Factor Analysis

(요인분석/인자분석)

H. Park

HUFS

Factor Analysis

- 측정하기 어려운 "보이지 않는 변수" =
- (잠재)요인 =
- (e.g.) 애국심, 회사의 이미지, 업무태도 ...
- (e.g.) 장거리 주행능력, 단거리 주행능력, ...
- latent variable
- Charles Spearman (1904) in Psychology :

-학생들의 과목 성적들의 상관관계를 이용하여 "수리능력" 과 "언어능력" 이라는 2개의 인자를 발굴함-

- can be used in Marketing (product evaluation, client segmentation)

Developing Customer Profiles

The most obvious role for factor analysis is to analyze customers. Customer satisfaction or attitude and usage surveys typically barrage respondents with a large battery of attribute statements and buying practices questions, like "I prefer national brands" or "I frequently use coupons." Rather than trying to discern a pattern from all the responses, some of which are likely interrelated, you can execute a factor analysis to "bundle up" the individual variables into groups, clarifying relationships that describe underlying motivations for how and why customers gave their particular responses in your survey. For example, you may develop a "price-sensitive" factor versus a "quality-conscious" or "prestige shopping" factor to describe customer behavior.

Product Development

By reducing individual variables to underlying characteristics, factor analysis lets you prioritize which factors or issues to focus on as you develop new products, improve existing offerings, or expand product lines. You can then determine which characteristic needs the most development, rather than trying to highlight specific attributes from dozens of choices in a battery of questions.

FA 의 두 가지 종류

- (EFA) : Examine and explore the interdependence among the observed variables
- (CFA): Test specific hypotheses about the factorial structure of observed variables.

PCA vs. FA

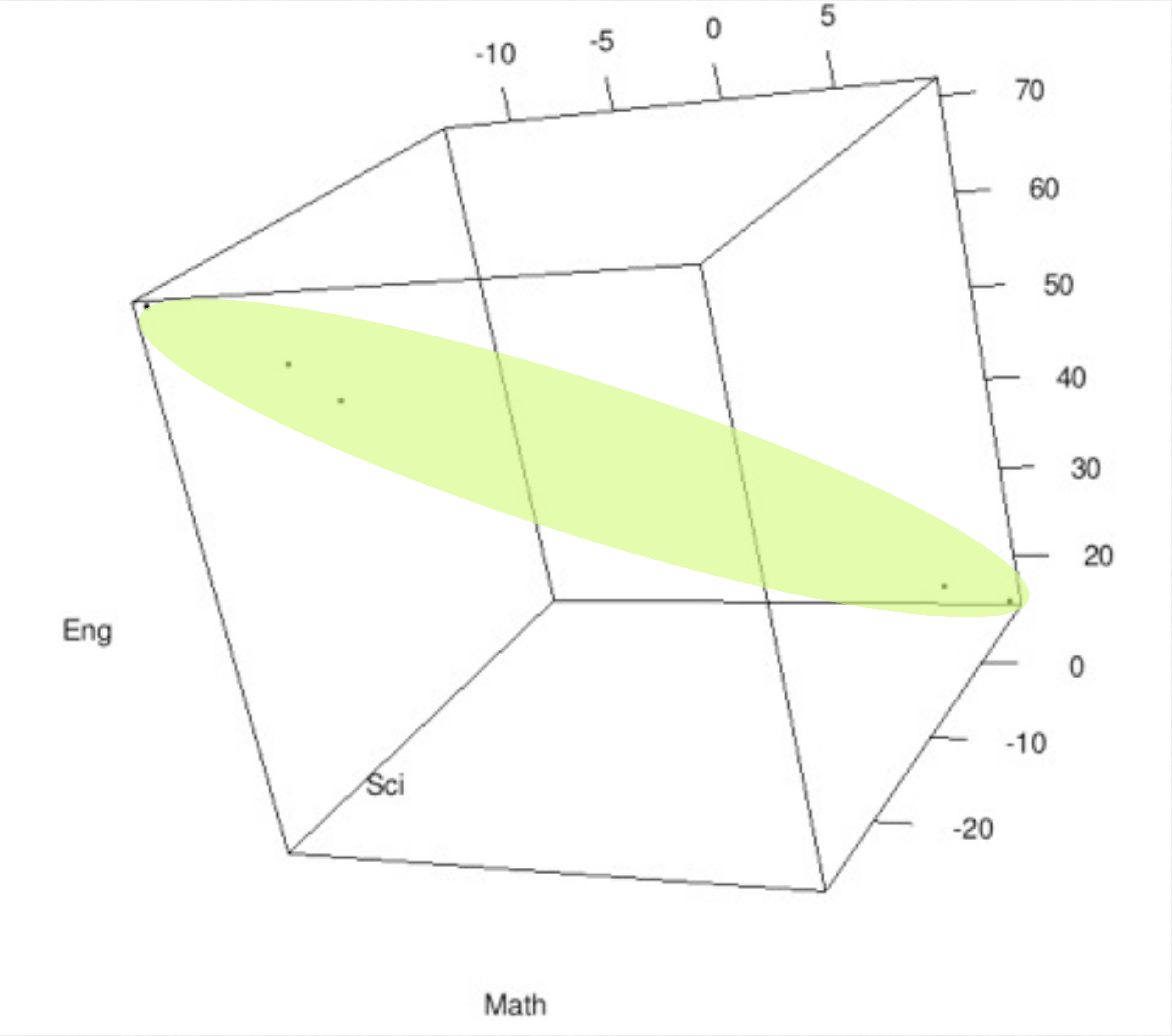
	PCA	FA
목적		
관심		
scale		
orthogonal		
구성		
score		
component 해석		

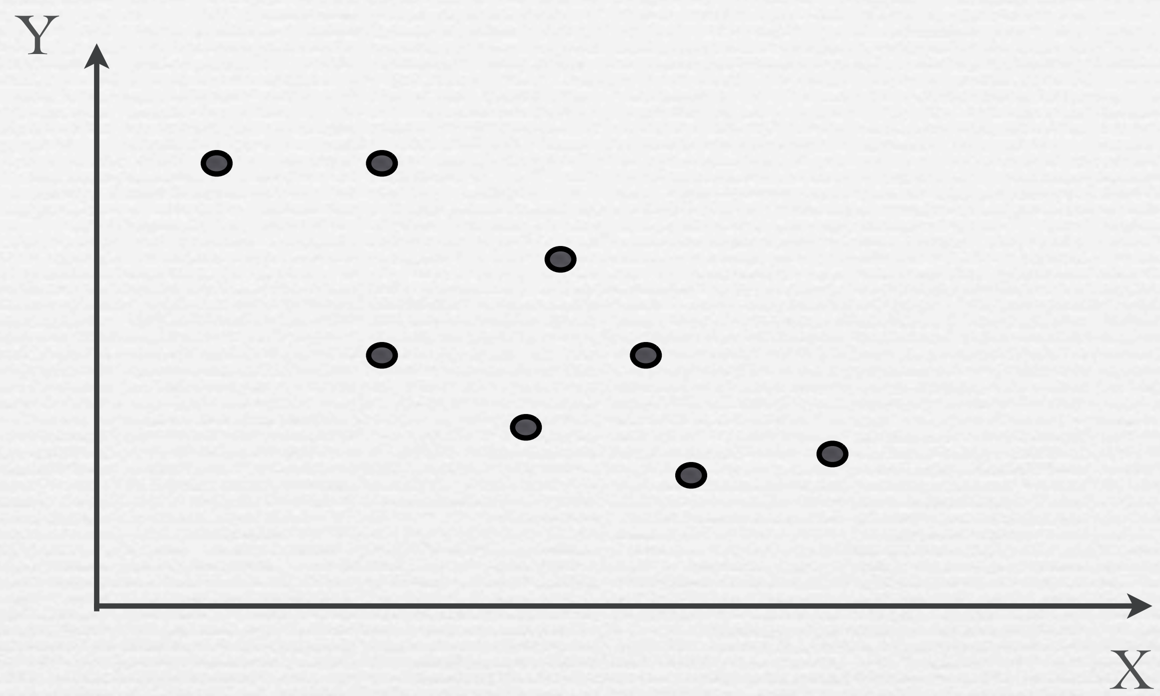
artificial FA data

Math(M) = -X + Y
English(E) = 5X + Y
Science(S) = -2X + Y

X=
Y=

x	y	M	E	S
14	1	-13	71	-27
12	2	-10	62	-22
...	...	...	...	...
11	2	-9	57	-20
1	10	9	15	8
2	9	7	19	5





Given X, Y >> we can get Math, Science, English scores

그러나, 현실에선...

Given Math, Science, English scores, >>

$$M = b_{11}X + b_{12}Y$$

$$E = b_{21}X + b_{22}Y$$

$$S = b_{31}X + b_{32}Y$$

M: 수학성적, P=물리성적, C=화학성적, E=영어성적, H: 역사성적, F=프랑스어성적

$\boxed{\text{Intelligence}}$

성적	$\boxed{M}$	$\boxed{P}$	$\boxed{C}$	$\boxed{E}$	$\boxed{H}$	$\boxed{F}$
적성	$\boxed{A_M}$	$\boxed{A_P}$	$\boxed{A_C}$	$\boxed{A_E}$	$\boxed{A_H}$	$\boxed{A_F}$

$$M = 0.8\,I + A_M$$

$$P = 0.7\,I + A_P$$

....

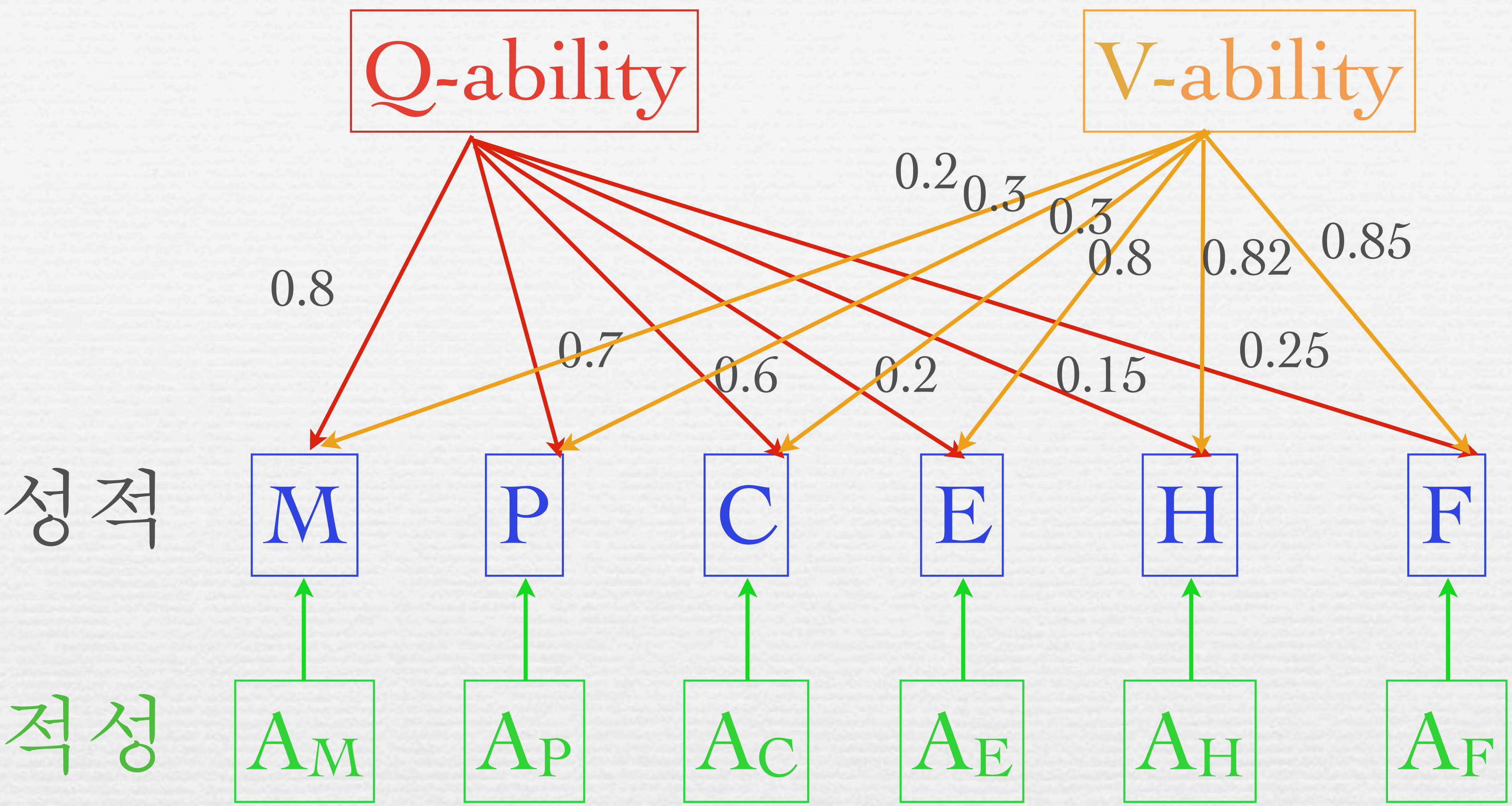
$$F = 0.3\,I + A_F$$

공통성 & 고유분산

M= 0.8 I + A_M $\longrightarrow$ $V(M) = 0.64 + V(A_M) = 1$

- 공통성 =
-

Two Factor Model



$$M = 0.8\ Q + 0.2\ V + A_M$$

$$P = 0.7\ Q + 0.3\ V + A_P$$

$$C = 0.6\ Q + 0.3\ V + A_C$$

$$E = 0.2\ Q + 0.8\ V + A_E$$

$$H = 0.15\ Q + 0.82\ V + A_H$$

$$F = 0.25\ Q + 0.85\ V + A_F$$

Factor 의 해석(공통성 이용)

	Communality			Unique Variance
	Q	V	Q+V	
M	0.64	0.04	0.68	
P	0.49	0.09	0.58	
C	0.36	0.09	0.45	
E	0.04	0.64	0.68	
H	0.023	0.672	0.695	
F	0.063	0.723	0.786	
total				

- Factor Q : 1.616 중에 (0.64+0.49+0.36) 이 92%

  >> Factor Q =

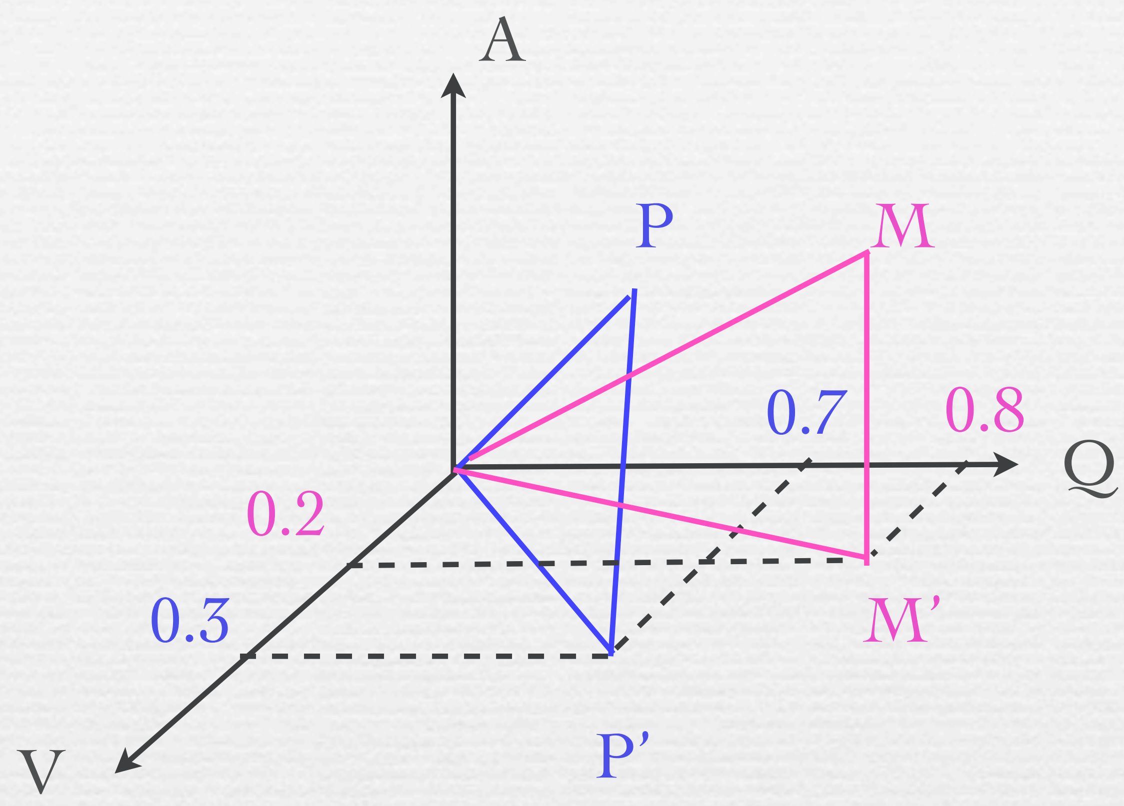
- Factor V : 2.255 중에 (0.64+0.672+0.723)이 90%

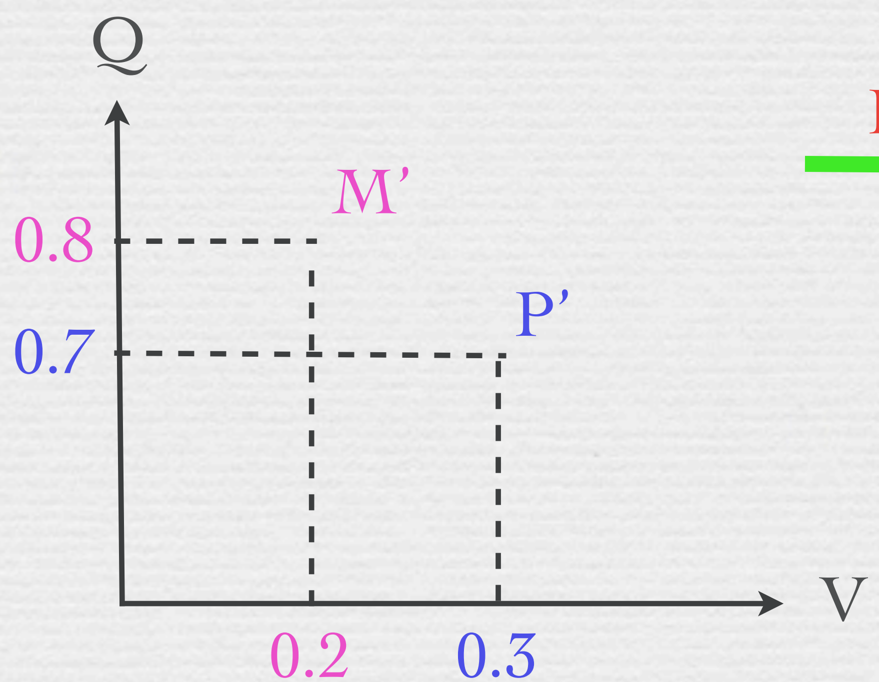
  >> Factor V =

Factor Rotation

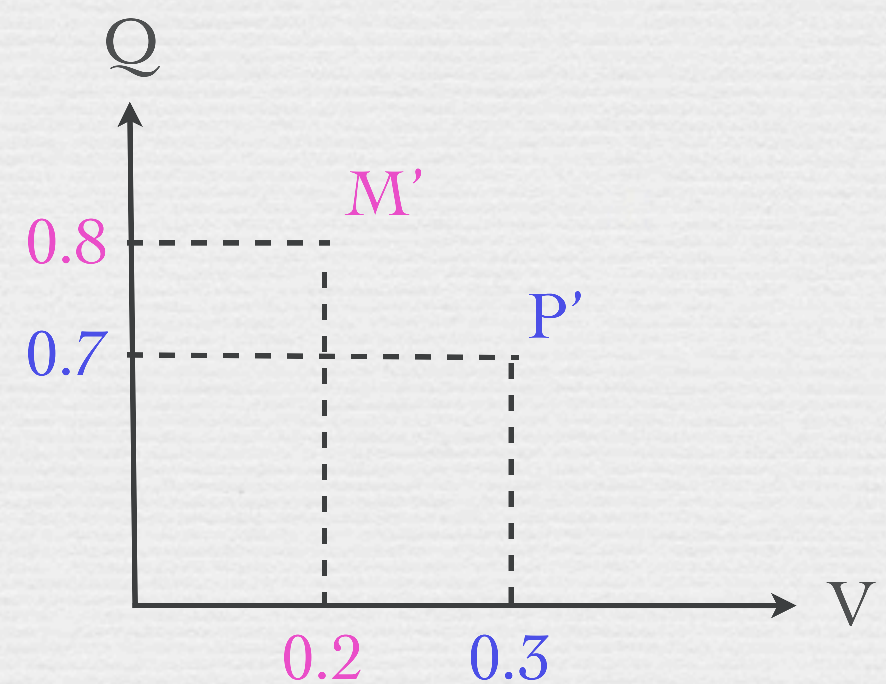
$M = 0.8\,Q + 0.2\,V + A_M$

$P = 0.7\,Q + 0.3\,V + A_P$





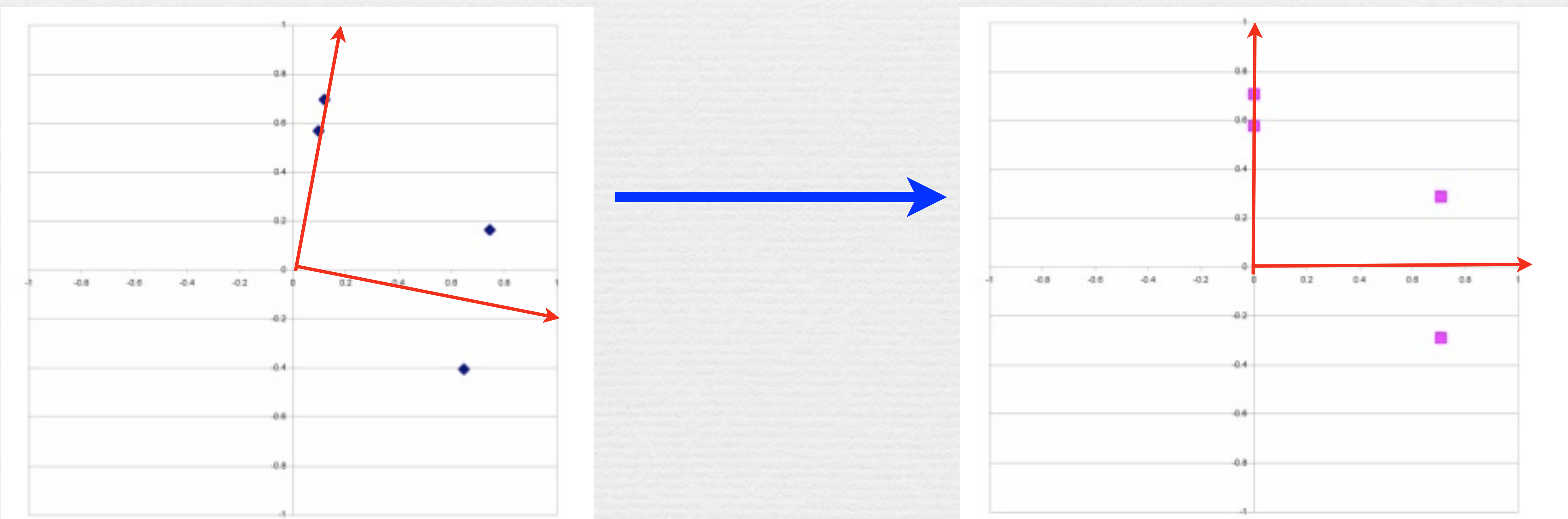
Rotation



V*

Varimax Technique

- Factor 를 회전시킬때 주어진 변수들이 한 개의 요인에 high factor loading 을 갖고 나머지 요인에 대해서는 0에 가까운 factor loading 을 갖도록 회전시키는 방법



How to get... factors ??

"Principal Components" : 대표적 요인추정법

1. 변수 M, P, C, E, H, F 에 대한 P.C. 를 구한다.

   $PC_1 = a_{11}\,M + a_{12}\,P + a_{13}\,C + a_{14}\,E + a_{15}\,H + a_{16}\,F$

   $PC_2 = a_{21}\,M + a_{22}\,P + a_{23}\,C + a_{24}\,E + a_{25}\,H + a_{26}\,F$

   $PC_3 = a_{31}\,M + a_{32}\,P + a_{33}\,C + a_{34}\,E + a_{35}\,H + a_{36}\,F$

   $PC_4 = a_{41}\,M + a_{42}\,P + a_{43}\,C + a_{44}\,E + a_{45}\,H + a_{46}\,F$

   $PC_5 = a_{51}\,M + a_{52}\,P + a_{53}\,C + a_{54}\,E + a_{55}\,H + a_{56}\,F$

   $PC_6 = a_{61}\,M + a_{62}\,P + a_{63}\,C + a_{64}\,E + a_{65}\,H + a_{66}\,F$

2. 식 6개 문자 6개 >> 연립방정식 M,P,C,E,H,F 를 PC1~PC6로 표현가능함.

We assume

therefore,

$$\boldsymbol{PC}_{p\times 1} = A_{p\times p}\,\boldsymbol{x}_{p\times 1} \longrightarrow \boldsymbol{x}_{p\times 1} =$$

However,

$$\boldsymbol{PC}_{p\times 1} = A_{p\times p}\,\boldsymbol{x}_{p\times 1} \longrightarrow ?$$

Factor Analysis (이론)

$$x_i = \mu + b_{i1} f_1 + b_{i2} f_2 + \cdots + b_{ik} f_k + \epsilon_i, \quad i = 1, 2, \cdots, p$$

We assume

$(A1)$

$(A2)$

$(A3)$

$(A4)$

PC를 이용한 FA

$$V(\boldsymbol{x}) = \Sigma_{p\times p} =$$

$$= \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_p \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0\cdots & 0 & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sqrt{\lambda_p} \end{pmatrix} \begin{pmatrix} \quad \end{pmatrix} \begin{pmatrix} \quad \end{pmatrix}$$

$$= \begin{pmatrix} \sqrt{\lambda_1}\boldsymbol{e}_1 & \sqrt{\lambda_1}\boldsymbol{e}_2 & \cdots & \sqrt{\lambda_1}\boldsymbol{e}_p \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1}\boldsymbol{e}_1^T \\ \sqrt{\lambda_1}\boldsymbol{e}_2^T \\ \vdots \\ \sqrt{\lambda_1}\boldsymbol{e}_p^T \end{pmatrix}$$

$$\approx \begin{pmatrix} \quad \end{pmatrix} \begin{pmatrix} \quad \end{pmatrix} + \begin{pmatrix} \quad \end{pmatrix}$$

$$=$$

$$\hat{\Sigma} = S \approx LL^T + \Psi$$

therefore,

$$\boldsymbol{x} = \boldsymbol{\mu} + L\boldsymbol{f} + \boldsymbol{\epsilon}$$

$$= \boldsymbol{\mu} + \qquad \left(\quad \right) + \boldsymbol{\epsilon}$$

MLE 를 이용한 FA

+

$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma),$

$l(\boldsymbol{\mu}, \Sigma) = \log L(\boldsymbol{\mu}, \Sigma) =$

로그우도함수를 최대로 하는

$\hat{L}$ =

In SAS,

proc factor ;

proc factor ;

PCA vs. Factor Analysis

- PCA : data 의 variation 을 잘 설명해 주는 변수들의 선형조합을 구하는 것이 목적

- FA : data 간의 상관관계 구조를 잠재변수(요인)를 사용하여 구명하는 것이 목적

In SAS

```
proc factor method=                                        ;

run;
```

	PCA	FA
목적	Data Reduction	Statistical Modeling
관심	Variance of the data	Correlation Pattern
scale	dependent (S or R)	free
orthogonal	always	maybe/maybe not
구성	L.C. of variables	L.C. of common parts of variables
score	PC score is calculated exactly	Factor score is estimated
component 해석	해석불가능도 OK	해석해야 의미있음

example : customer survey

- 새로운 상품의 고객 선호도 조사
- 선호정도를 1-7 까지 나타냄
- 질문1 : 맛 있습니까?
- 질문2 : 가격은 적절합니까?
- 질문3 : 풍미는 좋습니까?
- 질문4 : 스낵으로 좋습니까?
- 질문5: 에너지 식품으로 좋습니까?
- 5개의 질문을 수행하여 반응에 대한 상관행렬 계산

Is there a difference between flavor and taste?

Answered by Discovery Channel

Group

Discovery Channel

Taste and flavor are not the same thing, although they often are confused. Taste is one of your senses; it's picked up by receptor cells in your taste buds. Flavor is a combination of a few senses: Taste (gustatory), smell (olfactory), touch (tactile) and temperature (thermal) stimuli make up flavor. When spicy food is concerned, sometimes pain also is part of what makes up your flavor perception. Taste is a subjective sense; some people have more heightened palates than others and the same foods can taste differently to different people.

Attribute (Variable)		1	2	3	4	5
Taste	1	1	.02	**.96**	.42	.01
Good buy for money	2		1	.13	.71	**.85**
Flavor	3			1	.50	.11
Suitable for snack	4				1	**.79**
Provides lots of energy	5					1

- 아래 단계를 따라가며, 5개 질문 응답의 결과에 숨어있는 요인을 구해 보시오.

(1) 5개 변수로 구성되는 5개의 PC를 구한다

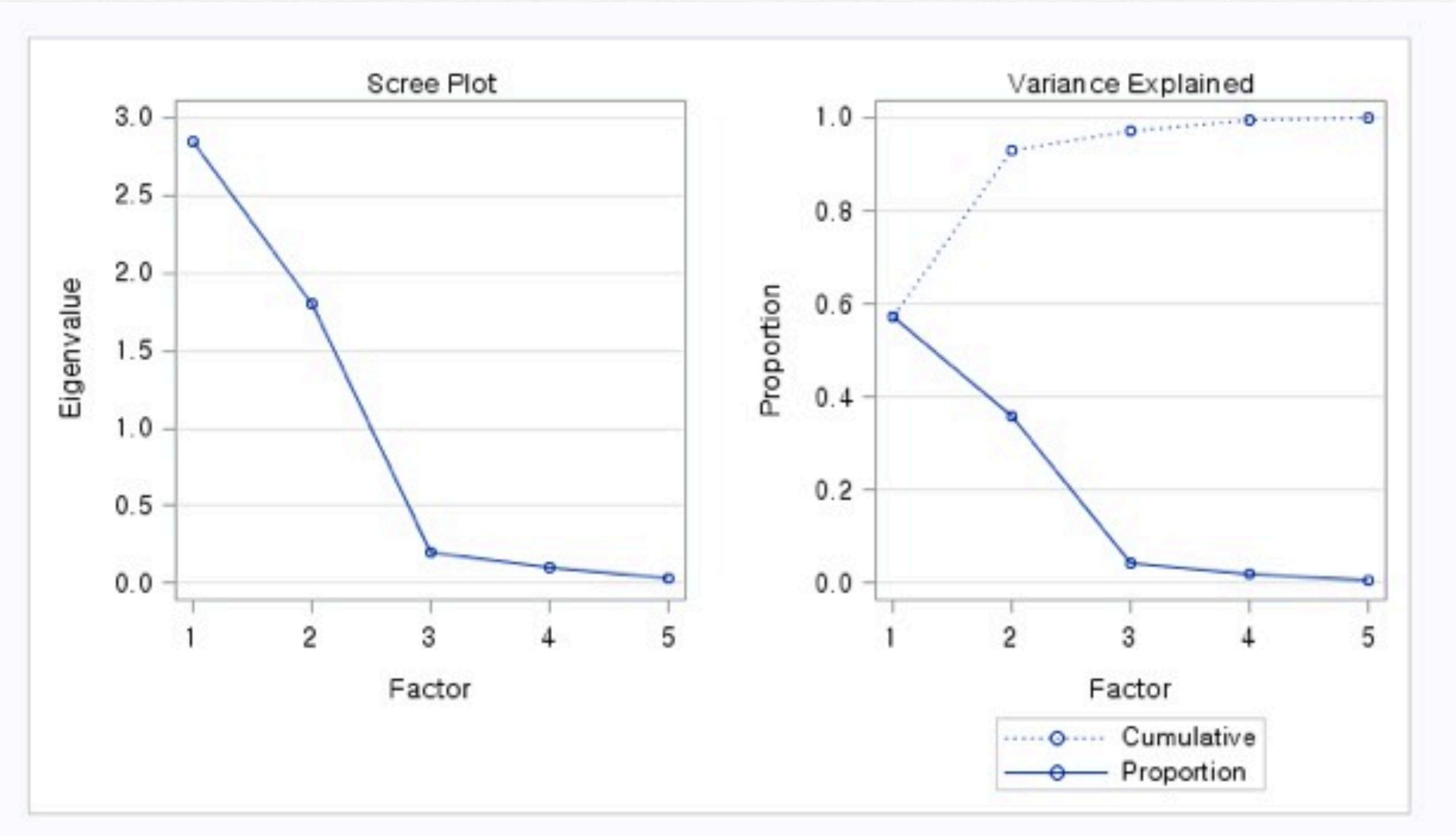
(2) 적절한 PC 개수를 결정한다

(3) 각 변수를 PC 들의 선형조합으로 표시한다

(4) Factor Loading 을 보고 요인의 이름을 결정한다

SAS code

```
data a(type=corr); _type_ = ‘corr’;
input _name_ $ taste worth flavor gsnack penergy;
cards;
taste 1.00 0.02 0.96 0.42 0.01
worth 0.02 1.00 0.13 0.71 0.85
flavor 0.96 0.13 1.00 0.50 0.11
gsnack 0.42 0.71 0.50 1.00 0.79
penergy 0.01 0.85 0.11 0.79 1.00
;
proc factor data=a                run;
```



The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 5 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	2.85309042	1.04675797	0.5706	0.5706
2	1.80633245	1.60184223	0.3613	0.9319
3	0.20449022	0.10208076	0.0409	0.9728
4	0.10240947	0.06873203	0.0205	0.9933
5	0.03367744		0.0067	1.0000

proc factor 의 default 옵션은 1보다 큰 고유값까지 유지

2 factors will be retained by the NFACTOR criterion.

Eigenvectors		
	1	2
taste	0.33145	0.60722
worth	0.46016	-0.39003
flavor	0.38206	0.55651
gsnack	0.55598	-0.07806
penergy	0.47256	-0.40419

PC1 = 0.331*taste + 0.460*worth+0.382*flavor + 0.555*gsnack+0.472*penergy

=

PC2 = 0.607*taste -0.390*worth+0.556*flavor -0.07*gsnack-0.404*penergy

=

Factor Pattern		
	Factor1	Factor2
taste	0.55986	0.81610
worth	0.77726	-0.52420
flavor	0.64534	0.74795
gsnack	0.93911	-0.10492
penergy	0.79821	-0.54323

Variance Explained by Each Factor	
Factor1	Factor2
2.8530904	1.8063325

Final Communality Estimates: Total = 4.659423				
taste	worth	flavor	gsnack	penergy
0.97946135	0.87892002	0.97588288	0.89292750	0.93223112

taste = 0.55986*F1 + 0.81610*F2

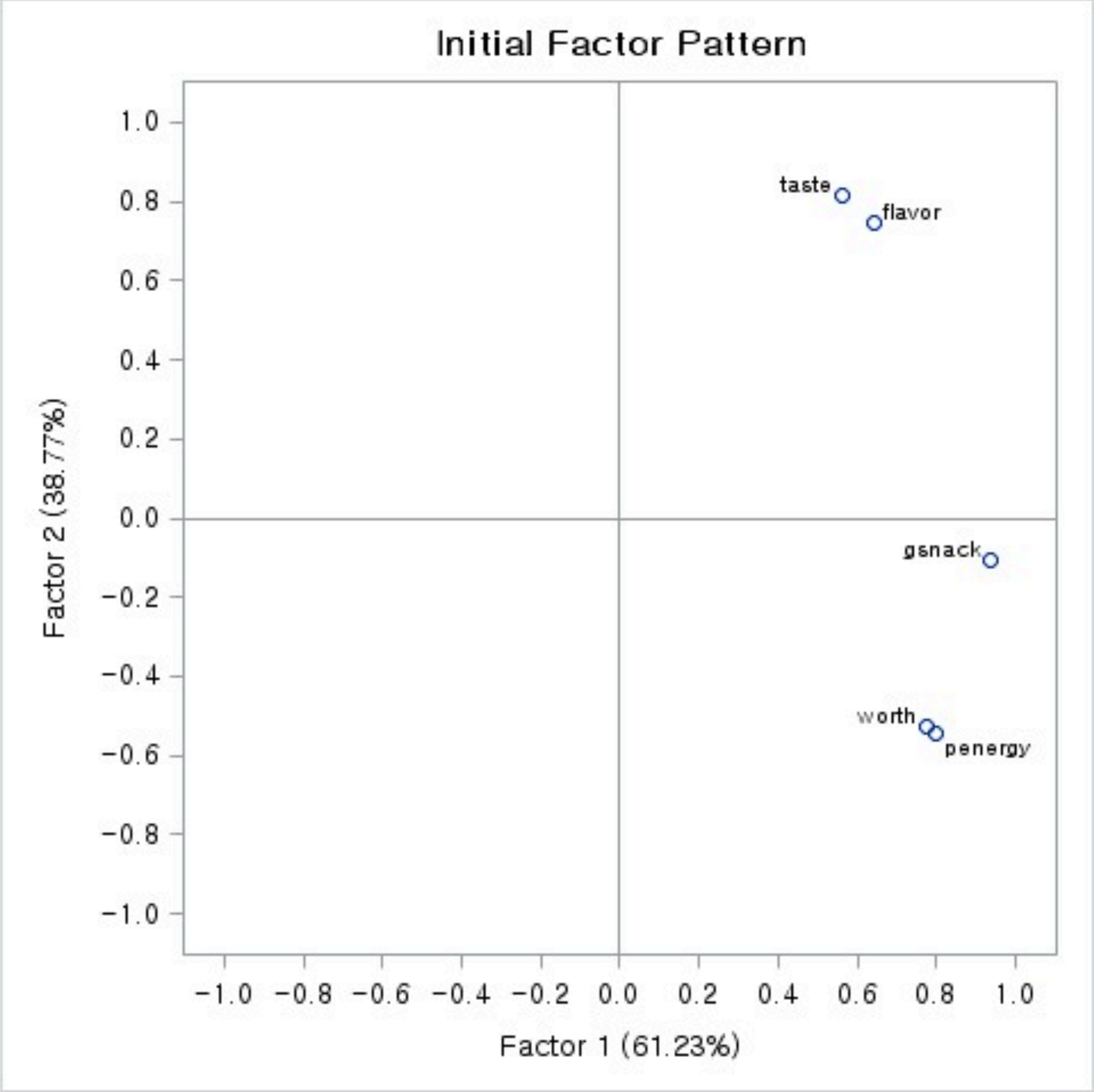
worth = 0.77726*F1 -0.52420*F2

flavor = 0.64534*F1 +0.74795*F2

gsnack = 0.93911*F1 -0.10492*F2

penergy =0.79821*F1 -0.54323*F2

The FACTOR Procedure
Initial Factor Method: Principal Components



The FACTOR Procedure
Rotation Method: Varimax

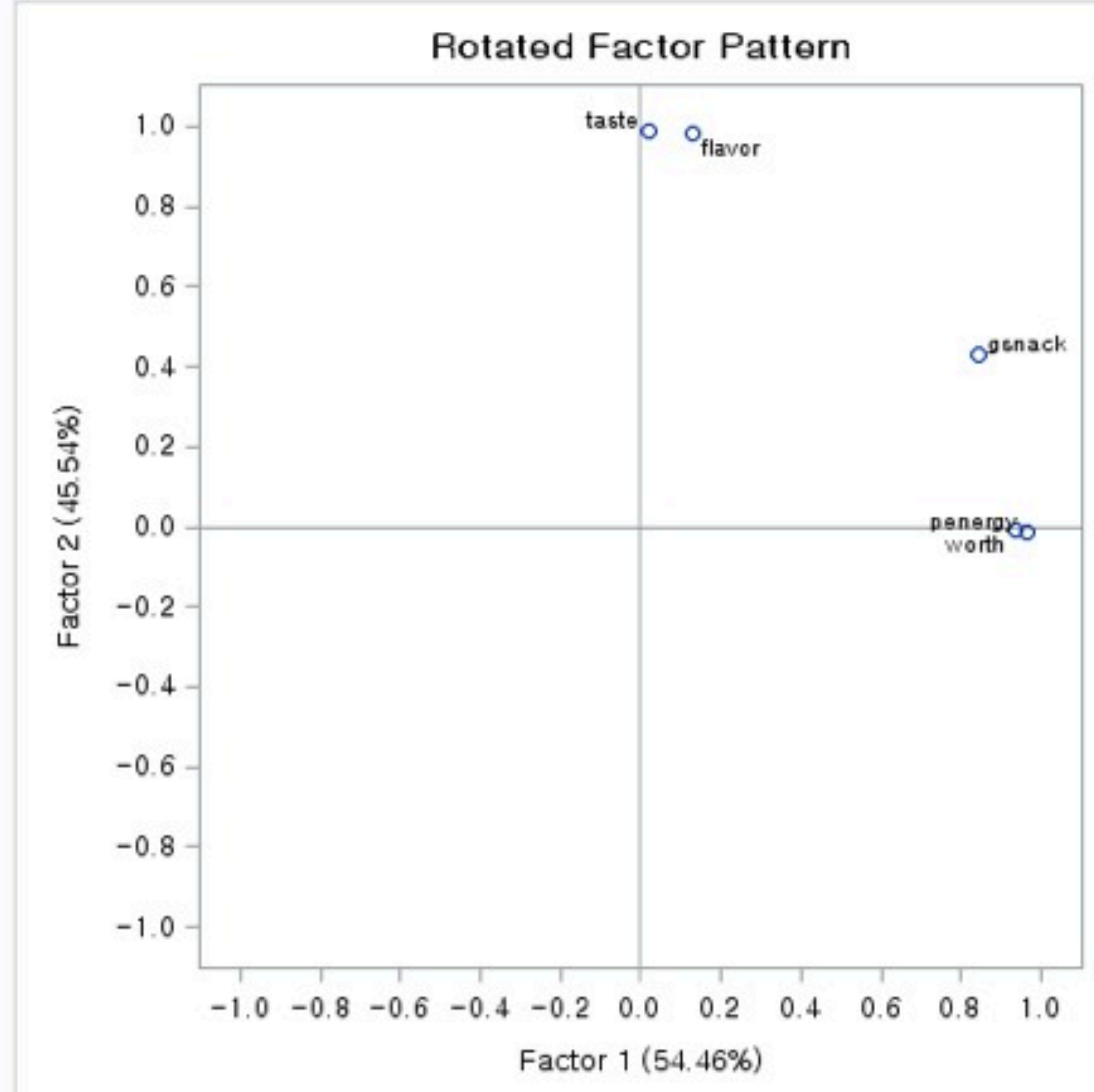
Orthogonal Transformation Matrix		
	1	2
1	0.83571	0.54917
2	-0.54917	0.83571

Rotated Factor Pattern		
	Factor1	Factor2
taste	0.01970	0.98948
worth	0.93744	-0.01123
flavor	0.12856	0.97947
gsnack	0.84244	0.42805
penergy	0.96539	-0.01563

Variance Explained by Each Factor	
Factor1	Factor2
2.5373960	2.1220269

Final Communality Estimates: Total = 4.659423				
taste	worth	flavor	gsnack	penergy
0.97946135	0.87892002	0.97588288	0.89292750	0.93223112

The FACTOR Procedure
Rotation Method: Varimax



taste = 0.019*F1’ + 0.989*F2’

worth = 0.937*F1’ - 0.01*F2’

flavor = 0.128*F1’ + 0.979*F2’

gsnack = 0.842*F1’ + 0.428*F2’

genergy = 0.965*F1’ - 0.015*F2’