eliminating indirect between-attribute interactions of other attributes not being tested. In present study I first proposed three correlation measures, revised Dice coefficient, overlap coefficient, and proportion correlation. In addition, I proposed partial correlation measures for some correlation measures, of which Jaccard correlation, revised Dice coefficient, overlap coefficient, and point correlation are for binary attributes; Spearman rank correlation and proportion correlation are for interval value attributes. The full algorithm and Matlab codes (Pearson linear correlation is included also) are given. Users can add other general correlation measures in the Matlab codes.

Keywords partial correlation; correlation measures; proportion correlation; statistic test; data analysis.

Selforganizology
ISSN 2410-0080
URL: http://www.iaees.org/publications/journals/selforganizology/online-version.asp
RSS: http://www.iaees.org/publications/journals/selforganizology/rss.xml
E-mail: selforganizology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction
To eliminate indirect between-attribute interactions of other attributes not being tested, partial linear correlation of Pearson linear correlation measure has been widely used. However for most other general correlation measures, partial correlation measures are not available. In present study, I proposed some correlation measures and partial correlation measures for some correlation measures. The full algorithm and Matlab codes were given. Users can add other general correlation measures in the Matlab codes.

2 Methods
2.1 Methods
2.1.1 Correlation measures
There are a variety of correlation measures (Zhang, 2007, 2011, 2012a, 2012b; Zhang et al., 2014). Here I propose three correlation measures, revised Dice coefficient, overlap coefficient, and proportion correlation, and use more correlation measures to derive their partial correlations, in exception of Pearson linear correlation

which has been affiliated with partial linear correlation (Zhang, 2011, 2012a; Zhang and Li, 2015), among which Jaccard correlation, revised Dice coefficient, overlap coefficient, and point correlation are for binary attributes (nodes, taxa, variables, etc), and Spearman rank correlation and proportion correlation are for interval value attributes (nodes, taxa, variables, etc).

First, assume there are *m* attributes (nodes, taxa, variables, etc) and *n* samples (components, members, etc), and the raw data is a matrix (x_{ij}), i=1, 2, …,m; j=1, 2, …, n.

Jaccard correlation is:

$$r_{ij}=(e-(c+b))/(e+c+b) \qquad i, j=1, 2,\cdots,m$$

where $-1\le r_{ij}\le 1$, c is the number of sample pairs of 1 for attribute i but not for j; b is the number of sample pairs of 1 for attribute j but not for i; e is the number of sample pairs of 1 for both attribute i and attribute j.

Revised Dice coefficient is:

$$r_{ij}=4*e/(b+c)-1 \qquad i, j=1, 2,\cdots,m$$

where $-1\le r_{ij}\le 1$, b is the number of sample of 1 for attribute i; c is the number of sample of 1 for attribute j; e is the number of sample pairs of 1 for both attribute i and attribute j.

Point correlation is (Zhang, 2007)

$$r_{ij}=(ad-bc)/((a+b)(c+d)(a+c)(b+d))^{1/2} \qquad i, j=1, 2,\cdots,m$$

where $-1\le r_{ij}\le 1$, both attribute i and attribute j take values 0 or 1. a is number of both attribute i and attribute j take value 0, b is number of attribute i takes 0 and attribute j takes 1, c is number of attribute i takes 1 and attribute j takes 0, and d is number of both attribute i and attribute j take value 1.

Overlap coefficient is:

$$r_{ij}=2e/(e+b)-1 \qquad i, j=1, 2,\cdots,m$$

where $-1\le r_{ij}\le 1$, b is the number of sample pairs with different values; is the number of sample pairs of 1 for both attribute i and attribute j.

Proportion correlation is:

$$r_{ij}= 2\sum_{k=1}^{n} (p_{ik}\, p_{jk})/(\sum_{k=1}^{n} p_{ik}^{2} \sum_{k=1}^{n} p_{jk}^{2})^{1/2}-1 \qquad i, j=1, 2,\cdots,m$$

where $-1\le r_{ij}\le 1$, and $p_{il}=x_{il,}/\sum_{k=1}^{n} x_{ik,}$, $p_{jl}=x_{jl,}/\sum_{k=1}^{n} x_{jk,}$, l=1, 2, …, n. It is used specifically to the situations of $r_{ij}\ge 0$, i=1, 2, …,m; j=1, 2, …, n.

Spearman rank correlation is (Spearman, 1904; Schoenly and Zhang, 1999):

$$r_{ij}=1-6*\sum d^2/[n(n^2-1)] \qquad i, j=1, 2,\cdots,m$$

where $-1\le r_{ij}\le 1$, $d=r(i)-r(j)$, and $r(i)$ and $r(j)$ are rank of an element in the attribute i and attribute j, from the smaller to the larger values in n samples.

For the correlation measures, calculate $t=|r_{ij}|/[(1-r_{ij}^2)/(n-2)]^{1/2}$, and if $t>t_\alpha(n-2)$, the correlation (either positive or negative correlation) between attributes *i* and *j* is statistically significant.

2.1.2 Partial correlations

The partial correlation between two attributes eliminates the indirect effects produced by the remaining attributes. Let $R=(r_{ij})$, where r_{ij} are correlations calculated above. Calculate

$$\sum_{k=1}^{m} r_{ik}\, r_{kj}^{-}=0 \qquad i, j=1, 2,\cdots,m;\ i\neq j$$
$$\sum_{k=1}^{m} r_{ik}\, r_{ki}^{-}=1 \qquad i=1, 2,\cdots,m$$

That is, calculate $R^{-}=(r_{ij}^{-})$ such that $RR^{-}=I$ and I is the unit matrix. Partial correlation between attributes *i* and *j* is

$$parr_{ij}=-r_{ij}^{-}/(r_{ii}^{-}*r_{jj}^{-})^{1/2} \qquad i, j=1, 2,\cdots,m;\ i\neq j$$

where $-1\leq parr_{ij}\leq 1$. It should be noted that a singular or ill-condition matrix, *R*, will result in wrong results.

For the partial correlation measures, calculate $t=|parr_{ij}|/[(1-parr_{ij}^{2})/(n-m)]^{1/2}$, and if $t>t_{\alpha}(n-m)$, the partial correlation (either positive or negative correlation) between attributes *i* and *j* is statistically significant, and there is a direct interaction/interdependence between attributes *i* and *j* in terms of the correlation measure being used.

2.2 Matlab codes

The Matlab codes of the full algorithm above and of finding interactions are listed as follows, of which the Pearson linear correlation measure (for interval value attributes) is included also:

```
%Reference: Zhang WJ. 2015. Calculation and statistic test of partial correlation of general correlation measures.
% Selforganizology, 2(4): 65-77
str=input('Input the file name of raw data matrix (e.g., raw.txt, raw.xls, etc. The file has m rows (taxa) and n columns (samples)): ','s');
X=load(str);
measure=input('Input a number for correlation measure (1: Pearson linear correlation; 2: Spearman rank correlation; 3: Proportion correlation; 4: Point correlation; 5: Jaccard correlation; 6: Revised Dice coefficient; 7: Overlap coefficient)');
sig=input('Input significance level(e.g., 0.01)');
%X is m*n raw data matrix. m: number of attributes (taxa, etc); n: number of samples.
dim=size(X);
m=dim(1); n=dim(2);
r=zeros(m);
switch measure
    case 1
        for i=1:m-1; for j=i+1:m; r(i,j)=linearcorre(X(i,:)',X(j,:)');end;end;
    case 2
        for i=1:m-1; for j=i+1:m; r(i,j)=spearman(X(i,:)',X(j,:)');end;end;
    case 3
        for i=1:m-1; for j=i+1:m; r(i,j)=propcorre(X(i,:)',X(j,:)');end;end;
    case 4
        for i=1:m-1; for j=i+1:m; r(i,j)=pointcorre(X(i,:)',X(j,:)');end;end;
    case 5
        for i=1:m-1; for j=i+1:m; r(i,j)=jaccard(X(i,:)',X(j,:)');end;end;
```

```
    case 6
        for i=1:m-1; for j=i+1:m; r(i,j)=reviseddice(X(i,:)',X(j,:)');end;end;
    case 7
        for i=1:m-1; for j=i+1:m; r(i,j)=overlap(X(i,:)',X(j,:)');end;end;
end
for i=1:m-1; for j=i+1:m; r(j,i)=r(i,j);end;end;
for i=1:m; r(i,i)=1;end;
disp('Correlation matrix')
r
tvalues=abs(r)./sqrt((1-r.^2)/(n-2));
alpha=(1-tcdf(tvalues,n-2))*2;
sigmat=alpha<sig;
sigmat=sigmat.*r-eye(m);
sigmatr= sigmat;
disp('Pairs with statistically significant correlation')
if (sigmat~=ones(m))
[pairx,pairy,rvalues]=find(sigmat);
temp1=pairx; temp2=pairy;
pairxs=pairx(temp1<temp2);
pairys=pairy(temp1<temp2);
rvaluess=rvalues(temp1<temp2);
PairsAndCorrelations=[pairxs pairys rvaluess]
else
disp('No significant pairs')
end
inversr=inv(r);
for i=1:m-1; for j=i+1:m; parr(i,j)=-inversr(i,j)/sqrt(inversr(i,i)*inversr(j,j));end;end;
for i=1:m-1; for j=i+1:m; parr(j,i)=parr(i,j);end;end;
for i=1:m; parr(i,i)=1;end;
disp('Partial correlation matrix')
parr
if (n>m)
tvalues=abs(parr)./sqrt((1-parr.^2)/(n-m));
alpha=(1-tcdf(tvalues,n-m))*2;
else
disp('The number of samples is not enough to support the required statistic test (DF=n-m) of partial correlations. Here use the statistic test with DF=n-2 (not recommended). Please input the proportion of statistically significant pairs based on DF=n-m vs. statistically significant pairs based on DF=n-2 (y, %) as the following. For Pearson linear correlation measure, the estimation formula, y=88.748 exp(-0.045m), is suggested for use, and for Spearman rank correlation measure, y=120.687exp(-0.045m), is suggested, where m is the number of attributes (taxa, etc). If it is hard to be estimated, the full percent, 100, can be input. ')
y=input('Input the proportion (a value between 0 and 100)')
tvalues=abs(parr)./sqrt((1-parr.^2)/(n-2));
alpha=(1-tcdf(tvalues,n-2))*2;
end
```

```
sigmat=alpha<sig;
sigmat=sigmat.*parr-eye(m);
if (n<=m) threshr=rrank(sigmat,y); sigmat=sigmat>=threshr; sigmat=sigmat.*parr; end;
sigmatparr= sigmat;
disp('Pairs with statistically significant partial correlation')
if (sigmat~=ones(m))
[pairx,pairy,rvalues]=find(sigmat);
temp1=pairx; temp2=pairy;
pairxs=pairx(temp1<temp2);
pairys=pairy(temp1<temp2);
rvaluess=rvalues(temp1<temp2);
PairsAndPartialCorrelations=[pairxs pairys rvaluess]
else
disp('No significant pairs')
end
x=sigmatparr & (~sigmatr);
y=(~sigmatparr) & sigmatr;
z=sigmatparr & sigmatr;
for i=1:3;
switch i
    case 1
        mat=x; s='Significant partial correlation but insignificant linear correlation';
    case 2
        mat=y; s='Significant linear correlation but insignificant partial correlation';
    case 3
        mat=z; s='Significant both partial correlation and correlation';
end;
[pairx,pairy]=find(mat);
temp1=pairx; temp2=pairy;
pairxs=pairx(temp1<temp2);
pairys=pairy(temp1<temp2);
disp([s])
SignificantPairs=[pairxs pairys]
end;
```

Eight M function files, linearcorre.m, spearman.m, propcorre.m, pointcorre.m, jaccard.m, reviseddice.m, and overlap.m are as follows:

```
function linearcor = linearcorre(x,y)       %x and y: two column vectors to be tested.
m=max(size(x));
if (m~=max(size(y)))
  error('Array sizes do not match.');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
  error('Both x and y are vectors');
```

```
end
xbar=mean(x);
ybar=mean(y);
aa=sum((x-xbar).*(y-ybar));
bb=sum((x-xbar).^2);
cc=sum((y-ybar).^2);
linearcor=aa/sqrt(bb*cc);


function spearm =spearman(x,y)          %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
  error('Array sizes do not match.');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
  error('Both x and y are vectors');
end
n=max(size(x));
for i=1:n
rx(i)=0;ry(i)=0;xx(i)=0;yy(i)=0;
end
for j=1:n
nx=1;ny=1;
for i=1:n
if (x(i)<x(j)) nx=nx+1; end
if (y(i)<y(j)) ny=ny+1; end
end
rx(j)=nx;
ry(j)=ny;
end
for j=1:n
if (rx(j)==(n+1)) continue; end
nx=rx(j);
ntie=-1;
for i=1:n
if (rx(i)~=nx) continue; end
ntie=ntie+1;
xx(i)=rx(i);
rx(i)=0;
end
for i=1:n
if (rx(i)~=0) continue; end
xx(i)=xx(i)+(ntie*0.5);
rx(i)=n+1;
end
end
```

```
for j=1:n
if (ry(j)==(n+1)) continue; end
ny=ry(j);
ntie=-1;
for i=1:n
if (ry(i)~=ny) continue; end
ntie=ntie+1;
yy(i)=ry(i);
ry(i)=0;
end
for i=1:n
if (ry(i)~=0) continue; end
yy(i)=yy(i)+ntie*0.5;
ry(i)=n+1;
end
end
rs=0;
rs=sum((xx-yy).^2);
spearm=1-((6*rs)/(n*(n^2-1)));


function prop = propcorre(x,y)        %x and y: two column vectors to be tested.
m=max(size(x));
if (m~=max(size(y)))
  error('Array sizes do not match.');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
  error('Both x and y are vectors');
end
xp=x/sum(x);
yp=y/sum(y);
aa=sum(xp.*yp);
bb=sum(xp.^2);
cc=sum(yp.^2);
prop=2*aa/sqrt(bb*cc)-1;


function pointcor = pointcorre(x,y)                    %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
  error('Array sizes do not match.');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
  error('Both x and y are vectors');
end
aa=sum((x==0) & (y==0));
```

```
bb=sum((x==0) & (y~=0));
cc=sum((x~=0) & (y==0));
dd=sum((x~=0) & (y~=0));
pointcor=(aa*dd-bb*cc)/sqrt((aa+bb)*(cc+dd)*(aa+cc)*(bb+dd));

function jac = jaccard(x,y)                    %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
  error('Array sizes do not match.');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
  error('Both x and y are vectors');
end
bb=sum((x==0) & (y~=0));
cc=sum((x~=0) & (y==0));
dd=sum((x~=0) & (y~=0));
jac=(dd-(cc+bb))/(dd+cc+bb);


function dic = reviseddice(x,y)                    %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
  error('Array sizes do not match.');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
  error('Both x and y are vectors');
end
dd=sum((x~=0) & (y~=0));
bb=sum(x~=0);
cc=sum(y~=0);
dic=4*dd/(bb+cc)-1;


function overl = overlap(x,y)                    %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
  error('Array sizes do not match.');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
  error('Both x and y are vectors');
end
bb=sum(x~=y);
dd=sum((x~=0) & (y~=0));
overl=2*dd/(dd+bb)-1;


function threshr = rrank(mat,percent)
dim=size(mat); m=dim(1);
```

```
len=(m*m-m)/2;
vec=zeros(1,len);
n=0;
for i=1:m-1;
for j=i+1:m;
if (mat(i,j)~=0) n=n+1; vec(n)=mat(i,j); end;
end;
end;
num=round(percent/100*n);
vecc=sort(vec,'descend');
if (num~=0) threshr=vecc(num);
else threshr=1;
end;
```

3 Application Examples

3.1 Interval value measures

An example data matrix of interval values, with 8 attributes and 15 samples, is

8.086	9.237	4.679	7.099	3.494	4.031	7.806	8.979	9.522	8.973	4.173	4.657	6.372	2.269	2.666
9.263	3.459	9.396	8.262	0.779	4.744	8.577	3.713	10.097	2.649	8.391	6.628	5.198	6.423	9.539
4.801	2.963	8.367	0.335	1.715	8.500	3.183	6.501	5.818	9.800	6.831	1.292	0.000	3.548	2.698
9.029	7.547	10.988	4.680	8.904	8.480	7.240	2.471	1.818	3.454	5.803	0.432	3.340	7.943	8.667
6.929	10.521	7.038	7.137	7.891	2.905	7.887	8.060	1.722	2.432	9.061	5.016	2.952	0.418	5.178
1.730	5.232	0.000	2.721	3.334	5.610	2.029	3.682	3.150	3.825	3.399	8.984	8.920	7.983	10.128
3.489	8.489	8.913	7.107	10.699	4.618	1.490	5.042	8.660	6.084	8.527	9.167	3.053	5.302	1.999
5.579	1.624	9.796	0.000	0.981	6.859	5.168	7.071	3.519	7.133	8.417	1.106	0.738	2.701	3.171

3.1.1 Spearman rank correlation

I used Spearman rank correlation measure, the calculated Spearman rank correlations between attributes are

1	0.0036	0.1679	-0.4357	0.175	-0.3964	-0.0036	0.1393
0.0036	1	0.0143	0.1143	-0.1714	-0.3	-0.1571	0.1929
0.1679	0.0143	1	0.1714	-0.1357	-0.2821	0.0321	0.9214
-0.4357	0.1143	0.1714	1	0.1643	-0.3071	-0.1	0.2107
0.175	-0.1714	-0.1357	0.1643	1	-0.325	0.1429	0.0536
-0.3964	-0.3	-0.2821	-0.3071	-0.325	1	-0.175	-0.3393
-0.0036	-0.1571	0.0321	-0.1	0.1429	-0.175	1	-0.0464
0.1393	0.1929	0.9214	0.2107	0.0536	-0.3393	-0.0464	1

The statistically significant attribute pair and corresponding Spearman rank correlation are as follows (p<0.01)

3 8 0.9214

Partial correlation of Spearman rank correlation is

1	0.0634	0.314	-0.7077	0.3017	-0.5402	-0.357	-0.2754
0.0634	1	-0.6399	0.0795	-0.6029	-0.3583	0.0274	0.6346
0.314	-0.6399	1	0.268	-0.7346	-0.0889	0.3137	0.9652
-0.7077	0.0795	0.268	1	0.2899	-0.4961	-0.3631	-0.2199
0.3017	-0.6029	-0.7346	0.2899	1	-0.2061	0.2744	0.6994
-0.5402	-0.3583	-0.0889	-0.4961	-0.2061	1	-0.317	0.0476
-0.357	0.0274	0.3137	-0.3631	0.2744	-0.317	1	-0.3126
-0.2754	0.6346	0.9652	-0.2199	0.6994	0.0476	-0.3126	1

and the statistically significant attribute pair and corresponding partial correlation of Spearman rank correlation is (p<0.01)

3	8	0.9652

3.1.2 Proportion correlation

Using proportion correlation measure, the calculated proportion correlations between attributes are

1	0.6866	0.602	0.5116	0.7181	0.408	0.6786	0.5388
0.6866	1	0.5153	0.6725	0.612	0.5036	0.5921	0.5836
0.602	0.5153	1	0.5541	0.4076	0.1481	0.5426	0.9324
0.5116	0.6725	0.5541	1	0.6947	0.3927	0.5809	0.599
0.7181	0.612	0.4076	0.6947	1	0.312	0.705	0.5056
0.408	0.5036	0.1481	0.3927	0.312	1	0.4377	0.078
0.6786	0.5921	0.5426	0.5809	0.705	0.4377	1	0.4533
0.5388	0.5836	0.9324	0.599	0.5056	0.078	0.4533	1

The statistically significant attribute pairs and corresponding proportion correlations are as follows (p<0.01)

1	2	0.6866
2	4	0.6725
1	5	0.7181
4	5	0.6947
1	7	0.6786
5	7	0.7050
3	8	0.9324

Partial correlations of proportion correlations are

1	0.6115	0.7121	-0.4884	0.751	0.0237	-0.3648	-0.6351
0.6115	1	-0.612	0.4499	-0.4902	0.3324	0.4215	0.6454
0.7121	-0.612	1	0.4058	-0.7929	0.1334	0.6794	0.9611
-0.4884	0.4499	0.4058	1	0.5797	0.1557	-0.228	-0.2963
0.751	-0.4902	-0.7929	0.5797	1	-0.0116	0.6816	0.7432
0.0237	0.3324	0.1334	0.1557	-0.0116	1	0.0612	-0.2382
-0.3648	0.4215	0.6794	-0.228	0.6816	0.0612	1	-0.6406
-0.6351	0.6454	0.9611	-0.2963	0.7432	-0.2382	-0.6406	1

and the statistically significant attribute pair and corresponding partial correlation of proportion correlation is (p<0.01)

3 8 0.9611

3.2 Binary value measures

The example data matrix of binary values, with 8 attributes and 15 samples, is

1	1	0	1	0	0	1	1	1	1	0	0	1	0	0
1	0	1	0	0	0	1	0	1	0	1	1	0	1	1
0	0	0	1	0	1	0	1	0	1	1	0	0	0	0
1	1	1	0	1	0	1	0	0	0	1	0	0	1	1
0	1	0	1	1	0	1	1	0	0	1	1	0	0	0
1	0	1	0	0	0	0	0	0	1	0	0	1	1	1
0	1	0	0	0	0	0	1	0	0	1	1	0	0	0
1	0	1	0	0	1	0	1	0	0	1	0	1	0	0

I chosen point correlation measure for example use. The calculated point correlations between attributes are

1.0000	-0.3393	0.0945	-0.3393	0.0714	-0.0546	-0.0403	-0.0546
-0.3393	1.0000	-0.4725	0.4643	-0.1964	0.2182	-0.0403	-0.0546
0.0945	-0.4725	1.0000	-0.4725	0.1890	-0.2887	0.2132	0.2887
-0.3393	0.4643	-0.4725	1.0000	0.0714	0.2182	-0.0403	-0.0546
0.0714	-0.1964	0.1890	0.0714	1.0000	-0.7638	0.6447	-0.2182
-0.0546	0.2182	-0.2887	0.2182	-0.7638	1.0000	-0.4924	0.1667
-0.0403	-0.0403	0.2132	-0.0403	0.6447	-0.4924	1.0000	0.1231
-0.0546	-0.0546	0.2887	-0.0546	-0.2182	0.1667	0.1231	1.0000

The statistically significant attribute pairs and corresponding point correlations are as follows (p<0.01)

5 6 -0.7638
5 7 0.6447

Partial correlations of point correlations are

1.0000	-0.1882	-0.1497	-0.3168	0.1984	0.1549	-0.1232	0.0246
-0.1882	1.0000	-0.2813	0.3081	-0.2678	-0.1202	0.2186	-0.0692
-0.1497	-0.2813	1.0000	-0.3430	0.1231	-0.0425	-0.0123	0.3384
-0.3168	0.3081	-0.3430	1.0000	0.5152	0.4144	-0.2422	0.1883
0.1984	-0.2678	0.1231	0.5152	1.0000	-0.6919	0.5796	-0.3316
0.1549	-0.1202	-0.0425	0.4144	-0.6919	1.0000	0.1016	-0.0049
-0.1232	0.2186	-0.0123	-0.2422	0.5796	0.1016	1.0000	0.3555
0.0246	-0.0692	0.3384	0.1883	-0.3316	-0.0049	0.3555	1.0000

No significant attribute pairs were found with partial correlations of point correlations.

3.3 Comparison of results of statistic test of partial correlations with DF=*n*-*m* and DF=*n*-2

Comparisions between statistic tests of partial correlations with DF=n-m and DF=n-2, based on the datasets from Zhang (2011), were made (p<0.01). It is obvious that the results of statistic tests with different degrees of freedom (DFs) are somewhat different. However, the statistically significant pairs with DF=n-m fall in the scope of that with DF=n-2. Therefore, for non-strict uses only, e.g., coarse screening of interactions (Zhang, 2015), we may permit the number of samples less than the number of attributes (nodes, taxa, variables, etc), i.e., n<m, and the statistc test with DF=n-2 can be carefully used in this situation.

The proportion of statistically significant pairs based on DF=n-m *vs*. statistically significant pairs based on DF=n-2 (y, %) is related to significance level p (in present study, p<0.01) and number of attributes (nodes, taxa, variables, etc) (m). For this case study, the regression relationships are as follows

y=88.748e$^{-0.045m}$, r^2=0.579, p=0.047 (for Pearson partial linear correlation)
y=120.687e$^{-0.045m}$, r^2=0.956, p=<0.00001 (for partial Spearman rank correlation)

For example, if Pearson linear correlation is used and m=50, then y=9.4%. Therefore, the first approximately 9.4% of statistically significant pairs based on DF=n-2, which have the greatest partial correlations, should be the statistically significant pairs based on DF=n-m. Table 1 lists the y-m relationship based on the two regression equations. Matlab algorithm in present study has provided both situations.

Table 1 Relationship between the proportion of statistically significant pairs based on DF=n-m *vs*. statistically significant pairs based on DF=n-2 (y, %) and number of taxa (m).

No. Taxa (m)	10	20	30	40	50	60	70	80	90	100	150	200
For Pearson partial linear correlation (y, %)	56.59	36.08	23.01	14.67	9.35	5.96	3.80	2.42	1.55	0.99	0.10	0.01
For Spearman partial rank correlation (y, %)	76.52	48.79	31.11	19.84	12.65	8.06	5.14	3.28	2.09	1.33	0.14	0.01

Acknowledgment

I am thankful to the support of High-Quality Textbook *Network Biology* Project for Engineering of Teaching Quality and Teaching Reform of Undergraduate Universities of Guangdong Province (2015.6-2018.6), from Department of Education of Guangdong Province, and Project on Undergraduate Teaching Reform (2015.7-2017.7), from Sun Yat-sen University, China.

References

Schoenly KG, Zhang WJ. 1999. IRRI Biodiversity Software Series. V. RARE, SPPDISS, and SPPANK: programs for detecting between-sample difference in community structure. IRRI Technical Bulletin No.5. International Rice Research Institute, Manila, Philippines

Spearman C. 1904. The proof and measurement of association between two things. American Journal of Psychology, 15: 72-101

Zhang WJ. 2007. Computer inference of network of ecological interactions from sampling data. Environmental Monitoring and Assessment, 124: 253-261

Zhang WJ. 2011. Constructing ecological interaction networks by correlation analysis: hints from community sampling. Network Biology, 1(2): 81-98

Zhang WJ. 2012a. Computational Ecology: Graphs, Networks and Agent-based Modeling. World Scientific, Singapore

Zhang WJ. 2012b. How to construct the statistic network? An association network of herbaceous plants constructed from field sampling. Network Biology, 2(2): 57-68

Zhang WJ. 2015. A hierarchical method for finding interactions: Jointly using linear correlation and rank correlation analysis. Network Biology, 5(4): 137-145

Zhang WJ, Qi YH, Zhang ZG. 2014. Two-dimensional ordered cluster analysis of component groups in self-organization. Selforganizology, 1(2): 62-77

Zhang WJ, Li X. 2015. Linear correlation analysis in finding interactions: Half of predicted interactions are undeterministic and one-third of candidate direct interactions are missed. Selforganizology, 2(3): 39-45