Chapter 7

PATHSAS: Path Coefficient Analysis of Quantitative Traits

Christopher S. Cramer
Todd C. Wehner
Sandra B. Donaghy

Purpose

To calculate path coefficients (direct effects) and indirect effects between independent (x) and dependent (y) variables.

Definitions

Path coefficient analysis: the correlation between two traits is a function of the direct relationship between two traits and the indirect relationships of related traits (Wright, 1934).

$$r_{10} = \rho_{01} + \rho_{02}r_{12} + \rho_{03}r_{13} + \rho_{04}r_{14}$$

where r_{10} = the correlation between X_1 and Y; ρ_{01} = the path coefficient between X_1 and Y; ρ_{02} = the path coefficient between X_2 and Y; r_{12} = the correlation between X_1 and X_2; $\rho_{02}r_{12}$ = the indirect effect of X_2 on the correlation between X_1 and Y; ρ_{03} = the path coefficient between X_3 and Y; r_{13} = the correlation between X_1 and X_3; $\rho_{03}r_{13}$ = the indirect effect of X_3 on the correlation between X_1 and Y; ρ_{04} = the path coefficient between X_4 and Y; r_{14} = the correlation between X_1 and X_4; $\rho_{04}r_{14}$ = the indirect effect of X_4 on the correlation between X_1 and Y.

Originator

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics* 5:161-215.

Software Available

Cramer, C.S., Wehner, T.C., and Donaghy, S.B. (1999). PATHSAS: A SAS computer program for path coefficient analysis of quantitative data. *Journal of Heredity* 90:260-262 (free of charge).

Some References Where the Software Has Been Used

Cramer, C.S. and Wehner, T.C. (1998). Fruit yield and yield component means and correlations of four slicing cucumber populations improved through six to ten cycles of recurrent selection. *Journal of American Society of Horticulture Science* 123:388-395.

Cramer, C.S. and Wehner, T.C. (1999). Little heterosis for yield and yield components in hybrids of six cucumber inbreds. *Euphytica* 110:101-110.

Cramer, C.S. and Wehner, T.C. (2000). Path analysis of the correlation between fruit number and plant traits of cucumber populations. *HortScience* 35(4):708-711.

Contact

Dr. Todd Wehner, Department of Horticultural Science, North Carolina State University, Raleigh, NC 27695-7609, USA. E-mail: <todd_wehner@ncsu.edu>.

EXAMPLE

Data to be analyzed:

Plot	Replication	Cycle	Plant number	Pistillate flowers	Branch number	Leaf number	Total fruit number	Culled fruit number	Early fruit number
001	01	1	29	022	040	0240	01	00	00
002	02	1	21	017	034	0120	17	02	01
003	03	1	31	052	032	0440	29	15	03
004	04	1	30	049	077	0550	25	09	05
005	04	3	30	058	071	0810	30	10	05

006	03	3	23	039	040	0460	32	03	13
007	02	3	26	044	047	0460	27	07	08
008	01	3	22	023	027	0330	12	04	00
009	01	2	19	025	054	0510	27	05	04
010	02	2	27	035	038	0510	34	03	05
011	03	2	23	050	027	0290	01	00	00
012	04	2	32	035	055	0560	47	10	08
013	05	1	28	088	140	1315	58	13	27
014	06	1	28	162	105	0986	39	09	14
015	07	1	25	026	070	0741	36	05	10
016	08	1	31	074	125	0803	55	08	35
017	08	2	33	048	127	0870	57	11	13
018	07	2	25	069	038	0639	26	04	09
019	06	2	32	041	105	0878	31	04	13
020	05	2	30	021	098	0982	53	02	06
021	05	3	26	012	064	0622	31	03	09
022	06	3	31	024	111	1133	26	01	07
023	07	3	24	046	082	0879	33	04	04
024	08	3	28	048	161	1122	63	05	07

SAS Program

```
DATA DST1;
INPUT PLOT REP CYC PLANTNO PISTFLOW BRANCHNO LEAFNO TOTALNO CULLNO
     EARLYNO;
MARK=TOTALNO-CULLNO;
BRANPLAN=BRANCHNO/PLANTNO;
NODEBRAN=LEAFNO/(BRANCHNO+PLANTNO);
TOTFEMND=PISTFLOW+TOTALNO;
PERFENOD=(TOTFEMND/LEAFNO);
FRTSET=TOTALNO/PISTFLOW;
FRTPLANT=TOTALNO/PLANTNO;
MARKPLAN=MARK/PLANTNO;
EARPLAN=EARLYNO/PLANTNO;
CARDS;
001 01 1 29 022 040 0240 01 00 00
002 02 1 21 017 034 0120 17 02 01
003 03 1 31 052 032 0440 29 15 03
004 04 1 30 049 077 0550 25 09 05
005 04 3 30 058 071 0810 30 10 05
006 03 3 23 039 040 0460 32 03 13
007 02 3 26 044 047 0460 27 07 08
```

```
008 01 3 22 023 027 0330 12 04 00
009 01 2 19 025 054 0510 27 05 04
010 02 2 27 035 038 0510 34 03 05
011 03 2 23 050 027 0290 01 00 00
012 04 2 32 035 055 0560 47 10 08
013 05 1 28 088 140 1315 58 13 27
014 06 1 28 162 105 0986 39 09 14
015 07 1 25 026 070 0741 36 05 10
016 08 1 31 074 125 0803 55 08 35
017 08 2 33 048 127 0870 57 11 13
018 07 2 25 069 038 0639 26 04 09
019 06 2 32 041 105 0878 31 04 13
020 05 2 30 021 098 0982 53 02 06
021 05 3 26 012 064 0622 31 03 09
022 06 3 31 024 111 1133 26 01 07
023 07 3 24 046 082 0879 33 04 04
024 08 3 28 048 161 1122 63 05 07
;

%macro path(data,indep,dep0,dep,bylist,printreg,printout);
    /*
      Parameters to macro are:
      data =name of dataset to analyze
      indep=list of independent variables
      dep0=primary dependent variable
      dep=other dependent variables
      bylist=by variable list
      printreg=print regression? ( value is either yes or no)
      printout=print results(direct,indirect effects)?
               (value is either yes or no)
   */

%local noind word nodep noby bylast printr;

  /*  create noind macro variable                                  */
  /*  noind is the number of independent variables in &indep       */
  %let noind=0;
  %if &indep ne %then %do;
      %let word=%scan(&indep,1);
      %do %while (&word ne );
          %let noind=%eval(&noind+1);
          %let word=%scan(&indep,&noind+1);
          %end;
      %end;

  /*  create nodep macro variable                                  */
  /*  nodep is the number of dependent variables in &dep           */
  %let nodep=0;
  %if &dep ne %then %do;
      %let word=%scan(&dep,1);
      %do %while (&word ne );
          %let nodep=%eval(&nodep+1);
          %let word=%scan(&dep,&nodep+1);
          %end;
      %end;

  /*  create noby  macro variable                                  */
```

```
  /*  noby is the number of by variables in &bylist          */
  %let noby=0;
  %if &bylist ne %then %do;
      %let word=%scan(&bylist,1);
      %do %while (&word ne );
          %let noby=%eval(&noby+1);
          %let word=%scan(&bylist,&noby+1);
          %end;
      %end;
   %let bylast=%scan(&bylist,&noby);

  /*   create printr macro variable                          */
  /*   printr has a blank value or the value NOPRINT         */
  /*   specifies whether to print regression output or not   */
   %if %upcase(&printreg)=YES %then %let printr=;
      %else %let printr=noprint;

data data1; set &data;
   keep &bylist &dep0 &dep &indep;
   run;

proc sort data=data1;
   by &bylist;
proc standard data=data1 mean=0 std=1 out=sdata2;
      by &bylist;
      var &indep &dep0 &dep;
   run;

proc reg data=sdata2  &printr
      outsscp=sscp(keep=&bylist intercep _type_)
      outest=estdep(drop=_model_  _type_  _rmse_  intercep);
      by &bylist;
      model &dep0=&indep;
   run;

   /*
   _type_='N' is the number of obs in the dataset;
   nobs, number of obs., is created
   needed for checking that there are enough obs.
   if not, the reg. coefficients are biased, and need to set to miss-
     ing
   */
data sscp; set sscp;
   if _type_='N';
   rename intercep=nobs;
   drop _type_;

/*   if no. of obs. is <= the no. of indep. variables, then
     set the regression coefficients to missing          */
data estdep; merge sscp estdep;
   by &bylist;
   array v &indep;
   look='no ';
   if nobs<=&noind then do;
      look='yes';
      do over v;
         v=.;
```

```
         end;
      end;
   run;

proc print data=estdep;
   where look='yes';
   var &bylist nobs;
title3
'The following identification levels do not have enough obs. for anal-
   ysis';
title4 '   and the regression coefficients were set to missing
   ';
   run;
title3 ' ' ;

proc reg data=sdata2 &printr
   outest=estindep(drop=_model_ _type_ _rmse_ intercep);
   by &bylist;
   model &dep=&dep0;
   run;

data estind2; set estindep;
   by &bylist;
   array r  regc1-regc&nodep;
   retain   regc1-regc&nodep;
   if first.&bylast then _i_=0;
   _i_+1;
   r=&dep0;
   if last.&bylast then do;
      output;
      do over r;
         r=.;
         end;
      end;
   drop &dep0 &dep _depvar_;
   run;

proc corr data=data1  outp=corr  noprint;
   by &bylist;
   var &indep;
   run;
data corr; set corr;
   if _type_='CORR';
   drop _type_;
   run;

data estdep; set estdep;
   array reg &indep;
   array r2  reg1-reg&noind;
   do over reg;
     r2=reg;
     end;
   drop &indep;
   run;

data tog;
   merge corr estdep;
```

```
   by &bylist;
   array dir  &indep;
   array corr &indep;
   array r2   reg1-reg&noind;
   if first.&bylast then do;
      totc=0;
      n=0;
      end;
   n+1;
   &dep0=.;
   do over dir;
      if n=_i_ then dir= r2;
         else dir=r2*corr;
       &dep0 + dir;
       end;
   drop n;
   keep &bylist--_name_ &indep  &dep0 _depvar_ nobs;
   format &indep &dep0 5.2;
   run;

data tog2; merge tog estind2; by &bylist;
   array r regc1-regc&nodep;
   array t &dep;
   do over r;
      t=&dep0 * r;
      end;
   format &dep &dep0 5.2;
   format regc1-regc&nodep 5.2;
  * drop regc1-regc&nodep;
   drop _depvar_;
   run;

%if %upcase(&printout)=YES %then
   %str(proc print data=tog2(drop=regc1-regc&nodep); run;);

%mend path;

%path(data=dst1,
      indep=branplan nodebran perfenod frtset,
      dep0=frtplant,
      dep=markplan earplan,
      bylist=cyc,
      printreg=no,
      printout=yes
      );

RUN;
```

SAS Output

```
                    B     N     P                 F     M
                    R     O     E                 R     A     E
                    A     D     R     F           T     R     A
              N     N     E     F     R           P     K     R
              A     P     B     E     T  N        L     P     P
```

OBS	CYC	ME	LAN	RAN	NOD	SET	OBS	ANT	LAN	LAN
001	1	BRANPLAN	0.72	0.18	-0.06	0.03	8	0.87	0.80	0.78
002	1	NODEBRAN	0.37	0.34	-0.10	0.04	8	0.65	0.61	0.59
003	1	PERFENOD	-0.17	-0.15	0.23	0.01	8	-0.08	-0.07	-0.07
004	1	FRTSET	0.07	0.05	0.01	0.30	8	0.42	0.39	0.38
005	2	BRANPLAN	0.72	-0.13	-0.41	0.42	8	0.61	0.54	0.31
006	2	NODEBRAN	-0.26	0.34	0.01	-0.01	8	0.08	0.07	0.04
007	2	PERFENOD	-0.58	0.01	0.50	-0.53	8	-0.60	-0.53	-0.30
008	2	FRTSET	0.40	-0.01	-0.34	0.77	8	0.82	0.72	0.41
009	3	BRANPLAN	1.06	-0.03	-0.37	0.12	8	0.78	0.75	0.28
010	3	NODEBRAN	-0.15	0.20	-0.31	-0.10	8	-0.36	-0.35	-0.13
011	3	PERFENOD	-0.46	-0.07	0.86	-0.22	8	0.10	0.09	0.04
012	3	FRTSET	0.28	-0.04	-0.42	0.46	8	0.28	0.26	0.10