A Novel Dummy Variable Approach to Preference Mapping of Likings from JAR Data

R. Xiong and J.-F. Meullenet

Food Science Department

University of Arkansas

Introduction

- Hedonic and JAR (just-about-right) scales are widely used together to provide directional information for product reformulation or optimization

Hedonic scale	JAR scale for flavor
Like extremely	Much to weak
Like very much	Too weak
Like moderately	Just about right
Like slightly	Too strong
Neither like nor dislike	Much too strong
Dislike slightly	
Dislike moderately	
Dislike very much	
Dislike extremely	

Introduction

- **Penalty analysis**
  - A graphical technique, understandable to managers
  - Ignoring correlations among JAR-scale attributes
  - Not a regression method
- **MARS** (multivariate adaptive regression splines)
  - Regression method
  - Not an apropriate method to deal with correlations among JAR-scale attributes
- **PLS** (partial least squares) **linear regression**
  - A good method to handle collinearity problems
  - Not appropriate using original JAR-scale attributes

Introduction

- **Dummy variables**
  - Regression model requires numerical variables
  - Categorical variables must be converted to numerical variables (called dummy variables) before using any regression models (for example, gender can represented by Z=0 for male and Z=1 for female, where Z is a dummy variable and can be used in regression models)
  - Dummy variables have not been used for JAR-scale attributes with PLS models

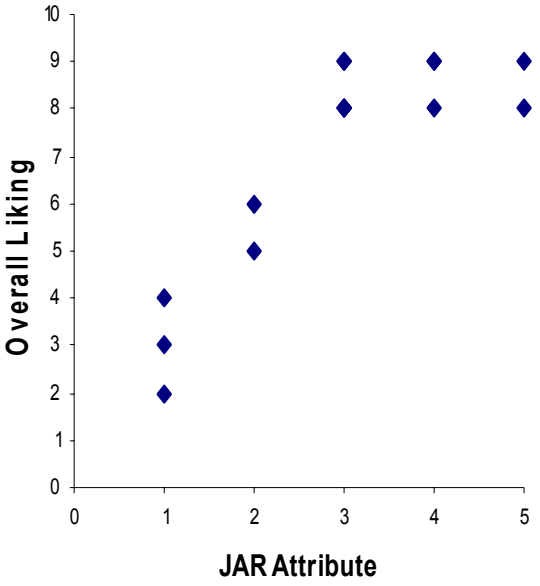
Objective

- To develop the dummy variable approach that transforms each of original JAR-scale variables into a pair of dummy variables that are then used to model the relationship between liking and JAR scores

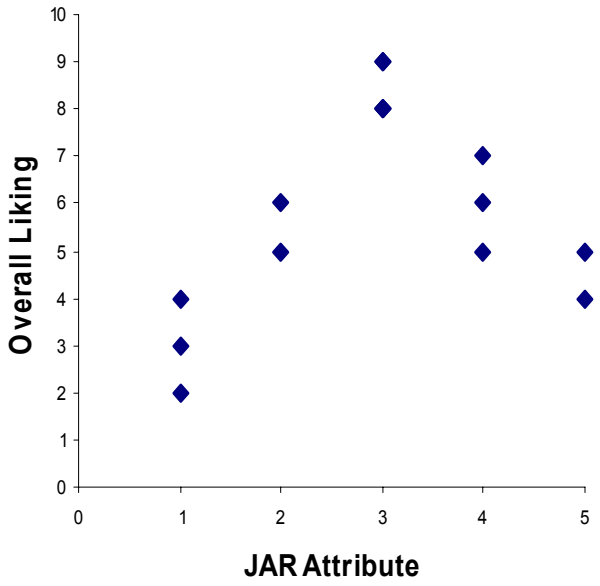
Likings VS JAR Data

- **Three patterns of likings vs JAR data**

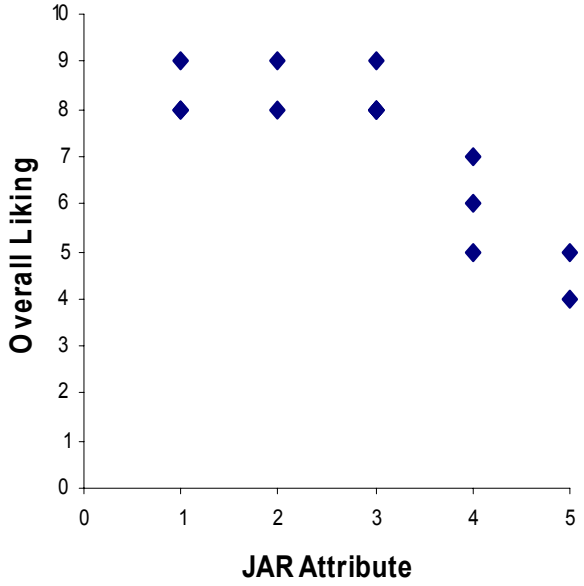
Tenderness



Salt

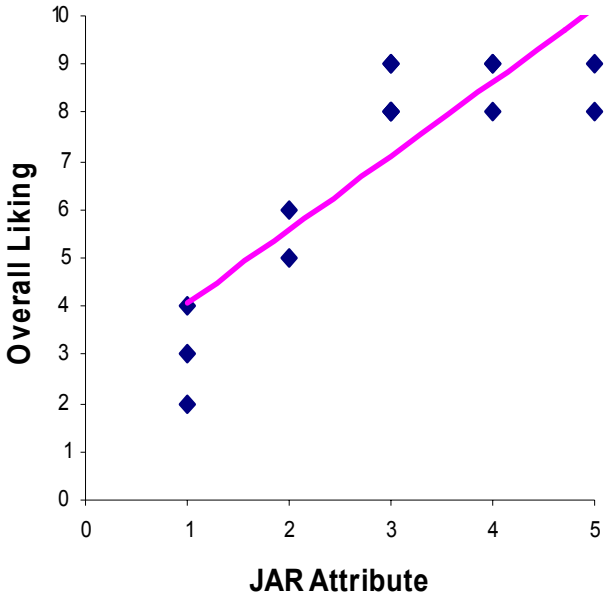


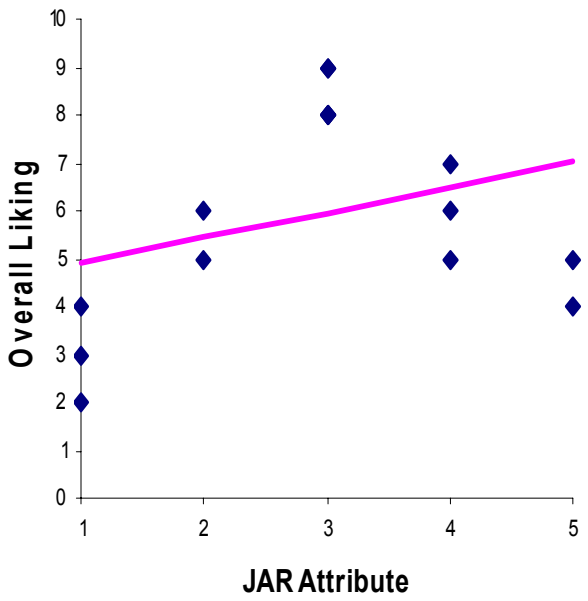
Heat

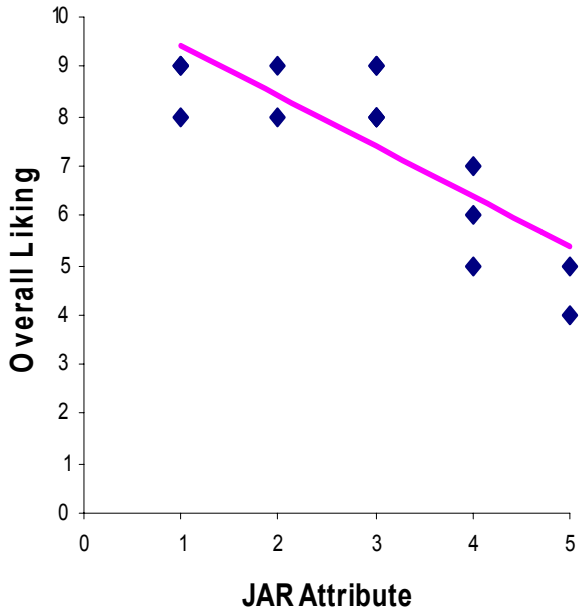


PLS Linear Regression

- Use of the original JAR-scale variables

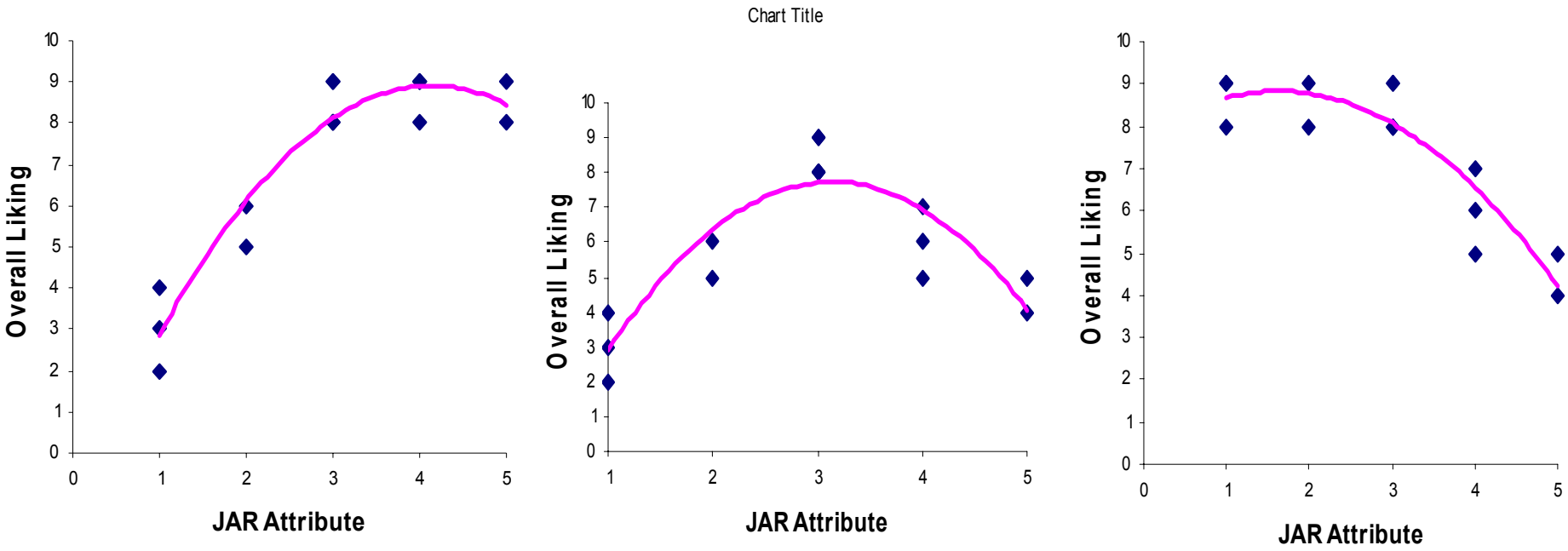






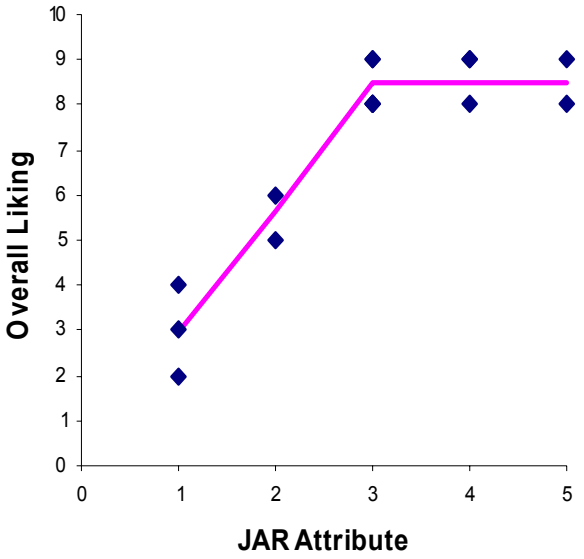
PLS Linear Regression

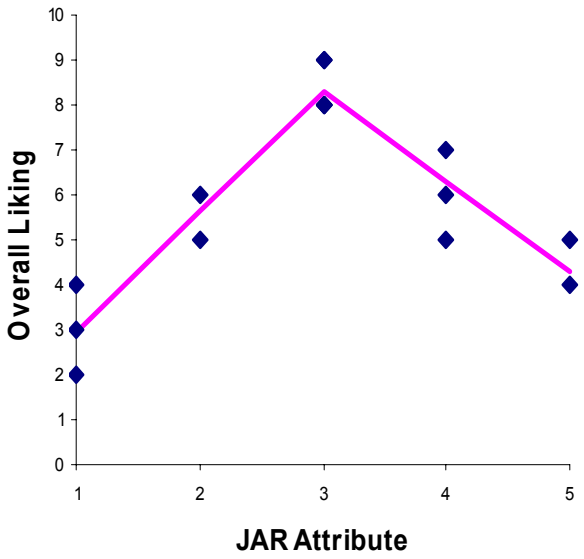
- Use of the original JAR-scale variables

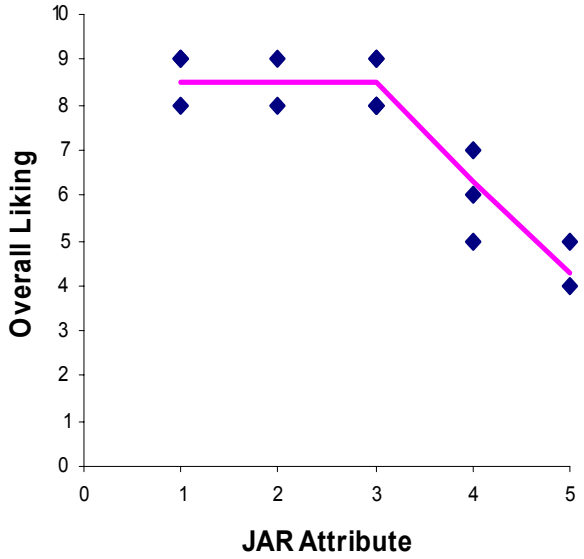


PLS Linear Regression

- Use of paired dummy variables

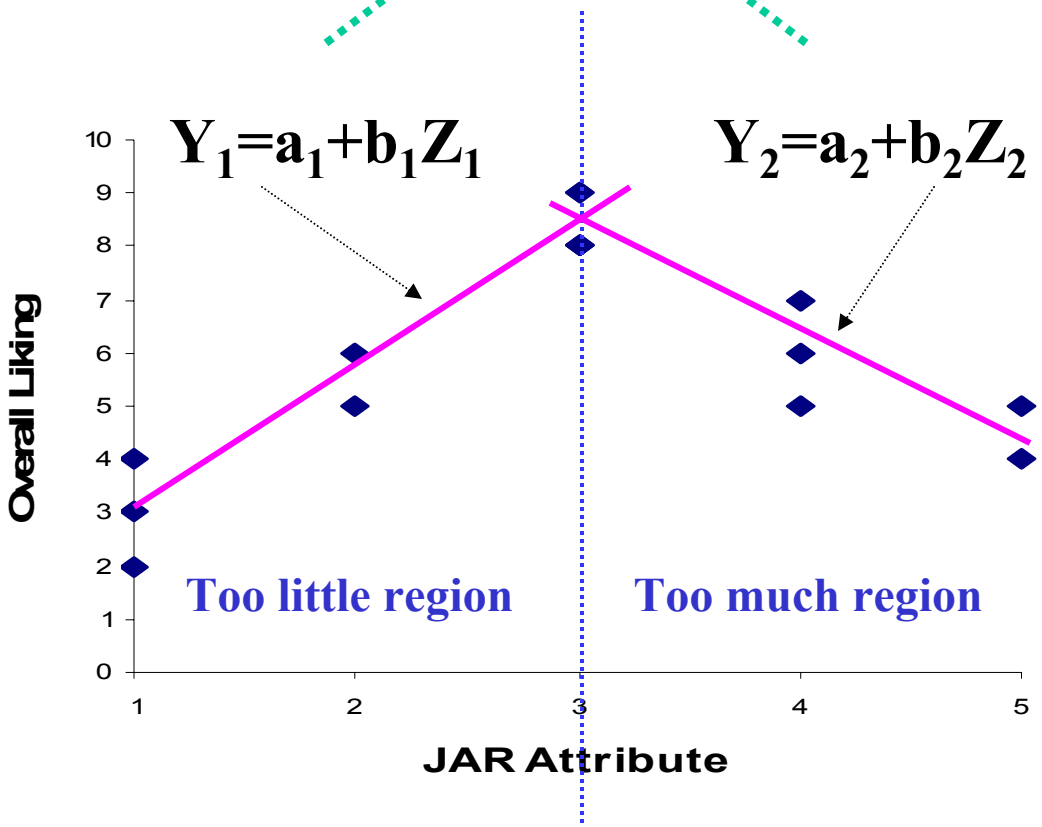






Dummy Variables

$$Y=b_0+b_1Z_1+b_2Z_2$$



Scheme for Two Dummy Variables Using a 5 point JAR Scale

Original variable	Dummy variables		
X	Z_1	Z_2	$Z=Z_1-Z_2$
1	-2	0	-2
2	-1	0	-1
3	0	0	0
4	0	1	-1
5	0	2	-2

Scheme for Two Dummy Variables Using 7 point JAR Scale

Original variable	Dummy variables		
X	Z_1	Z_2	$Z=Z_1-Z_2$
1	-3	0	-3
2	-2	0	-2
3	-1	0	-1
4	0	0	0
5	0	1	-1
6	0	2	-2
7	0	3	-3

Use of Dummy Variables

- **Paired dummy variables can be used with**
  - Regular linear regression
  - Analysis of covariance
  - PLS or PC regression
  - Logistic regression
  - Other regression models

Principle Components (PC)

- **Original variables**

$$PC_i = a_{i1}X_{i1} + a_{i2}X_{i2} + \ldots + a_{ip}X_{ip} \qquad \text{(i=1,2,3, …, p)}$$

- **Dummy variables**

$$PC_i = \left(b_{i11}Z_{i11} + b_{i12}Z_{i12}\right) + \left(b_{i21}Z_{i21} + b_{i22}Z_{i22}\right) + \ldots + \left(b_{ip1}Z_{ip1} + b_{ip2}Z_{ip2}\right)$$

(i=1,2,3, …, p)

PLS Regression Models

- **Original variables**

$$Y = a_0 + a_1 PC_1 + a_2 PC_2 + \ldots + a_q PC_q \qquad (q \le p)$$
$$= b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p$$

- **Dummy variables**

$$Y = a_0 + a_1 PC_1 + a_2 PC_2 + \ldots + a_q PC_q$$
$$= b_0 + (b_{11} Z_{11} + b_{12} Z_{12}) + (b_{21} Z_{21} + b_{22} Z_{22}) + \ldots + (b_{p1} Z_{p1} + b_{p2} Z_{p2})$$

$$\bar{Y} = b_0 + (b_{11} \bar{Z}_{11} + b_{12} \bar{Z}_{12}) + (b_{21} \bar{Z}_{21} + b_{22} \bar{Z}_{22}) + \ldots + (b_{p1} \bar{Z}_{p1} + b_{p2} \bar{Z}_{p2})$$

Data Set

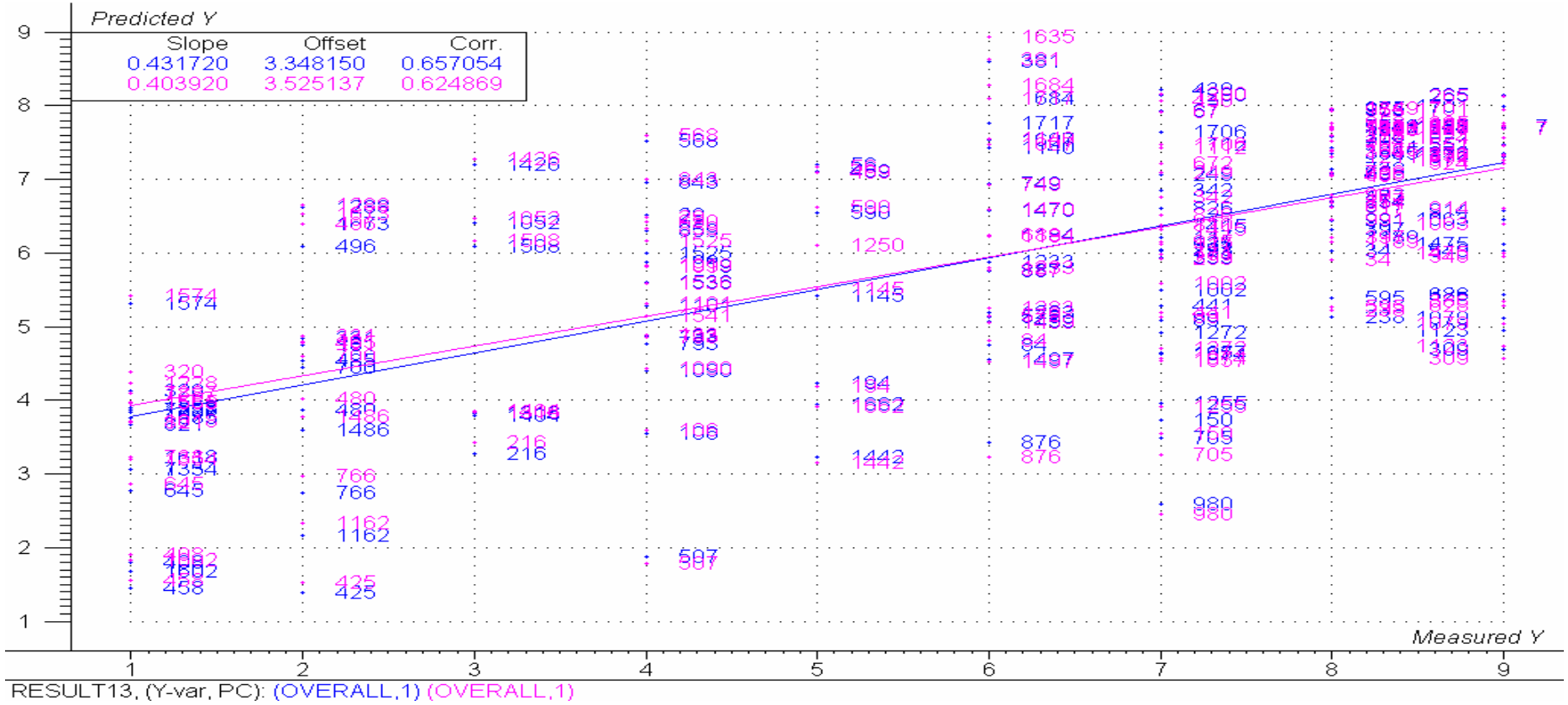
- **A published data set** provided by Popper (2003) for the JAR Data Workshop of the 5th Pangborn Sensory Science Symposium
- **157 consumers**
- **11 products** (A, B, C,…, K): A is competitor's product, B-K are prototypes.
- **Overall liking** (9-point hedonic scale)
- **16 JAR-scale attributes** (5-point scale)

Statistical Analysis

- PLS models using 16 original JAR variables
- PLS models using 32 paired dummy variables
- PLS models were fitted separately for each product
- Unscrambler and SAS statistical packages were used
- Jackknife optimization method in the Unscrambler was used to select important variables

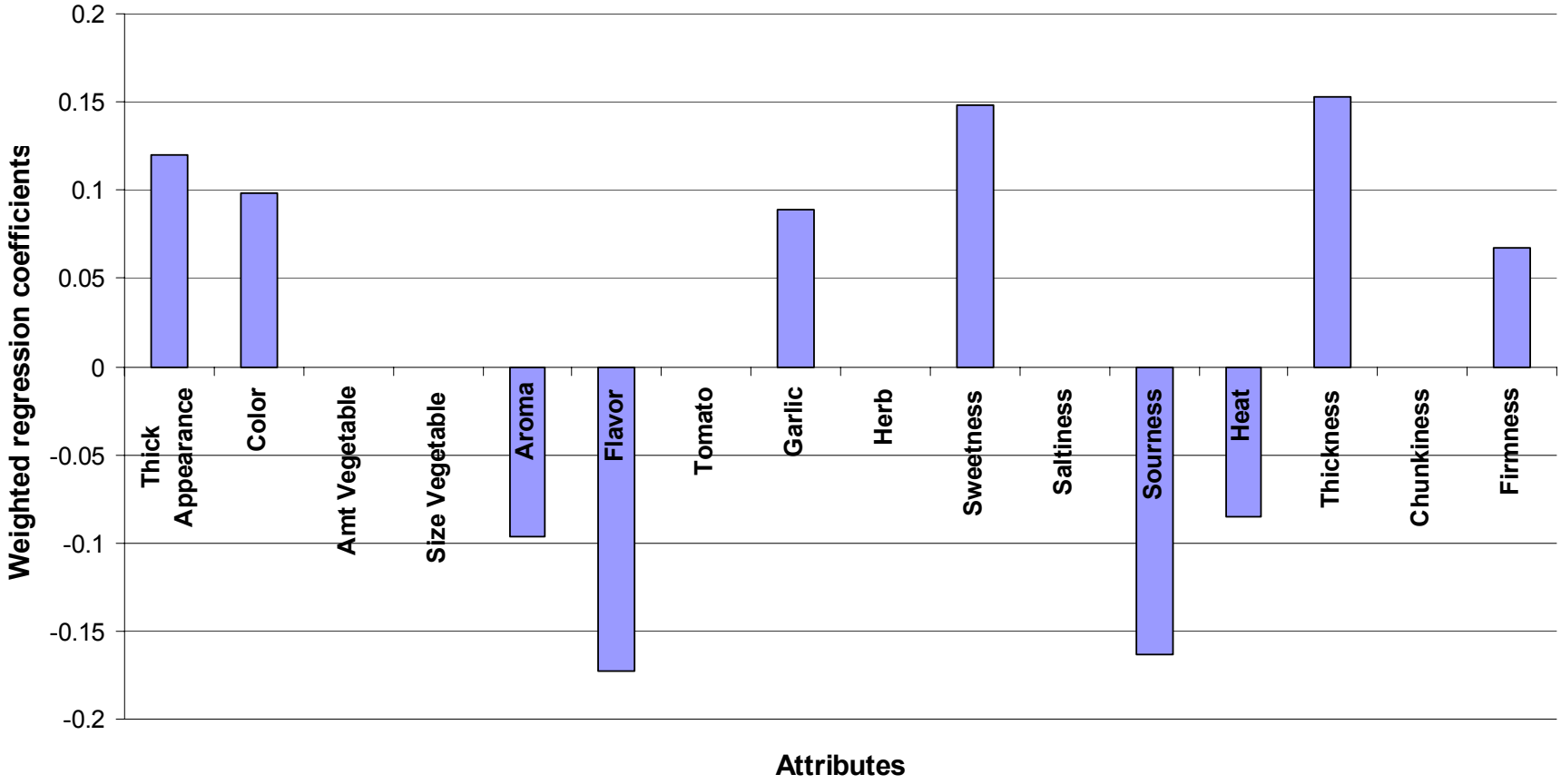
Results

- **PLS model (original JAR variables) for Product A**



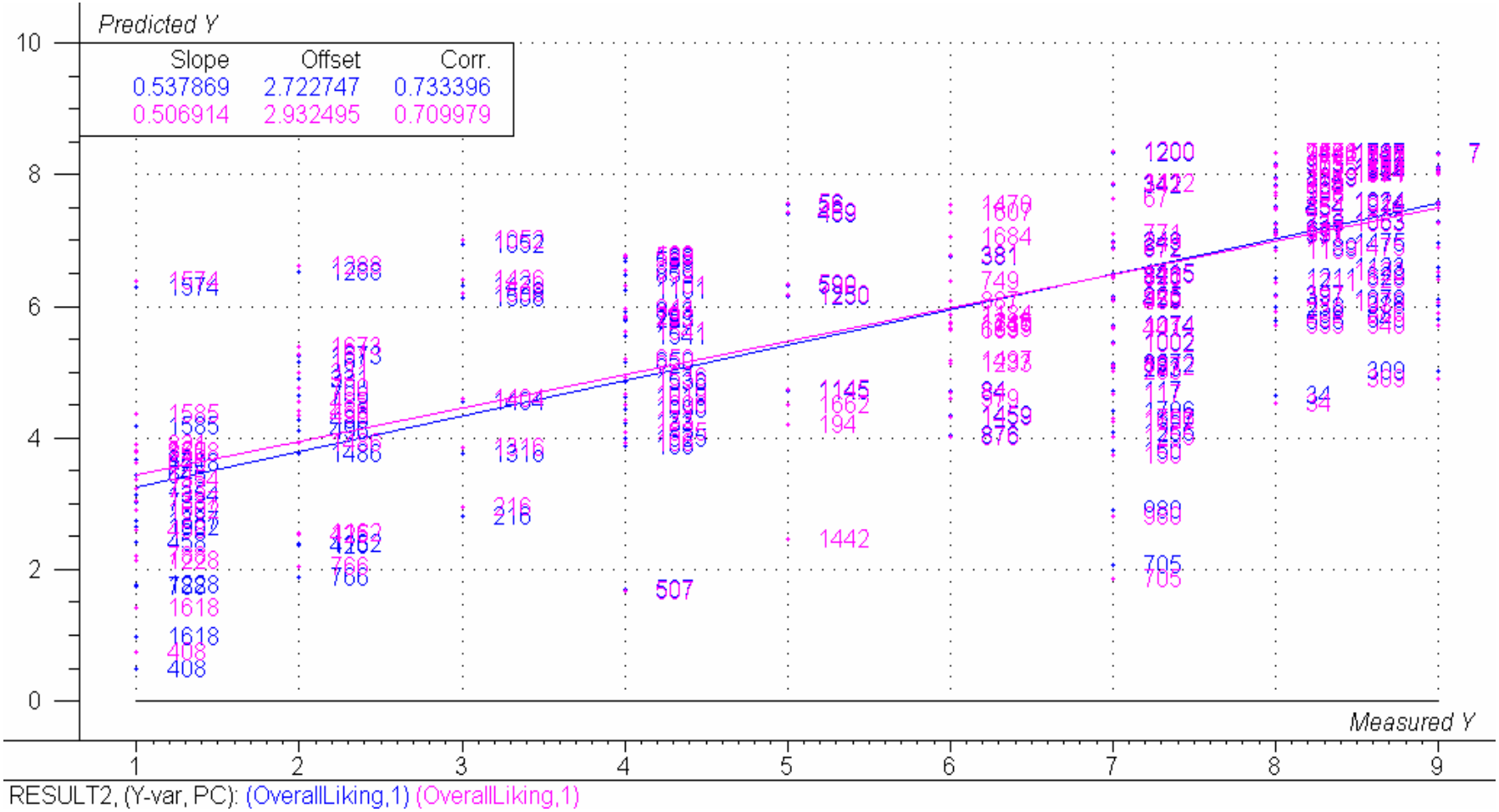
Results

- **Weighted regression coefficients of the PLS model (original JAR variables) for Product A**



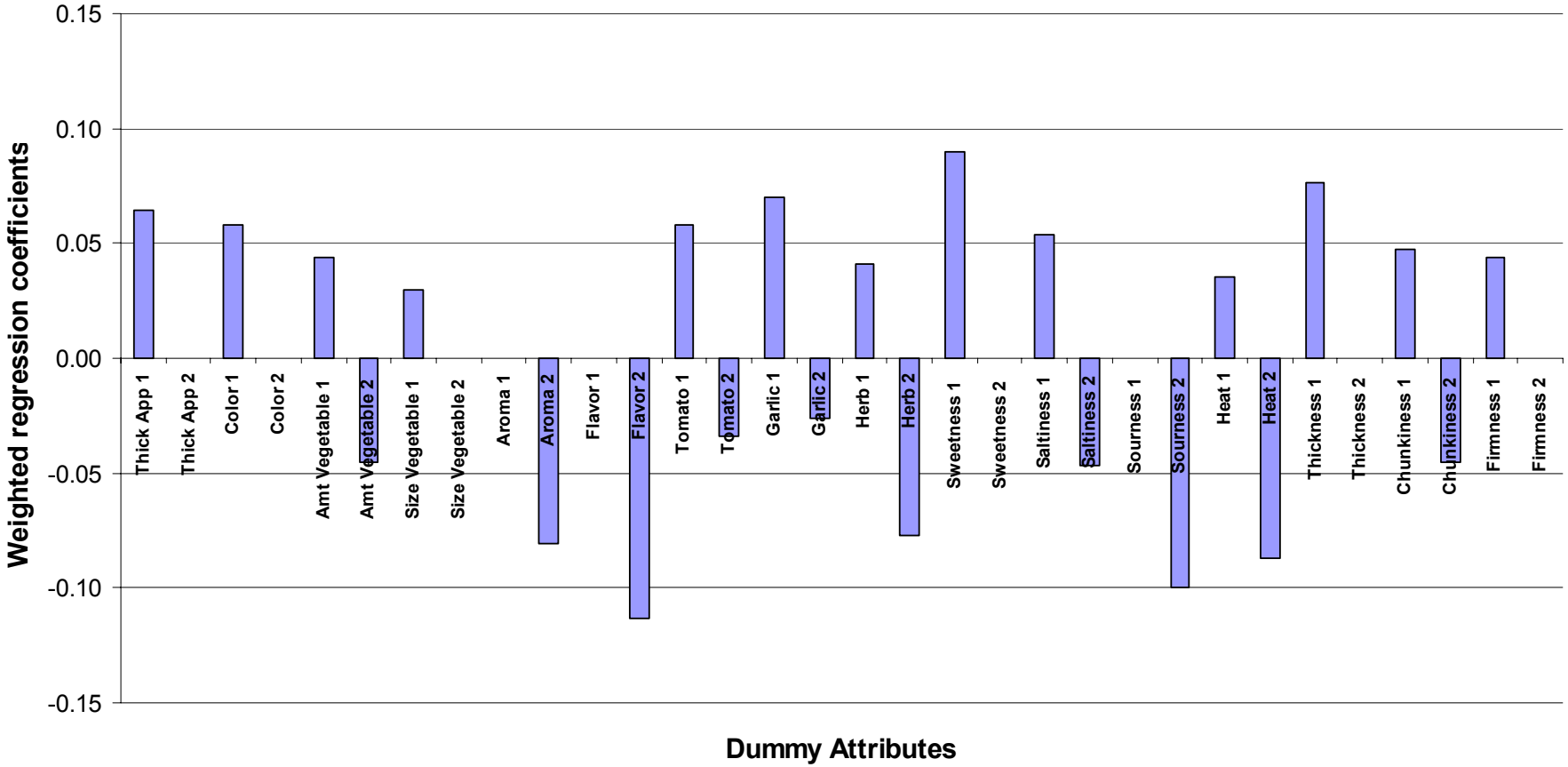
Results

- **PLS model (dummy variables) for Product A**



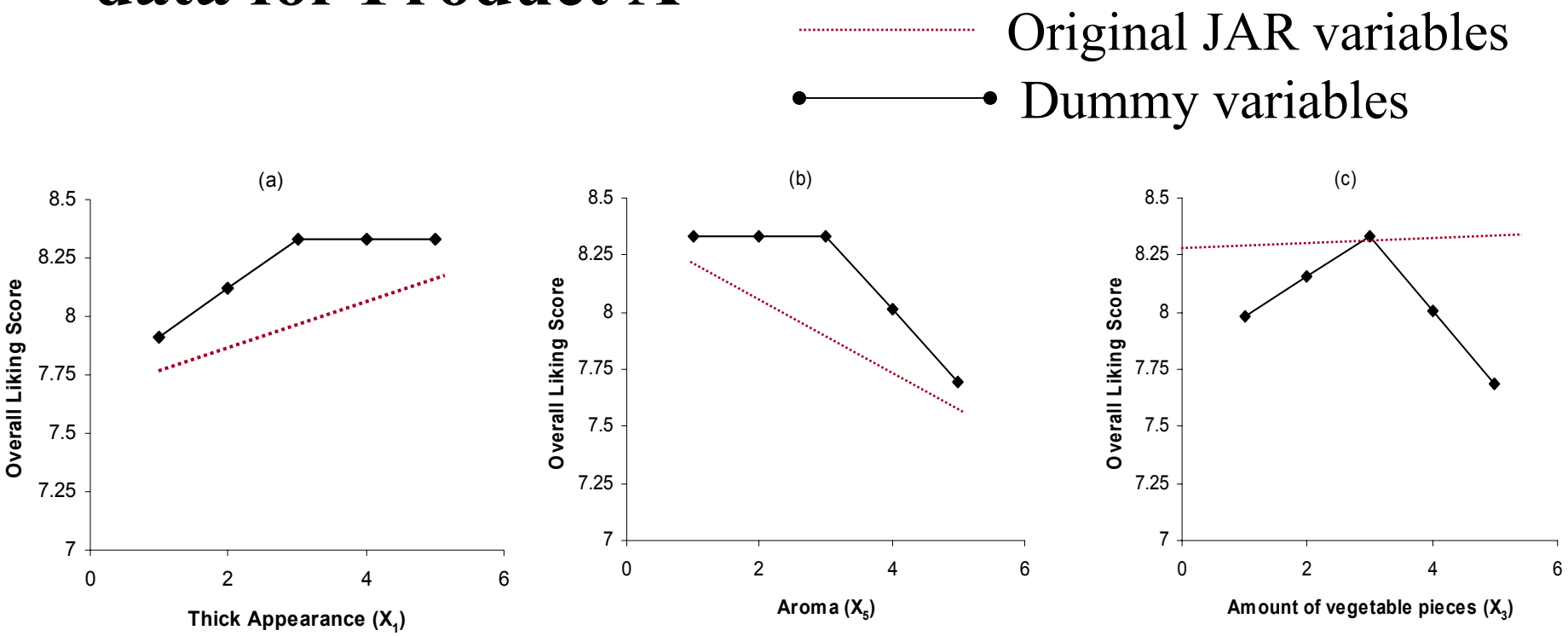
Results

- **Weighted regression coefficients of the PLS model (dummy variables) for Product A**



Results

- **Relationships between likings and JAR data for Product A**



Results

Plot of mean drop vs %consumers for Product A

$$\bar{Y} = b_0 + \left(b_{11}\bar{Z}_{11} + b_{12}\bar{Z}_{12}\right) + \left(b_{21}\bar{Z}_{21} + b_{22}\bar{Z}_{22}\right) + \ldots \left(b_{16,1}\bar{Z}_{16,1} + b_{16,2}\bar{Z}_{16,2}\right)$$

These attributes caused a mean drop of 1.12 out of the total of 2.44

Results

- **Predicted mean drops for 11 products due to JAR-scale attributes not being just-about-right**

	A	B	C	D	E	F	G	H	I	J	K
Intercept	8.33	8.29	7.79	8.28	8.14	8.39	8.16	8.33	8.37	8.24	8.38
Observed Mean	5.89	7.31	7.27	7.44	6.82	7.34	7.33	7.43	7.14	6.30	6.14
Predicted Mean	5.89	7.31	7.27	7.44	6.82	7.34	7.33	7.43	7.14	6.30	6.14
Mean Drop	2.44	0.98	0.52	0.84	1.32	1.05	0.83	0.90	1.23	1.94	2.24

Min

Max

Conclusions

- Dummy variables can be used with many regression models
- It provides penalty-analysis-like graphical presentation of relationships between JAR-scale and liking variables
- Unlike penalty analysis, it estimates the “true” mean drop of overall liking due to an attribute not being JAR
- Drawback is that you have to transform JAR variables into paired dummy variables

Acknowledgements

The authors would like to thank

Dr. Richard Popper

for allowing us to use the Pangborn workshop data set !