



물관리 분야 | 대국민 빅데이터 콘테스트

수돗물과 보건과의 영향 분석 :

다변량 통계기법을 이용한 수질과 질병들 간의 관계

충북대학교 정보통계학과 소속

김경민 임세은 조희연 채다현 탁현준

목차 :

01 서론

02 본론

기초분석

상관분석

주성분분석

군집분석

ANOVA

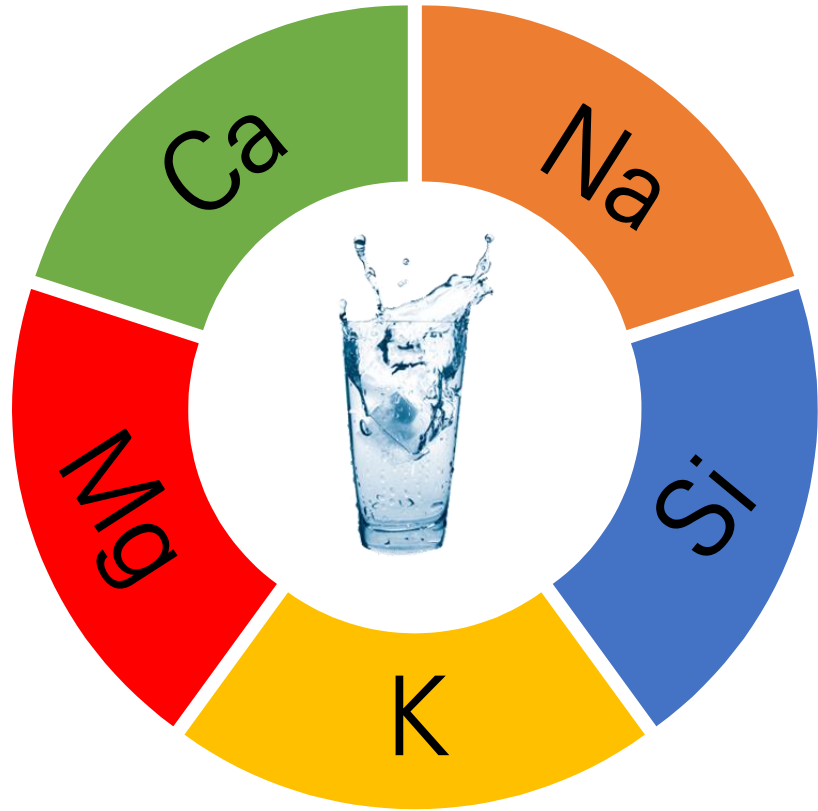
03 결론

서론:

배경 및 목적

01 서론

배경 및 목적



다섯 가지 미네랄



질병

본론:

기초분석 | 상관분석 | 주성분분석
군집분석 | ANOVA



기초분석

미네랄	평균	표준편차	상위 10% 정수장	하위 10% 정수장
Ca	17.637	15.058	황지 충주 단양 영춘 적성 용궁 지보 풍양	대야 청산 금당 보길 생일 회동 청룡 남동
Mg	3.474	1.907	충주 단양 적성 녘도 용문 지보 풍양 개포	동화 덕정 구천 밀양 양산 대야 금당 일운
Na	11.675	6.726	고령광역 반송 파주 다시 녀도 육동 관매도 감천	동화 별량 화순 운문 사천 밀양 양산 영춘
K	2.127	1.15	구미 고령광역 반송 파주 광주 함평 용문 감천	대야 약산 청산 소안 금당 회동 청룡 일운
Si	2.544	2.112	적성 다시 연지 예천 용문 지보 개포 고령지방	성남 수지 청주 천안 아산 덕정 반송 금일

● Group 1

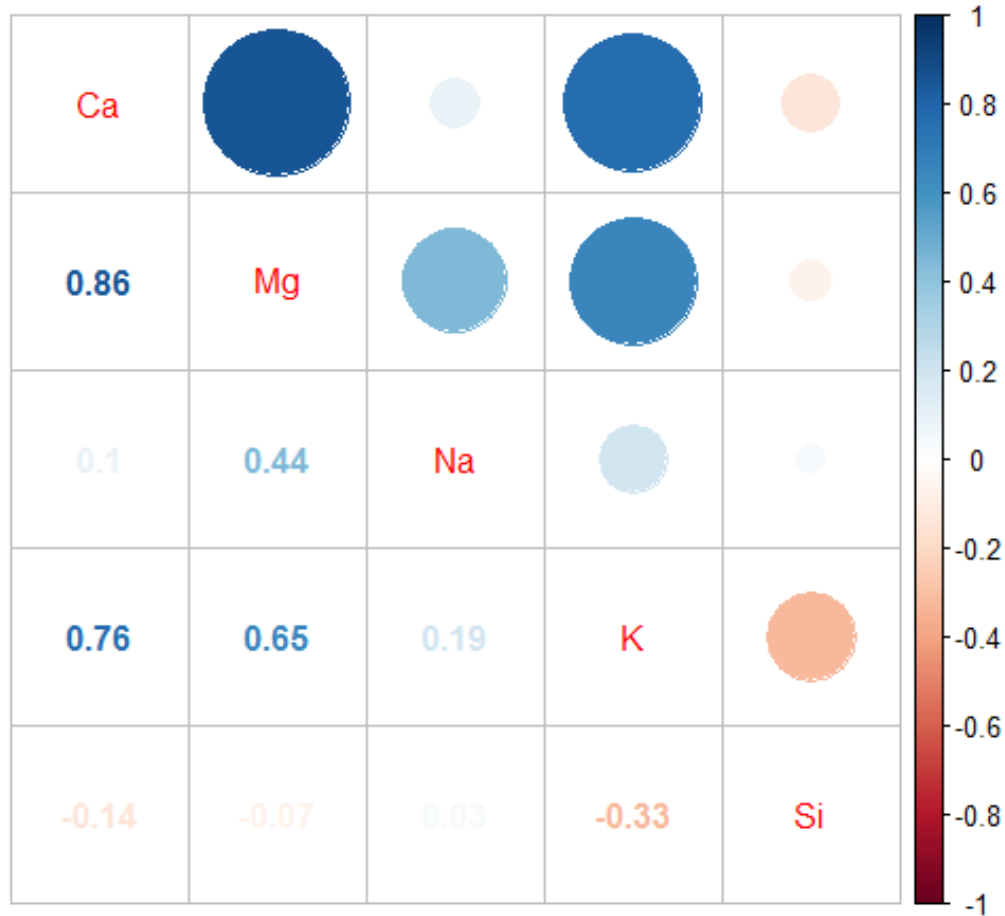
● Group 2

● Group 3

[표 1] 미네랄 기초분석표



상관분석 : 전체 그룹



- 높은 양의 상관관계 : (Ca, Mg), (Ca, K), (Mg, K)
- 음의 상관관계 : (K, Si)

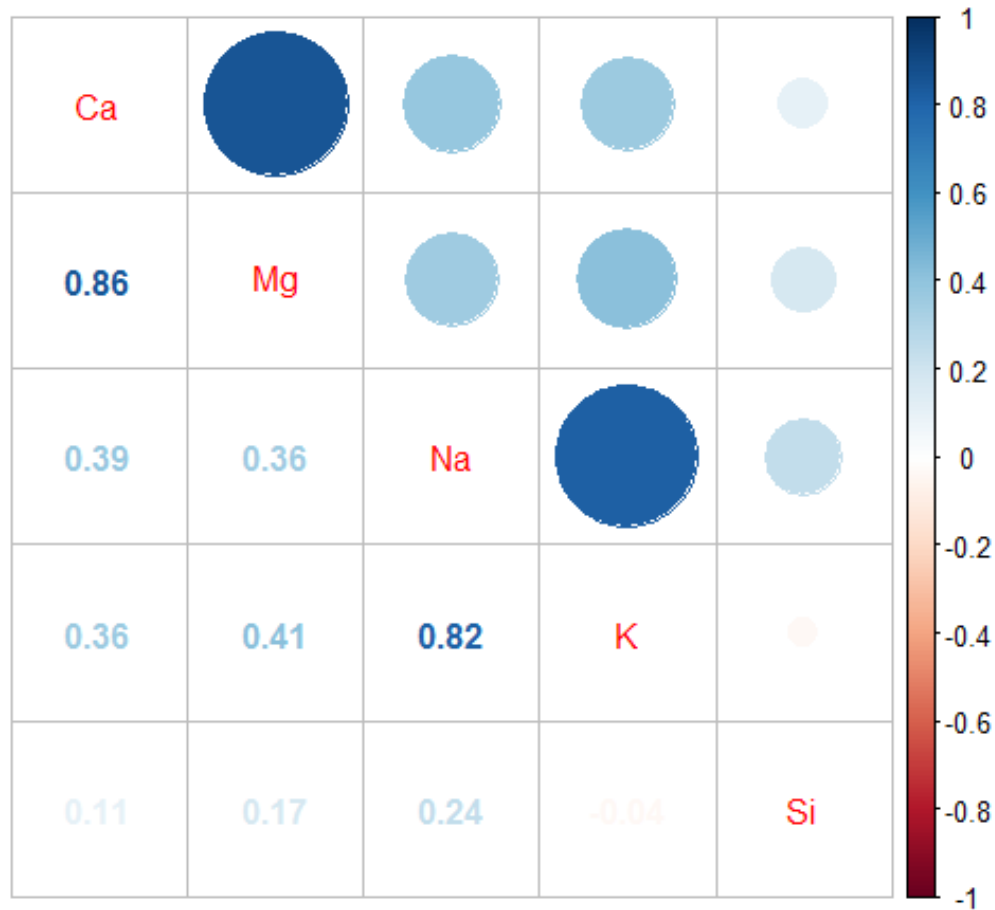
전체	Ca	Mg	Na	K	Si
Ca					
Mg	0.86***				
Na	0.10	0.44***			
K	0.76***	0.65***	0.19*		
Si	-0.14	-0.07	0.03	-0.33**	

Signif.codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

[그림 2] 전체 그룹의 상관관계



상관분석 : 그룹 1



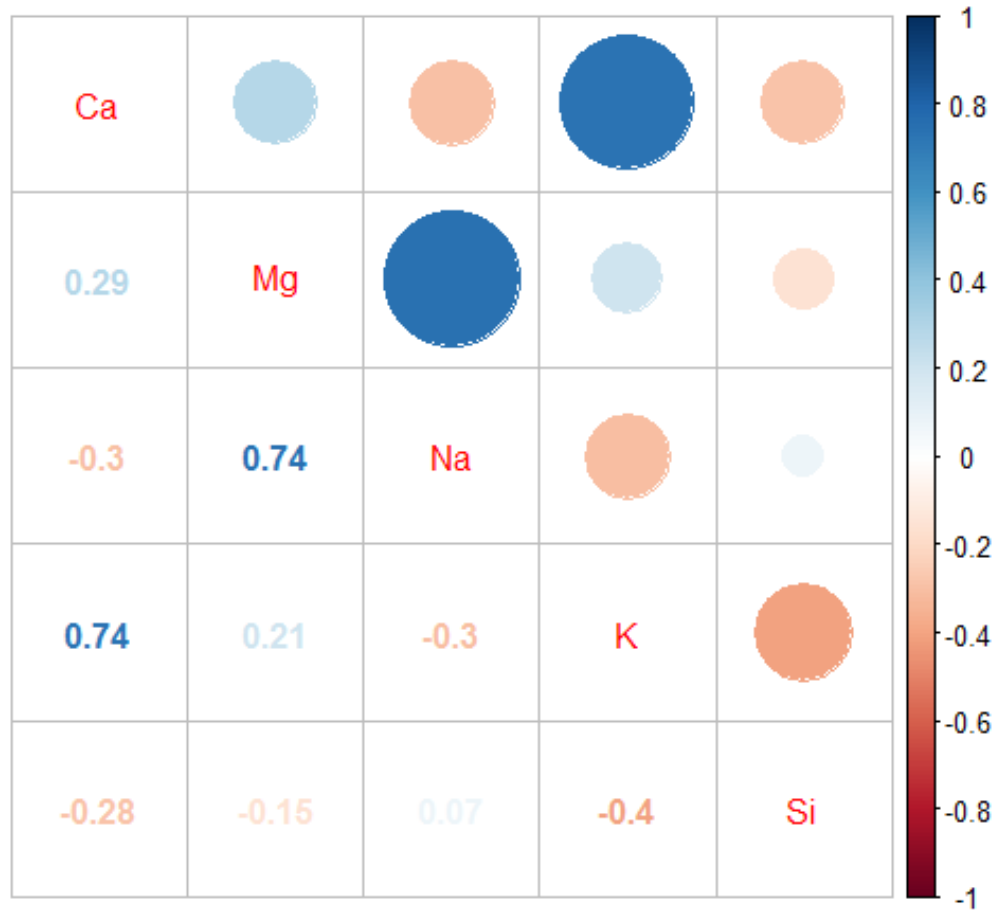
· 높은 양의 상관관계 : (Ca, Mg), (Na, K)

그룹1	Ca	Mg	Na	K	Si
Ca					
Mg	0.86***				
Na	0.39*	0.36*			
K	0.36*	0.41*	0.82***		
Si	0.11	0.17	0.24	-0.04	

[그림 3] 그룹 1의 상관관계



상관분석 : 그룹 2



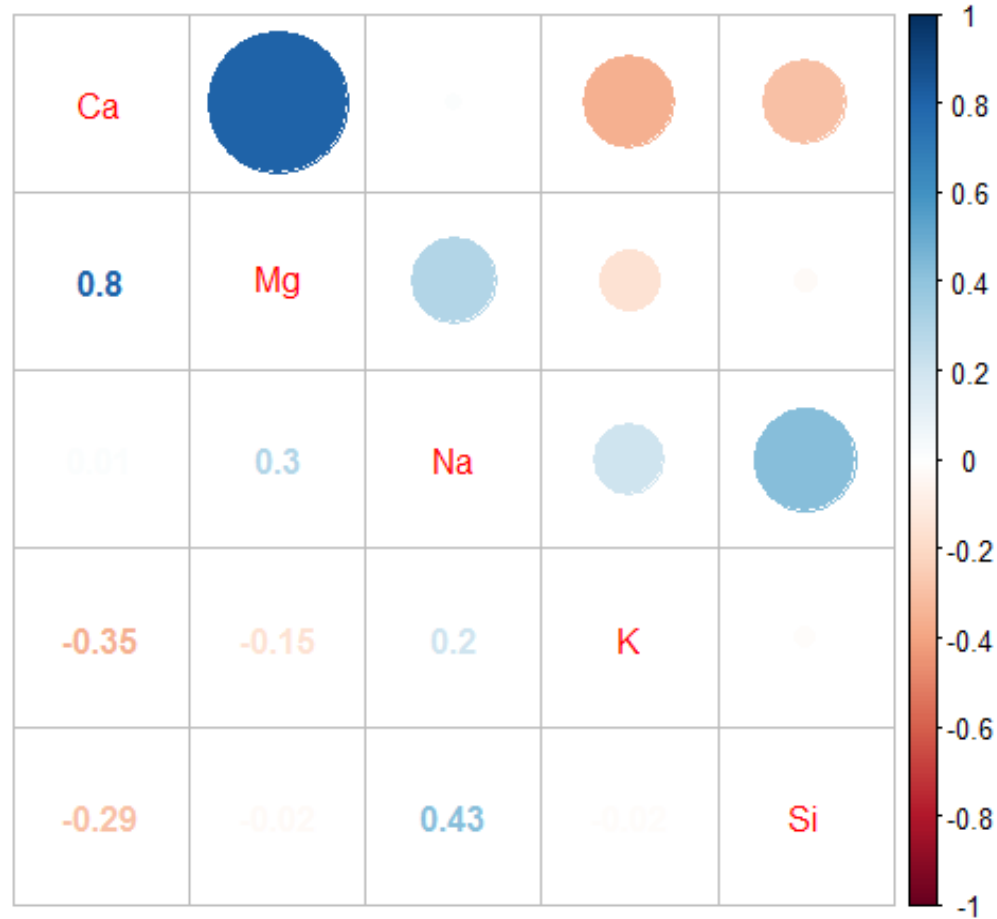
- 높은 양의 상관관계 : (Mg, Na), (Ca, K)
- 낮은 음의 상관관계 : (Ca, Na), (Na, K)

그룹2	Ca	Mg	Na	K	Si
Ca					
Mg	0.29 [*]				
Na	-0.30 [*]	0.74 ^{***}			
K	0.74 ^{***}	0.21	-0.30 [*]		
Si	-0.28	-0.15	0.07	-0.40 ^{***}	

[그림 4] 그룹 2의 상관관계



상관분석 : 그룹 3



· 높은 양의 상관관계 : (Ca, Mg)

그룹3	Ca	Mg	Na	K	Si
Ca					
Mg	0.80**				
Na	0.01	0.30			
K	-0.35	-0.16	0.20		
Si	-0.29	-0.02	0.43	-0.02	

[그림 5] 그룹 3의 상관관계



주성분 분석

[표 2] 주성분 변수

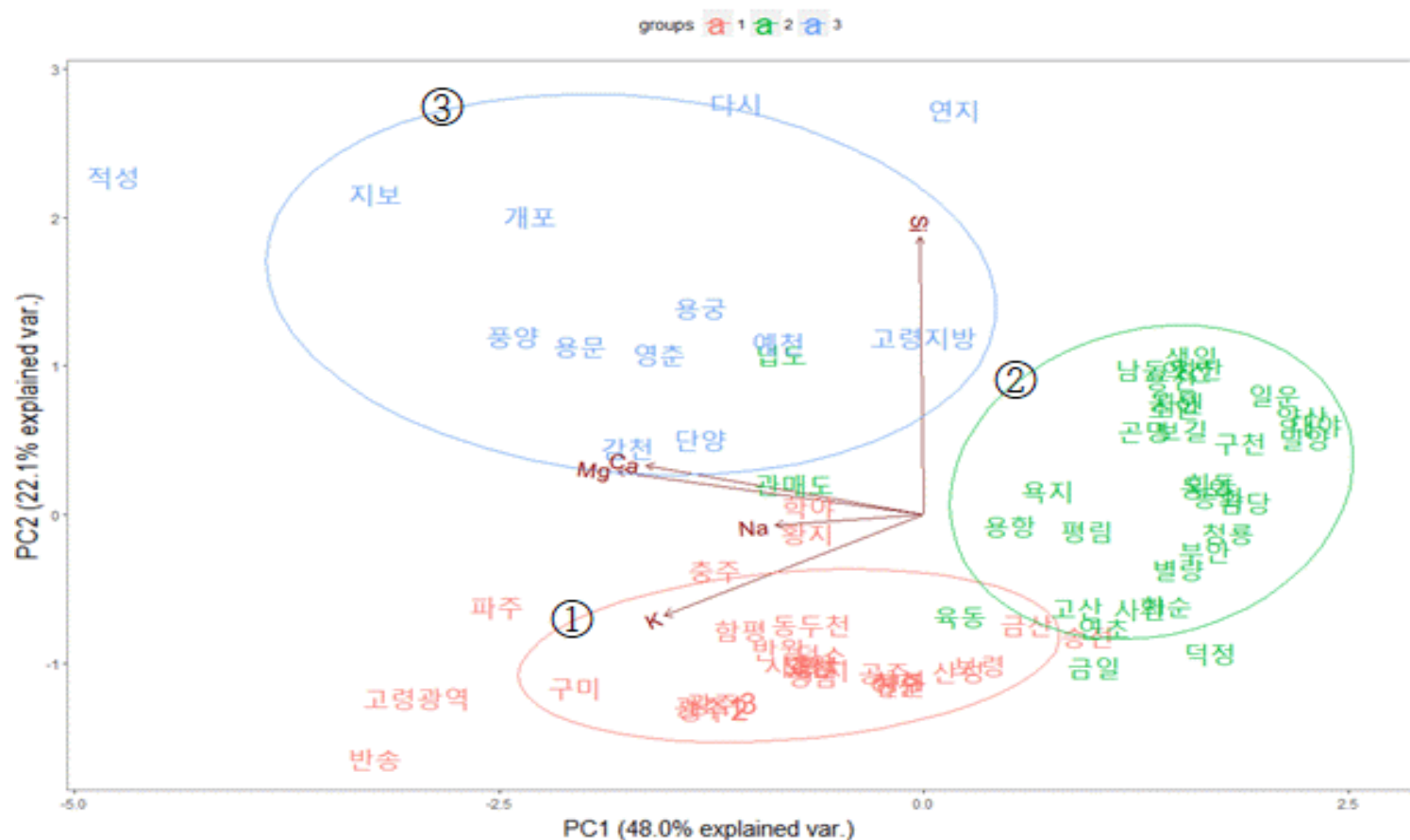
Importance of components	PC1	PC2	PC3	PC4	PC5
Standard deviation	1.550	1.052	0.974	0.636	0.377
Proportion of Variance	0.480	0.221	0.190	0.081	0.028
Cumulative Proportion	0.480	0.701	0.891	0.972	1.000

· 데이터의 총 분산의 70%를 설명하는 PC1, PC2 선택

[표 3] 주성분 변수의 상관관계

	PC1	PC2	PC3	PC4	PC5
Ca	-0.54	0.16	-0.41	0.31	-0.64
Mg	-0.60	0.14	-0.03	0.32	0.72
Na	-0.29	-0.03	0.90	0.16	-0.27
K	-0.51	-0.33	-0.04	-0.79	-0.01
Si	-0.01	0.92	0.10	-0.39	-0.01

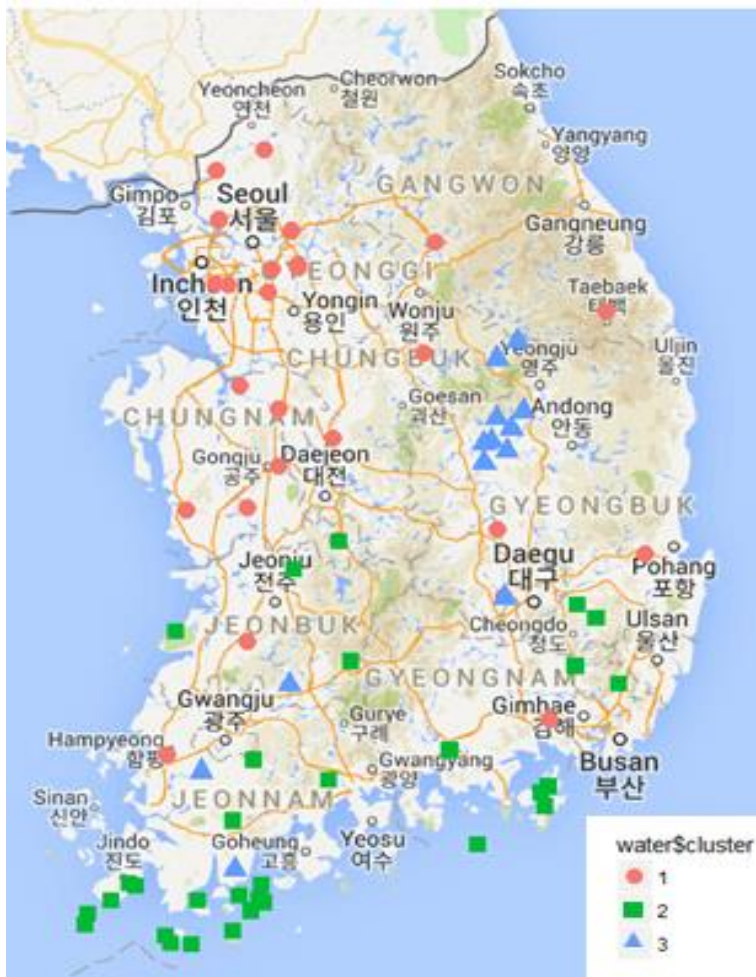
- PC1 : Ca, Mg, Na, K 음의 방향으로 영향력이 높음
- PC2 : Si 양의 방향으로 영향력이 높음



- PC1 : [그룹 1, 3] ↔ [그룹 2] 구별
- PC2 : [그룹 1] ↔ [그룹 3] 구별

- **그룹 1** : Si ↓
- **그룹 2** : Ca, Mg, Na, K ↓
- **그룹 3** : Si ↓ / Ca, Mg, Na, K ↑

[그림 6] 주성분 변수의 상관관계별 정수장



[그림 7-a] 그룹별 정수장 위치



[그림 7-b] k-water 권역부문별 지도

군집분석 결과
그룹별 정수장 위치
(동쪽, 서쪽, 남쪽)



k-water 권역부문과
상당히 일치



ANOVA : 미네랄비교

요인	F	P-value
칼슘(Ca)	53.8	$< 0.001^{***}$
마그네슘(Mg)	47.8	$< 0.001^{***}$
나트륨(Na)	0.28	0.76
칼륨(K)	61.4	$< 0.001^{***}$
규소(Si)	72.3	$< 0.001^{***}$

평균의 차이가 **있는** 그룹
: Ca, Mg, K, Si

평균의 차이가 **없는** 그룹
: Na

[표 4] 미네랄별 분산분석표



ANOVA : 미네랄비교

[표 5] 그룹 간 미네랄 평균

	칼슘(Ca)	마그네슘(Mg)	나트륨(Na)	칼륨(K)	규소(Si)
그룹1	22.2 ⑥	4.13 ⑥	12.4 ①	3.03 ①	0.74 ③
그룹2	6.18 ③	2.02 ③	11.1 ①	1.13 ⑥	2.90 ⑥
그룹3	37.4 ①	5.81 ①	11.6 ①	2.73 ①	5.63 ①

- Ca, Mg, Si : 모든 그룹 간에 차이
- K : (그룹1, 그룹2) , (그룹2, 그룹3) 에서 차이



ANOVA : 질병비교

요인	F	P-value
고혈압	3.98	0.023 [*]
당뇨병	8.2	0.0006 ^{***}

[표 6] 고혈압, 당뇨병에 대한 분산분석표

요인	chi-squared	P-value
관절염	10	0.003 ^{**}

[표 7] 관절염에 대한 분산분석표(Kruskal-Wallis)



ANOVA : 질병비교

	고혈압	당뇨병	관절염
그룹1	25.7 ⑥	10.5 ⑥	19.3 ⑥
그룹2	30.0 ①	12.3 ①	25.5 ①
그룹3	28.5 ①⑥	13.8 ①	22.1 ①⑥

[표 8] 그룹간 질병 경험률 평균 비교

- 고혈압, 관절염 : (그룹1, 그룹2) 에서 차이
- 당뇨병 : (그룹1, 그룹2) , (그룹1, 그룹3) 에서 차이

결론:

정리 | 시사점

03 결론

정리

Ca과 Mg 상관도 大

02

그룹 1 :
가장 낮은 질병률



상관분석



군집 특성



질병률 비교

01

- 1 한강
- 2 금강 영산강 섬진강
- 3 낙동강

03

03 결론

시사점

녹조에 과불화화합물까지...낙동강 식수 관리 '비상'

'과불화화합물 수치' 기존보다 높게 나타나...대형마트서 생수 품귀현상도
환경부 "국외 사례와 비교하면 안전 수치"...환경단체, 정부·지자체에 대책 촉구

대구 수돗물 발암물질 논란

사용 문제 없어 vs 못 믿겠다
수질개선 근본적인 대책 필요

대구 '발암물질 수돗물' 소식에 대형마트 생수 '품절' 사태

정수원을 파악하거나

유해한 성분들과 질병의 관계를 알 수 있을 것으로 기대

+ 출처

1) 수자원공사

지역별 수돗물 미네랄 함유량

2) KOSIS(국내통계-주제별통계-보건·복지-보건-지역사회건강조사-이환)

- 시군구별 고혈압 진단 경험률(30세 이상)
- 시군구별 당뇨병 진단 경험률(30세 이상)
- 시군구별 관절염 진단 경험률(50세 이상)

감사합니다