

보건의료 빅데이터를 활용한 보건행정 교육프로그램 실무수업에 관한 고찰

양옥렬*, 이연희
혜전대학교 보건의료행정과

A Study on Practical Classes for Healthcare Administration Education Program Using Health and Medical Big Data

Ok-Yul Yang*, Yeon-Hee Lee
Department of Healthcare & Medical Administration, Hyejeon University

요약 본 연구는 보건의료빅데이터를 활용한 보건의료행정 관련학과에서 빅데이터 관련 교육프로그램 활용 가능성에 대한 고찰이다. 본 논문에서는 크게 5가지 관점으로 보건의료빅데이터를 고찰하고자 한다. 첫째, '보건의료빅데이터개방시스템' 이외에도 '한국복지패널', '공공빅데이터', '서울시', '통계청' 등에서 공개하고 있는 공공빅데이터의 특징 및 활용기술에 대하여 살펴보고자 한다. 둘째, 해당되는 보건의료 빅데이터를 전문대학 보건의료행정 및 보건의료정보 관련학과의 정규과목 내에서 살아있는 데이터로써 활용이 가능한지에 대한 적정성 여부를 살펴보고자 한다. 셋째, 기존에 활용하고 있는 통계처리패키지 및 프로그래밍 언어를 활용한 실습실 강의 여부에 가장 적절한 툴을 선정하고자 한다. 넷째, 검증된 보건의료빅데이터와 적정 툴을 활용하여 수업에서 그래프 등의 표현가능성 여부 및 보고서 작성까지의 단계를 시험해보고자 한다. 마지막으로, 최종적으로 휴대성, 설치성, 비용성, 호환성, 빅데이터처리가능성 등을 만족할 수 있는 R언어의 상대적 이점을 기술하고자 한다.

Abstract This study is a study on the possibility of using big data-related education programs in health and medical administration-related departments using health and medical big data. This paper intends to examine the health and medical big data from five perspectives. 1st, in addition to the aforementioned 'Health and Medical Big Data Open System', I would like to examine the characteristics and application technologies of public big data disclosed by 'Korea Welfare Panel', 'Public Big Data', 'Seoul City Big Data', 'Statistical Office Big Data', etc. 2nd, it is intended to examine the appropriateness of whether the applicable health and medical big data can be used as living data in regular subjects of health and medical administration and health information related departments of junior colleges. 3rd, we want to select the most appropriate tool for classroom lectures using existing statistical processing packages and programming languages. Fourth, finally, by using verified health and medical big data and appropriate tools, we want to test the possibility of expressing graphs, etc. in class and the steps from writing a report. 4th, I would like to describe the relative advantages of R language that can satisfy portability, installability, cost effectiveness, compatibility, and big data processing potential.

Key Words Open, Health, Bigdata, Administrator, Education

1. 서론

보건복지부(MOHW.go.kr)와 건강보험심사평가원

가원(HIRA.or.kr)이 공동 주관한 지난 2022년 4월 19일부터 6월 7일까지 50일간에 걸쳐 '도전, 공공데이터 아이디어' 및 '제품 및 서비스 개발'이라는 부분에 걸친 공모전을 실시하고 있다. 이는 건강보험심사평가원(이하 심평원)은 구축

*Corresponding Author : Ok-Yul YANG(Hyejeon Univ.)

E-mail: cache@hj.ac.kr

Received June 29, 2022

Revised July 05, 2022

Accepted July 18, 2022

된 보건의료 빅데이터를 보다 획기적인 비즈니스 모델 개발 및 활용을 위해 ‘2022년 보건의료 빅데이터를 활용한 창업경진대회’의 세부 과제의 형식으로 개최된 것이다. 위의 2부문의 공모전의 주요 목적은 보건의료 빅데이터를 축적하는데 머무르지 않고 ‘사회문제 해결’, ‘실용화 서비스 제품개발’, ‘기존서비스 제품 개선’ 등 3가지이다. 모두 보건의료 빅데이터를 활용하여 기존 산업과의 경계를 허물고 융복합 서비스를 개발하고자 하는 정부의 의지를 담고 있다고 판단된다[1-3]. 해당 경진대회의 발표심사 및 시상식 등의 최종결과는 2022년 7월 말로 예정되어 있다.

국내 대표적인 보건의료 빅데이터 개방 사이트는 ‘보건의료 빅데이터 개방시스템 (opendata.hira.or.kr)’이라고 할 수 있을 것이다. 심평원에서 운영관리하는 해당 시스템은 운영관리에 있어서 국민의 개인정보와 민감정보를 모두 포함하고 있어서 개방을 위해서는 전처리 작업으로 민감정보를 비식별화하는 과정을 거쳐 개방하고 있다. 또한 현재까지는 대학, 연구소, 연구원, 석·박사학위 과정의 대학원생 등 특정한 목적을 가진 집단이나 개인에게 해당 연구의 목적성을 확인하고 일정한 비용을 지불한 후 사용할 수 있도록 하고 있다.

본 논문에서는 크게 3가지 관점으로 보건의료 빅데이터를 고찰하고자 한다.

첫째, 앞서 언급한 ‘보건의료빅데이터개방시스템’ 이외에도 ‘한국복지패널’, ‘공공빅데이터’, ‘서울시’, ‘통계청’ 등에서 공개하고 있는 공공빅데이터의 특징 및 활용기술에 대하여 살펴보고자 한다. 둘째, 해당하는 보건의료 빅데이터를 전문대학 보건의료행정 및 보건의료정보 관련학과의 정규과목 내에서 살아있는 데이터로써 활용할 수 있는지에 대한 적정성 여부를 살펴보고자 한다. 셋째, 기존에 활용하고 있는 통계처리패키지 및 프로그래밍 언어를 활용한 실습실 강의 여부에 가장 적절한 틀을 선정하고자 한다.

넷째, 마지막으로 검증된 보건의료 빅데이터와 적정 틀을 활용하여 수업에서 그래프 등의 표현

가능성 여부 및 보고서 작성까지의 단계를 시험해보고자 한다. 다섯째, 최종적으로 휴대성, 설치성, 비용성, 호환성, 빅데이터 처리 가능성을 만족할 수 있는 대학교육에서의 R언어의 상대적 이점을 기술하고자 한다.

2. 본론

2.1 관련연구

2.1.1 보건의료 빅데이터의 개념

보건의료 빅데이터는 의료기관을 이용한 환자에 대한 진료정보, 처방정보, 건강검진정보 등을 획득(수집), 저장, 관리되고 있는 데이터들의 집합이라고 정의할 수 있다[4]. 이렇게 구축된 보건의료 빅데이터는 인공지능(AI), 정보통신기술(ICT), 딥러닝(Deep Learning) 등의 기술을 융합하여 국민건강정보의 예측, 의료의 질평가, 보건관련 현장 전문가 및 공무원 등이 해당 빅데이터를 해석, 통계, 그래프를 활용한 시각화를 통해 전문가나 일반인이 이해할 수 있는 보고서 작성이 가능하게 해준다. 보건의료 빅데이터는 다양한 분야의 전문가 협업을 통해 최종적으로는 빅데이터가 국민건강 및 보건산업에 활용할 수 있는 객관성 보장, 사고력과 통찰력을 배양할 수 있게 한다. 본 논문에서는 이러한 전문가나 공무원이 되거나 이들을 지원할 수 있는 전문대학 보건관련학과 학생들이 최소한의 빅데이터의 개념과 실무 데이터 처리능력, 결과 데이터의 이해, 분석, 리포팅 능력을 배양할 수 있는 실무교육 프로그램에 대한 프로토타입을 개발하고자 한다.

2.1.2 보건의료빅데이터의 교육분야 활용

국가차원에서의 보건의료 빅데이터를 활용하려는 전략과 계획은 2017년 이후 의미있는 발걸음을 내딛기 시작했다. 2017년 5월, 개인정보보호법의 시행과 함께 ‘익명가공정보제도’를 통해 의료분야 데이터 이용 활성화에 기초작업이 만들어지지 시작했다. 이후, 과학기술정보통신부의 ‘빅데이터 플랫폼 및 센터구축’ 사업에서 10개의

산업분야별 100개의 연계 센터를 구축하는 사업을 진행하였다. 2021년 3월, 과학기술정보통신부는 향후 3년여에 걸쳐 100억원의 투자를 통한 ‘마이데이터(MyData)’ 사업을 추진하고 있다. 이는 보건의료 정보가 생성되는 정보주체를 중심으로 흩어져 있는 개인 데이터를 한 곳으로 모으고 개인이 직접 열람, 저장하고 통합관리하는 맞춤형 개인데이터 플랫폼서비스(PDS)를 개발하는 것이다. 여기에 포함되는 4대분야(금융/의료/생활소비/교통)에도 개인의 보건의료데이터 활용이 가능해질 것이다.

2019년 5월, ‘바이·헬스산업 혁신전략’을 발표하고 국민건강 및 생명보장을 위한 핵심선도사업 육성을 시작하였다. 해당 사업의 일환으로 ‘국가 바이오 빅데이터’ 구축관련 R&D 투자에 2025년까지 4조원 이상의 투자를 진행 중이다. 2021년 11월, 보건복지부(보건의료데이터진흥과)에서는 ‘미래지향형 보건의료데이터가 흐르는 혁신생태계 창출’이라는 추진전략 하에 ‘FLOW2025’를 발표하였다. 이는 향후 5년간 ‘보건의료데이터 표준화 로드맵’을 구축하는데 역점을 둔 사업이다.

이러한 국가적인 공공의료 빅데이터에 활용될 수 있는 데이터를 보유한 기관이나 플랫폼으로는 보건의료분야마이크로데이터(MDIS), 국민건강보험공단빅데이터(NHIS), 건강보험심사평가원빅데이터(HIRA), 암빅데이터플랫폼(NCC), 종합·상급종합병원 등이 보유한 임상빅데이터 등이 존재한다. 이밖에도 서울시열린데이터광장, 통계청자료(Kostat), 공공데이터포털(Data.go.kr), 한국복지패널(KOWEPS) 등에서 쉽게 데이터에 접근이 가능하다.

2.2 보건의료행정 전공교육용 플랫폼 제안

앞 절 언급한 바와 같이 공공의 이익과 편의를 위해 국가나 기업이 제공하는 다양한 빅데이터를 획득/수집/가공/저장/통계/분석/그래프/리포팅에 이르는 일련의 과정을 통해 의사결정권자(대표, 병원장 등)가 필요로 하는 의사결정 결과물을 생성할 수 있어야 한다. 그것이 공익적

빅데이터 개발의 취지라고 판단된다. 그렇다면 보건의료 관련 보험사나 손해사정회사 등의 기업이나 종합·상급종합병원, 국공립병원, 보건소 등의 의료기관, 보건복지부, 심평원, 한국보건복지인력개발원 등의 연구기관 등에서 데이터 운용에 필요한 인력을 대학에서 배출시켜야 할 것이다.

현재 전문대학 및 대학교의 보건행정 관련 학과에서의 보건의료빅데이터에 대한 교육은 최종학년이나 전공심화과정에서 학점을 이수하도록 교육과정이 설계된 경우가 대부분이다.

공공의료 빅데이터가 아무리 잘 구축되어 있어도 이를 활용할 수 있는 교육기관에서의 지속적인 인력 배출이 불가능하다면 그 의미를 상실한다고 판단된다. 따라서, 보건의료 빅데이터가 대학교육에서 가능하도록 하는 현실적인 교육 플랫폼은 반드시 필요하다. 또한, 보건행정관련 학과의 전공교육용 플랫폼을 위해서는 대학의 현실을 반영할 필요성도 있다. 대학의 사정에 따라 보건행정 관련 학과에는 전용 컴퓨터실(PC-LAB)을 운영하기도 하지만 공용컴퓨터실을 이용하기도 한다. 만일, 학과가 전용으로 사용하는 전공실습실이 아니라면 빅데이터 분석용 프로그램을 설치하고 사용하는 데에는 다소 어려움이 있다.

2.2.1 전공교육용 플랫폼 기본조건

기본적인 전공교육용 플랫폼 구축을 위한 항목을 나열해 보면 다음과 같이 7가지 핵심요소를 만족해야 한다고 판단된다.

첫째, 보건의료 빅데이터를 수집/획득/저장 등이 가능해야 하는데, 이는 전용실습실이나 공용실습실(PC-LAB)에서도 충분히 가능하다.

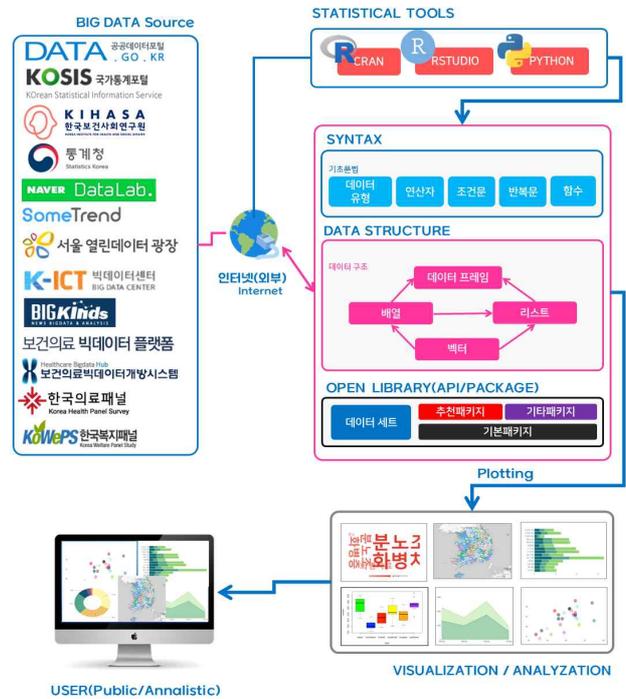
둘째, 공공으로 제공되는 보건의료 빅데이터의 배포 형식(포맷)이 확인 가능해야 한다. 수집/획득/저장이 완료된 데이터의 형식이 일반 프로그래밍 언어나 통계분석용 툴에서 적재(load)할 수 있어야 한다. 대부분의 공공 빅데이터는 크게 3가지 형식으로 데이터를 제공하는데 대표적인 통계분석용 툴인 STATA, SAS, SPSS의

형태이다.

셋째, 가공이 가능하고 재저장이 가능해야 한다. 획득된 빅데이터를 시각적으로 보면서 가공하여 재저장이 가능해야 한다. 데이터 포맷이 확인되고 가공 가능한 일반적인 통계분석용 틀에서 열리는 파일이라면 가공하고 재저장에는 큰 문제가 없다. 다만 앞서 언급한 통계분석용 틀인 STATA, SAS, SPSS 등의 구매비용이 크기 때문에 이를 전공실습실에 설치해 사용하는 학과는 많지 않은 실정이다.

넷째, 충분한 오픈 라이브러리를 제공받아야 한다. 이를 위해서는 오픈 라이선스이거나 셰어웨어나 프리웨어의 개념으로 사용할 수 있는 틀이나 프로그래밍 언어일 경우가 많다. 오픈API, 오픈라이브러리는 빅데이터를 분석하기 위한 분석기법을 사용자가 다른 사용자를 위해 제공하는 일종의 함수들의 집합이라고 할 수 있는데 최근에는 이러한 오픈라이브러리를 제공하는 사이트가 전 세계적으로 활발히 운영되고 있다. 파인리포트(FineReport), 차트블럭(Chartblocks, 시각화툴), Google Data Studio(시각화 및 대시보드툴), 오픈히트맵(Openheatmap), 파워비아이(Microsoft Power BI), 파이썬(Python, 대화식 데이터 시각화 툴), D3(Data Driven Document), 리플릿(Leaflet, 지도형 차트 기능), 오픈리파인(Open Refine), 태블로(Tableau, 오픈 시각화툴), 크랜(CRAN), 깃허브(GitHub) 등이 해당된다[10].

다섯째, 비용적 측면을 고려한 틀을 선정해야 한다. 저렴하거나 무료형식의 틀이어야 한다. 비용적인 측면에서 교육용버전, 무료버전으로 활용할 수 있는 틀이어야 한다. 빅데이터 분석은 쉽게 결과를 도출할 수 없기 때문에 학생들이 수업시간에 전공실습실에서 충분한 훈련학습이 이루어지지 않았거나 과제물을 집이나 기숙사 등에서 연속적으로 학습하고자 할 때, 개인용 노트북이나 PC 혹은 PC방 등의 공용장소에서도 설치 가능한 틀이어야만 제대로 된 연속학습이 가능하다고 판단된다.



[Fig. 1] Example of a platform for health and medical administration major education
 [그림 1] 보건의료행정 전공교육용 플랫폼 예시

여섯째, 결과분석을 위한 시각화 기능이 간편해야 한다. 2022년 초부터 ICT 개발업체에 불고 있는 새로운 패러다임 중에 하나가 코드없이 프로그램을 코딩하는 일명 ‘노코드(No-code)’ 이슈이다. 노코드는 직업적 특성의 이유로 점점 줄어들고 있는 프로그래머 수요와도 관련있다. 전문적인 소프트웨어 개발자 채용이 어려워지고 교육기관인 대학에서도 학생들이 프로그램 코딩 자체에 대한 거부감이 심화되고 있기 때문이다 [11]. 여기에 부응하는 새로운 코딩 패러다임이 코드 없이 쉽게 앱이나 프로그램을 개발하는 노코드(No-code)나 가급적 음성이나 행동을 통한 직관적 개발로 코딩을 최소화하는 로코드(Low-code)가 유행하고 있는 이유이다. 또한, 노코드나 로코드의 확산은 구인란 때문만은 아닌데, 다른 이유 중에 하나가 클라우드컴퓨팅(Cloud computing) 확산에 따른 시대적 환경변화 때문이라고 할 수 있다. 대표적인 노코드/로코드 플랫폼 개발 기업으로는 다국적 기업인 구

클(앱시트), 마이크로소프트(파워앱스), 아마존, SAP, ORACLE, SALESFORCE 등의 기업이 개발도구 툴을 판매하고 있다. 국내에서는 LG-CNS(테브온 NCD), 소프트파워(SmartMaker), 업스테이지사(AI Pack), 딥노이드사(Deep:PHI), 비아이메트릭스(AUD Platform), 아가도스(AGADOS-APPS Autobot Platform) 등이 배포되고 있다.

일곱째, 위의 6가지 조건을 만족하면서도 최종적으로는 다른 통계분석용 툴에서 데이터를 호환할 수 있는 결과자료의 호환성을 제공해야 한다. 수집/분석/가공 등의 과정을 거친 결과물을 SPSS, STATA, R, Python 등의 프로그램에서도 데이터 가져오기 및 공유가 가능해야만 공공성을 확보할 수 있다.

2.2.2 빅데이터 분석 언어 및 툴

앞 절에서 언급한 바와 같이 전공교육용 플랫폼의 7가지 기본조건을 정리하면 아래 [Table 1]과 같이 정리할 수 있다. 빅데이터 전공교육 프로그램을 위한 플랫폼의 7가지 기본조건인 데이터획득성(data availability), 데이터포맷 가독성(data format readability), 데이터 재사용성(data reusability), 데이터 확장성(data scalability), 데이터 비용성(data cost reduction), 데이터 시각화(data visibility), 데이터 호환성(data compatibility)을 만족하는 빅데이터 분석용 툴의 선정은 크게 프로그래밍 언어 방식과 통계용 패키지언어 방식으로 구분할 수 있을 것이다. 빅데이터 전공교육프로그램 교육 플랫폼을 위한 7가지 기본 요소에서 데이터 호환성(data compatibility)에 해당하는 부분을 크게 2가지로 분류할 수 있다. SPSS, STATA, R, Python과 같은 툴은 통계분석용 패키지 언어(혹은 툴)라는 방식이라고 할 수 있다. 자바(JAVA), 줄리아(JULIA), 스칼라(SCALA), 파이썬(Python) 등은 프로그래밍 언어 방식으로 구분할 수 있다.

그러나 최근에는 패키지방식이나 프로그래밍방식에서의 구분에 큰 의미를 두고 있지 않으며

목적과 사용자의 편의에 따라 툴이나 언어를 선택하고 있다. 다만, 2022년 현재, 빅데이터를 분석하기 위한 대표적인 툴(혹은 언어)로 자바(JAVA), R, 줄리아(JULIA), 파이썬(Python), 스칼라(SCALA), SAS, SQL 등 6가지에 대한 활용도가 높다고 할 수 있다.

[Table 1] 7 basic conditions for a platform for major education using big data in health care

[표 1] 보건의로 빅데이터 전공교육용 플랫폼 7가지 조건

Condition	Mean
데이터 획득성 data availability	- 보건의로 빅데이터를 데이터 원천으로부터 수집/획득/저장 등이 가능
데이터포맷 가독성 data format readability	- 공공으로 제공되는 보건의로 빅데이터의 배포 형식(포맷)이 확인 가능
데이터 재사용성 data reusability	- 가공이 가능하고 재저장이 가능. - 획득된 빅데이터를 시각적으로 보면서 가공하여 재저장이 가능 - 범용 통계분석용 툴(STATA, SAS, SPSS)에서 파일 오픈 및 재저장 가능
데이터 확장성 data scalability	- 충분한 오픈 라이브러리를 제공. - 지속적인 함수 패키지 지원 가능 - 사용자가 직접 오픈 함수/패키지 업로딩 공유 가능
데이터 비용성 data cost reduction	- 비용적 측면을 고려한 툴을 선정 - 무료/웨어/프리웨어 제공여부 - 소스코드 공개여부
데이터 시각화 data visibility	- 결과분석을 위한 시각화 기능이 간편성 제공 - 노코드, 로코드 등 비코딩 방식 제고
데이터 호환성 data compatibility	- 다른 통계분석용 툴에서 데이터를 호환할 수 있는 결과자료의 호환성을 제공

2.2.3 대학 및 학과 빅데이터 교육 플랫폼 설계

대부분의 보건행정관련학과는 대학 내에 보건계열이나 자연계열에 속해있는 경우가 대다수이다. 이공계열의 학과가 아닌 경우에 전용 컴퓨터 실습실을 보유하지 않거나 공용 컴퓨터 실습실을 이용하는 경우가 많다. 학과 전용 컴퓨터 실습실을 보유한 경우에도 실습용 프로그램이 기초전산실습(엑셀, 파워포인트, 워드, 한글), 병원정보시스템실무(HIS, EMR, OCS), 의료데이터베이스(MS-SQL Server, ACCESS) 등의 수업용으로 최적화되어 있기 때문에 보건의로 빅데이터 관련 수업을 수행하기에 적절한 통계 패키지나 프로그래밍 언어용 컴파일러를 구매하

여 구축하고 있는 경우는 많지 않다.

따라서 앞 절에서 언급한 바와 같이 각 대학 내에 보건행정 관련학과에서 빅데이터 분석을 위해서는 3가지 부분을 고려해야 한다. 첫 번째는 보건의료 데이터의 세부 주제, 제공처, 비용(유료/무료), 재가공, 보건통계, 분석, 시계열성, 시의적절성 등이 가능한 공공의료 빅데이터의 구체적인 데이터 선정을 고려해야 한다. 두 번째는 빅데이터 분석을 위한 언어 및 툴을 결정하기 위해서는 학과의 시설 및 예산 등을 고려할 필요가 있다. 세 번째는 [Table 1]에서 언급한 보건의료 빅데이터 전공교육용 플랫폼이 갖추어야 할 6가지 조건 중에 몇 가지를 필요충분 조건으로 만족할 수 있는가에 대한 고려가 필요하다.

아래 [Table 2]는 다양한 빅데이터 분석을 위해 사용되는 대표적인 7가지 툴의 특징을 분석한 표이다. 표의 각 점수는 1에서 5점까지의 점수를 부여하였는데, 이는 보건행정 관련학과의 현실과 프로그램 자체의 특성에 따른 주관적인 반영점수이다.

파인리포트(fineReport)는 DB의 접근성이 좋아서 데이터의 추가/삭제/수정이 용이하고, 멀티시트방식으로 엑셀파일로 저장된 데이터를 일괄적으로 불러올 수 있지만, SPSS/STATA/SAS 등의 공공 빅데이터의 포맷을 사용하는데에는 다소 어려움이 있다[13]. 1991년 귀도 반 로섬(Guido van Rossum)에 의해 개발된 파이썬(Python)은 빅데이터 분석툴로 널리 알려진 툴이다. 영어 문장이나 단어를 나열하는 것과 같은 스크립트(script) 방식의 인터프리터

[Table 2] Comparison of healthcare big data source data and tool selection

[표 2] 보건의료 빅데이터 원본 데이터 및 툴의 선정 비교1

Condition	FineReport	Python	Hadoop	Storm	Zoho Analytics	Sisense	R
툴 조작 용이성 easy of tool operation	2	4	4	1	5	3	5
처리 속도 processing speed	3	5	5	4	5	5	4
데이터 획득성 data availability	4	4	3	5	3	3	5
데이터포맷 가독성 data format readability	2	4	3	4	4	4	5
데이터 재사용성 data reusability	2	5	3	4	3	3	5
데이터 확장성 data scalability	3	4	3	3	3	2	5
데이터 비용성 data cost reduction	3	4	3	3	2	1	5
데이터 시각화 data visibility	5	4	3	3	5	4	4
데이터 호환성 data compatibility	4	5	3	4	4	4	5
가중치 합계 weight sum	28	39	30	31	34	29	43

(interpreter) 언어이기 때문에 간결하고 쉬운 문법, 다른 컴파일러 언어와 결합이 가능한 확장성이 좋다. 또한 2022년 현재, 가장 많은 사용자를 보유하고 있기 때문에 개발되고 배포된 라이브러리가 많다는 장점도 갖는다[14]. 구글, 인스타그램, 넷플릭스, 드롭박스 등 인터넷 플랫폼 기반 기업이 파이썬을 사용하지만 스크립트 언어가 갖는 속도가 느리다는 단점을 가지고 있다. 하둡(Hadoop)은 오픈소스를 제공하는 빅데이터 분석 툴이며 동영상, 오디오, 텍스트 등의 다양한 형식의 데이터를 분석할 수 있다는 장점이 있으나, 데이터의 중복 및 디스크 용량을 많이 차지한다는 단점을 갖고 있다. 따라서 고사양의 컴퓨터가 필요하다. 조호어널리틱스(Zoho Analytics)는 파인리포트와 함께 데이터 시각화가 매우 용이한 툴이다. 보고서 생성이나 데이터 불러오기나 내보내기, 시각적분석, 온라인 분석 등의 강점이 있으나 평가판은 무료이지만 서버방식으로 사용하기 위해서는 월 사용료를 지불해야 한다는 단점이 있다[15]. 또한 엑셀이나 CSV, HTML, 텍스트파일 등을 사용할 수 있으나 우리가 원하는 공공 빅데이터의 저장 포맷을 활용하는데 다소 제약이 있다. 사이센스(Sisense)는 좋은 빅데이터 분석 툴이지만 기업의 데이터를 분석하여 전략적 및 일상적 의사결정을 위한 프로세스를 제공하는 비즈니스 인텔리전스(BI, Business Intelligence) 프로세스를 제공하는 전문적인 툴이다. 따라서, 대학교육을 위한 학습 툴로는 적합하지 않다[16]. 마지막으로 알(R)은 가장 포괄적이고 보편적인 통계분석 패키지 툴이라고 할 수 있다. 오픈소스이면서 무료로 프로그램을 제공받을 수 있으며, 다중패러다임 및 동적 소프트웨어 환경을 제공한다는 점에서도 큰 강점을 갖는다. 파이썬에 비해 메모리관리나 속도, 시스템 보안성이 떨어진다는 단점을 가지고 있으나 산업계 및 학계에서 많은 사용자를 보유하고 있고 CRAN이나 GitHub 등에서 필요한 공개 패키지를 자유롭게 다운로드 받아 사용하고 수정하여 재사용할 수도 있다. 시각화 전문 툴인 파인리포트나 조호어널리틱스

만큼은 아니지만 통계분석에 필요한 다양한 그래픽 생성과 분석차트를 생성해주는 강력한 기능을 제공한다. 이러한 다양한 이유 때문에 대학이라는 제한적 환경과 보건행정 관련학과의 커리큘럼을 감안하였을 때, 알(R)이 가장 적합한 통계분석 및 빅데이터 분석 툴로 적합하다고 판단된다.

또한 [Table 3]은 [Fig. 1]에서 언급한 바와 같이 공공성을 지닌 보건의료 빅데이터 제공 사이트의 데이터별 특징을 비교하였으며, 적합한 데이터 여부를 역시 보건행정 관련학과에서 다룰 수 있는 현실성을 반영한 점수이다.

2010년에 개설된 공공데이터포털(open.go.kr)은 ‘공공데이터의 제공 및 이용에 관한 법률’에 의거 2013년부터 공식적으로 행정안전부에서 제공하는 공공데이터 통합시스템이다. 여기에는 16개 분류체계(공공행정, 과학기술, 의료, 보건의료 등)에 따라 7개 서비스유형(Link, LOD, REST, RSS/ATOM, SOAP, Download, 배송및 기타) 등으로 구분된 데이터를 14개 확장자(CSV, HWP, JPG, MP4 등)의 데이터포맷 파일로 제공한다[17]. 빅데이터 분석을 위한 데이터 포맷은 주로 CSV, XML 등으로 제공받을 수 있다. 데이터의 접근, 통계분석은 용이하지만 검색되는 보건의료 데이터에 대한 자세한 사전지식이 없다면 데이터를 시간의 흐름에 따라 배치하고 분석하기 위한 시계열성, 데이터 설명이 구체적이지 못하다는 단점을 갖는다.

국가통계포털(KOSIS, Korean Statistical Information Service)은 통계청이 개발하여 국내, 국외, 북한의 주요 통계를 한 번에 찾아볼 수 있도록 원스톱(one-stop)서비스를 제공하는 포털서비스이다[18]. 경제, 사회, 환경, 인구, 보건 등 30개 분야야 국내 통계자료를 제공하며, 공유서비스인 OpenAPI를 제공한다. 보건의료 빅데이터로 활용할 주제분야는 인구, 보건, 복지 등의 주제가 해당되지만 해당 검색 데이터를 보건의료 빅데이터로 보기에는 다소 데이터의 양이 부족하고, 오히려 통계처리된 결과 위주로 자료를 확인할 수 있기 때문에 보건행정 관련

빅데이터 원천으로 사용하기에 다소 부적합하다고 할 수 있다.

1989년 몇몇 연구원을 통합하여 설립된 한국보건사회연구원(KIHASA, Korea Institute for Health and Social Affairs)은 국민건강보험, 국민연금, 보건의료, 사회복지, 사회정책 등 5가지 주요 정책과제를 중심으로 연구 및 분석을 수행하는 공공기관이다[19]. 한국보건사회연구원에서 제공하는 공공데이터개방 데이터에는 ‘보건복지데이터포털’, ‘한국의료패널’, ‘한국복지패널’ 등 크게 3개의 공공데이터를 운영관리하고 있다. 한국복지데이터포털은 주요통계지표, 국내통계, 해외통계, 해외메타데이터 등 4개 주제를 중심으로 데이터를 제공하고 있다. 국내통계 데이터는 보건복지부/식품의약품안전처/질병관리청/한국보건사회연구원 등의 자료를 제공한다.

해외통계의 경우 IMF/World Bank/OECD/UN /WTO/ILO 등 전 세계 6대 주요지표 데이터를 활용할 수 있다. 이 중 ‘국내통계>보건의료질통계’의 급성질환 및 만성질환에 대한 시계열 데이터를 활용가능하다. 한국의료패널(Korea Health Panel Survey)은 보건의료체계의 접근성, 대응성 향상을 위한 정책수립을 위한 목적으로 조사되는 자료로 단기적으로는 국민 개인별 의료비 규모, 의료비 재원분석, 수요자의 의료이용 행태 등을 확인하고, 장기적으로는 국민의료비 산출과 변화를 장기적인 관점에서 관찰하고 모니터링하기 위해 조사되고 있다[21]. 한국의료패널의 자료를 사용하기 위해서는 활용동의서 작성 후 담당자에게 동의서를 제출한 후 제공 받을 수 있다. 제공 받는 자료는 SAS/SPSS/STATA 등의 포맷으로 선택 가능하다. 한국복지패널(Korea Welfare Panel Study)의 데이터는 2022년 현재 2006년부터 2021년까지 16차 웨이브 데이터를 시계열 방식으로 제공하고 있다[6-7]. 본 고찰에서 나열된 보건의료 관련 빅데이터에서 가장 오랜 기간동안 시계열에 따른 다양한 인구집단의 데이터가 축적되고 분석된 자료가 한국복지패널의 데이터라고 할 수 있다. 또한, 보건행정 관련학과의 빅

데이터 수업에 적합한 이유는 다음과 같다. 첫째, 데이터 설명이 매우 자세하다. 건강/의료/경제활동상태/주거/소득/생활비/재산/생활여건/사회보험 등 다양한 속성별 데이터를 결합하여 새로운 보건의료 분석이 가능하다. 둘째, 가구원용 데이터의 경우 사회보험/연금/근로/생활실태/사회적환경/교육/가족관계/정신건강 등 연계성을 포함한 속성정보의 결합이 매우 용이하다. 셋째, 시계열성을 갖는다. 2006년에서부터 2021년까지 16차에 걸쳐 연도별 데이터가 매우 자세히 축적되어 있기 때문에, 시계열성 보건의료 연관정보 추출이 용이하다. 넷째, 표본자료의 신뢰성이 좋다. 2021년(16차)에 수집된 데이터는 조사대상 가구 6,240가구 중에서 총 5,996가구의 조사자료를 사용하고 있는데, 이는 전국에 표본이 되는 가구의 자료로 신뢰성이 높을수 밖에 없다. 2015년의 경우 957개 속성(변수)별 16,664명의 전국의 표준 복지패널 데이터를 활용할 수 있으며, 2021년의 경우 1,050개의 속성에 13,144명의 데이터의 활용이 가능하다.

[Table 3] Comparison of health administration-related departments and utilization of public big data related to healthcare

[표 3] 보건의료 관련 공공 빅데이터의 보건행정 관련학과 활용성 비교

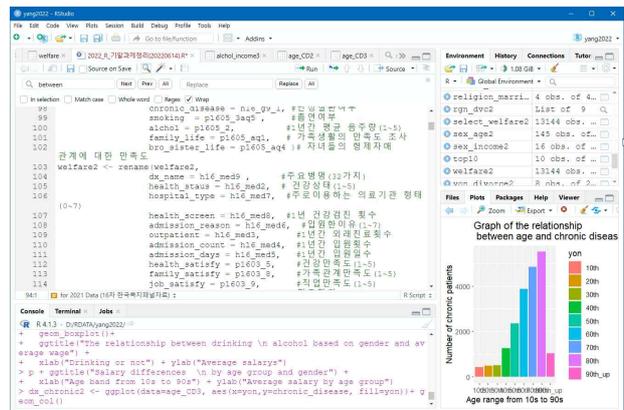
Condition	통계청		HIRA	한국보건사회연구원		
	공공 데이터 포털	국가통계 포털	보건의료빅데이터 개방 시스템	보건복지 데이터 포털	한국의료패널	한국복지패널
접근 용이성	5	4	1	4	4	5
데이터포맷 다양성	4	2	3	3	5	5
통계분석 가능성	4	2	4	2	5	5
데이터 사용비용	5	5	1	5	5	5
분석 용이성	3	2	3	2	4	4
시계열성	3	2	3	4	5	5
데이터설명	3	3	3	2	5	5
가중치 합계 weight sum	24	20	18	20	28	29

통계청(KOSTAT)에서는 국가통계포털(KOSIS), 마이크로데이터(MDIS, Micro Data Integrated Service), 통계데이터센터(SDC), 통계지리정보(SGIS) 등을 운영관리하고 있다[20]. 여기에서 보건의료 관련 교육용 빅데이터를 활용할 수 있는 사이트는 앞서 설명한 국가통계포털(KOSIS)과 마이크로데이터(MDIS)이다. 마이크로데이터(MDIS)에서는 통계청 및 주요부처, 지자체, 연구기관 등 통계작성기관의 마이크로데이터를 통합관리하고 있으며, 주제별/기관별 자료 제공이 가능하다. MDIS는 18개 주제별로 자료를 제공하는데 ‘보건’관련 주제에는 사망원인보완조사, 사망원인통계, 국민체력측정통계, 한국인인체치수조사, 산후조리실태조사, 의료서비스경험조사, 급성심장정지조사, 의료기기산업실태조사, 한의약산업실태조사 등 대부분에 해당하는 9개 주제별로 시계열 데이터를 활용가능하다. 물론 이 모든 데이터의 통합 연동은 국가통계포털(KOSIS)를 통해 연결된 자료의 형태로 활용하게 된다.

보건의료빅데이터개방시스템은 건강보험심사평가원(HIRA)에서 보유한 공공데이터 개방시스템이다. 제공되는 데이터의 형태는 데이터셋(OpenAPI) 형태의 공공데이터, 빅데이터센터의 의료 빅데이터분석 데이터, 의료기관, 질병 및 의약품과 관련한 의료통계분석 등의 공공데이터를 제공한다. 그러나, 이 데이터는 다소 복잡한 신청 및 공공데이터 제공에 따른 수수료 명목의 비용이 발생하기 때문에 교육용 데이터로서의 활용도는 다소 떨어진다[7].

결론적으로 보건의료 빅데이터를 보건행정 관련 학과에서 활용하는데 가장 적절한 프로그래밍 언어나 툴은 가중치 점수가 가장 높은 알(R)을 활용하고, 빅데이터 원천은 무료이면서 자료획득이 용이하고, 속성(변수) 간 연관분석이 가능한 한국의료패널이나 한국복지패널의 데이터가 적절하다고 판단된다. 그 중에서도 한국복지패널의 데이터가 시계열성, 주제지향성, 속성간 연관분석 가능성, 표본자료의 신뢰성, 표본자료의 분량 등을 고려했을 때 가장 적절하다고 판단된

다.



[Fig. 2] Examples of using R tools for big data analysis [그림 2] 빅데이터 분석용 R 툴의 사용 예

2.2.4 R 언어의 장단점

보건의료 빅데이터 교육용 툴로 가장 적절하다고 판단한 알(R)의 경우 앞 절에서 언급한 바와 같이 다양한 장점을 포함하고 있다. 첫째, 프로그램을 무료로 제공받을 수 있다는 점이다. 동적 소프트웨어 환경을 제공하며, 전 세계적으로 현재 가장 많이 사용되는 10대 프로그래밍 언어에 속하는 사용자가 보편화되었다는 것이 장점이다. 둘째, 학계 및 산업계의 다양한 사용자를 보유하고 있다는 점이다. 셋째, 오픈소스 방식이기 때문에 필요한 라이브러리를 패키지 형식으로 개발하여 사용자에게 무상으로 제공하고 제공받을 수 있다. 넷째, 설치가 쉽고 프로그램이 무겁지 않아서 일반적인 공용 컴퓨터나 노트북에서도 사용 가능하다. 뿐만아니라 윈도우즈 운영체제나 맥북 등의 iOS에서도 동일하게 동작하고 코드 호환이 가능하다는 점이 매우 큰 장점이다. 다섯째, 시각화 전문 툴인 파인리포트나 조호어널리틱스 만큼은 아니지만 통계분석에 필요한 다양한 그래픽 생성과 분석차트를 생성해주는 강력한 기능을 제공한다. 이러한 다양한 장점으로 인해 대학이라는 제한적 환경과 보건행정 관련학과의 커리큘럼을 감안하였을 때, 알(R)이 가장 적합한 통계분석 및 빅데이터 분석 툴로 적합하다고 판단된다. 여섯째, CRAN과 마

이크로소프트사에서 프로그램 개발을 위한 통합 환경(IDE)를 제공한다. RStudio 통합환경을 통해 학생들은 손쉽게 프로그램을 코딩하고 결과를 동시에 한 화면에서 확인할 수 있다.

알(R)의 단점은 파이썬에 비해 메모리관리나 속도, 시스템 보안성이 떨어진다는 단점을 가지고 있기도 하다.

2.3 보건의료 빅데이터 교육의 현황 및 문제점

본 고찰에서는 앞 절에서 언급한 바와 같이 대학의 보건행정관련 학과를 대상으로 하는 보건의료 빅데이터 교육의 현황을 살펴보았다. 이 공계와는 다른 컴퓨터를 활용한 실습환경이 그다지 적정하지 못하다는 점이다. 또한, 의료기관이나 국민건강보험공단, 건강보험심사평가원 등과 같이 개인에게서 발생하는 보건의료 데이터를 일선의 교육용으로 사용하기 어렵다는 점도 제약사항이 되고 있다. 현재 전문대학의 경우에는 전공심화 과정에서 보건의료 빅데이터에 관한 기초이론, 기초실습 등이 이루어지고 있는데 해당되는 교육용 자료를 구하기도 어렵고 분석할 수 있는 툴이나 프로그래밍 언어를 선택하는데 있어도 어려움이 따른다. 이는 보건계열 학과의 학생들은 특별히 정규 커리큘럼 상에 프로그래밍 교육을 이수하는 대학이 거의 없기 때문이다.

2.4 보건의료 빅데이터 교육의 대안

앞서 언급한 바와 같이 보건의료 빅데이터 교육의 현황과 문제점을 확인하였다. 따라서, ‘2.2.3 대학 및 학과 빅데이터 교육 플랫폼 설계’에서 조사한 바와 같이 교육에 필요한 통계분석용 툴은 알(R)을 선정하였고, 통합환경을 제공하는 CRAN R3.3.x 버전과 MS RStudio Build+492 버전을 활용하고자 한다. 교육용 분석 데이터는 ‘한국복지패널’의 1,050개의 속성에 13,144명의 데이터를 활용하는 것이 가장 적합한 대안이라고 판단된다.

2.4.1 공공 빅데이터에서 데이터 추출

앞서 언급한 공공 빅데이터 중에서 표본이 전국적인 분포의 데이터이면서 자료의 신뢰성이 높고, 빅데이터 분석을 위한 시계열성을 고려하여 한국복지패널의 공공 데이터를 활용하고자 한다. 2021년 한국복지패널 데이터 불러오기 및 변수변환, 데이터 추출은 다음과 같다. 2021년(16차)에 수집된 데이터는 조사대상 가구 6,240 가구 중에서 총 5,996가구의 조사자료를 사용하고 있는데, 이는 전국에 표본이 되는 가구의 자료로 신뢰성이 높다. 최종적으로는 2015년의 경우 957개 속성(변수)별 16,664명의 전국의 표준 복지패널 데이터를 활용할 수 있으며, 2021년의 경우 1,050개의 속성에 13,144명의 데이터를 활용하였다.

2.4.2 R을 활용한 빅데이터 가공 및 결과생성

본 고찰에서는 한국복지패널 빅데이터를 활용하여 가공하였다. 이 알(R)툴을 사용하여 SPSS 형식의 데이터를 다운로드하고, 해당파일을 메모리에 적재한다. 적재된 1,050개의 속성에서 필요한 속성 명칭을 변경하여 13,144명의 데이터를 성별, 출생연도, 혼인여부, 월급, 직종, 만성질환여부, 흡연여부, 주요병명, 건강상태, 의료기관형태, 건강검진횟수, 외래진료횟수, 입원횟수, 입원일수, 건강만족도 등 일반적인 특성정보와 보건의료정보를 분석하였다. 아래 [Fig. 3]은 해당 빅데이터를 적재하여 결측데이터(NA)를 소거한 후 시각화하는 대표적인 R script의 내용 예이다.

```
install.packages("foreign")
install.packages("ggplot2")
install.packages("dplyr")
install.packages("readxl")
library(foreign)
library(dplyr)
library(ggplot2)
library(readxl)
#-----2022, data upload-----
raw_welfare2 <-
  read.spss(file.choose(), reencode='utf-8',
to.data.frame = T)
welfare2 <- raw_welfare2

welfare2 <- rename(welfare2,
sex = h16_g3, # 성별
birth = h16_g4, # 태어난 연도
```

```

marriage = h16_g10,      # 혼인 상태
religion = h16_g11,     # 종교
income = p1602_8aq1,    # 월급
code_job = h16_eco9,    # 직종 코드
code_region = h16_reg7) # 지역 코드
chronic_disease = h16_g9_1, # 만성질환여부
smoking = p1605_3aq5,  # 흡연여부
alchol = p1605_2,      # 1년간 평균 음주량(1~5)
family_life = p1605_aq1, # 가족생활의 만족도
bro_sister_life = p1605_aq4 # 형제자매 관계 만족도
dx_name = h16_med9,    # 주요병명(32가지)
health_staus = h16_med2, # 건강상태(1~5)
hospital_type = h16_med7, # 의료기관형태(0~7)
health_screen = h16_med8, # 1년 건강검진 횟수
admission_reason = h16_med6, # 입원한이유(1~7)
outpatient = h16_med3,  # 1년간 외래진료횟수
admission_count = h16_med4, # 1년간 입원횟수
admission_days = h16_med5, # 1년간 입원일수
health_satisfy = p1603_5, # 건강만족도(1~5)
family_satisfy = p1603_8, # 가족관계만족도(1~5)
job_satisfy = p1603_9,  # 직업만족도(1~5)
education = p1607_3aq1, # 최종학력(1~5)
arbit = c1605_21 )     # 아르바이트의 종류(1~20)

select_welfare2 <- welfare2 %>%
select(sex, birth, marriage, religion, income,
code_job, code_region, chronic_disease,
smoking, alchol, family_life, bro_sister_life,
dx_name, health_staus, hospital_type,
health_screen, admission_reason, outpatient,
admission_count, admission_days, health_satisfy,
family_satisfy, job_satisfy, education, arbit
)
View(select_welfare2)

welfare2 <- welfare2 %>%
mutate(yon = ifelse(age < 20, "10th",
ifelse(age < 30, "20th", ifelse(age < 40, "30th",
ifelse(age < 50, "40th", ifelse(age < 60, "50th",
ifelse(age < 70, "60th", ifelse(age < 80, "70th",
ifelse(age < 90, "80th", "90th_up"))))))))

sex_income2 <- welfare2 %>%
filter(!is.na(income)) %>%
group_by(sex, yon) %>%
summarise(mean_income2 = mean(income))

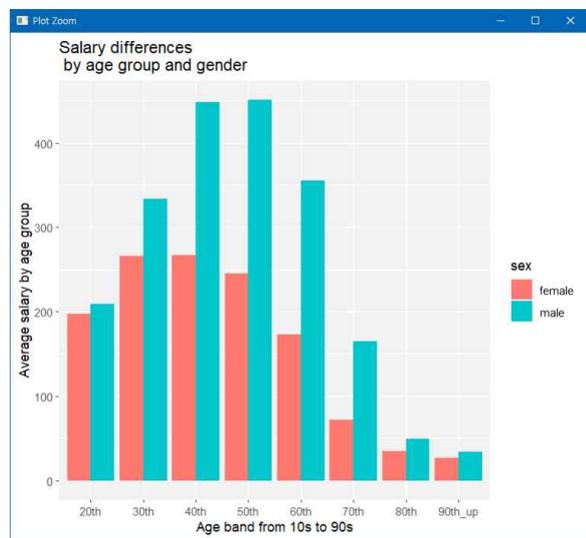
#-----[Fig. 4] Graph -----
p <- ggplot(data = sex_income2,
aes(x = yon, y = mean_income2, fill = sex)) +
geom_col(position = "dodge")
p + ggtitle("Salary differences by age group and
gender") + xlab("Age band from 10s to 90s") +
ylab("Average salary by age group")

#-----[Fig. 7] Graph -----
welfare2$chronic_disease <-
ifelse(welfare2$chronic_disease==0,NA,
welfare2$chronic_disease)
age_CD2 <- welfare2 %>%
filter(!is.na(chronic_disease)) %>%
group_by(ageeg) %>%
summarise(chronic_disease)
dx_chronic2 <- ggplot(data=age_CD2,
aes(x=yon,y=chronic_disease, fill=yon)) + geom_col()
dx_chronic2 + ggtitle("Graph of the relationship between
age and chronic diseases") +
xlab("Age range from 10s to 90s") +
ylab("Number of chronic patients")
    
```

[Fig. 3] 2021 Korea Welfare Panel Average Pay Analysis by Age/Gender Example
 [그림 3] 2021년도 한국복지패널 나이대별/성별 평균 급여 분석 예시

2.4.3 R을 활용한 빅데이터 결과분석

앞 절에서 R툴을 통해 분석된 한국복지패널의 데이터를 통해 산점도, 막대그래프, 선차트, 박스플롯, 히스토그램 형식으로 표현가능하다. 알(R) 툴을 이용해서는 머신러닝 모델링, 텍스트마이닝, 소셜미디어 분석, 지리정보 시각화, 주식데이터분석, 이미지분석, 소리분석, 웹사이트분석, 앱분석 등 다양한 분석이 가능하다. 아래 그림은 2021년도 16차 한국복지패널 빅데이터를 활용하여 시각화한 예시이다.



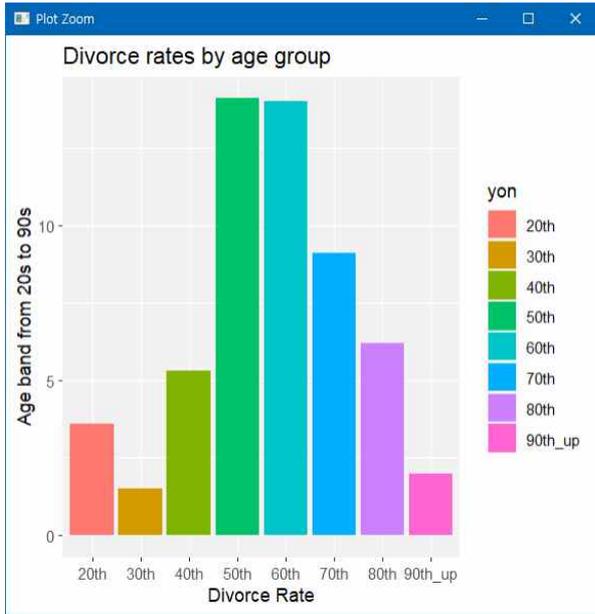
[Fig. 4] 2021 Korea Welfare Panel Average Pay Analysis by Age/Gender Example
 [그림 4] 2021년도 한국복지패널 나이대별/성별 평균 급여 분석 예시

[Fig. 4]는 13,144명의 한국복지패널 데이터에서 나이를 10부터 90대까지로 그룹화하여 남녀 성별에 따른 평균 급여의 차이를 그래프화 하였다. 여성은 30대말에서 40대초반에 임금의 피크를 보이며 남성은 50대에 임금의 피크를 보였다.

[Fig. 5]는 연령대별 이혼율을 분석한 결과로 30대의 이혼율이 가장 낮으며, 50대에 이혼율이 가장 높은 것을 확인할 수 있었다.

[Fig. 6]은 성별과 음주여부에 따른 평균급여와의 관계를 분석한 예이다. 해당 자료는 박스플

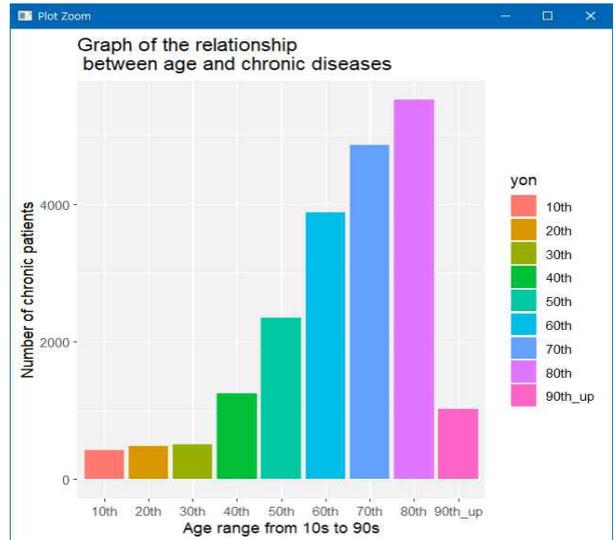
를 사용하여 1, 2, 3분위값, 중앙값, 평균값 등을 확인한 예이다.



[Fig. 5] 2021 Korea welfare panel divorce rate analysis by age category example

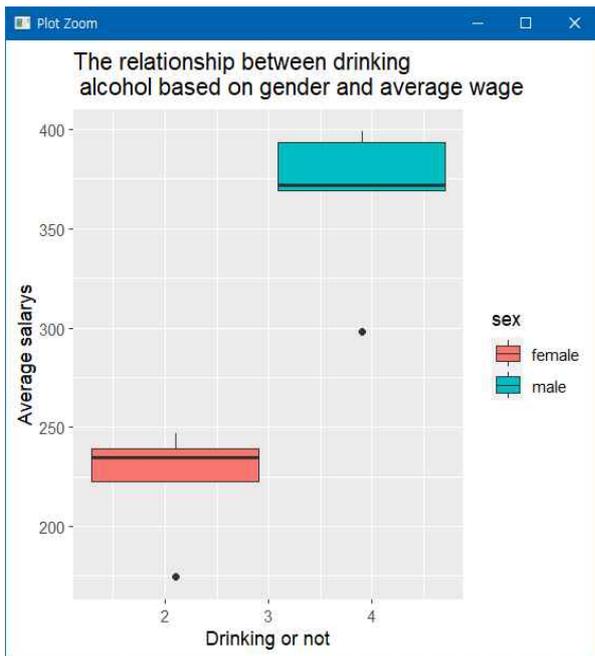
[그림 5] 2021년도 한국복지패널 연령대별 이혼율 분석 예시

[Fig. 7]은 성별과 연령별에 따른 만성질환비율의 결과이다. 분석결과를 살펴보면 80대의 만성질환자가 가장 높은 수치를 보이고 있음을 확인할 수 있었다.



[Fig. 7] Example of chronic rate analysis by gender/age

[그림 7] 성별/연령별 만성질환율 분석 예시



[Fig. 6] Example of an analysis of the average salary relationship based on gender/alcohol consumption

[그림 6] 성별/음주여부에 따른 평균급여 관계 분석 예시

2.4.4 분석보고서 작성

앞 절은 알(R) 툴과 한국복지패널 빅데이터를 사용하여 13,144명의 1,050개의 속성(변수)중에 일반적특성과 보건의료정보를 확인하고 그래프로 표현할 수 있음을 확인하였다. 이렇게 확인된 그래프의 내용은 그 해석이 매우 중요하다고 할 수 있다. 수강학생들은 최종적으로 생성된 그래프를 기반으로 해당 내용을 분석하여 발표자료(handout)나 보고서(report) 파일 형식으로 작성까지 완성한다면 보건의료 빅데이터를 분석하는 가장 기초적인 실습이 가능하다고 판단된다.

3. 결론

본 고찰은 보건의료행정 관련학과에서 보건의료 빅데이터 관련 교육프로그램 활용 가능성에 대한 분석 및 제안하였다. 본 논문에서는 크게 5가지 관점으로 보건의료 빅데이터의 활용을 제

안하였다. 첫째, 앞서 언급한 '보건의료빅데이터 개방시스템' 이외에도 '한국복지패널', '공공빅데이터', '서울시', '통계청' 등에서 공개하고 있는 공공빅데이터의 특징 및 활용기술에 대하여 살펴보았다. 둘째, 관련되는 보건의료빅데이터를 전문대학 보건의료행정 및 보건의료정보 관련학과의 정규과목 내에서 살아있는 데이터로써 활용이 가능한지에 대한 적정성 여부를 살펴보았다. 셋째, 기존에 활용하고 있는 통계처리패키지 및 프로그래밍 언어를 활용한 실습실 강의 여부에 가장 적절한 툴을 선정하였다. 넷째, 검증된 보건의료 빅데이터와 적정 툴을 활용하여 수업에서 그래프 등의 표현가능성 여부 및 보고서 작성까지의 단계를 시험해보았다. 다섯째, 최종적으로 휴대성, 설치성, 비용성, 호환성, 빅데이터처리가능성 등을 만족할 수 있는 알(R) 툴의 상대적 이점을 기술하였다.

이로써 보건의료빅데이터분석과 관련한 보건행정 관련학과의 정규 교육과정에 보편적활용, 설치용이성, 오픈소스제공, 무료사용여부, 활발한 라이브러리 개발 및 보급, 폭넓은 사용자층 등을 고려하여 선정하였다. 사용가능한 보건의료 관련 빅데이터는 공공데이터로 제공되는 데이터의 요건인 접근 용이성, 데이터포맷 다양성, 통계분석 가능성, 데이터 사용비용, 분석 용이성, 시계열성, 데이터설명 등이 명확하게 제공되는 '한국의료패널' 및 '한국복지패널'의 공공데이터를 활용하였다. 따라서, 알(R)을 기반으로 하는 CRAN, RStudio를 설치하고, 공공 빅데이터는 우선적으로 '한국복지패널'의 2021년(16차)에 정리된 1,050개의 속성(변수)에 13,144명의 데이터를 활용하였다.

향후 본 고찰의 완성을 위해서는 보건의료 빅데이터 교육을 위한 주차별 강의 세부계획서를 개발하고, 수강 학생들에게 개인별, 팀별 PBL(Problem Based Learning) 방식의 과제를 출제와 제출과 발표에 이르는 일련의 과정까지의 개발이 필요하다고 판단된다.

References

- [1] HIRA, Startup competition using health and medical big data in 2022, <https://www.hira.or.kr/bbs/Dummy.do?pgmid=HIRAA020041000100&brdScnBltno=4&brdBltno=10564&pageIndex=1>
- [2] Y.W. Kim, 'Big Data Education Contents for Healthcare Officials', p. 236-242, Journal of the Korea contents association. 2020.05.28.
- [3] M.E. Oh, 'Strategies and Tasks for Big Data in Health and Welfare', Health and welfare policy forum. p.29-40, KIHASA, 2019.08.01.
- [4] K-MOOC, http://www.kmooc.kr/courses/course-v1:KOH1+KOH1005+2021_001/about
- [5] Korea Welfare Panel Study, <https://www.koweps.re.kr:442/research/report/view.do>
- [6] Eun-Hee, Jeong, 'Korea Welfare Panel Study Report : 2021 Korea Welfare Panel Survey and Analysis Report', KIHASA, 2021.
- [7] HIRA Opendata, <https://opendata.hira.or.kr/op/opc/selectOpenDataAplInfoView.do>
- [8] J.H. Kim, Daddy's ggplot2+, Book&Edu Publish, 2020.2.
- [9] Y.W. Kim, Easy to learn R data analysis, EasyPublishing, 2020.9.
- [10] Open Library, <https://www.finereport.com/kr/10-data-visualization-tools-open-source-that-you-cannot-miss-in-2020/>
- [11] No-Code, <https://www.etnews.com/20220527000067>
- [12] Big Data Analytic Tools, <https://www.finereport.com/kr/%EB%AC%B4%EB%A3%8C-%ED%95%9C%EA%B8%80%ED%8C%90-%EC%9E%90%EB%A3%8C-top-6-data-science-programming-languages-for-2020/>
- [13] fineReport, <https://www.finereport.com/kr/finereport-vs-oz-report/>
- [14] Python, <https://library.gabia.com/contents/9256/>
- [15] Zoho Analytics, <https://4w-3.tistory.com/26>
- [16] Sisense, <https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=zalesia2020&logNo=222168913309>
- [17] Public Data Portal, <https://www.data.go.kr/>

- [18] KOSIS, <https://kosis.kr/serviceInfo/kosisIntroduce.do>
- [19] KIHASA, <https://www.kihasa.re.kr/openinfo/public/data>
- [20] KOSTAT, <https://kostat.go.kr/portal/korea/index.action>
- [21] Korea Health, Panel Survey, <https://www.khp.re.kr:444/web/introduction/introduction.do>