

데이터 분석 포트폴리오

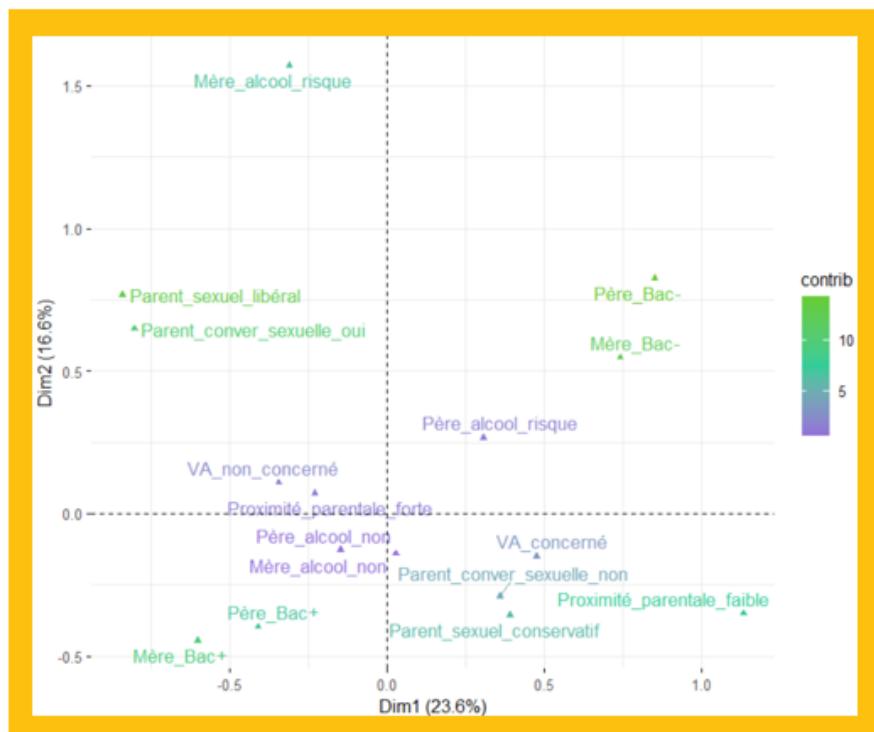
주요 분석	- 고차원 데이터 : 약 27,300 행 x 4,000 열 (약 1 억 셀 규모)
데이터셋	- 고차원 데이터 : 약 25,000 행 x 1,100 열 (약 2,750 만 셀 규모) - 중간규모 데이터(자체 구축) : 약 1,000 행 x 320 열 (약 32 만 셀 규모)
주요 언어	- R, Python, SQL

포트폴리오의 분석 사례들은 응용데이터분석의 논리적 구조를 따름(데이터 탐색 → 변수 식별 및 유형화 → 상관분석 → 다중회귀분석)

분석 사례 1) 다중대응분석(MCA)을 통한 다변량 패턴 식별

- biplot 차원 상에서 다양한 범주형 변수들 간의 연관성과 분포를 식별
- 각 차원에서 근사하게 위치한 변수그룹들의 통계적 유의성 식별
- 분석의 응용 범위:
 - 구매행동과 관련성이 높은 소비자 행동패턴 또는 소비자 특성의 군집화를 통한 유형별 전략 수립
 - 비용절감과 관련된 기술·아키텍처·인프라수준 등 다변량 요인의 조합 식별
 - ROI 와 QCD 에 영향을 미치는 다양한 성과지표 간의 복합적인 조합 탐색

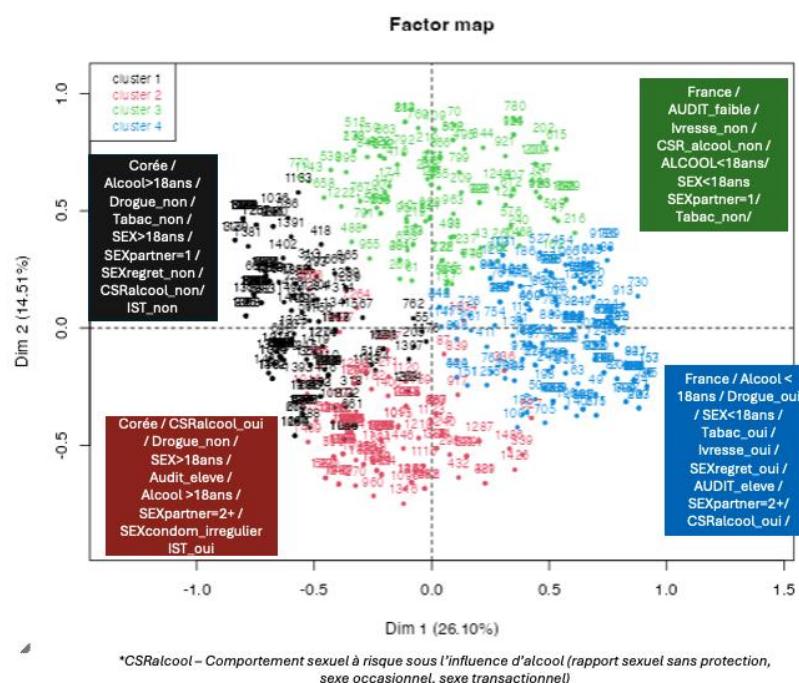
(출처 : 박사과정 졸업논문, 불문)



분석 사례 2) 위계적 클러스터링 분석을 통한 유형화(프로파일링) 수행

- biplot 차원 상에서 유사한 특성을 공유하는 관측치들을 클러스터링
- 각 클러스터가 공유하는 특성 중 통계적으로 유의도가 높은 변수들을 기준으로 유형화
- 분석의 응용 범위
 - 다변수를 활용한 세그먼테이션 수행, 유형별로 차별화된 마케팅 전략 또는 KPI 개발

(출처 : 프랑스국립과학연구원 CNRS 제출용 박사과정 연구계획서, 불문)



분석사례 3) 영향변수를 통제한 상태에서 변수 간 선형상관관계 분석 (편상관분석)

- 통제변인의 영향을 제한한 상태에서 두 계량 데이터 간의 선형관계를 측정
- 분석의 응용 범위:
 - 다양한 영향 변수를 통제한 상태에서 투자와 매출 증대의 상관관계 측정
 - ROI 평가도구 개발 시 다중회귀모델 구축 전단계에서 통계적 상관관계가 높은 성과지표 선정에 활용

(출처 : 국제학회 논문 프로시딩 및 포스터, 영문)

	Age of A.I.	AUDIT-C score	Number of S.D.	Age of A.I.	AUDIT-C score	Number of S.D.	
Spearman Correlation	-1.0 0.0 1.0						
Age at FSX	0.17***	-0.14***	-0.14***	Age at FSX	0.28***	-0.16***	-0.10***
Number of S.P.	-0.02	0.19***	0.26***	Number of S.P.	-0.12***	0.19***	0.23***
Inconsistent CU	-0.07***	0.15***	0.12***	Inconsistent CU	-0.07**	0.06*	0.14***

-p^*<0.05, p**<0.01, p***<0.001, Control variables: age, gender, occupation, level of education, subjective socioeconomic status

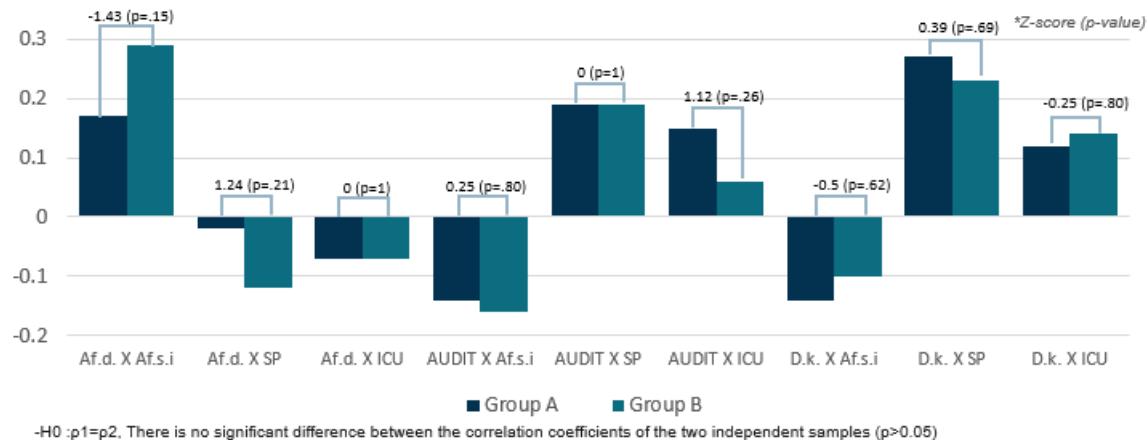
분석 사례 4) 두 개의 독립된 집단에서 측정한 상관계수의 차이 비교 (z-test)

- 두 집단 간 편상관분석 계수에 대한 표준화된 z-score 비교

■ 분석의 응용 범위:

- 신규 시스템 도입이 불량률 감소에 미치는 효과가 모든 공장에서 동일한지 분석
- 시스템 도입에 대한 표준 솔루션 수립 판단

(출처 : 국제학회 논문 프로시딩 및 포스터, 영문)



분석 사례 5) 다변량 예측변수 및 상호작용항을 포함한 다중회귀모델 개발

- 통제변인의 영향을 제한한 상태에서 다양한 예측변수들과 결과변수의 상관관계에 대한 다중 회귀 분석
- 결과변수가 수치형 또는 범주형일 때의 예측값 계산, 상호작용항을 추가하여 카테고리별 예측값 도출

■ 분석의 응용 범위:

- D2C 플랫폼 내에서 다양한 디지털 행동데이터에 따른 구매 여부·재방문 여부 등에 대한 예측값 도출
- 투자금의 증가에 따른 매출의 '순수' 증대 예측값 도출
- QCD 범주형 성과지표(예, 품질 양호-불량)에 따른 투자가 매출 증대에 미치는 효과 산출

(출처 : 박사과정 졸업논문, 불문)

