

A glance at Drake data on peanut preferences
Jason A. Osborne, Dept. of Statistics, NCSU, May 2002

Sensory measurement

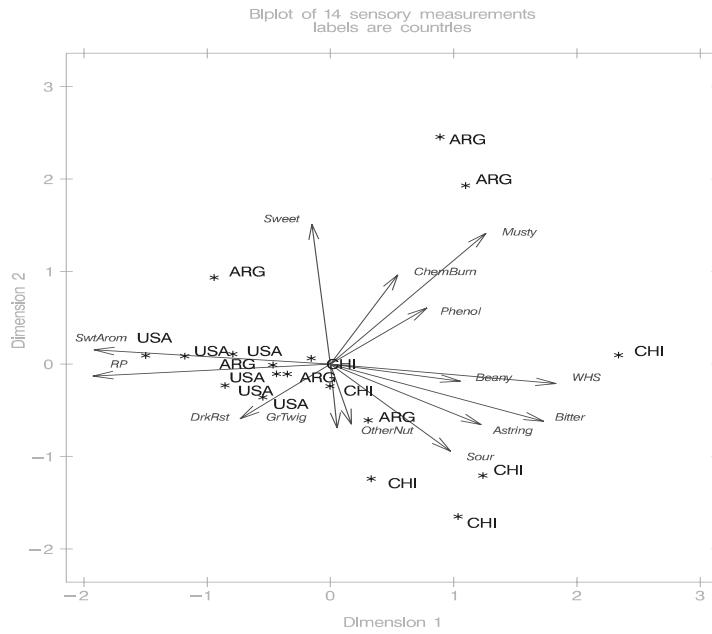
Principal components analysis of sensory measurements on 18 peanut samples (varieties). Six of these samples come from each of three countries: Argentina, Chile, USA. The samples from Chile score high on the PC1 and a little low on PC2, the samples from USA score low on PC1, the samples from Argentina score high on PC2. This can be seen from inspection of the means and figure below:

The SAS System

1

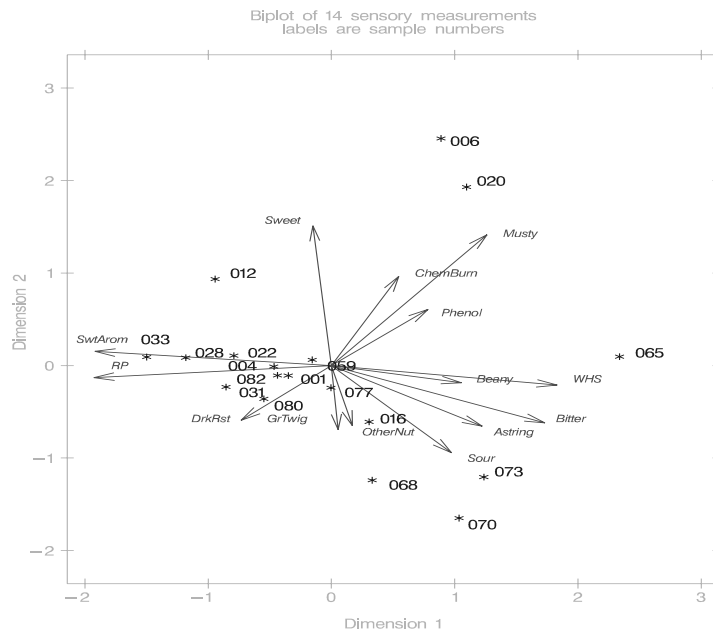
The MEANS Procedure

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



PC1 accounted for 54% of the total sample variance in the 14 descriptive attributes, PC2 accounted for 16% for a cumulative total of 70%.

The graph below depicts the contributions or "loadings" of the 14 attributes to the first two principal components:



Note that some of the attributes of the peanuts measured exhibit almost no variability and may not explain much of the total sample variance. In particular, *Chem Burn* was identically 0 for all but one of the 15 samples, *Phenol* was identically 0 for all but two samples, and *GrTwig*, *Musty* and *OtherNut* were 0 in thirteen samples of the 16 samples. However, when these 5 attributes were excluded from the principal components analysis, the configuration of the samples and loadings on the principal components did change appreciably. The final analysis includes all 14 attributes.

Data from consumer survey

For the sensory dataset of $n \approx 2000$ consumers, responses of dimension 14 were taken:

1. overall liking
2. overall flavor
3. intensity of color
4. intensity of roasted peanut flavor
5. intensity of swtaste
6. intensity of bitter
7. intensity of fresh peanut flavor
8. intensity of texture
9. liking of color
10. liking of roasted peanut flavor
11. liking of swtaste
12. liking of bitter
13. liking of fresh peanut flavor
14. liking of texture

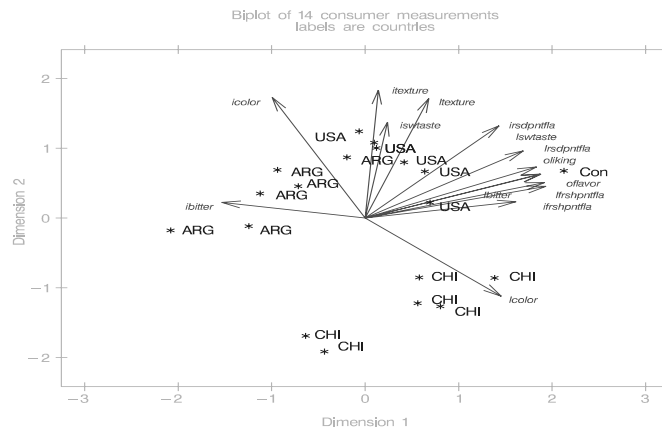
The mean overall liking for each sample are given below. The F-ratio for a test of equality of these means was highly significant ($p < 0.0001$).

sample	country	oliking LSMEAN
001	ARG	4.34042553
004	ARG	4.54347826
006	ARG	4.67676768
012	ARG	5.07216495
016	ARG	4.38383838
020	ARG	3.71717172
059	CHI	5.12500000
065	CHI	4.47000000
068	CHI	5.10416667
070	CHI	5.24175824
073	CHI	4.28282828
077	CHI	5.65656566
con	Con	6.41086587
022	USA	5.28865979
028	USA	5.29473684
031	USA	5.51485149
033	USA	5.39393939
080	USA	5.52688172
082	USA	5.16326531

The means by country are given below. These observed means also differed significantly ($p < 0.0001$).

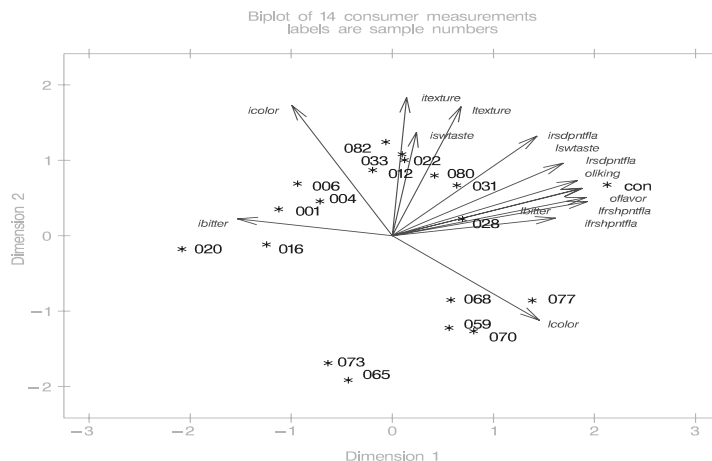
Level of country	N	-----oliking----- Mean	Std Dev
ARG	580	4.45344828	2.03753343
CHI	581	4.97418244	1.93364616
Con	589	6.41086587	1.56631769
USA	583	5.36363636	1.83649014

A principal components analysis of all 2000 14-dimensional responses can be carried out in principle, but a biplot with so many points is beyond interpretation. Instead, multivariate mean vectors were computed for each sample and used in the principal component analysis. The first two principal components explained 56% and 39% of the total sample variance respectively, for a cumulative total of 95%. Biplots are given below:



Inspection of this plots reveals clustering by country. The samples from the USA tend to lie in or near the upper right quadrant of the biplot, scoring positively on PC1 and PC2. Not that overall liking and overall flavor point in the direction of the upper right quadrant as well. The samples from Argentina lie to the left of the USA cluster, scoring lower on PC1. The samples from Chile score low on and about equally with the USA on PC1.

Plots with samples for labels are given below:



To investigate the associations between the descriptive/sensory measurements (*predictors*) and the hedonic consumer/acceptance measurements (*responses*), correlation coefficients were computed on sample means. Predictors which exhibited significant correlations with overall liking included WHS $r = -0.54$, SwtArom $r = 0.47$, Phenol $r = -0.66$, Musty $r = -0.48$, Bitter $r = -0.51$ and ChemBurn $r = -0.57$. These correlation coefficients were significantly different from 0 using level of significance $\alpha = 0.05$ without any adjustment for multiplicity of comparisons.

To further investigate the association between the multivariate response and multivariate predictors, partial least squares regression analysis was carried out. Partial least squares maximizes the covariance between the reduced-dimension predictor factors and response factors.

The loadings of the **sensory** attributes on the first two principal components extracted for the partial least squares analysis are plotted below. The proportions of variation accounted for for the two factors are 37% and 19% respectively, for a cumulative total of 56%.



The loadings of the **acceptance** attributes on the first two principal components extracted for the partial least squares analysis are plotted below. The proportions of variation accounted for for the two factors are 40% and 17% respectively, for a cumulative total of about 56%.

