

TRƯỜNG ĐẠI HỌC BÁCH KHOA TP. HCM
Khoa KTXD - Bộ môn Kỹ thuật & Quản lý Tài nguyên nước

PHÂN TÍCH DỮ LIỆU



Giảng viên: PGS. TS. NGUYỄN THÔNG
E-mail: nthong56@yahoo.fr or nguyenthong@hcmut.edu.vn
Web: <http://www4.hcmut.edu.vn/~nguyenthong/index>
PGS. TS. Nguyễn Thông | Tél. (08) 38 691 592- 098 99 66 719

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

NỘI DUNG MÔN HỌC

Chương 1. Thống kê mô tả (ôn).
Chương 1bis. Xác suất & phân phối thống kê (ôn).
Chương 2. Khoảng tin cậy.
Chương 2bis: Kỹ thuật chọn mẫu.
Chương 3. Kiểm định thống kê.
Chương 4. Phân loại (khúc) dữ liệu (Segmentation).
Chương 5. Phân nhóm dữ liệu (Cluster).
Chương 6. Phân tích thành phần chính (PCA).
Chương 7. Phân tích chuỗi thời gian.
Chương 8. Hồi quy tuyến tính.
Chương 9. Xử lý số liệu thực nghiệm.
Chương 10. Giới thiệu phần mềm SPSS or R.

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

GIỚI THIỆU VÀ ĐẶT VẤN ĐỀ

Phân loại dữ liệu áp dụng rộng rãi nhằm Phân khúc thị trường : phương pháp cho phép chia thị trường thành nhiều « tiểu thị trường », với mỗi tiểu thị trường có các đặc thù tương đối đồng nhất, và từ đó chúng ta sẽ định hướng được trong công tác thâm nhập.

Có nhiều phương pháp phân tích dữ liệu cho phép thực hiện sự phân chia này, trong số đó phương pháp phân khúc được xem là một phương pháp cơ bản.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

- Ví dụ : Xác định thành phần khách hàng sẽ tiêu thụ một sản phẩm mới, nhờ vào các thuộc tính như : giới tính, nghề nghiệp, tuổi, nơi cư trú, thu nhập
- Loại khách hàng tiêu thụ sản phẩm là biến cần được giải thích, ví dụ dưới dạng : loại khách tiêu thụ nhiều, loại khách tiêu thụ ít, loại khách không tiêu thụ.
- Ngoài ra, biến này còn có thể xuất hiện dưới dạng định lượng như số lượng tiêu thụ.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

→ Áp dụng phương pháp phân khúc, sẽ cho phép liên kết lại các loại hành vi người tiêu dùng định trước, cụ thể hoá bởi giá trị biến cần giải thích.

→ Nó cho phép nhận biết được các nhóm khách hàng phù hợp với loại hành vi tiêu dùng này.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

GIỚI THIỆU PHƯƠNG PHÁP

→ Cho một biến định lượng cần được giải thích Y, có m thể thức (đồng ý hoặc không đồng ý : m=2, tiêu thụ nhiều, tiêu thụ ít, không tiêu thụ : m=3...).

→ Giả sử có k biến giải thích X_k , mỗi biến giải thích có c_k thể thức.

→ Giá trị n_{ij} cho biết số lượng cá nhân thỏa đồng thời thể thức i của biến cần giải thích Y và thể thức j của biến giải thích X_k .

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
SỐ LIỆU: Bảng số số ban đầu				
Biến cần giải thích Y	Biến giải thích X_k			Số số
	Class 1	Class j...	Class ck	
Class 1	.		.	$N_1.$
...				
Class i	.	N_{ij}	.	$N_i.$
...				
Class m	.	.	.	$N_m.$
Số số	$n_{.1}$	$n_{.i}$	$n_{.ck}$	N

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

➔ Mục đích của sự nhóm lại này là ta sẽ tạo ra 2 nhóm S1 và S2 với các thuộc tính liên kết. Hai nhóm này sẽ cho ta sự « tách biệt » là lớn nhất.

➔ Trên cơ sở lập nhóm này, sẽ cho chúng ta dễ dàng dự báo tính cách của cá nhân khi chúng ta biết cá nhân này sắp vào nhóm nào.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
BẢNG SỐ SỐ ĐƯỢC PHÂN KHÚC				
Biến cần giải thích Y	Biến giải thích X_k		Số số	
	S1	S2		
Class 1	.	.	$N_1.$	
...				
Class i	N_{i1}	N_{i2}	$N_i.$	
...				
Class m	.	.	$N_m.$	
Số số	$n_{.1}$	$n_{.2}$	$N=N_{.}$	

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

CÁC BƯỚC CƠ BẢN TRONG KỸ THUẬT PHÂN KHÚC

Các kỹ thuật phân khúc khác nhau chủ yếu ở các tiêu chuẩn sử dụng để phân biệt sự khác biệt của 2 nhóm.

➔ Tuy nhiên, các bước tiến hành phân khúc là giống nhau. Quy trình này, về cơ bản là một quy trình tính lặp và thử dần, bao gồm các bước cơ bản sau đây :

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Bước 1: Phân khúc lần lượt cho từng bảng số số trong số p bảng (một bảng cho mỗi biến giải thích X) và chúng ta sẽ chọn bảng nào cho kết quả về phân nhóm là « rõ ràng » nhất theo tiêu chuẩn sử dụng. Có nghĩa là giá trị đánh giá sự phân khúc là lớn nhất.

➔ Hai nhóm S1 và S2 lần lượt với các số số N1 và N2 sẽ được thành lập.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Bước 2: Với mỗi bảng mới thành lập, thiết lập (p-1) bảng số số, trong bảng này ta sẽ xác định các thể thức của biến Y và các thể thức của (p-1) các biến giải thích X chưa sử dụng.

➔ Mỗi bảng này sẽ tiếp tục được phân khúc. Bảng nào cho kết quả phân khúc tốt nhất sẽ được chọn : kết quả này sẽ làm xuất hiện 4 phân khúc : S11, S12 bắt nguồn từ S1 và S21, S22 bắt nguồn từ S2.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Bước 3: Quy trình sẽ lặp lại như trên, cho mỗi phân khúc được tạo ra cho mỗi bước phân khúc, với các biến còn lại chưa xét.

→ Ta sẽ dừng tính toán khi sĩ số trong bảng nhỏ hơn một giá trị ngưỡng tối thiểu, hoặc trong trường hợp sự phân khúc không còn rõ ràng giữa 2 nhóm (trên cơ sở bài toán kiểm định giả thiết theo tiêu chuẩn KHI-Square) hoặc là trường hợp ta không còn biến giải thích X để xét.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

KẾT QUẢ PHÂN KHÚC

- Kết quả về quá trình phân khúc thường được trình bày dưới dạng bởi một đồ thị dưới dạng phân nhánh. Đồ thị này cho phép chúng ta thấy diễn biến của N cá thể ban đầu được lần lượt đưa vào phân khúc. Đồ thị sau trình bày một ví dụ cho dạng kết quả này. Các nhánh trong đồ thị sẽ được cụ thể hoá theo một số thuộc tính liên kết, thông qua các chỉ tiêu :

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

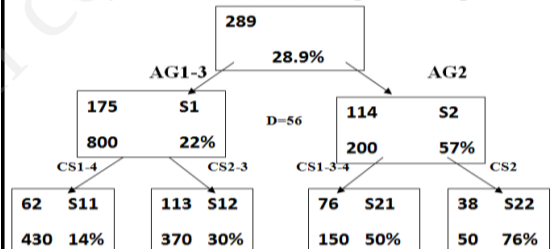
- Tỷ lệ phần trăm của mỗi phân khúc đại diện cho mỗi loại thuộc tính, có nghĩa là đại biểu cho thể thức nào đó của biến cần giải thích Y (khi biến này là loại biến định tính).
- Giá trị trung bình, khi biến Y là định lượng.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Đồ thị 1 : Sơ đồ nhánh của quá trình phân khúc



2: Ký hiệu phân khúc
1: Số cá nhân đồng ý mua tạp chí
3: Tỷ lệ phần trăm cá nhân đồng ý mua tạp chí
4: Tổng cá nhân của phân khúc tương ứng.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

CÁC TIÊU CHUẨN PHÂN KHÚC

→ Mục tiêu của một tiêu chuẩn là nhằm tìm ra, trong số các phân khúc có thể được, phân khúc nào cho ta sự « tách biệt » rõ ràng nhất (lớn nhất) khi sát nhập lại còn 2 nhóm.

→ Điều này trong thực tế có nghĩa tạo ra 2 tập hợp « đối nghịch » nhất, cho phép ta dễ dàng nhất trong việc tìm hiểu cung cách khác biệt của 2 nhóm được tách biệt

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

TIÊU CHUẨN BELSON

Đây là tiêu chuẩn sử dụng phổ biến nhất trong kỹ thuật phân khúc. Xét dạng số liệu trong bảng trên. Tiêu chuẩn D của Belson được tính như sau :

$$D = \sum_i \sum_j |N_{ij} - N_{i.} N_{.j} / N|$$

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Trong tất cả các phương án phân khúc (sát nhập), phân khúc nào cho D → max sẽ là phương án chọn. Trong đó :

N_{ij} : số quan trắc có trong ô (i,j)

$N_{i.} = N_{i1} + N_{i2} + ... + N_{ik}$

$N_{.j} = N_{1j} + N_{2j} + ... + N_{mj}$

$N = N_{..}$: tổng các quan trắc

Chú ý : Giá trị $N_{i.}N_{.j}/N$ là giá trị lý thuyết để 2 thể thức liên quan hoàn toàn độc lập.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

➔ Vấn đề áp dụng tiêu chuẩn Belson sẽ rất dễ dàng trong trường hợp biến cần giải thích Y chỉ có 2 thể thức và trong trường hợp chúng ta tiến hành phân khúc dạng « chẻ đôi », giống như số liệu trong bảng ví dụ sau.

➔ Trong trường hợp tính toán « thủ công », đây là phương pháp dễ dàng áp dụng nhất. Một hạn chế quan trọng của phương pháp này là nó không cho phép chúng ta có được giới hạn, ở đó chúng ta không cần tiến hành tiếp tục phân khúc.

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

VÍ DỤ TÍNH D THEO BELSON

Quan trắc

	S1	S2	Sĩ số
Class 1	10	20	30
Class 2	30	20	50
Sĩ số	40	40	80

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Lý thuyết: $N_{ij}=N_{i.}*N_{.j}/N$

	S1	S2	Sĩ số
Class 1	15	15	30
Class 2	25	25	50
Sĩ số	40	40	80

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Sự khác biệt giữa quan trắc & lý thuyết:

	S1	S2	Sĩ số
Class 1	-5	5	0
Class 2	5	-5	0
Sĩ số	0	0	0

Giá trị D theo tiêu chuẩn Belson :

D =4|±5| =20

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

PHƯƠNG PHÁP AID

Phương pháp AID có nguồn gốc từ nghiên cứu phương sai (variance analysis). Biến nghiên cứu có thể là biến nhị nguyên (có 2 thể thức) hoặc biến định lượng. Trong trường hợp biến nhị nguyên, tiêu chuẩn phân khúc như sau :

$$E = N_1(f_1)^2 + N_2(f_2)^2$$

with $f_1 = \frac{N_{21}}{N_{.1}}$ & $f_2 = \frac{N_{22}}{N_{.2}}$

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Trong trường hợp biến định lượng:

$$V_F = N_1(Y_1 - Y_0)^2 + N_2(Y_2 - Y_0)^2 \\ = N_1Y_1^2 + N_2Y_2^2 - NY_0^2$$

trong đó Y_1, Y_2 chỉ giá trị trung bình của mỗi phân khúc, và Y_0 là giá trị trung bình của toàn bộ N cá thể nghiên cứu.

Do NY_0^2 là hằng số → tiêu chí AID trở thành:

$$E = N_1Y_1^2 + N_2Y_2^2$$

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Phương pháp AID thực hiện sự tìm kiếm các phương án chẻ đôi tốt nhất có thể từ tập hợp « mẹ », dựa trên khái niệm phân tích phương sai : phương sai tổng, phương sai giải thích được và phương sai thặng dư.

→ Một ưu thế của áp dụng tiêu chuẩn này so với tiêu chuẩn Belson là cho phép ta tiếp tục hay kết thúc sự phân khúc trên cơ sở so sánh giữa phương sai giải thích và phương sai tổng (thường chỉ tiến hành khi tỷ số này lớn hơn 5%).

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

PHƯƠNG PHÁP KHÁC

→ Theo tiêu chuẩn khoảng cách KHI-Square.

→ Theo tiêu chuẩn hệ số « redondance »

$$E = \sum_i \sum_j \frac{N_{ij}}{N} \text{Log} \left[\frac{N_{ij}N}{N_i N_j} \right]$$

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Ví dụ: Xác định đối tượng tiếp thị

Công ty TVPlus chủ nhân của một chương trình TV chuyên đề thể thao, dự kiến sẽ đưa vào thị trường một tạp chí thể thao mới. Một thăm dò được thực hiện trên 1000 khán giả trung thành với chương trình TV thể thao và kết quả là 289 trong số khán giả trả lời sẽ là độc giả thường xuyên của tạp chí (mua).

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Bảng sau trình bày kết quả tổng hợp lấy ra từ thăm dò, trong đó biến Y biểu thị trả lời đồng ý mua (biến Y có 2 thể thức : đồng ý hoặc không đồng ý mua). Các thông tin ghi nhận từ các khán giả (các biến giải thích) là :

Tuổi AG(AG1,AG2,AG3)

Lĩnh vực hoạt động PCS(CS1,CS2,CS3,CS4)

Nơi cư ngụ LO(LO1,LO2,LO3).

Câu hỏi sẽ được trả lời là đối tượng nào cho tỷ lệ trả lời đồng ý mua thuận lợi nhất ?

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Mua	AG1			
	LO1	LO2	LO3	Sous total
Tổng				
khảo sát				
CS1	1 39	1 51	0 20	2 110
CS2	3 6	5 11	12 63	20 80
CS3	4 5	9 10	14 55	27 70
CS4	2 10	5 18	0 12	7 40
Total	10 ^(***) 60	20 90	26 150	56 ^(**) 300

Pi

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
AG2				
	LO1	LO2	LO3	Sous total
CS1	25	18	3	46
	30	38	22	90
CS2	1	7	30	38
	5	8	37	50
CS3	2	2	9	13
	4	2	24	30
CS4	0	1	16	17
	1	2	27	30
Total	28 ^(***)	28	58	114
	40	50	110	200

PHÂN TÍCH DỮ LIỆU					
Chương 4: Phân loại dữ liệu (Segmentation)					
AG3					
	LO1	LO2	LO3	Sous total	Total (total)
CS1	33	10	0	43	91 ^(*)
	109	41	0	150	350
CS2	10	10	4	24	82
	12	28	80	120	250
CS3	9	18	15	42	82
	10	43	47	100	200
CS4	3	7	0	10	34
	19	48	63	130	200
Total	55 ^(***)	45	19	119	289
	150	160	190	500	1000

PHÂN TÍCH DỮ LIỆU	
Chương 4: Phân loại dữ liệu (Segmentation)	
Ghi chú:	
1	→ 1 người sẽ mua
30	→ 30 tổng khảo sát (có thuộc tính theo hàng, cột tương ứng)
Total (total): Cột tổng hợp cho 3 biến AG, PCS, LO	
Total: Tổng theo cột cho từng « lớp »	
Sous total: Tổng theo hàng cho từng « lớp »	
PGS. TS. Nguyễn Thông	

PHÂN TÍCH DỮ LIỆU	
Chương 4: Phân loại dữ liệu (Segmentation)	
THỰC HIỆN NHÓM CÁC BIẾN THEO THỨ TỰ ƯU TIÊN (THEO TIÊU CHÍ CHỌN) SẼ LẦN LƯỢT TẠO THÀNH CÒN LẠI 2 « TIỂU NHÓM »	
→ Xem sơ đồ cây ví dụ theo sau:	
PGS. TS. Nguyễn Thông	

PHÂN TÍCH DỮ LIỆU	
Chương 4: Phân loại dữ liệu (Segmentation)	
Đồ thị 1 : Sơ đồ nhánh của quá trình phân khúc	
289 AG1-3 175 800 CS1-4 62 430 14% S11 28.9% D=56 CS2-3 113 370 30% S12 114 200 57% CS1-3-4 76 150 50% S21 AG2 S2 CS2 38 50 76% S22	
1 2 4 3 2: Ký hiệu phân khúc 1): Số cá nhân đồng ý mua tạp chí 3: Tỷ lệ phần trăm cá nhân đồng ý mua tạp chí 4: Tổng cá nhân của phân khúc tương ứng.	
PGS. TS. Nguyễn Thông	

PHÂN TÍCH DỮ LIỆU

Chương 4: Phân loại dữ liệu (Segmentation)

Chú ý:

Khi thực hiện khảo sát tìm phương án nhóm lại cho 1 biến X (ví dụ có 4 thể thức) nào đó → trước hết cần gom các biến còn lại để nhận được bảng trước khi phân tích có dạng, ví dụ:

	X_1	X_2	X_3	X_4
Đồng ý	91	82	82	34
Không đồng ý	259	168	118	166
Total	350	250	200	200

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
NHÓM CÁC BIẾN LÀN 1				
→ Biến PCS: Sẽ thực hiện để thành 2 tiểu nhóm:				
	CS1	CS2	CS3	CS4
Đồng ý	91 ^(*)	82	82	34
(%)	26	32.8	41	17
Không đồng ý	259	168	118	166
Total	350	250	200	200
Chú ý: 91 = 2 + 46 + 43 ; 350 = 110 + 90 + 150 PGS. TS. Nguyễn Thông 259 = 350 - 91				

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Nhóm PCS còn lại 2 nhóm			
	CS1-4	CS2-3	Total
Đồng ý	125	164	289
(%)	22.72	36.44	28.90
Không đồng ý	425	286	711
Total	550	450	1000
→ $D = 550 \cdot 289 / 1000 - 125 = 34$ (thực ra tính đủ phải là $D = 4 \cdot 34 = 136$) PGS. TS. Nguyễn Thông			

PHÂN TÍCH DỮ LIỆU	
Chương 4: Phân loại dữ liệu (Segmentation)	
Nhóm PCS còn lại 2 nhóm	
Ghi chú : → Đây là phương án sát nhập còn lại 2 nhóm cho giá trị D lớn nhất. Với CS1,CS2,CS3,CS4 ta có nhiều phương án để còn lại 2 nhóm. → Ý nghĩa kiểm định Belson là sự lựa « chấp » nhóm này sẽ làm cho 2 tiểu nhóm mới là có sự « tách biệt » lớn nhất (giá trị D thể hiện sự khác với bảng lý thuyết ở đó bảng 2D có 2 biến (mỗi biến có 2 thể thức) là « độc lập » nhau.	
PGS. TS. Nguyễn Thông	

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Ví dụ: Phương án nhóm CS1-2 & CS3-4			
	CS1-2	CS3-4	
Đồng ý	173	116	289
Không đồng ý	427	284	711
Tổng	600	400	1000
→ $D = 600 \cdot 289 - 173 = 0.4 < 34!$ PGS. TS. Nguyễn Thông			

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Biến AG: Thực hiện để thành 2 tiểu nhóm:			
	AG1	AG2	AG3
Đồng ý	56 ^(**)	114	119
(%)	18.67	57	23.8
Không đồng ý	244	86	381
Total	300	200	500
PGS. TS. Nguyễn Thông			

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Biến AG: Thực hiện để thành 2 tiểu nhóm:			
	AG1-3	AG2	Total
Đồng ý	175	114	289
(%)	21.88	57	28.9
Không đồng ý	625	86	711
Total	800	200	1000
→ $D = 800 \cdot 289 / 1000 - 175 = 56$ PGS. TS. Nguyễn Thông			

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
Biến LO: Thực hiện để thành 2 tiểu nhóm:				
	LO1	LO2	LO3	Total
Đồng ý	93 ^(***)	93	103	289
(%)	37.2	31	22.89	28.9
Không đồng ý	157	207	347	711
Total	250	300	450	1000

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Biến LO: Thực hiện để thành 2 tiểu nhóm:			
	LO1-2	LO3	Total
Đồng ý	186	103	289
(%)	33.82	22.89	28.9
Không đồng ý	364	347	711
Total	550	450	1000

→ $D=550*289/1000-186 = 27$

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU	
Chương 4: Phân loại dữ liệu (Segmentation)	
Kết luận: Từ kết quả phân khúc bước 1, ta thấy « phương án » nhập AG thành 2 nhóm AG1-3 và AG2 cho chỉ tiêu Belson tốt nhất (D=56). Kết quả này sẽ được chọn. Ta tiếp tục cho các biến còn lại LO và CS.	
→ Với phương án chọn AG1-3 và AG2, từ bảng số liệu ban đầu sẽ biến đổi như sau :	

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
NHÓM CÁC BIẾN LẦN 2				
	AG1-3			
	LO1	LO2	LO3	Total
CS1	34	11	0	45 ⁽¹⁾
	148	92	20	260 ⁽³⁾
CS2	13	15	16	44
	18	39	143	200
CS3	13	27	29	69
	15	53	102	170
CS4	5	12	0	17 ⁽²⁾
	29	66	75	170 ⁽⁴⁾
Total	65	65	45	175
	210	250	340	800

F

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
	AG2			
	LO1	LO2	LO3	Total
CS1	25	18	3	46
	30	38	22	90
CS2	1	7	30	38
	5	8	37	50
CS3	2	2	9	13
	4	2	24	30
CS4	0	1	16	17
	1	2	27	30
Total	28	28	58	114
	40	50	110	200

P

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Xem xét AG1-3			
Phương án « nhập » CS tốt nhất			
	CS1-4	CS2-3	Total
Đồng ý	62 ^(*)	113	175
(%)	14.4	30.5	21.8
Không đồng ý	368	257	625
Total	430	370	800

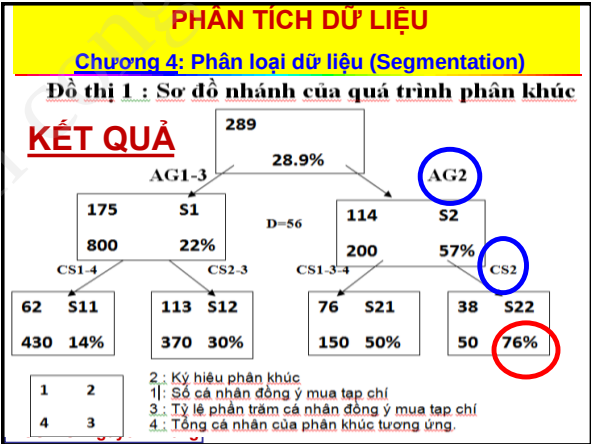
→ $D=430*175/800 - 62 = 32$

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Từ đó: → CS1-4:	62 ^(*)	S11	
	430 ^(**)	62/430=14.4%	
(*) = (1)+(2) (**) = (3)+(4)			
→ CS2-3:	113	S12	
	370	30.5%	
PGS. TS. Nguyễn Thông			

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Xem xét AG2			
Phương án « nhập » CS tốt nhất:			
	CS1-3-4	CS2	Total
Đồng ý	76	38	114
(%)	50.6	76	57
Không đồng ý	74	12	86
Total	150	50	200
→ D = 150*114/200-76 = 150*86/200-74 = 9.5			
PGS. TS. Nguyễn Thông			

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Từ đó :	76	S21	
→ CS1-3-4	150	50.6%	
→CS2:	38	S22	
	50	76%	
Kết luận : Từ kết quả phân tích trên, ta thấy mục tiêu để tiến hành một chiến dịch Marketing trực tiếp và có « tập trung » hợp lý nhất sẽ là đối tượng có đồng thời tính chất AG2 và CS2			
PGS. TS. Nguyễn Thông			



PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
Bài tập 1 : Nhằm nhận biết khả năng chậm tiến độ xây dựng công trình cao ốc VP (Y:chậm, không chậm), một khảo sát 100 công trình trong quá khứ với các yếu tố tiềm năng gây chậm: Tổng đầu tư ĐT (1: nhỏ, 2: vừa, 3: lớn), Nhà thầu NT (1: trong nước, 2: ngoài nước, 3: mixte), Kỹ thuật KT (1:Bình thường, 2: Cao, 3: Đặc biệt).			
Anh(Chị) hãy dùng phương pháp phân khúc để xác định «yếu tố» không chậm tiến độ nhất.			
PGS. TS. Nguyễn Thông			

PHÂN TÍCH DỮ LIỆU			
Chương 4: Phân loại dữ liệu (Segmentation)			
ĐT1			
Không trễ	NT1	NT2	NT3
Tổng khảo sát	2	1	1
KT1	4	3	2
KT2	2	1	1
KT3	5	3	3
	3	2	2
	3	4	3
Total	7 ^(***)	10	8
Pi	12		30 ^(**)

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
Không trẻ	ĐT2			
Tổng khảo sát	NT1	NT2	NT3	Sous total
KT1	1 5	0 4	2 6	
KT2	1 3	2 3	2 4	
KT3	2 4	4 6	0 5	
Total	4 ^(***) 12			(**) 40

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
Không trẻ	ĐT3			
Tổng khảo sát	NT1	NT2	NT3	Sous total
KT1	2 3	1 2	2 3	
KT2	3 4	1 3	0 3	
KT3	0 5	2 4	1 3	
Total	5 ^(***) 12			(**) 30

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
Bài tập 2 : Một Công ty kinh doanh giày dự kiến đưa ra thị trường một tạp chí tháng, giới thiệu về thời trang của thị trường giày. Để có thể đánh giá tình hình thuê bao tiềm năng của khách hàng cho tạp chí, một thăm dò trên một mẫu gồm 1000 khách hàng ở Tp. HCM, của Công ty được chọn ngẫu nhiên trong hồ sơ khách hàng của Công ty. Anh(Chị) hãy dùng phương pháp phân khúc để xác định « đối tượng » cho phép tiến hành Marketing « hy vọng » có hiệu quả nhất.				
PGS. TS. Nguyễn Thông				

PHÂN TÍCH DỮ LIỆU				
Chương 4: Phân loại dữ liệu (Segmentation)				
Các biến khảo sát: Gia đình: GD (1 : độc thân, 2 : có gia đình, 3 : ly hôn) Nghề nghiệp: PRO (1 : giám đốc-phó giám đốc, 2 : công nhân viên, 3 : tự do, 4 : thương mại-dịch vụ) Tuổi: AGE (1: 15-25 ;2: 26-35 ;3: 36-45 ;4: >45) Tổng hợp kết quả thăm dò như sau:				
PGS. TS. Nguyễn Thông				

PHÂN TÍCH DỮ LIỆU					
Chương 4: Phân loại dữ liệu (Segmentation)					
Đồng ý Số khảo sát	GD_1				Tổng
	AGE_1	AGE_2	AGE_3	AGE_4	
PRO_1	10 25	2 20	6 15	7 10	25 75
PRO_2	3 30	6 15	3 15	8 20	20 80
PRO_3	5 15	7 20	5 20	10 15	27 70
PRO_4	2 20	4 15	10 25	2 15	18 75
Tổng	20 90	19 70	24 80	27 60	90 300

PHÂN TÍCH DỮ LIỆU					
Chương 4: Phân loại dữ liệu (Segmentation)					
	GD_2				Tổng
	AGE_1	AGE_2	AGE_3	AGE_4	
PRO_1	4 25	2 20	5 15	5 20	16 105
PRO_2	3 30	6 15	3 25	2 15	14 80
PRO_3	1 20	2 20	5 20	3 20	11 80
PRO_4	2 20	4 25	1 25	2 25	9 95
Tổng	10 95	14 80	14 80	12 105	50 360

PHÂN TÍCH DỮ LIỆU						
Chương 4: Phân loại dữ liệu (Segmentation)						
	GD_3				Tổng	
	AGE_1	AGE_2	AGE_3	AGE_4		
PRO_1	2	2	4	5	13	100
		25	20	20	35	
PRO_2	5	6	3	2	16	80
		25	15	15	25	
PRO_3	2	1	5	6	14	80
		20	10	20	30	
PRO_4	3	7	1	6	17	80
		15	20	25	20	
Tổng	12	16	13	19	60	
		85	65	80	110	340

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU						
Chương 4: Phân loại dữ liệu (Segmentation)						
Hướng dẫn: Xem PKTT_Vidu.xls						
→ Tổ hợp nhóm cho GD						
→ Tổ hợp nhóm cho PRO						
→ Tổ hợp nhóm cho AG						
→ Chọn D → max → Lập bảng số liệu cho nhóm lần 1						
→tiếp tục ... nhóm lần 2....						
PGS. TS. Nguyễn Thông						

PGS. TS. Nguyễn Thông

PHÂN TÍCH DỮ LIỆU						
Chương 4: Phân loại dữ liệu (Segmentation)						
HẾT CHƯƠNG						
PGS. TS. Nguyễn Thông						

PGS. TS. Nguyễn Thông