

PATHSAS: A SAS Computer Program for Path Coefficient Analysis of Quantitative Data

C. S. Cramer, T. C. Wehner, and S. B. Donaghy

PATHSAS is a SAS computer program for path coefficient analysis of quantitative data that is compatible with any computer system that can run SAS (Mac OS, MS-DOS, MS-Windows, OS/2, UNIX, MVS, CMS, VSE, and OpenVMS). The program includes two parts: a startup file for specifying the variables and the path analysis macro. A file explaining the operation of the program and an example dataset are also included. A copy of the program can be obtained from the World Wide Web (<http://cuke.hort.ncsu.edu/index.html>) or on diskette from T.C.W.

Path analysis was first described by Wright (1921, 1934) as a means of determining the influence of independent factors on dependent factors, while also calculating the simple correlation between pairs of independent factors. The total correlation between an independent factor, A, and a dependent factor, E, is described as the direct effect of A on E (a) plus the indirect effects of B, C, and D on E ($e \times b$, $f \times c$, $g \times d$) where e, f, g are the simple correlations between A and B, C, D, respectively; and b, c, d are the direct effects between E and B, C, D, respectively (Figure 1).

The current method for calculating the total correlation between independent and dependent factors through path analysis is to standardize the data, determine

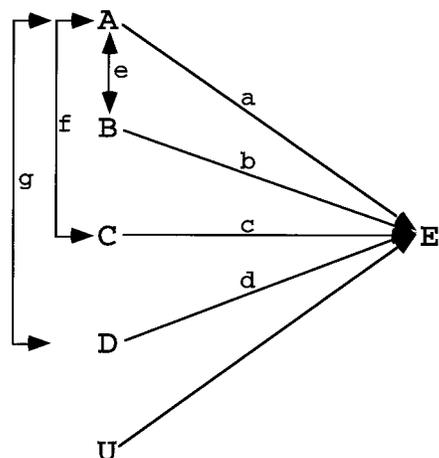


Figure 1. Sample path diagram with independent variables (A, B, C, D), dependent variables (E), error (U), direct effects (a, b, c, d) and simple correlations (e, f, g).

Correlation coefficients for Independent variables

OBS	POP	CYN	_NAME_	BRANPLAN	NODEBRAN	PERFENOD	FRTSET
1	NCWBP	0	BRANPLAN	1.00	0.06	-0.60	0.65
2	NCWBP	0	NODEBRAN	0.06	1.00	-0.24	-0.37
3	NCWBP	0	PERFENOD	-0.60	-0.24	1.00	-0.68
4	NCWBP	0	FRTSET	0.65	-0.37	-0.68	1.00
5	NCWBP	3	BRANPLAN	1.00	-0.27	-0.33	-0.09
6	NCWBP	3	NODEBRAN	-0.27	1.00	-0.09	0.14
7	NCWBP	3	PERFENOD	-0.33	-0.09	1.00	-0.30
8	NCWBP	3	FRTSET	-0.09	0.14	-0.30	1.00
9	NCWBP	5	BRANPLAN	1.00	0.00	-0.25	0.61
10	NCWBP	5	NODEBRAN	0.00	1.00	0.05	-0.29
11	NCWBP	5	PERFENOD	-0.25	0.05	1.00	0.16
12	NCWBP	5	FRTSET	0.61	-0.29	0.16	1.00

Correlation coefficients for dependent variables

OBS	POP	CYN	_NAME_	FRTPLANT	MARKPLAN	EARPLAN
1	NCWBP	0	FRTPLANT	1.00	0.97	0.12
2	NCWBP	0	MARKPLAN	0.97	1.00	0.07
3	NCWBP	0	EARPLAN	0.12	0.07	1.00
4	NCWBP	3	FRTPLANT	1.00	0.89	0.62
5	NCWBP	3	MARKPLAN	0.89	1.00	0.77
6	NCWBP	3	EARPLAN	0.62	0.77	1.00
7	NCWBP	5	FRTPLANT	1.00	0.92	0.80
8	NCWBP	5	MARKPLAN	0.92	1.00	0.87
9	NCWBP	5	EARPLAN	0.80	0.87	1.00

Direct Effects, Indirect Effects and Total Correlations

	O	P	C	Y	N	A	M	E	R	O	P	F	R	M	E
1	NCWBP	0	BRANPLAN	0.79	0.04	-0.65	0.42	15	0.61	0.59	0.07				
2	NCWBP	0	NODEBRAN	0.05	0.71	-0.26	-0.24	15	0.26	0.25	0.03				
3	NCWBP	0	PERFENOD	-0.48	-0.17	1.07	-0.43	15	-0.01	-0.01	-0.00				
4	NCWBP	0	FRTSET	0.52	-0.26	-0.72	0.64	15	0.17	0.16	0.02				
5	NCWBP	3	BRANPLAN	0.44	-0.13	-0.27	-0.03	15	0.00	0.00	0.00				
6	NCWBP	3	NODEBRAN	-0.12	0.51	-0.07	0.04	15	0.36	0.32	0.22				
7	NCWBP	3	PERFENOD	-0.15	-0.05	0.81	-0.09	15	0.53	0.47	0.33				
8	NCWBP	3	FRTSET	-0.04	0.07	-0.24	0.30	15	0.09	0.08	0.06				
9	NCWBP	5	BRANPLAN	-0.05	0.00	-0.11	0.15	13	-0.01	-0.01	-0.01				
10	NCWBP	5	NODEBRAN	-0.00	0.43	0.02	-0.07	13	0.38	0.35	0.31				
11	NCWBP	5	PERFENOD	0.01	0.02	0.43	0.04	13	0.51	0.47	0.41				
12	NCWBP	5	FRTSET	-0.03	-0.12	0.07	0.24	13	0.15	0.14	0.12				

Figure 2. Sample output from PATHSAS.

the simple correlations between independent factors, and regress all independent factors on each dependent factor separately in order to obtain the direct effects in the form of partial regression coefficients (path coefficients) (Li 1975). The indirect effects are calculated by multiplying the path coefficients by the simple correlation coefficients. The total correlation is calculated by adding the direct and indirect effects. The entire procedure is tedious and prone to computational errors for experiments with many factors or large datasets.

To assist researchers interested in studying components using path analysis, we developed a SAS (SAS Institute 1988) computer program, PATHSAS, that calculates path coefficients, simple correlations, indirect effects, and total correlations for a set of data. SAS programs run on most computer operating systems with only minor modifications by the user. Currently SAS runs on the following computer operating systems: Macintosh, MS-DOS, MS-Windows, Windows NT, UNIX, OS/2, MVS, CMS, OpenVMS, and VSE. In PATHSAS we utilized the SAS macro language to

Bootstrap 95% confidence intervals - using EC method
 Random Seed= 4768902
 Number of Resamples=15

OBS	POP	CYN	Independent Variables	Name	Approximate Lower Confidence Limit	Observed Statistic	Approximate Upper Confidence Limit
1	NCWBP	0	BRANPLAN	EARPLAN	-0.16	0.10	0.78
2	NCWBP	0	BRANPLAN	FRTPLANT	0.18	0.79	1.29
3	NCWBP	0	BRANPLAN	MARKPLAN	0.16	0.77	1.18
4	NCWBP	0	FRTSET	EARPLAN	-0.30	0.08	0.86
5	NCWBP	0	FRTSET	FRTPLANT	-0.31	0.64	1.30
6	NCWBP	0	FRTSET	MARKPLAN	-0.28	0.62	1.18
7	NCWBP	0	NODEBRAN	EARPLAN	-0.26	0.09	0.54
8	NCWBP	0	NODEBRAN	FRTPLANT	-0.10	0.71	1.02
9	NCWBP	0	NODEBRAN	MARKPLAN	-0.10	0.69	0.98
10	NCWBP	0	PERFENOD	EARPLAN	-0.30	0.13	1.02
11	NCWBP	0	PERFENOD	FRTPLANT	0.91	1.07	1.46
12	NCWBP	0	PERFENOD	MARKPLAN	0.90	1.04	1.40
13	NCWBP	3	BRANPLAN	EARPLAN	0.08	0.27	0.43
14	NCWBP	3	BRANPLAN	FRTPLANT	0.29	0.44	0.60
15	NCWBP	3	BRANPLAN	MARKPLAN	0.26	0.39	0.56
16	NCWBP	3	FRTSET	EARPLAN	-0.31	0.19	0.45
17	NCWBP	3	FRTSET	FRTPLANT	-0.50	0.30	0.60
18	NCWBP	3	FRTSET	MARKPLAN	-0.41	0.27	0.56
19	NCWBP	3	NODEBRAN	EARPLAN	0.03	0.31	0.58
20	NCWBP	3	NODEBRAN	FRTPLANT	0.33	0.51	0.86
21	NCWBP	3	NODEBRAN	MARKPLAN	0.28	0.45	0.73
22	NCWBP	3	PERFENOD	EARPLAN	0.05	0.50	0.72
23	NCWBP	3	PERFENOD	FRTPLANT	0.78	0.81	0.92
24	NCWBP	3	PERFENOD	MARKPLAN	0.48	0.72	0.79
25	NCWBP	5	BRANPLAN	EARPLAN	-0.84	0.04	0.17
26	NCWBP	5	BRANPLAN	FRTPLANT	-1.24	0.05	0.23
27	NCWBP	5	BRANPLAN	MARKPLAN	-1.08	0.05	0.22
28	NCWBP	5	FRTSET	EARPLAN	-0.51	0.19	0.57
29	NCWBP	5	FRTSET	FRTPLANT	-0.60	0.24	0.74
30	NCWBP	5	FRTSET	MARKPLAN	-0.56	0.22	0.61
31	NCWBP	5	NODEBRAN	EARPLAN	0.01	0.34	0.55
32	NCWBP	5	NODEBRAN	FRTPLANT	0.00	0.43	0.71
33	NCWBP	5	NODEBRAN	MARKPLAN	0.00	0.39	0.62
34	NCWBP	5	PERFENOD	EARPLAN	-0.12	0.35	0.46
35	NCWBP	5	PERFENOD	FRTPLANT	-0.19	0.43	0.66
36	NCWBP	5	PERFENOD	MARKPLAN	-0.16	0.40	0.51

Figure 2. Continued.

standardize the data, calculate simple correlation and regression coefficients, calculate total correlations, and present the results in a table format that is easy to read (Figure 2).

To use PATHSAS, the user adjusts the startup.sas file (Figure 3) to their needs. The user enters the names of the desired dataset and input variables, the number and names of independent, dependent, and BY (sorting) variables. In addition, the user can output full regression analysis, simple correlations of independent and dependent variables, path coefficients, indirect effects, total correlations, and bootstrap correlation estimates (along with 95% confidence intervals for those estimates). In addition, PATHSAS will notify the user when the number of observations for regression analysis is insufficient to perform the analysis.

Error variation can be estimated by us-

ing PATHSAS to calculate the confidence interval of each total correlation estimate. PATHSAS uses BOOTSTRAP analysis from the SAS statistical package (SAS Institute 1988) to calculate a total correlation estimate and a confidence interval for each estimate. The number of resampling events is set at the recommended level of 1000, but can be set to any number by the user. The user must specify a random number seed to initiate the bootstrap analysis. BOOTSTRAP analysis can take several h to complete and also can use an extensive amount of computer random access memory (RAM) depending upon the number of resampling events. The user will need to access jackboot.sas from the SAS Institute Web page (<http://www.sas.com/service/techsup/stat/jackboot.html>). This file must be stored with PATHSAS (pathsas.sas) in order for BOOTSTRAP analysis to function.

Currently pathsas.sas uses two macros, analyze and path, to analyze the data. The analyze macro is used to standardize the data, to run regression analysis, and to calculate simple correlations. In addition, the analyze macro organizes the data and groups it into tables. The path macro also conducts regression analysis, and calculates simple correlations, in addition to conducting the bootstrap analysis.

To our knowledge, there are no other SAS programs for calculating path coefficients for quantitative genetic data. SAS currently offers a path analysis procedure, CALIS, which is intended more for social sciences than for natural sciences. In addition, SEGPATH, has been developed for segregation and path analysis of genetic epidemiological models (Province and Rao 1995). PATHSAS has been used successfully in cucumber (*Cucumis sativus* L.) to calculate the path coefficients and correlations between yield components and fruit yield (Cramer and Wehner 1998).

PATHSAS is distributed free of charge. A sample file is available that illustrates the use of the two macros. The program, startup, readme, and sample files can be obtained from the Web at <http://cuke.hort.ncsu.edu/index.html>, or by sending a blank diskette to T.C.W.

From the Departments of Horticultural Science (Cramer and Wehner) and Statistics (Donaghy), North Carolina State University, Raleigh, NC 27695. C. S. Cramer is currently at the Department of Agronomy and Horticulture, New Mexico State University, Las Cruces, New Mexico. The research reported herein was funded in part by the North Carolina Agricultural Research Service. Use of trade names in this publication does not imply endorsement by the NCARS or the products named nor criticism of similar ones not mentioned. Address correspondence to T. C. Wehner at the address above.

© 1999 The American Genetic Association

References

- Cramer CS and Wehner TC, 1998. Fruit yield and yield component means and correlations of four slicing cucumber populations improved through six to ten cycles of recurrent selection. *J Amer Soc Hort Sci* 123: 388-395.
 - Li CC, 1975. Path analysis—a primer. Pacific Grove, California: Boxwood Press.
 - Province MA and Rao DC, 1995. General purpose model and a computer program for combined segregation and path analysis (SEGPATH): automatically creating computer programs from symbolic language model specifications. *Gen Epidemiol* 12:203-219.
 - SAS Institute, 1988. SAS/STAT user's guide, release 6.03. Cary, North Carolina: SAS Institute.
 - Wright S, 1921. Correlation and causation. *J Agric Res* 20:557-585.
 - Wright S, 1934. The method of path coefficients. *Ann Math Stat* 5:161-215.
- Received January 21, 1998
 Accepted June 30, 1998
 Corresponding Editor: Bruce S. Weir

```

title1 'PATHSAS startup file example';
options linesize=78 ps=54 pageno=1 nodate;

data crdst1;
  infile 'orig.dat';
  input density $ popn $ cyn branplan nodebran perfenod frtset
         frtplant markplan earplan;
run;

filename in 'orig.dat';
%inc 'pathsas.sas';
%inc 'jackboot.sas';
%path(data=crdst1,
      indep=branplan nodebran perfenod frtset,
      dep0=frtplant,
      dep=markplan earplan,
      bylist=popn cyn,
      printreg=no,
      printout=yes,
      corrind=yes,
      corrdep=yes,
      boot=yes,
      random=4768902,
      samples=200
      );

```

Figure 3. Example of startup.sas file for PATHSAS output in Figure 2.