Deqiang Zou | School of Management | Fudan University

Market Segmentation with Latent Class Regression

Applications of the package “FlexMix”

November 14, 2010 @ Shanghai University of Finance and Economics

Segment and Segmentation

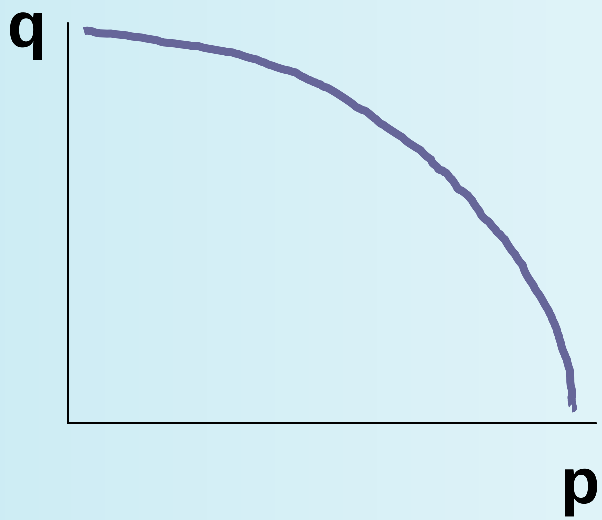
- A **<u>segment</u>** is a group of end-users that share a unique set of wants/needs and/or purchase behaviors
- **<u>Segmentation</u>** is the process that companies use to divide large heterogeneous markets into small markets that can be reached more efficiently and effectively with products and services that match their unique needs

Segmentation Bases

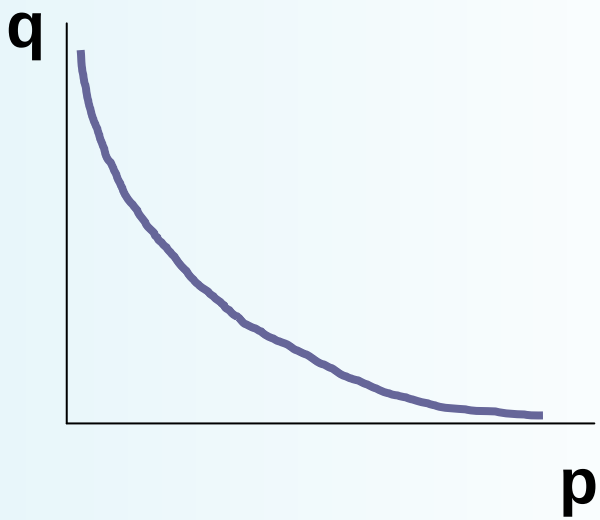
	General	Product-specific
Observable	Cultural, geographic, demographic and socio-economic variables	User status, usage frequency, store loyalty and patronage, situations
Unobservable	Psychographics, values, personality and life-style	Psychographics, benefits, perceptions, elasticities, attributes, preferences, intention

Wedel and Kamakura (2000), p. 7

Pricing Segments

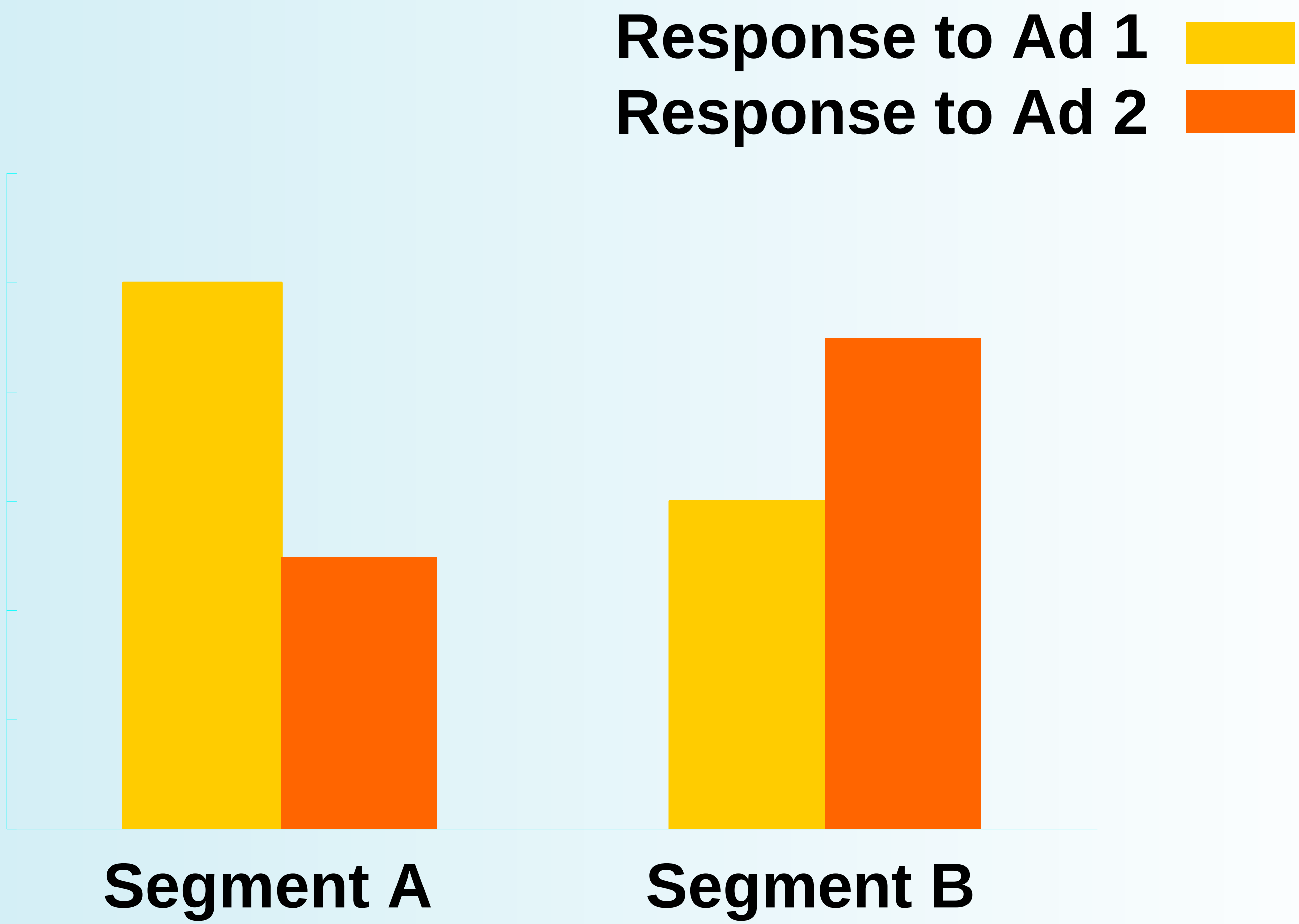


Brand Loyals



Deal Prone

Advertising Segments



Segmentation Methods

	a priori	Post hoc
Descriptive	Contingency tables, Log-linear models	Clustering methods: Nonoverlapping, overlapping, Fuzzy techniques, ANN, mixture models
Predictive	Cross-tabulation, Regression, logit and Discriminant analysis	AID, CART, Clusterwise regression, ANN, mixture models

Wedel and Kamakura (2000), p. 17

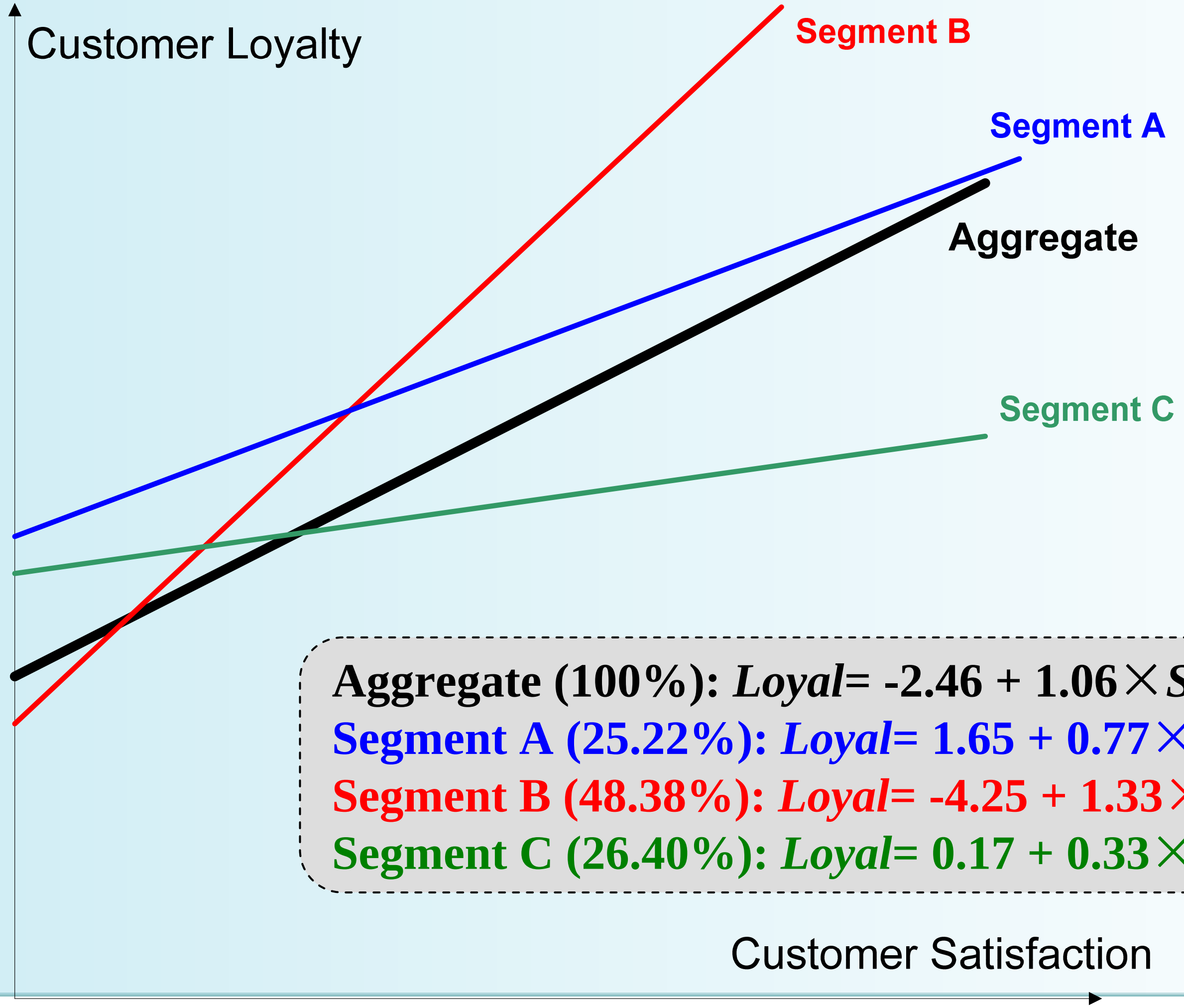
Segmentation: Distance?

- Segmentation is the process of clustering consumers on basis of distances between them?

Income \ Education	Low	High
Low	Segment 1	Segment 2
High	Segment 3	Segment 4

Repurchase \ Satisfaction	Low	High
Low	Segment 1	Segment 2
High	Segment 3	Segment 4

Segmentation with Cause-Effect Relation



Aggregate (100%): ***Loyal*** **= -2.46 + 1.06 ×** ***Satisfaction***

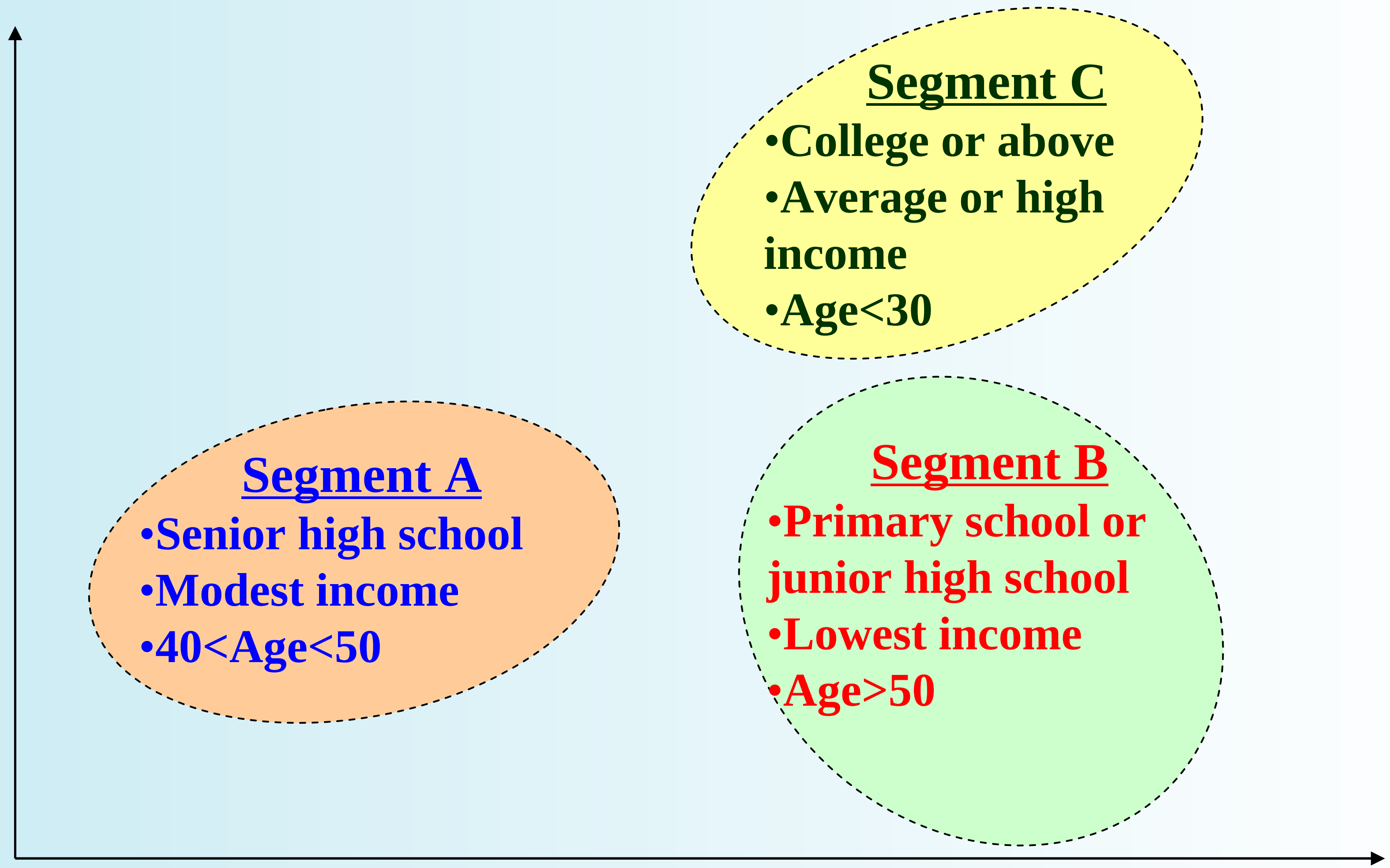
Segment A (25.22%): ***Loyal*** **= 1.65 + 0.77 ×** ***Satisfaction***

Segment B (48.38%): ***Loyal*** **= -4.25 + 1.33 ×** ***Satisfaction***

Segment C (26.40%): ***Loyal*** **= 0.17 + 0.33 ×** ***Satisfaction***

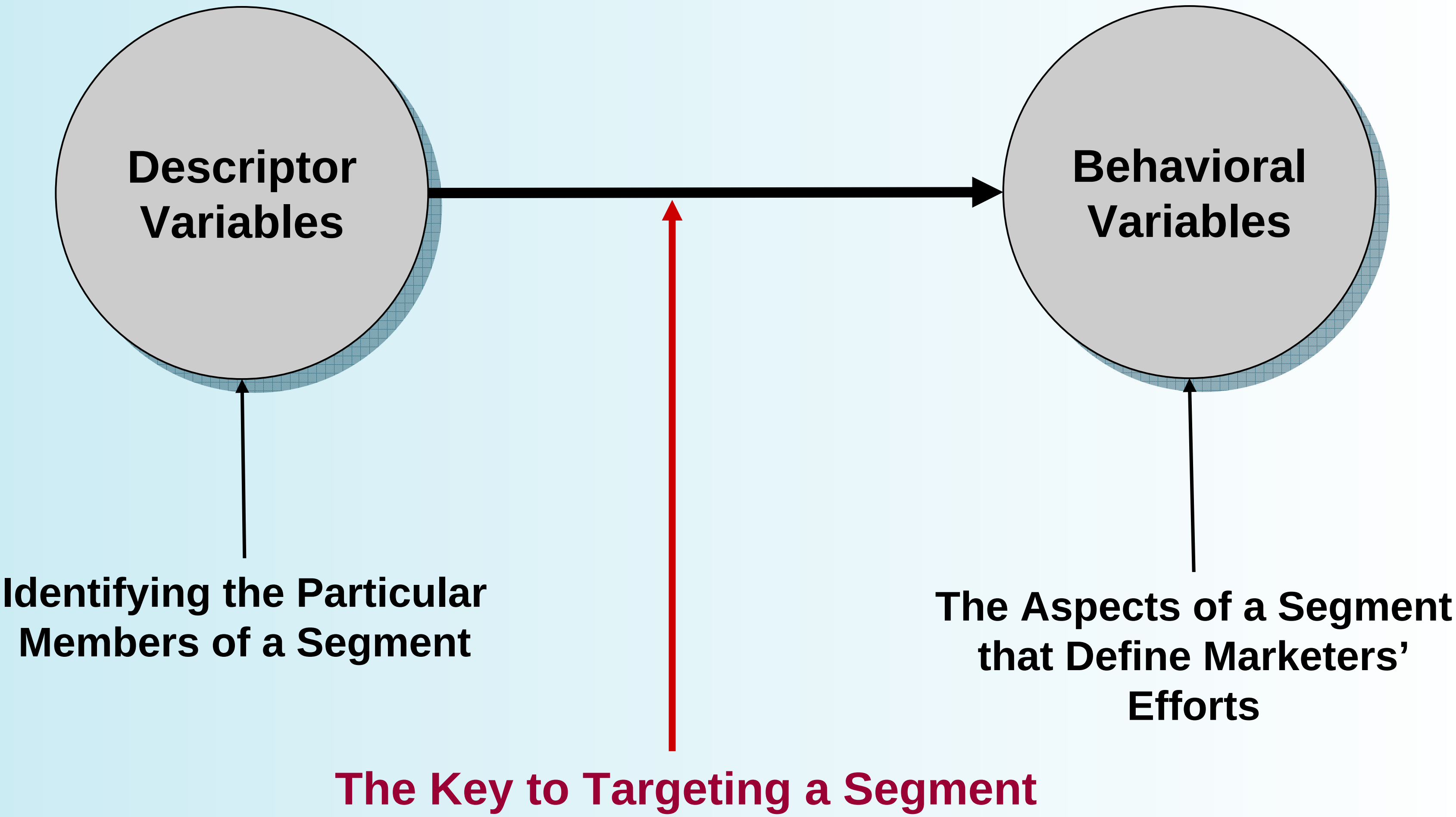
王霞, 赵平, 王高, 刘佳 (2005)

Multiple Correspondence Analysis



王霞, 赵平, 王高, 刘佳 (2005)

Relationship for Segment Targeting



Latent Class Regression

- Clusterwise / (finite) mixture regression
  - Consider finite mixture models with K components of form

$$h(y \mid x, \psi) = \sum_{k=1}^{K} \pi_k f(y \mid x, \theta_k) \qquad (1)$$

  - $\pi_k \geq 0, \ \sum_{k=1}^{K} \pi_k = 1$
  - where y is a (possibly multivariate) dependent variable with conditional density h, x is a vector of independent variables, π_k is the prior probability of component k, θ_k is the component specific parameter vector for the density function f, and $\psi = (\pi_1, \ldots, \pi_K, \theta'_1, \ldots, \theta'_K)'$ is the vector of all parameters

Latent Class Regression

- If f is a univariate normal density with component-specific mean $\beta'_k x$ and variance σ_k^2, we have $\theta_k = (\beta'_k, \sigma_k^2)$ and Equation (1) describes a mixture of standard linear regression models
- If f is a member of the exponential family, we get a mixture of generalized linear models

Posterior Probability

- The posterior probability that observation (x, y) belongs to class j is given by

$$P(j \mid x, y, \psi) = \frac{\pi_i f(y \mid x, \theta_j)}{\sum_k \pi_k f(y \mid x, \theta_k)}$$

- The posterior probabilities can be used to segment data by assigning each observation to the class with maximum posterior probability
- Individual-level predictions of finite mixture models are a weighted combination of the segment-level regression functions, weighted with the posterior membership probabilities (DeSarbo, Kamakura, and Wedel 2006)

Parameter Estimation

- The log-likelihood of a sample of N observations $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ is given by

$$\log L = \sum_{n=1}^{N} \log h(y_n | x_n, \psi) = \sum_{k=1}^{K} \log(\sum_{k=1}^{K} \pi_k f(y_n | x_n, \theta_k))$$

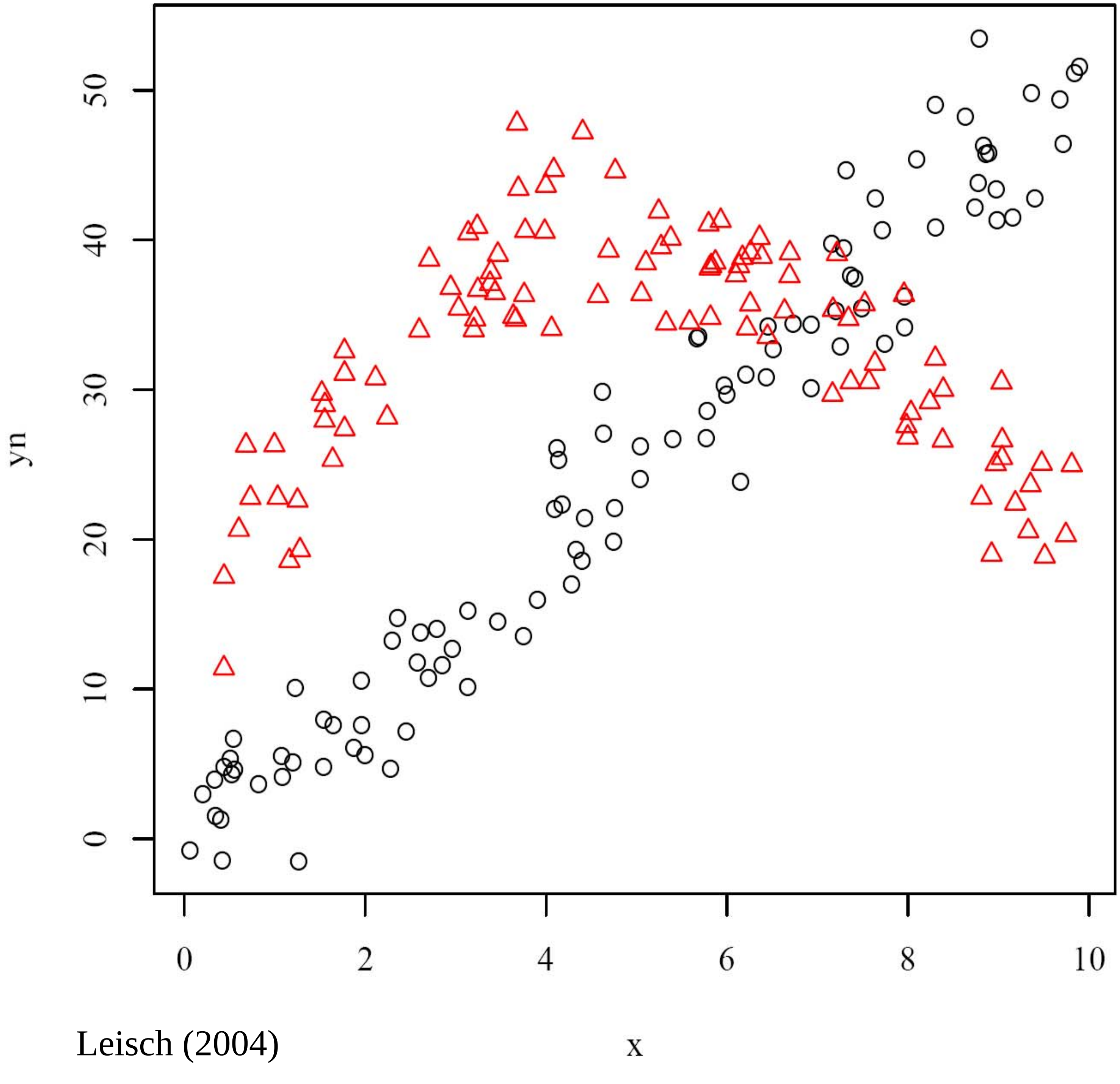
- The most popular method for maximum likelihood estimation of the parameter vector ψ is the iterative EM algorithm (Leisch 2004)

Using FlexMix

- As a simple example we use artificial data with two latent classes of size 100 each:
  - Class 1: $y = 5x + \varepsilon$
  - Class 2: $y = 15 + 10x - x^2 + \varepsilon$
  - with $\varepsilon \sim N(0, 9)$ and prior class probabilities $\pi_1 = \pi_2 = 0.5$
- We can fit this model in R using the commands

```
> library(flexmix)
> data(NPreg)
> m1 = flexmix(yn ~ x + I(x^2), data = NPreg, k =
2)
> m1
```

Leisch (2004)



Leisch (2004)

```
Call:
flexmix(formula = yn ~ x + I(x^2), data = NPreg,
k = 2)
Cluster sizes:
  1   2
100 100
convergence after 15 iterations
> parameters(m1, component = 1)
$coef
(Intercept)           x      I(x^2)
-0.20989331 4.81782414 0.03615728
$sigma
[1] 3.47636
> parameters(m1, component = 2)
$coef
(Intercept)         x      I(x^2)
14.7168295 9.8466698 -0.9683534
$sigma
[1] 3.479809
```

Leisch (2004)

Using FlexMix

```
> summary(m1)
Call:
flexmix(formula = yn ~ x + I(x^2), data = NPreg,
k = 2)
        prior size post>0 ratio
Comp.1  0.494 100  145    0.690
Comp.2  0.506 100  141    0.709
`log Lik.' -642.5453 (df=9)
AIC: 1303.091 BIC: 1332.775
> table(NPreg$class, m1@cluster)
   1 2
1 95 5
2 5 95
```

Leisch (2004)

Significance Test

```
> rm1 = refit(m1)
> summary(rm1)
Call:
refit(m1)
Component 1 :
            Estimate  Std. Error t value Pr(>|t|)
(Intercept) -0.208996 0.673900   -0.3101  0.7568
x           4.817015  0.327447   14.7108  <2e-16
I(x^2)      0.036233  0.032545   1.1133   0.2669
-------------
Component 2 :
            Estimate  Std. Error t value  Pr(>|t|)
(Intercept) 14.717541 0.890843   16.521   < 2.2e-16
x           9.846148  0.390385   25.222   < 2.2e-16
I(x^2)      -0.968304 0.036951   -26.205  < 2.2e-16
```

Leisch (2004)

Automated Model Search

- In real applications the number of components is unknown and has to be estimated
- Fit models with an increasing number of components and compare them using AIC or BIC

```
> m7 = stepFlexmix(yp ~ x + I(x^2), data = NPreg,
control = list(verbose = 0), K = 1:5, nrep = 5)
> sapply(m7, BIC)
       1        2        3        4        5
946.7477 925.9972 942.1553 960.0626 960.9347
```

- Choose the number of components minimizing the BIC

Leisch (2004)

Finite Mixtures with Concomitant Variables

- If the weights depend on further variables, these are referred to as concomitant variables
- The model class is given by

$$h(y \mid x, \omega, \psi) = \sum_{k=1}^{K} \pi_k(\omega, \alpha) f_k(y \mid x, \theta_k)$$

  - Where *w* denotes the concomitant variables, α are the parameters of the concomitant variable model

$$\sum_{k=1}^{K} \pi_k(\omega, \alpha) = 1 \qquad \pi_k(\omega, \alpha) > 0, \forall k$$

Grün and Leisch (2008)

Segmenting Newspaper Readers

变量类型	变量名称	细分市场1		细分市场2		细分市场3	
		系数	T值	系数	T值	系数	T值
感知变量	截距	5.022*	5.454	0.621*	2.136	0.355	1.004
	新闻栏目评价	-0.268	-1.5544	0.125*	2.051	0.418*	6.280
	经济栏目评价	0.059	0.46541	0.074	1.471	0.052	1.117
	娱乐栏目评价	0.069	0.468	0.074	1.542	0.040	0.717
	北京栏目评价	0.270	1.845	-0.032	-0.703	0.137*	2.580
	版面设计评价	0.360*	2.071	0.060	0.913	0.161*	2.556
	印刷质量评价	-0.025	-0.178	0.002	0.043	0.188*	4.022
	广告评价	0.153	1.385	0.077*	1.968	0.071	1.694
	购买便利性评价	-0.402*	-3.103	0.198*	4.978	-0.109*	-2.842
	感知价格	0.119	1.337	0.280*	9.170	-0.032	-0.991
	样本量	119		495		290	
	市场份额（%）	13.16		54.76		32.08	

Note: The dependent variable is “Customer Satisfaction” (N= 904), *: p<0.05

王燕, 赵平 (2009)

变量类型	变量名称	细分市场1		细分市场2		细分市场3	
		系数	T值	系数	T值	系数	T值
个人特征变量	截距			0.081	0.075	-0.578	-0.503
	阅读频率[a]						
	每天阅读			-0.069	-0.162	0.188	0.378
	每次读报用时[b]						
	半小时以下			0.644	1.572	1.327*	2.658
	阅读地点[c]						
	家中			0.492	0.705	-0.966	-1.451
	上班			0.125	0.193	-0.001	-0.002
	性别[d]						
	男			1.138*	2.796	0.643	1.414
	教育程度[e]						
	高中及以下			1.016*	2.091	1.319*	2.638
	年龄[f]						
	25岁以下			0.163	0.292	-0.772	-1.228
	25-35岁			-0.235	-0.344	0.542	0.899
	家庭月收入[g]						
	2000-4000元			-1.689*	-3.055	-0.746	-1.291
	4000元以上			0.725	1.036	0.948	1.185

Recap

- The underlying basis of customer heterogeneity (i.e., discrete market segments) is unknown *a priori*
- The objective is to *simultaneously* estimate the number of market segments, their size and composition, and the segment specific regression coefficients
- Concomitant variable mixtures allow for demographic variables to explain segment membership simultaneously
- This class of methods enables marketers to engage in *response*-based segmentation, i.e., from descriptive to *predictive* segmentation

References (I)

- 王霞, 赵平, 王高, 刘佳 (2005), "基于顾客满意和顾客忠诚关系的市场细分方法研究," 南开管理评论, 8 (5), 26-30.
- 王燕, 赵平 (2009), "伴生变量混合模型在市场细分中的应用," 营销科学学报, 5 (1), 27-34. (http://www.jms.org.cn/read/15/3.pdf)
- Grün, Bettina and Friedrich Leisch (2008), "FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters," *Journal of Statistical Software*, 28 (4), (http://www.jstatsoft.org/v28/i04)
- Leisch, Friedrich (2004), "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R," *Journal of Statistical Software*, 11 (8), (http://www.jstatsoft.org/v11/i08)

References (II)

- Desarbo, Wayne S., Wagner A. Kamakura, and Michel Wedel (2006), "Latent Structure Regression," in Rajiv Grover and Marco Vriens (Eds.), *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*, Thousand Oaks: Sage Publications.
- McLachlan, Geoffrey and David Peel (2000), *Finite Mixture Models*, Now York: John Wiley & Sons, Inc.
- Wedel, Michel and Wagner A. Kamakura (2001), *Market Segmentation: Conceptual and Methodological Foundations (2nd Edition)*, Boston: Kluwer Academic Publishers.

Q & A

- Your comments are appreciated