

데이터 03 마이크로데이터와 코호트

(주)씨에스리 조경미 기술사
(kmicho@cslee.co.kr)

마이크로데이터와 코호트

KeyWord

마이크로데이터, 원시데이터, 매크로데이터, 코호트(cohort), 코호트분석, 기간공유, 경험공유, 특정요인 경험

인공지능과 빅데이터 기술이 활용되는 분야에 대한 기사와 논문, 관련된 연구 들이 어떻게 진행되고 있는지 알아보면서, 노령사회의 의료 정책 개선, 사회보장을 위한 공공 데이터 오픈 등의 연구 과제 등에서 **코호트 분석과 마이크로데이터**라는 용어를 새롭게 알게 되었으며, 이에 그 기본 개념을 알아보고자 한다.

1. 개별 단위의 정제 데이터, 마이크로데이터(MicroData)

- 통계조사 원자료에서 조사,입력 오류 등을 수정한 개인, 가구, 사업체 등의 개별단위의 자료이다.

원시 데이터(Raw Data)	마이크로 데이터(Micro Data)	매크로 데이터(Macro Data)
입력 첫 단계이 자료로서 입력오류나 조사오류 등이 걸러지기 이전 단계의 자료	오류를 수정한 자료로, 통계표 작성 등 데이터 가공의 바탕이 되는 자료	마이크로데이터를 임의의 기준에 따라 집계한 자료로 집계의 정도에 따라 세분된 자료부터 통합된 자료까지 다양하게 제공

마이크로데이터는 공공이용을 위해 일반 이용자에게 제공하기 이해 응답자가 식별, 노출되지 않도록 자료처리를 한 **공공이용 마이크로데이터(Public use microdata)**와 자료관리기관의 승인을 얻어야 이용이 가능한 **승인된 마이크로데이터(Licensed microdata)**로 분류되기도 한다.

통계청에서는 정부 3.0 추진 과제의 하나로 마이크로데이터 통합시스템(<http://mdis.kostat.go.kr>) 서비스포털을 구축하여 국가 주요 정책의 수립, 기업 경영전략 수립, 학술 논문 등 심층 연구분석에 활용되는 마이크로데이터를 효율적으로 활용하기 위해, 정부 각 부처, 지자체, 연구기관 등 타 통계작성기관의 마이크로데이터를 하나의 서비스로 제공하고 있다.



이와 같이 구축된 마이크로데이터를 이용하거나 별도로 특정 요인에 노출된 집단의 주제별로 데이터를 구축한 코호트(Cohort)와 코호트 연구(Cohort Study)에 대해 알아보자.

2. 주제 중심 특성 분석을 위한 코호트(Cohort)와 코호트연구(Cohort Study)

코호트(Cohort)	코호트 연구(Cohort Study)
-조사 연구와 인구학적 연구에서 특별한 기간 내에 출생하거나 조사하는 주제와 관련된 특성을 공유하는 대상의 집단	- 특정 요인에 노출된 집단과 노출되지 않은 집단을 추적하고 연구 대상 질병의 발생률을 비교하여 요인과 질병 발생 관계를 조사하는 연구 방법
정의를 보고 생각나는 코호트는 <u>386 세대, X 세대, Y 세대 등이 생각한다</u> Ⓢ	역시 정의를 보고 생각나는 코호트 연구는 <u>일본 후쿠시마 원전과의 거리별 질병 발생률을 조사할 수 있을 것 같다</u> Ⓢ

코호트 연구는 노출에 대한 정보를 수집하는 시점에 따라 전향적 코호트 연구와 후향적 코호트 연구로 나뉜다.

3. 코호트 연구 (prospective cohort study)의 유형

가. 전향적 코호트 연구 (prospective cohort study)

- 현재 시점, 즉 코호트가 정의된 시점에서 노출에 대한 자료를 수집하여 가장 최신의 자료를 기반으로 하여 왜곡(misclassification bias)을 최소화하는 연구 기법으로 노출시점부터 질병이 발생하는 시점까지의 기간이 너무 긴 경우에는 한계점이 있다.

나. 후향적 코호트 연구(historical/retrospective/resconstructed cohort study)

- 노출시점부터 질병이 발생하는 시점까지의 기간이 긴 경우 활용되는 연구로, 연구가 계획되기 훨씬 이전에 이미 노출 여부를 측정된 과거의 의무기록, 고용기록 등의 자료를 이용하여 노출이 질병 위험에 효과를 낼 때까지 기다려야 하는 시간을 줄이는 방법을 택하기도 한다.

일반인 집단이나 질환군 대상의 코호트 사업과 유전,환경 상호작용 연구 등을 위한 코호트 사업은 질병 발생 원인이 되는 위험요인 규명을 위한 정보를 제공하고, 맞춤 의학 구현과 국민건강 증진을 목적으로 질병관리본부와 국립보건연구원에서 추진하고 있다.

4. 질병관리본부, 국립보건연구원의 코호트 사례

일반인대상 코호트	지역사회 기반 코호트 (안산-안성)	농촌 기반 코호트	도시 기반 코호트
환자군 코호트	HCV에 의한 간질환 코호트	HIV/AIDS 코호트	결핵 고위험군 코호트
	HPV 코호트	결핵 고위험군 코호트	알레르기 비염 코호트
	노인 천식 코호트	알레르기 비염 코호트	심부전 질환 코호트
	급성심근경색 질환 코호트	심부전 질환 코호트	
특수집단 및 모델 코호트	국내 이주자 및 국제 협력 코호트 II	국제 협력 코호트	
	쌍둥이 및 가족 코호트	소아호흡기-알레르기질환 장기추적 코호트	
	소아비만 및 대사질환 코호트	노인질환 예방관리 코호트	
	갱년기 여성건강 코호트	간호사건강연구 코호트	

질병관리본부는 고령화에 따라 급증하고 있는 치매의 진단 마커 및 위험요인을 발굴하고, 예방, 진단, 치료 기술을 개발하기 위한 치매환자를 대상으로 코호트 연구를 진행하고 있으며, 2010 년터 65 세 이상 알츠하이머 치매 노인을 대상으로 혈장, DNA, 뇨 등 생물 자원을 수집하여 코호트를 구축하였다.

이와 같이 구축된 마이크로데이터와 코호트는 의료질병연구에 적극 활용되며, 인공지능 알고리즘과 융합하여 보다 정밀하고 빠른 의료진단시스템을 개발하는데 활용된다.

[참조] 질병관리본부 코호트 연구

<http://cdc.go.kr/CDC/contents/CdcKrContentView.jsp?cid=24543&viewType=CDC&menuIds=HOME001-MNU1215-MNU2515-MNU1217-MNU1308>

[참조] 지능정보사회 중장기 종합대책 (미래창조과학부) 2017.02

Contents connect communications!!

아이리포에 오시면 더 많은 지식을 가져가실 수 있습니다.

아이리포 IT지식창고 : <https://www.ilifo.co.kr/boards/knowledge>

아이리포 지덤시리즈 : <http://www.jidum.com>

아이리포 기술사/감리사 카페 : <http://cafe.naver.com/itlf>

서울시 마포구 상암동 1610번지, DDMC 3층 아이리포 교육센터

TEL: 02-303-9997 | MAIL: edu@ilifo.co.kr