SAS code and plotting coordinates for analysis of
Drake data on peanut preferences
Jason A. Osborne, Dept. of Statistics, NCSU, May 2002

The first batch of code generates biplot for sensory data:

```
libname drake ".";  /* creates a SAS library called drake */

/*
creates a permanent dataset in SAS library called drake.sensory
"nikki2.txt" is a textfile of comma delimited data I exported from
excel file
*/

/*data drake.sensory;
   infile "nikki2.txt" firstobs=3 delimiter=',';
   input Country $ Sample $ RP Beany DrkRst WHS SwtArom Phenol GrTwig Musty OtherNut Sweet Sour Bitter
run;*/

data sensory;
   set drake.sensory;
run;

/* run PROC PRINCOMP for principal components analysis: */

proc princomp data=sensory covariance out=two ;
   var RP Beany DrkRst WHS SwtArom Phenol GrTwig Musty OtherNut Sweet Sour Bitter Astring ChemBurn;
run;

/* PROC MEANS to look at the mean PC1 and PC2 by country */

proc means data=two mean n std;
   class country;
   var prin1 prin2;
run;

/* run PROC PRINQUAL to produce biplot (same coordinates as
those obtained by PROC PRINCOMP, just prettier) */

proc prinqual data=sensory out=results n=2 covariance mdpref;
   id country sample;
   transform identity(RP Beany DrkRst WHS SwtArom Phenol GrTwig Musty OtherNut Sweet Sour Bitter Astri
   *transform identity(RP Beany DrkRst WHS SwtArom Sweet Sour Bitter Astring);
run;

/* print coordinates for biplot for use in sigma-plot or excel or other
plotting software: */

data coord1;
   set results;
```

```
   if _type_="SCORE";
run;

proc print;
   title "coordinates for individual samples";
   var _name_ prin1 prin2;
run;

data coord2;
   set results;
   if _type_="CORR";
   rprin1=2*prin1;
   rprin2=2*prin2;
   label rprin1="Rescaled prin1";
   label rprin2="Rescaled prin2";
   attribute=sample;
run;

proc print labels;
   title "coordinates for directions of attributes";
   var prin1 prin2 rprin1 rprin2 attribute;
run;

/* following code takes temporary dataset called results
and generates biplots with labels= (1) country (2) sample
using the fancy PLOTIT macro written by W.F. Kuhfeld
graphics written to an encapsulated postscript file
called  "sasgraph.eps"
*/

goptions dev=pslepsf colors=(black) gsfmode=replace;

title1 "Biplot of 14 sensory measurements";
title2 "labels are countries";
%plotit(Data=results,labelvar=country, datatype=mdpref2 2);

title1 "Biplot of 14 sensory measurements";
title2 "labels are sample numbers";
%plotit(Data=results,labelvar=sample, datatype=mdpref2 2);
```

Some of the output from the preceding code appears below:

The MEANS Procedure

Country	N Obs	Variable	Mean	N	Std Dev
ARG	6	Prin1	0.2117440	6	1.9353419
		Prin2	1.0060721	6	1.6093834
CHI	6	Prin1	1.9074326	6	2.2421465
		Prin2	-0.9152467	6	0.9963916
USA	6	Prin1	-2.1191766	6	0.9506008
		Prin2	-0.0908254	6	0.2587561

The PRINCOMP Procedure

Total Variance 10.691339869

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	5.73096004	4.00166762	0.5360	0.5360
2	1.72929242	0.91503814	0.1617	0.6978
3	0.81425428	0.08994114	0.0762	0.7739
4	0.72431314	0.23916705	0.0677	0.8417
5	0.48514609	0.13081060	0.0454	0.8871
6	0.35433548	0.07696516	0.0331	0.9202
7	0.27737033	0.06504623	0.0259	0.9462
8	0.21232409	0.07825599	0.0199	0.9660
9	0.13406811	0.01846592	0.0125	0.9786
10	0.11560218	0.04233299	0.0108	0.9894
11	0.07326919	0.04811626	0.0069	0.9962
12	0.02515293	0.01295425	0.0024	0.9986
13	0.01219868	0.00914580	0.0011	0.9997
14	0.00305289		0.0003	1.0000

coordinates for individual samples 7

Obs	_NAME_	Prin1	Prin2
1	ARG	-0.34974	-0.10846
2	ARG	-0.46694	-0.01061
3	ARG	0.88922	2.45445
4	ARG	-0.94339	0.93446
5	ARG	0.30609	-0.60851
6	ARG	1.09546	1.92903
7	CHI	-0.15521	0.06106
8	CHI	2.33714	0.09611
9	CHI	0.33038	-1.23975
10	CHI	1.03561	-1.64846
11	CHI	1.23578	-1.20501
12	CHI	-0.00305	-0.23990
13	USA	-0.79047	0.10742
14	USA	-1.18176	0.08507
15	USA	-0.85465	-0.23196
16	USA	-1.49881	0.09185
17	USA	-0.54811	-0.36037
18	USA	-0.43755	-0.10641

coordinates for directions of attributes 8

Obs	Prin1	Prin2	Rescaled prin1	Rescaled prin2	attribute
1	-0.96355	-0.06525	-1.92709	-0.13049	RP
2	0.52743	-0.09292	1.05487	-0.18585	Beany
3	-0.36640	-0.29592	-0.73280	-0.59183	DrkRst
4	0.91474	-0.10624	1.82949	-0.21247	WHS
5	-0.95932	0.07696	-1.91864	0.15392	SwtArom
6	0.39178	0.30333	0.78356	0.60666	Phenol
7	0.02640	-0.34609	0.05280	-0.69219	GrTwig
8	0.63009	0.70688	1.26018	1.41377	Musty
9	0.08463	-0.32578	0.16926	-0.65156	OtherNut
10	-0.07590	0.75517	-0.15180	1.51035	Sweet
11	0.48696	-0.47146	0.97392	-0.94292	Sour
12	0.86569	-0.30950	1.73137	-0.61900	Bitter
13	0.61164	-0.32827	1.22329	-0.65654	Astring
14	0.27339	0.48142	0.54678	0.96284	ChemBurn

To read the data in and create the permanent sas dataset *drake.hedonic* which will be used in the analysis of the preferences from the consumer survey. data were written to comma-delimited textfiles separately for each day (e.g. *day2.txt*). These were read in using separate datasteps and then merged:

```
options ls=75 nodate;

libname drake ".";

/* Acceptance data */
data day1;
   infile "day1.txt" firstobs=12 delimiter=',' missover dsd;
   drop v1 v2;
   day=1;
   input v1 v2 panelist trtmnt oliking oflavor icolor irsdpntfla iswtaste ibitter
ifrshpntfla itexture lcolor lrsdpntfla lswtaste lbitter lfrshpntfla ltexture;
   if trtmnt=1 then do;  trt="804"; sample="con"; end;
   else if trtmnt=2 then do; trt="125"; sample="028"; end;
   else if trtmnt=3 then do; trt="349"; sample="068"; end;
   else if trtmnt=4 then do; trt="201"; sample="004"; end;
run;
data day2;
   infile "day2.txt" firstobs=12 delimiter=',' missover dsd;
   drop v1 v2;
   day=2;
   input v1 v2 panelist trtmnt oliking oflavor icolor irsdpntfla iswtaste ibitter
ifrshpntfla itexture lcolor lrsdpntfla lswtaste lbitter lfrshpntfla ltexture;
   if trtmnt=1 then do; trt="091";  sample="con"; end;
   else if trtmnt=2 then do; trt="580";  sample="080"; end;
   else if trtmnt=3 then do; trt="477";  sample="070"; end;
   else if trtmnt=4 then do; trt="051";  sample="001"; end;
run;
data day3;
   infile "day3.txt" firstobs=12 delimiter=',' missover dsd;
   drop v1 v2;
   day=3;
   input v1 v2 panelist trtmnt oliking oflavor icolor irsdpntfla iswtaste ibitter
ifrshpntfla itexture lcolor lrsdpntfla lswtaste lbitter lfrshpntfla ltexture;
   if trtmnt=1 then do; trt="567";  sample="Con"; end;
   else if trtmnt=2 then do; trt="314";  sample="022"; end;
   else if trtmnt=3 then do; trt="845";  sample="059"; end;
   else if trtmnt=4 then do; trt="122";  sample="012"; end;
run;
data day4;
   infile "day4.txt" firstobs=12 delimiter=',' missover dsd;
   drop v1 v2;
   day=4;
   input v1 v2 panelist trtmnt oliking oflavor icolor irsdpntfla iswtaste ibitter
ifrshpntfla itexture lcolor lrsdpntfla lswtaste lbitter lfrshpntfla ltexture;
   if trtmnt=1 then do; trt="973";  sample="Con"; end;
```

```
   else if trtmnt=2 then do; trt="482";  sample="031"; end;
   else if trtmnt=3 then do; trt="668";  sample="065"; end;
   else if trtmnt=4 then do; trt="015";  sample="006"; end;
run;
data day5;
   infile "day5.txt" firstobs=12 delimiter=',' missover dsd;
   drop v1 v2;
   day=5;
   input v1 v2 panelist trtmnt oliking oflavor icolor irsdpntfla iswtaste ibitter
ifrshpntfla itexture lcolor lrsdpntfla lswtaste lbitter lfrshpntfla ltexture;
   if trtmnt=1 then do; trt="739";  sample="Con"; end;
   else if trtmnt=2 then do; trt="306";  sample="082"; end;
   else if trtmnt=3 then do; trt="275";  sample="073"; end;
   else if trtmnt=4 then do; trt="142";  sample="020"; end;
run;
data day6;
   infile "day6.txt" firstobs=12 delimiter=',' missover dsd;
   drop v1 v2;
   day=6;
   input v1 v2 panelist trtmnt oliking oflavor icolor irsdpntfla iswtaste ibitter
ifrshpntfla itexture lcolor lrsdpntfla lswtaste lbitter lfrshpntfla ltexture;
   if trtmnt=1 then do; trt="623";  sample="Con"; end;
   else if trtmnt=2 then do; trt="809"; sample="033"; end;
   else if trtmnt=3 then do; trt="946"; sample="077"; end;
   else if trtmnt=4 then do; trt="441"; sample="016"; end;
run;


data drake.hedonic;
   set day1 day2 day3 day4 day5 day6;
   trtnum=1*trt;
   if trtmnt=2 then country="USA";
   else if trtmnt=3 then country="CHI";
   else if trtmnt=4 then country="ARG";
   else if trtmnt=1 then country="Control";
run;
```

The code below will generate means for the overall liking response by sample and by country:

```
data hedonic;
   set drake.hedonic;
   if sample="Con" then sample="con";
   *if sample="con" then delete;
   keep oliking oflavor sample country;
run;

proc glm;
   class sample country;
   model oliking=country;
   means country;
run;

proc glm;
   class sample country;
   model oliking=country sample(country);
   lsmeans country sample(country);
run;
```

The code below will generate the coordinates and pretty biplot of these consumer preference data:

```
options ls=75 nodate;

libname drake ".";

data hedonic;
   set drake.hedonic;
   if sample="Con" then sample="con";
run;

proc sort;
   by sample;
run;

proc means noprint;
   id sample country;
   class sample;
   var oliking oflavor icolor irsdpntfla iswtaste ibitter ifrshpntfla itexture lcolor lrsdpntfla lswta
   output out=two mean=;
run;

data two;
   set two;
   if _type_=1;
   drop _type_;
   if sample="con" then delete;
```

```
run;

proc princomp cov ;
   var oliking oflavor icolor irsdpntfla iswtaste ibitter ifrshpntfla itexture lcolor lrsdpntfla lswta
run;

proc prinqual data=two out=results n=2 covariance mdpref;
   id country sample;
   transform identity(oliking oflavor icolor irsdpntfla iswtaste ibitter ifrshpntfla itexture lcolor l
run;

/* print coordinates for biplot for use in sigma-plot or excel or other
plotting software: */

data coord1;
   set results;
   if _type_="SCORE";
run;

proc print;
   title "coordinates for samples (consumer data)";
   var _name_ prin1 prin2;
run;

data coord2;
   set results;
   if _type_="CORR";
   rprin1=2*prin1;
   rprin2=2*prin2;
   label rprin1="Rescaled prin1";
   label rprin2="Rescaled prin2";
   attribute=sample;
run;

proc print labels;
   title "coordinates for directions of attributes (consumer data)";
   var prin1 prin2 rprin1 rprin2 attribute;
run;

filename hedonic "hedonic.eps";
goptions gsfname=hedonic dev=pslepsf colors=(black);

title1 "Biplot of 14 consumer measurements";
title2 "labels are countries";
%plotit(Data=results,labelvar=country, datatype=mdpref2 2);

title1 "Biplot of 14 consumer measurements";
title2 "labels are sample numbers";
%plotit(Data=results,labelvar=sample, datatype=mdpref2 2);
```

The output from PROC PRINCOMP gives the proportion variance explained by the PCS:

The SAS System 3

The PRINCOMP Procedure

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	2.37838948	0.46545291	0.5207	0.5207
2	1.91293656	1.80736202	0.4188	0.9395
3	0.10557454	0.04810701	0.0231	0.9626
4	0.05746753	0.02173706	0.0126	0.9752
5	0.03573047	0.00847843	0.0078	0.9830
6	0.02725204	0.01105814	0.0060	0.9890
7	0.01619391	0.00379723	0.0035	0.9925
8	0.01239667	0.00464009	0.0027	0.9952
9	0.00775658	0.00173595	0.0017	0.9969
10	0.00602063	0.00286929	0.0013	0.9982
11	0.00315135	0.00075735	0.0007	0.9989
12	0.00239400	0.00056896	0.0005	0.9995
13	0.00182503	0.00113989	0.0004	0.9999
14	0.00068514		0.0001	1.0000

The coordinates for the biplot appear in the output below:

coordinates for samples (consumer data) 2

Obs	_NAME_	Prin1	Prin2
1	ARG	-1.12380	-0.34201
2	ARG	-0.80397	0.01558
3	ARG	-1.11589	0.08060
4	ARG	-0.52933	0.71727
5	ARG	-1.00228	-0.82587
6	ARG	-1.73533	-1.42570
7	USA	-0.29758	1.05649
8	USA	0.62617	0.77813
9	USA	0.34445	1.11212
10	USA	-0.35631	1.10780
11	CHI	1.21679	-0.55473
12	CHI	0.64231	-1.81855
13	CHI	1.05507	-0.21561
14	CHI	1.46377	-0.42798
15	CHI	0.34539	-1.75965
16	CHI	1.78953	0.29762
17	USA	0.07053	1.07042
18	USA	-0.58952	1.13409

Obs	Prin1	Prin2	Rescaled prin1	Rescaled prin2	attribute
1	0.60825	0.77765	1.21651	1.55530	oliking
2	0.60493	0.77754	1.20986	1.55509	oflavor
3	-0.87953	0.47029	-1.75905	0.94058	icolor
4	0.17396	0.95491	0.34791	1.90982	irsdpntfla
5	-0.42539	0.53561	-0.85078	1.07122	iswtaste
6	-0.67204	-0.20045	-1.34407	-0.40089	ibitter
7	0.69141	0.71202	1.38281	1.42405	ifrshpntfla
8	-0.47429	0.78685	-0.94859	1.57370	itexture
9	0.89209	-0.19388	1.78417	-0.38777	lcolor
10	0.55360	0.81477	1.10720	1.62953	lrsdpntfla
11	0.39136	0.88520	0.78272	1.77040	lswtaste
12	0.57663	0.45446	1.15326	0.90892	lbitter
13	0.66793	0.73670	1.33585	1.47341	lfrshpntfla
14	-0.23550	0.88752	-0.47100	1.77504	ltexture

Lastly, to obtain correlations among the descriptive/sensory and consumer/preference sample means, we use PROC CORR. To carry out the partial least squares analysis and obtain a plot with the loadings of both the sensory/descriptive attributes and the consumer preference attributes on the first two extracted factors, we need to merge the data sets and use PROC PLS and the PLOTIT macro again. The code starts below:

```
options ls=75;    libname drake ".";

data desc_sense;
   set drake.sensory;
run;

data hedonic;
   set drake.hedonic;
   if country="Con" then delete;
run;

proc means noprint;
   class sample;
   id country;
   var oliking oflavor icolor irsdpntfla iswtaste ibitter ifrshpntfla
itexture lcolor lrsdpntfla lswtaste lbitter lfrshpntfla ltexture;
   output out=ymeans mean=;
run;

proc sort data=ymeans;
   by sample;
run;
proc sort data=desc_sense;
   by sample;
```

```
run;

data both;
   merge ymeans desc_sense;
   by sample;
   if _type_=1;
   drop _type_;
run;

proc corr;
   var oliking oflavor RP Beany DrkRst WHS SwtArom Phenol GrTwig Musty OtherNut Sweet Sour Bitter Astr
run;

ods output xweights=xweights;     ods output yweights=yweights;

proc pls nfac=2 details;
   model oliking oflavor icolor irsdpntfla iswtaste ibitter ifrshpntfla
itexture lcolor lrsdpntfla lswtaste lbitter lfrshpntfla ltexture =
RP Beany DrkRst WHS SwtArom Phenol GrTwig Musty OtherNut Sweet Sour Bitter
Astring ChemBurn;
   output out=plsout xscore=xscore yscore=yscore;
run;

proc transpose data=xweights(drop=numberoffactors innerregcoef) out=xweights;

data xweights;
   set xweights;
   rename col1=factor1 col2=factor2;
   _name_=upcase(_name_);
run;

proc transpose data=yweights out=yweights;

data yweights;
   set yweights;
   rename col1=factor1 col2=factor2;
   if col1<1;
run;

data plotpls;
   set xweights yweights;
run;
proc print;
   title "coordinates for PLS plot";
run;

goptions dev=pslepsf colors=(black);

title1 "Plot of both sensory and acceptance loadings on first two factors";
%plotit(data=plotpls, plotvars=factor2 factor1,labelvar=_name_);
```

The PLS output and the coordinates for the PLS are printed below:

The SAS System 5

The PLS Procedure

Percent Variation Accounted for
by Partial Least Squares Factors

Number of Extracted Factors	Model Effects Current	Model Effects Total	Dependent Variables Current	Dependent Variables Total
1	36.8869	36.8869	39.5314	39.5314
2	19.1849	56.0718	16.7970	56.3284

coordinates for PLS plot 8

Obs	_NAME_	factor1	factor2	_LABEL_
1	RP	0.433759	-0.050142	
2	BEANY	-0.128683	0.241112	
3	DRKRST	0.093143	-0.140660	
4	WHS	-0.399875	0.141997	
5	SWTAROM	0.364453	-0.148284	
6	PHENOL	-0.327519	-0.337279	
7	GRTWIG	0.138017	0.272876	
8	MUSTY	-0.265386	-0.283443	
9	OTHERNUT	-0.100340	0.073988	
10	SWEET	-0.079076	-0.551273	
11	SOUR	-0.120055	0.350102	
12	BITTER	-0.418383	0.212197	
13	ASTRING	-0.272030	0.173860	
14	CHEMBURN	-0.251015	-0.379621	
15	oliking	0.295623	0.259436	
16	oflavor	0.290297	0.256895	
17	icolor	0.176465	-0.384165	
18	irsdpntf	0.376074	0.071904	
19	iswtaste	0.196346	-0.198744	
20	ibitter	-0.082381	-0.366400	
21	ifrshpnt	0.273924	0.266117	
22	itexture	0.282871	-0.237746	
23	lcolor	-0.037735	0.399158	
24	lrsdpntf	0.316706	0.225229	
25	lswtaste	0.359351	0.148374	
26	lbitter	0.174054	0.268201	
27	lfrshpnt	0.281268	0.276536	
28	ltexture	0.339277	-0.179949	