

# 단계적 군집분석에 기반한 고객세분화 방안

한국CRM협회 수석 부회장  
고객전략연구센터 센터장

김형수 교수

# 목 차

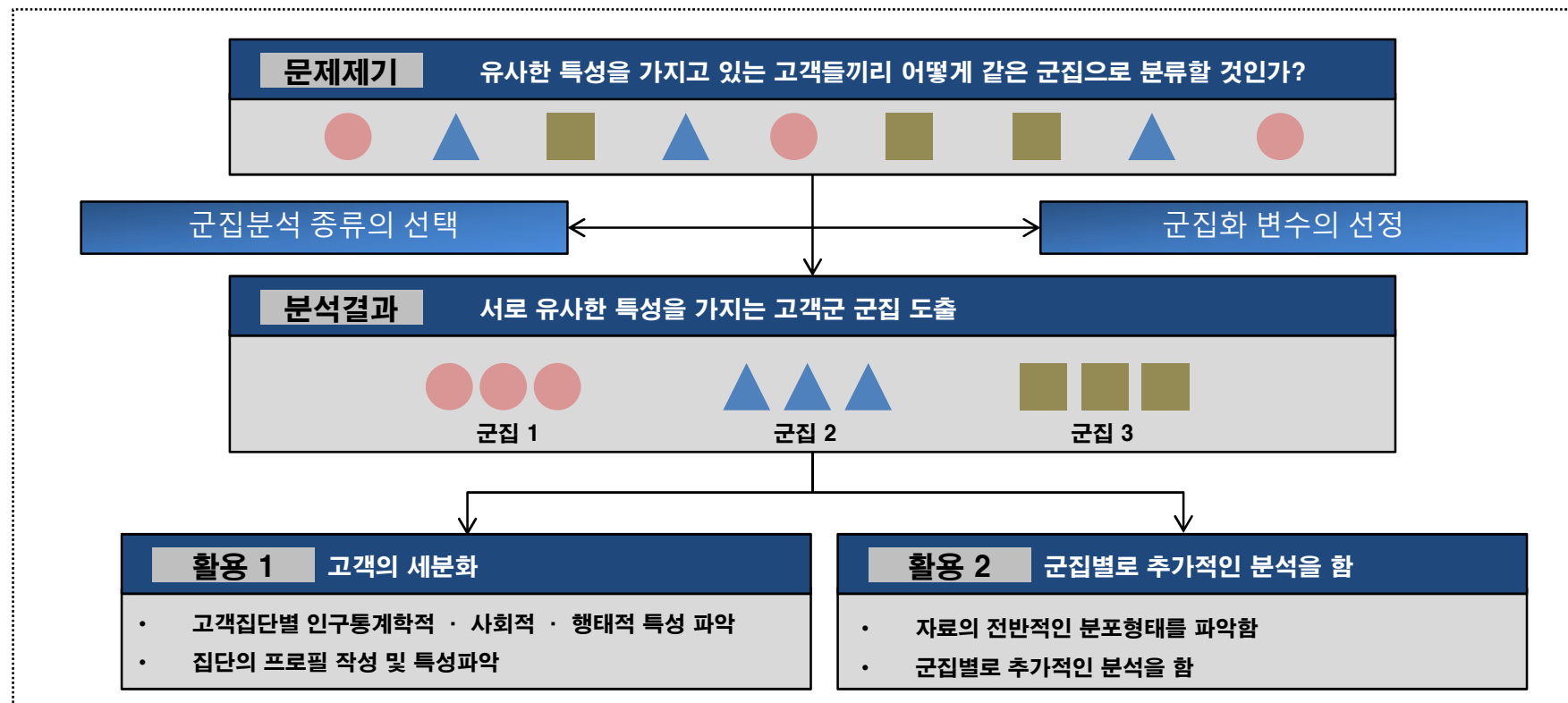
1. 군집분석을 이용한 고객세분화
2. 군집분석의 종류
3. 군집분석을 이용한 고객세분화의 한계점
4. 단계적 군집분석에 기반한 고객세분화
  - 1) 단계적 군집분석에 기반한 고객세분화 절차
  - 2) Step1. 주요 변수 도출
  - 3) Step2. 최적 군집 수 도출
  - 4) Step3. 군집 타당성 검증
  - 5) Step4. 군집 분류
5. 단계적 군집분석에 기반한 고객세분화 적용사례
  1. 단계적 군집분석에 기반한 고객세분화 절차
  2. Step1. 주요 변수 도출
  3. Step2. 최적 군집 수 도출
  4. Step3. 군집 타당성 검증
  5. Step4. 군집 분류

# 1. 군집분석을 이용한 고객세분화

## ❖ 군집분석의 정의

- 군집분석(cluster analysis)은 개인 또는 여러 개체 중에서 유사한 속성을 지닌 대상을 몇 개의 집단으로 그룹화한 다음, 각 집단의 성격을 파악함으로써 데이터 전체의 구조에 대해 이해하고자 하는 탐색적인 분석방법
- 군집분석은 거리정보를 이용해서 서로 가까운 거리에 있는 것들끼리 그룹화하여 분석대상을 몇 개의 집단으로 분류함.

## ❖ 군집분석을 이용한 고객세분화



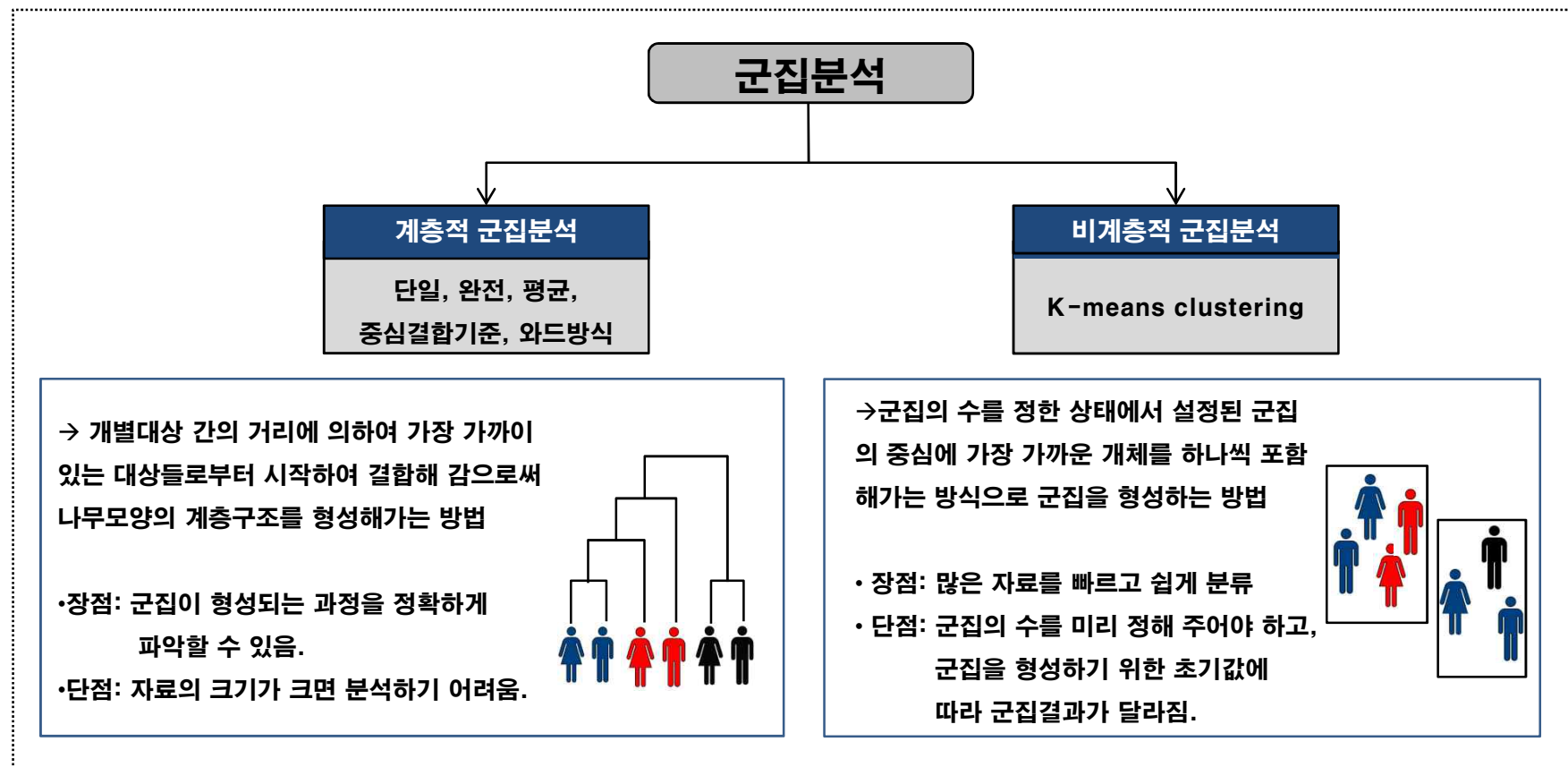
# 목 차

1. 군집분석을 이용한 고객세분화
2. 군집분석의 종류
3. 군집분석을 이용한 고객세분화의 한계점
4. 단계적 군집분석에 기반한 고객세분화
  - 1) 단계적 군집분석에 기반한 고객세분화 절차
  - 2) Step1. 주요 변수 도출
  - 3) Step2. 최적 군집 수 도출
  - 4) Step3. 군집 타당성 검증
  - 5) Step4. 군집 분류
5. 단계적 군집분석에 기반한 고객세분화 적용사례
  1. 단계적 군집분석에 기반한 고객세분화 절차
  2. Step1. 주요 변수 도출
  3. Step2. 최적 군집 수 도출
  4. Step3. 군집 타당성 검증
  5. Step4. 군집 분류

## 2. 군집분석의 종류

### ❖ 군집분석의 종류

- 군집화 방법에 따라서 군집분석은 크게 계층적 군집분석과 비계층적 군집분석으로 분류됨.



# 목 차

1. 군집분석을 이용한 고객세분화
2. 군집분석의 종류
3. 군집분석을 이용한 고객세분화의 한계점
4. 단계적 군집분석에 기반한 고객세분화
  - 1) 단계적 군집분석에 기반한 고객세분화 절차
  - 2) Step1. 주요 변수 도출
  - 3) Step2. 최적 군집 수 도출
  - 4) Step3. 군집 타당성 검증
  - 5) Step4. 군집 분류
5. 단계적 군집분석에 기반한 고객세분화 적용사례
  1. 단계적 군집분석에 기반한 고객세분화 절차
  2. Step1. 주요 변수 도출
  3. Step2. 최적 군집 수 도출
  4. Step3. 군집 타당성 검증
  5. Step4. 군집 분류

### 3. 군집분석을 통한 고객세분화의 한계점

#### ❖ 군집분석을 통한 고객세분화의 한계점



#### ❖ 한계점 해결방안

- 군집분석에서는 분석 자체에서 군집의 수를 도출하는 방법이 없음.
  - 군집의 수는 계층적 군집분석의 덴드로그램을 보고 연구자가 결정하는 것이 가장 바람직함.
  - But**, 데이터의 양이 많은 경우에는 계층적 군집분석을 수행하기가 어려움.
  - 표본을 추출한 다음, 이를 이용하여 계층적 군집분석을 시행 후 전체를 대상으로 비계층적 군집분석을 수행하는 것이 바람직함.

# 목 차

1. 군집분석을 이용한 고객세분화
2. 군집분석의 종류
3. 군집분석을 이용한 고객세분화의 한계점
4. 단계적 군집분석에 기반한 고객세분화
  - 1) 단계적 군집분석에 기반한 고객세분화 절차
  - 2) Step1. 주요 변수 도출
  - 3) Step2. 최적 군집 수 도출
  - 4) Step3. 군집 타당성 검증
  - 5) Step4. 군집 분류
5. 단계적 군집분석에 기반한 고객세분화 적용사례
  1. 단계적 군집분석에 기반한 고객세분화 절차
  2. Step1. 주요 변수 도출
  3. Step2. 최적 군집 수 도출
  4. Step3. 군집 타당성 검증
  5. Step4. 군집 분류

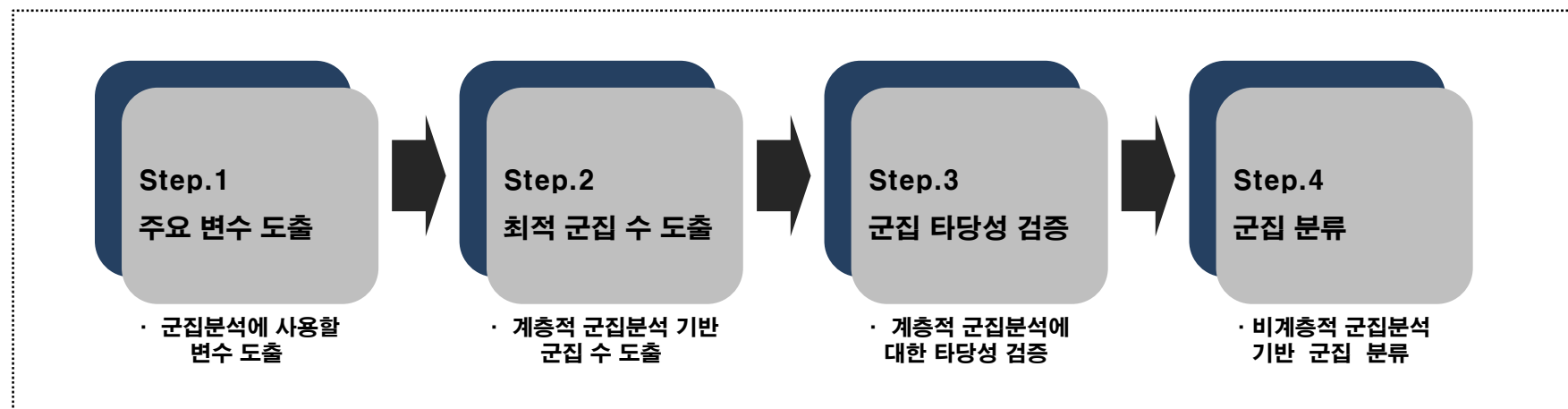


## 4. 단계적 군집분석에 기반한 고객세분화 (절차)

### ❖ 단계적 군집분석에 기반한 고객세분화

- 계층적 군집분석과 비계층적 군집분석을 단계적으로 적용함으로써 최적의 군집 수를 도출하여 고객세분화를 실시함.

### ❖ 단계적 군집분석에 기반한 고객세분화 절차



- Step1 : 단계적 군집분석에 사용할 수 있는 변수를 도출하는 과정을 설명한다.
- Step2 : 계층적 군집분석을 이용함으로써 최적의 군집 수를 도출하는 과정을 설명한다.
- Step3 : 비계층적 군집분석을 시행하기 전 계층적 군집분석을 통해 도출된 군집 수의 타당성을 검증한다.
- Step4 : 비계층적 군집분석을 시행하여 군집분류 및 그에 따른 군집 별 특성을 설명한다.

## 4. 단계적 군집분석에 기반한 고객세분화 (Step1. 주요 변수 도출)












### ❖ 단계적 군집분석의 주요 변수 선정의 중요성

- 단계적 군집분석에 기반한 고객세분화는 어떠한 변수를 선정하느냐에 따라 세분화의 방향성이 달라지게 되며 그에 따른 군집의 특성 또한 달라지게 됨. → **고객세분화의 가장 첫 단계인 변수선정 과정에 신중을 더해야 함.**

### ❖ 단계적 군집분석의 주요 변수 도출 과정

변수 선정 방법 : 변수의 선택

- 군집분석에 사용할 수 있는 변수의 척도는 일반적으로 거리 측정에 사용할 수 있는 변수이어야 함. → **등간척도, 비율척도로 측정된 변수**

구분	의미	예시	설명
명목척도	연구대상을 구분하거나 분류할 목적으로 라벨을 부여한 척도	 3번 선수  88번 선수	선수들의 고유 숫자를 부여한다.
서열척도	서로 비교할 수 있고 순위를 매길 수 있는 척도	 1등  2등  3등	선수의 기량에 따라 순위를 정한다.
등간척도	측정값이 차이(거리)가 의미를 갖는 척도	 섭씨 0도 미만  섭씨 0도 ~10 미만  섭씨 0도 ~20도 미만	관찰치 속성 차이를 의도적으로 양적 차이로 측정하기 위해 균일한 간격을 두고 분할하여 측정한다.
비율척도	절대 0값을 가짐으로써 비율의 성격을 갖는 척도	 9.8점  8.4점  0점	심사자가 발레 점수를 10점 만점으로 평가한다.

#### 명목척도 변수

- 직접적인 사용은 불가능
- 더미코딩하여 여러 개의 이산 변수로 변경 후 사용가능

- 단계적 군집분석의 변수 선정과정에서 상관분석이 필요함.
- 이분형 변수는 제거 → 군집분석은 변수 간 거리를 이용한 방법이기 때문임.

## 4. 단계적 군집분석에 기반한 고객세분화 (Step2. 최적 군집 수 도출)

### ❖ 계층적 군집분석을 통한 군집 수 도출의 중요성

- 계층적 군집분석은 최종적으로 군집의 수를 도출하는데 있어 쉽고 빠르게 이해할 수 있으며, 또한 그 과정이 복잡하지 않아 실무적으로 많이 쓰이는 방법

군집 수 도출 방법 : 계층적 군집분석

변수 선정



군집화 일정표

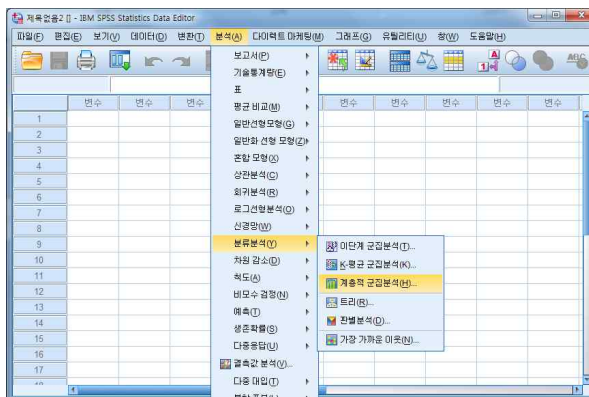
덴드로그램



최적의 K수 도출

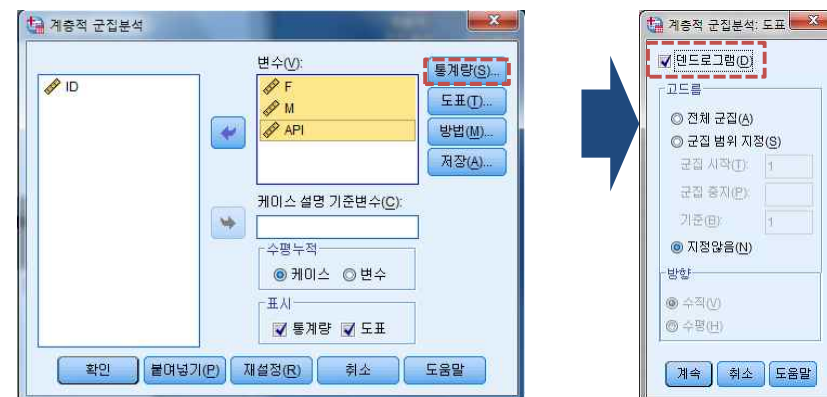
### ❖ 계층적 군집분석을 통한 군집 수 도출과정 (1/2)

① : 군집분석 대화상자 경로



- 분석 → 분류분석 → 계층적 군집분석

② : 군집분석 실행



- 계층적 군집분석 실행 -> 변수선택 -> 통계량(덴드로그램)

## 4. 단계적 군집분석에 기반한 고객세분화 (Step2. 최적 군집 수 도출)

### ❖ 계층적 군집분석을 통한 군집 수 도출과정 (2/2)

#### ③ : 군집화 일정표 및 덴드로그램

군집화 일정표

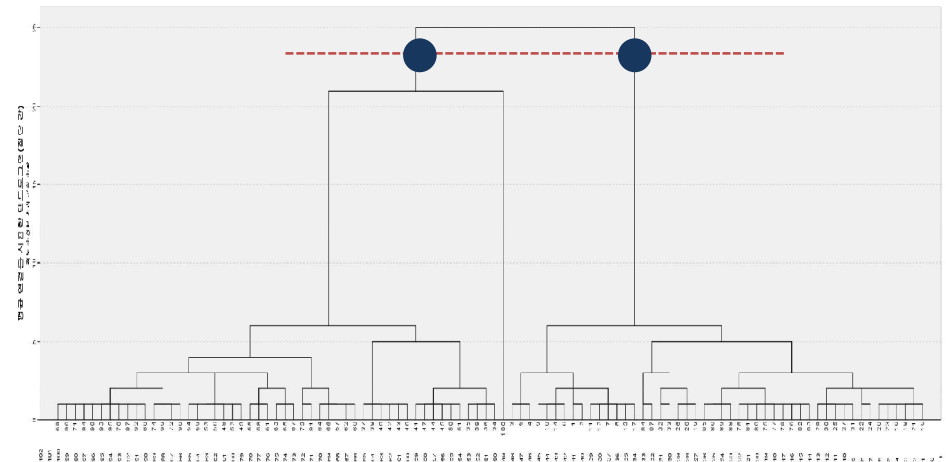
단계	결합 군집		계수	처음 나타나는 군집의 단계		다음 단계
	군집 1	군집 2		군집 1	군집 2	
1	69	96	.050	0	0	8
2	75	77	.050	0	0	13
3	64	66	.050	0	0	68
4	85	86	.090	0	0	14
5	74	99	.090	0	0	45
6	9	10	.100	0	0	42
7	78	81	.100	0	0	16
...						
90	15	75	9.269	85	81	95
91	48	68	10.290	88	87	93
92	1	3	11.393	83	49	96
93	48	57	13.899	91	82	97
94	34	37	15.849	86	71	97
95	15	19	17.129	90	89	96
96	1	15	19.865	92	95	99
97	34	48	22.646	94	93	98
98	34	100	79.508	97	0	99
99	1	34	95.918	96	98	0

#### • 군집 수의 결정

-단계 별 계수의 변화 폭이 크게 나타나는 구간에서 군집 수 결정  
(\* 97 ~ 98 단계가 계수의 폭이 가장 넓음)

덴드로그램에서 98단계인 34 고객과 100고객이 군집화되는 부분 위에 수평선을 그려 최적 군집 수를 도출해야 한다.

덴드로그램



#### • 군집 수의 결정

-2개의 군집이 최적의 군집 수로 나타나는 것을 볼 수 있다.

## 4. 단계적 군집분석에 기반한 고객세분화 (Step3. 군집 타당성 검증)

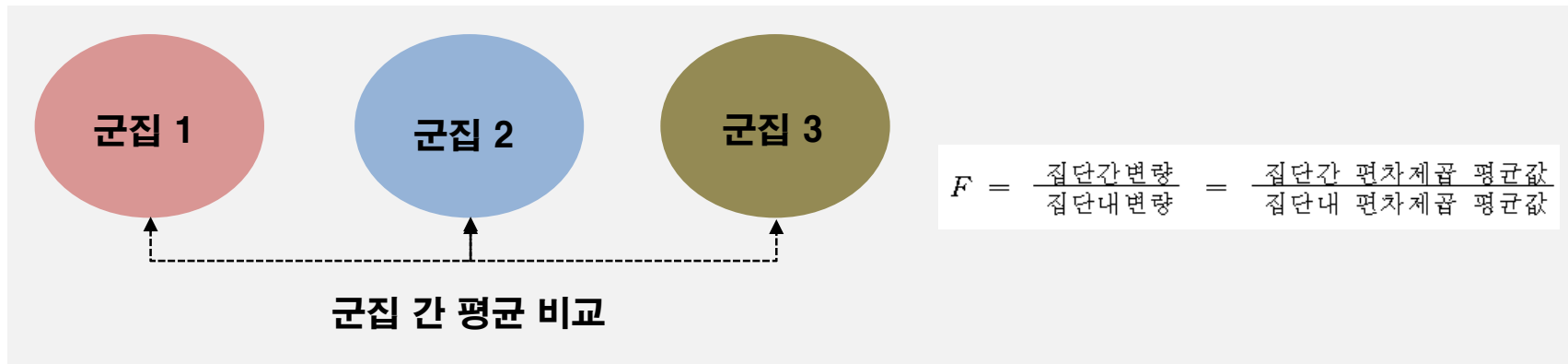
### ❖ 계층적 군집분석을 통해 도출된 군집 수 타당성 검증의 중요성

- 계층적 군집분석을 통한 군집 수 도출 후 비계층적 군집분석을 시행하기 전, 계층적 군집분석을 통해 도출된 군집의 수의 타당성을 검증할 필요가 있음.
- 계층적 군집분석의 타당성 검증을 위해 일원배치 분산분석을 이용할 수 있음.

### ❖ 계층적 군집분석을 통해 도출된 군집 수 타당성 검증 과정

타당성 검증 방법 : ANOVA분석 (1)

- 평균값을 기초로 여러 집단을 비교하고 이들 집단간에 차이점이 있는지 가설 검증을 통해 관계를 파악하는 통계분석 기법
- 집단간 분산과 집단내 분산의 비율인 F값을 F분포의 임계값과 비교하여 집단간의 차이 정도를 파악하는 통계분석 기법



## 4. 단계적 군집분석에 기반한 고객세분화 (Step3. 군집 타당성 검증)

### ❖ 계층적 군집분석을 통해 도출된 군집 수 타당성 검증 과정

타당성 검증 방법 : ANOVA분석 (2)

#### ▪ 분산분석 표

일원배치 분산분석

	제공합	df	평균 제공	F	유의확률
집단-간	41961.800	2	20980.900	78.975	.000
집단-내	7173.000	27	265.667		
합계	49134.800	29			

일원배치 분산분석

	제공합	df	평균 제공	F	유의확률
집단-간	38698.133	3	12899.378	32.135	.000
집단-내	10436.667	26	401.410		
합계	49134.800	29			

일원배치 분산분석

	제공합	df	평균 제공	F	유의확률
집단-간	41142.626	4	10285.657	32.174	.000
집단-내	7992.174	25	319.687		
합계	49134.800	29			

#### ▪ ANOVA 결과

군집수3개, 4개, 5개의 각 각 분산분석 결과  
모두 집단 간 차이가 유의하나 F값이 군집이  
3개일 때 가장 큼.

→ **군집 수 3개일 때 군집 간 차이가 가장 큼.**

## 4. 단계적 군집분석에 기반한 고객세분화 (Step3. 군집 타당성 검증)

### ❖ 계층적 군집분석을 통해 도출된 군집 수 타당성 검증 과정

타당성 검증 방법 : ANOVA분석 (2)

#### ▪ 사후분석 표

(I) 군집3개 (J) 군집3개		평균차(I-J)	표준오차	유의확률	95% 신뢰구간	
					하한값	상한값
1	2	-22.30000*	7.28926	.018	-41.1794	-3.4206
	3	-88.10000*	7.28926	.000	-106.9794	-69.2206
2	1	22.30000*	7.28926	.018	3.4206	41.1794
	3	-65.80000*	7.28926	.000	-84.6794	-46.9206
3	1	88.10000*	7.28926	.000	69.2206	106.9794
	2	65.80000*	7.28926	.000	46.9206	84.6794

(I) 군집4개 (J) 군집4개		평균차(I-J)	표준오차	유의확률	95% 신뢰구간	
					하한값	상한값
1	2	-14.66667*	10.34615	.578	-45.5763	16.2430
	3	-50.66667*	10.82026	.001	-82.9927	-18.3406
	4	-100.33333*	11.56734	.000	-134.8914	-65.7753
2	1	14.66667*	10.34615	.578	-16.2430	45.5763
	3	-36.00000*	9.50354	.009	-64.3923	-7.6077
	4	-85.66667*	10.34615	.000	-116.5763	-54.7570
3	1	50.66667*	10.82026	.001	18.3406	82.9927
	2	36.00000*	9.50354	.009	7.6077	64.3923
	4	-49.66667*	10.82026	.001	-81.9927	-17.3406
4	1	100.33333*	11.56734	.000	65.7753	134.8914
	2	85.66667*	10.34615	.000	54.7570	116.5763
	3	49.66667*	10.82026	.001	17.3406	81.9927

(I) 군집5개 (J) 군집5개		평균차(I-J)	표준오차	유의확률	95% 신뢰구간	
					하한값	상한값
1	2	-6.83333	12.64292	.990	-48.8315	35.1649
	3	-23.73333	11.76992	.418	-62.8315	15.3648
	4	-74.47619*	12.33823	.000	-115.4622	-33.4902
	5	-108.58333*	13.65592	.000	-153.9466	-63.2201
2	1	6.83333	12.64292	.990	-35.1649	48.8315
	3	-16.90000	9.23308	.514	-47.5711	13.7711
	4	-67.64286*	9.94740	.000	-100.6869	-34.5989
	5	-101.75000*	11.54136	.000	-140.0889	-63.4111
3	1	23.73333	11.76992	.418	-15.3648	62.8315
	2	16.90000	9.23308	.514	-13.7711	47.5711
	4	-50.74286*	8.81126	.000	-80.0127	-21.4730
	5	-84.85000*	10.57783	.000	-119.9882	-49.7118
4	1	74.47619*	12.33823	.000	33.4902	115.4622
	2	67.64286*	9.94740	.000	34.5989	100.6869
	3	50.74286*	8.81126	.000	21.4730	80.0127
	5	-34.10714	11.20675	.085	-71.3346	3.1203
5	1	108.58333*	13.65592	.000	63.2201	153.9466
	2	101.75000*	11.54136	.000	63.4111	140.0889
	3	84.85000*	10.57783	.000	49.7118	119.9882
	4	34.10714	11.20675	.085	-3.1203	71.3346

#### ▪ 사후분석 결과

사후분석 결과 군집 수 3개일 때는 모든 집단이 다 유의  
하나 나머지에선 몇 개 집단은 유의하지 않음.

## 4. 단계적 군집분석에 기반한 고객세분화 (Step4. 군집 분류)

### ❖ 비계층적 군집분석을 통한 군집 분류의 중요성

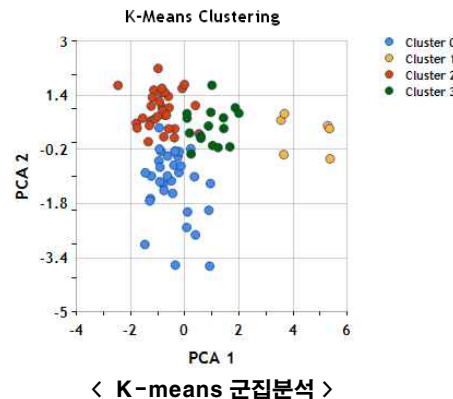
- K-means 군집분석은 대용량의 데이터 처리가 용이하다는 특징을 가지고 있지만, 군집을 수를 미리 지정하여야 한다는 어려움이 있음.
- 앞서 계층적 군집분석을 통해 이미 최적의 군집 수가 도출(step2)되었으므로 K-means 군집분석에서 군집의 수 결정 시 용이하게 진행할 수 있음.

#### 군집 분류 방법 : K-means 군집분석

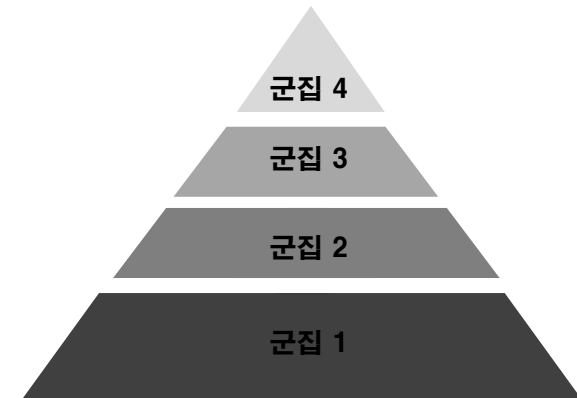
계층적 군집분석을  
통해 도출된 K수

K-means 군집분석  
K수 = ?

- 비계층적 군집분석에 사용할 K수는 계층적 군집분석을 통해 도출된 K수를 사용



- K-means 군집분석을 시행



- 정해진 군집 수에 대한 통계량이 도출됨



## 4. 단계적 군집분석에 기반한 고객세분화 (Step4. 군집 분류)

### ❖ 비계층적 군집분석을 통한 군집 분류 과정

#### ① : 데이터 준비

테이블 (5개 필드, 1,000개 레코드) #1

ID	총매출액	누적방문빈도	교차구매지수	평균구매주기
1	367900	15	3	14
2	64000	3	1	109
3	467400	10	8	12
4	61000	3	1	70
5	128000	4	2	45
6	353620	6	5	22
7	1416230	24	11	5
8	535970	14	5	5
9	112760	4	2	25
10	704540	33	10	6
11	1120170	43	8	7
12	132260	3	1	130
13	988320	63	8	6
14	125000	2	1	3
15	1021120	89	13	4
16	147600	8	3	40
17	40600	2	1	80
18	118980	3	3	41
19	277780	9	4	18
20	2296300	52	12	3
21	186900	2	2	14
22	248150	3	4	8
23	916470	22	10	8
24	2180870	76	21	4
25	196800	2	1	24
26	745150	20	7	8

#### ② : 변수 지정

유형

필드: 속성: 값: 누락: 확인: 역할:

필드	속성	값	누락	확인	역할
ID	연속형	[1,1000]	없음	없음	인용
총매출액	연속형	[1,3000,20]	없음	없음	인용
누적방문빈도	연속형	[2,199]	없음	없음	인용
교차구매지수	연속형	[1,26]	없음	없음	인용
평균구매주기	연속형	[0,439]	없음	없음	인용

☒ 현재 필드 보기 ☐ 사용되지 않은 필드 설정 보기

확인 취소 적용(A) 재설정(R)

• 군집분석에 사용될 변수를 지정

#### ③ : K-means 군집분석

K-평균

모델: 모델 고급 주석

모델 이름:  자동(O) 사용자 정의(M)

☒ 파티션된 데이터 사용

군집 수:

☐ 거리 함수 설정

군집 설명: ☐ 문자열 ☐ 숫자

설명 접두어:

최적화: ☐ 속도 ☒ 메모리

확인 실행 취소 적용(A) 재설정(R)

• 계층적 군집분석을 통해 도출된 k수를 군집 수에 적용하여 k-means 군집분석 실행

# 목 차

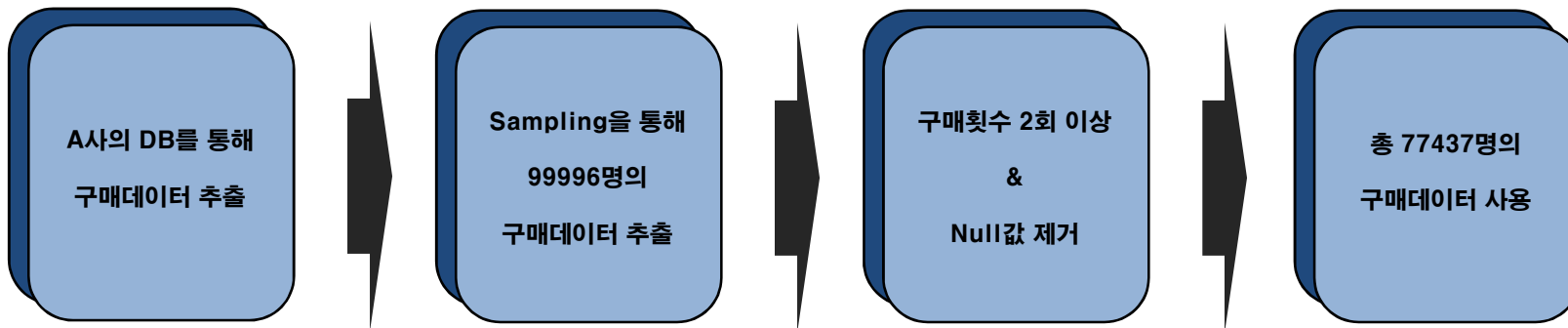
1. 군집분석을 이용한 고객세분화
2. 군집분석의 종류
3. 군집분석을 이용한 고객세분화의 한계점
4. 단계적 군집분석에 기반한 고객세분화
  - 1) 단계적 군집분석에 기반한 고객세분화 절차
  - 2) Step1. 주요 변수 도출
  - 3) Step2. 최적 군집 수 도출
  - 4) Step3. 군집 타당성 검증
  - 5) Step4. 군집 분류
5. 단계적 군집분석에 기반한 고객세분화 예시
  1. 단계적 군집분석에 기반한 고객세분화 절차
  2. Step1. 주요 변수 도출
  3. Step2. 최적 군집 수 도출
  4. Step3. 군집 타당성 검증
  5. Step4. 군집 분류

## 5. 단계적 군집분석에 기반한 고객세분화 예시 (절차)

### ❖ 단계적 군집분석에 기반한 고객세분화

- 앞서 제시한 단계적 군집분석에 기반한 고객세분화 방법론을 토대로 **실제 A기업의 고객을 대상으로 고객세분화 실시**
- 데이터 분석을 위하여 SPSS Statistic 19v와 SPSS MODELER 14v를 사용

#### A사 고객 데이터 정제



## 5. 단계적 군집분석에 기반한 고객세분화 예시 (Step1. 주요 변수 도출)

### ❖ 단계적 군집분석의 주요 변수 도출 과정 (1/2)

#### 변수 선정 방법 ① : 군집분석에 사용할 변수 선정

변수 명	변수 설명	변수 설명
1	cust_no	고객ID
2	base_ymd	매출일자
3	dept_cd	부서코드
4	head_gb	본부구분
5	site_gb	바로드림여부
6	sell_no	판매번호
7	Jo_cd	조코드
8	maechul_cnt	매출건수
9	maechul_qty	매출권수
10	maechul_kum	매출금액
11	mail_recv_yn	메일수신여부
12	maechul_mile_save	매출마일리지적립금액
13	maechul_mile_use	매출마일리지사용금액
14	coupon_kum	쿠폰금액

변수 명	변수 설명	사용 여부 근거
15	book_tae_g b	도서대구분
16	cust_gb	고객그룹
17	entr_date	가입일자
18	entr_site_cd	가입처
19	Age	연령
20	Sex_cd	성별
21	Job_cd	직업
22	sms_recv_yn	SMS수신여부
23	고객자산가치	
24	교차구매지수	
25	API (평균구매주기)	
26	방문빈도	
27	로열티프로그램 참여여부	

군집분석에 사용할 변수 :

• 첫번째 변수 선정 방법에서 인구통계학적 · 행동적 변수를 포함한 총 27개의 변수들 중 **군집분석에 사용할 변수 12개를 추출**

## 5. 단계적 군집분석에 기반한 고객세분화 예시 (Step1. 주요 변수 도출)

### ❖ 단계적 군집분석의 주요 변수 도출 과정 (2/2)

#### 변수 선정 방법 ② : 상관 분석

- 앞서 선정된 12개의 변수로 상관분석 실시하여 상관계수의 값이 높은 변수 추출

변수1	변수2	상관계수
매출권수	매출금액	0.897
매출권수	매출건수	0.965
매출권수	방문빈도	0.84
매출권수	교차구매지수	0.73
매출금액	매출건수	0.878
매출금액	방문빈도	0.77
매출금액	교차구매지수	0.677
매출건수	방문빈도	0.872
매출건수	교차구매지수	0.762
방문빈도	교차구매지수	0.78

- 상관계수값이 0.80이상은 변수1, 변수2 중 하나를 제거  
(전체적인 결과를 보고 제거해야 할 변수를 선택)

→ 매출권수, 매출건수 제거

#### 변수 선정 방법 ③ : 이분형 변수 제거

- 매출권수와 매출건수를 제거한 10개의 변수 중 이분형 변수를 제거

- 1) 매출액
- 2) 나이
- 3) 성별
- 4) ~~매일수신여부~~
- 5) ~~SMS수신여부~~
- 6) 누적 방문빈도
- 7) API
- 8) 교차구매지수
- 9) ~~로열티프로그램참여여부~~
- 10) CLV

이분형 변수  
제거

총 5개의 변수를 사용

## 5. 단계적 군집분석에 기반한 고객세분화 예시 (Step2. 최적 군집 수 도출)

### ❖ 계층적 군집분석을 통한 군집 수 도출과정 (1/2)

변수 수 도출 방법 단계

샘플링



분포비교



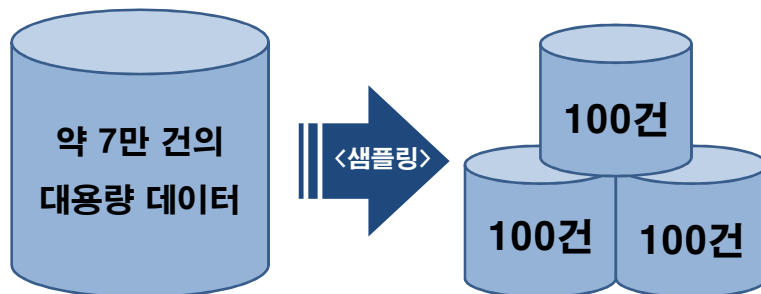
군집화 일정표



덴드로그램

#### 군집 수 도출 방법 ① : 샘플링

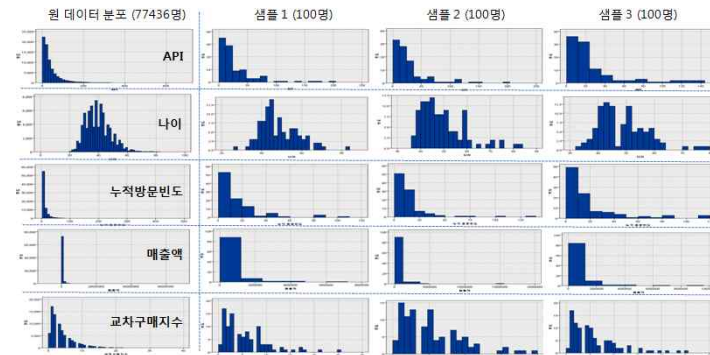
- 기업의 77437명 대용량 구매데이터는 계층적 군집분석에 사용할 수 없으므로 100건의 데이터를 Sampling



- 샘플링 데이터의 정확도를 높이기 위해 100건 씩 3번 샘플링을 하여 진행함.

#### 군집 수 도출 방법 ② : 분포비교

- 100건의 샘플링 과정을 거친 후 원 데이터와 비교하여 분포 비교  
(→ 샘플은 원 데이터를 반영해야 함.)



- 100건의 샘플링 데이터와 원 데이터의 분포를 비교한 결과 두 데이터의 분포간 차이가 없으므로 원 데이터를 잘 반영한 샘플링 데이터라 볼 수 있다.

## 5. 단계적 군집분석에 기반한 고객세분화 예시 (Step2. 최적 군집 수 도출)

### ❖ 계층적 군집분석을 통한 군집 수 도출과정 (2/2)

군집 수 도출 방법 ③ : 군집화 일정표

샘플 1 (100명) 군집화 일정표									
단계	군집1	군집2	계수	군집1	군집2	다음 단계	계수 LISI는 군집의 단계		
90	1	19	3.424	87	89	92			
91	16	20	3.708	0	79	93			
92	1	73	3.920	90	90	95			
93	16	49	4.152	91	0	96			
94	7	48	4.218	99	0	95			
95	1	7	4.831	92	94	97			
96	16	21	5.011	93	88	98			
97	1	74	5.722	95	0	99			
98	1	16	9.871	97	96	99			
99	1	79	24.778	99	0	0			
군집2, 군집3			15.105	군집3, 군집4			군집4, 군집5		
			4.66				0.18		

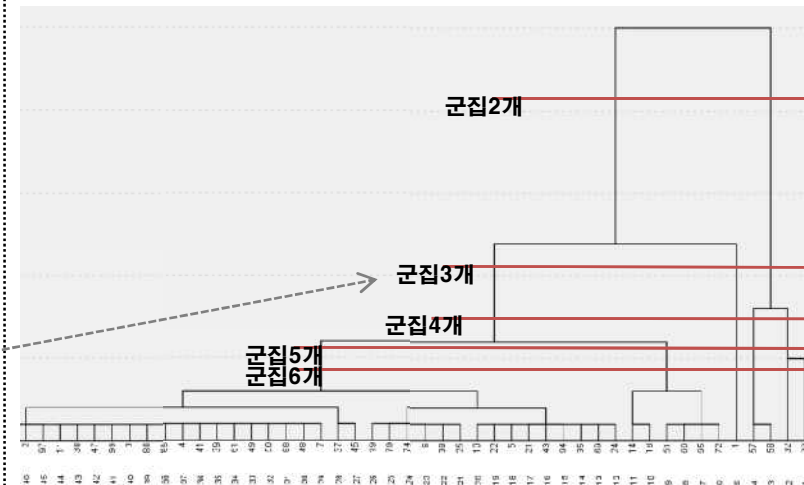
샘플 2 (100명) 군집화 일정표									
단계	군집1	군집2	계수	군집1	군집2	다음 단계	계수 LISI는 군집의 단계		
88	1	40	1.848	87	81	92			
89	11	12	1.878	0	67	91			
90	5	32	1.988	85	80	92			
91	11	43	2.330	89	0	94			
92	1	5	2.371	88	90	93			
93	1	52	2.584	92	0	96			
94	11	14	2.888	91	0	98			
95	88	75	3.538	0	0	97			
96	1	11	3.982	93	94	97			
97	1	88	5.246	95	95	98			
98	1	2	8.355	97	0	99			
99	1	28	38.975	99	0	0			
군집2, 군집3			28.621	군집3, 군집4			군집4, 군집5		
			3.109				1.708		

샘플 3 (100명) 군집화 일정표									
단계	군집1	군집2	계수	군집1	군집2	다음 단계	계수 LISI는 군집의 단계		
96	9	44	2.338	0	89	92			
97	1	13	2.339	94	0	98			
98	58	88	2.411	0	0	92			
99	1	24	2.485	97	94	93			
90	4	95	2.584	0	0	95			
91	12	87	2.622	95	0	94			
92	9	58	2.749	98	88	94			
93	1	3	3.029	99	62	96			
94	9	12	3.208	92	91	99			
95	4	17	3.552	90	0	99			
96	1	9	4.188	93	94	99			
97	42	88	4.718	0	0	99			
98	1	4	4.972	98	95	99			
99	1	42	12.048	99	97	0			
군집2, 군집3			8.132	군집3, 군집4					
			3.109						

- 100건의 샘플링으로 계층적 군집분석을 시행한 결과 군집화 일정표 계수값이 군집2에서 군집3으로 변화할 때 차이가 가장 크므로 **군집 수는 3개로 도출**

군집 수 도출 방법 ④ : 덴드로그램



#### 결과

- 군집화 일정표와 덴드로그램(dendrogram)을 통해 최적의 군집 수는 3개로 도출됨.

## 5. 단계적 군집분석에 기반한 고객세분화 예시 (Step3. 군집 타당성 검증)

### ❖ 계층적 군집분석을 통해 도출된 군집 수 타당성 검증 과정

#### 타당성 검증 방법 ① : ANOVA분석

- 6개의 변수로 ANOVA분석을 시행.

	군집3개		군집4개		군집5개	
	F값	유의확률	F값	유의확률	F값	유의확률
매출금액	62.852	0.000	42.172	0.000	35.436	0.000
나이	174.708	0.000	116.606	0.000	95.15	0.000
방문빈도	111.295	0.000	74.283	0.000	63.411	0.000
개인평균구매주기	625.976	0.000	420.284	0.000	314.774	0.000
교차구매지수	127.224	0.000	84.929	0.000	70.291	0.000
고객생애가치	1.779	0.169	2.079	0.101	1.552	0.184

F값이 가장 큰 것 :

#### 결과

- 6개의 변수로 ANOVA분석을 한 결과 6개의 변수들 중 5개의 변수가 3개의 군집 수에서 F값이 가장 크게 나타났으며 이는 군집을 3개로 나누었을 때가 군집 수 4개, 군집 수 5개일 때보다 집단을 뚜렷하게 분류하고 있음이 검증됨.



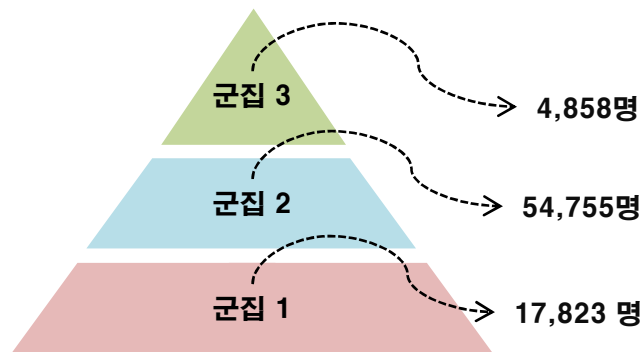
## 5. 단계적 군집분석에 기반한 고객세분화 예시 (Step4. 군집 분류)

### ❖ 비계층적 군집분석을 통한 군집 분류 과정

군집 분류 방법 : K-means 군집분석

K-평균 군집 별 특징

▪ K-평균 군집결과



▪ 군집 통계량

	평균 연령	매출액	평균구매주기	주 구매 품목
군집3	46세	1,006,731	13	중고 학습
군집2	38세	183,719	39	아동, 유아
군집1	40세	47,045	249	외국어

▪ 군집 특징 및 마케팅 방향성

- **군집 3**: 책벌레 고객 군. 평균매출액이 100만원 대로 상당히 높으며 평균구매주기가 13일로 상당히 자주 이용하는 고객임을 알 수 있음.  
→ “VIP도서기부” 전략 = VIP 고객들의 구매액에 비례하여 일정 포인트 이상 누적 시 자신이나 자녀의 명의로 자신이 지정한 단체에 책을 기부
- **군집 2**: 캥거루 군. 유아코드 비율이 높은 것을 보아 아동과 중고생의 자녀를 둔 학부모집단임을 추정할 수 있음.  
→ “고객자녀careservice” 전략 = 고객 자녀의 성장단계에 따라 적합한 도서를 선별해 추천해 주고 지속적인 관리를 통한 관계 유지 및 강화
- **군집 1**: 구름 고객 군. 주 구매 품목이 외국어이고, 평균구매주기가 242일로 한 두 번 방문 후 오지 않는 고객임을 알 수 있음.  
→ “재 구매 유도” 전략 = 고객이 가장 최근 구매한 도서를 분석하여 추천도서 선정 후 할인쿠폰을 통해 재 구매를 유도

---

END

---