



데이터 과학은 어떻게 과학적 법칙을 찾는가?

「과학사」와 「e-비즈니스 전략」강의페어링

e-Business 학과, 김동욱, 201221470, 송하석교수님 지도

목적

데이터 과학이라는 말은 1960년대 컴퓨터 과학의 대체 용어로 쓰이기 시작하면서 출현했다. 1990년대 IT의 급격한 발달과 함께 많은 기업가와 연구가들이 다량의 데이터를 통해 지식(Knowledge)와 통찰(Insight)을 도출하면서 성장했다. 최근 빅데이터와 딥러닝이 많은 관심을 받으면서 데이터 과학이 현대 과학의 중심으로 들어오기 시작했다. 하지만 이러한 데이터 과학의 인기에도 불구하고 철학적인 관점에서 데이터 과학을 살펴보는 연구는 많이 부족한 실정이다. 그러므로 이번 강의 페어링을 통해 과학사적 관점에서 데이터 과학을 해석하는 기회를 마련해 보고자 한다.

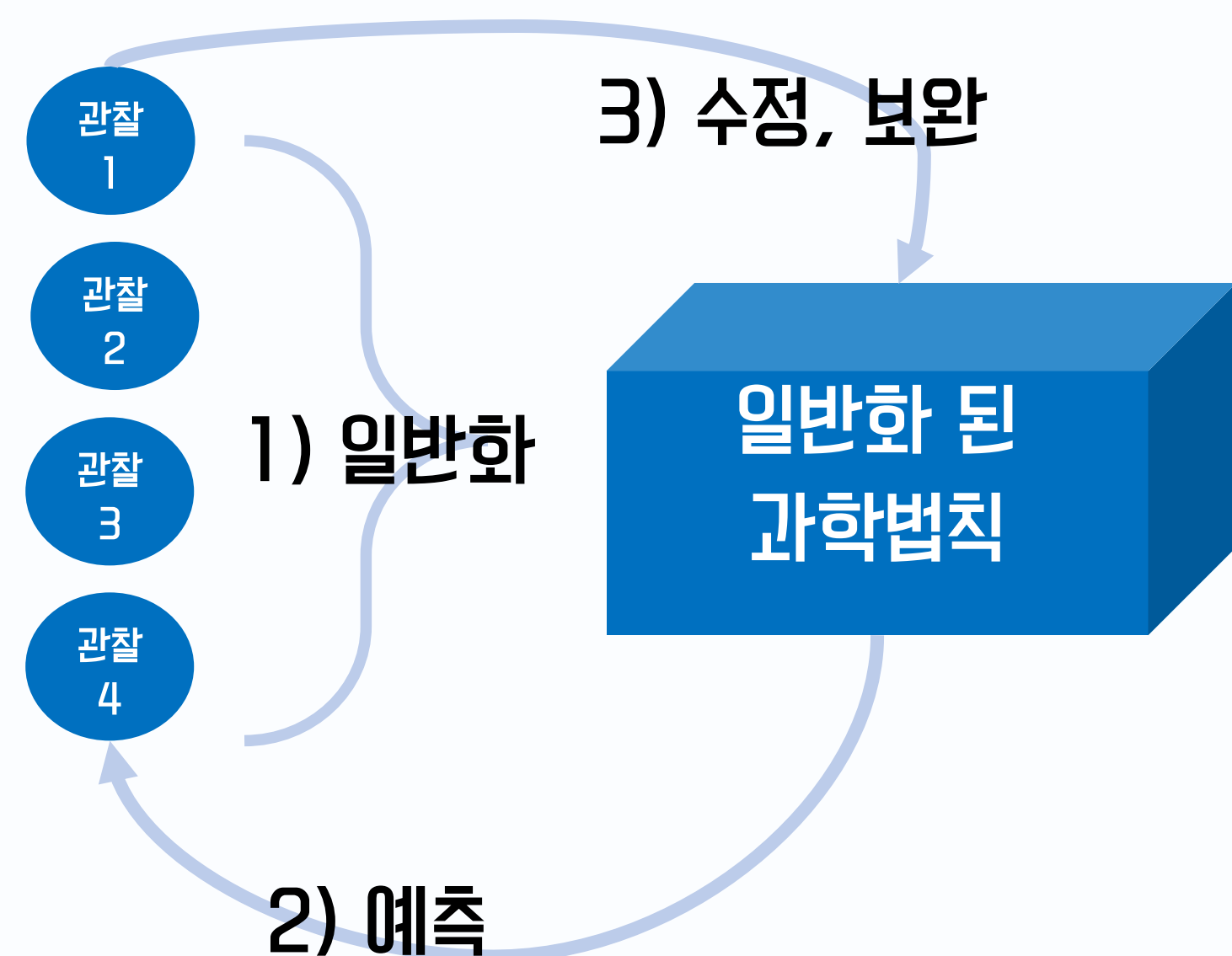
전통적인 과학관

[관찰에 대한 이해]

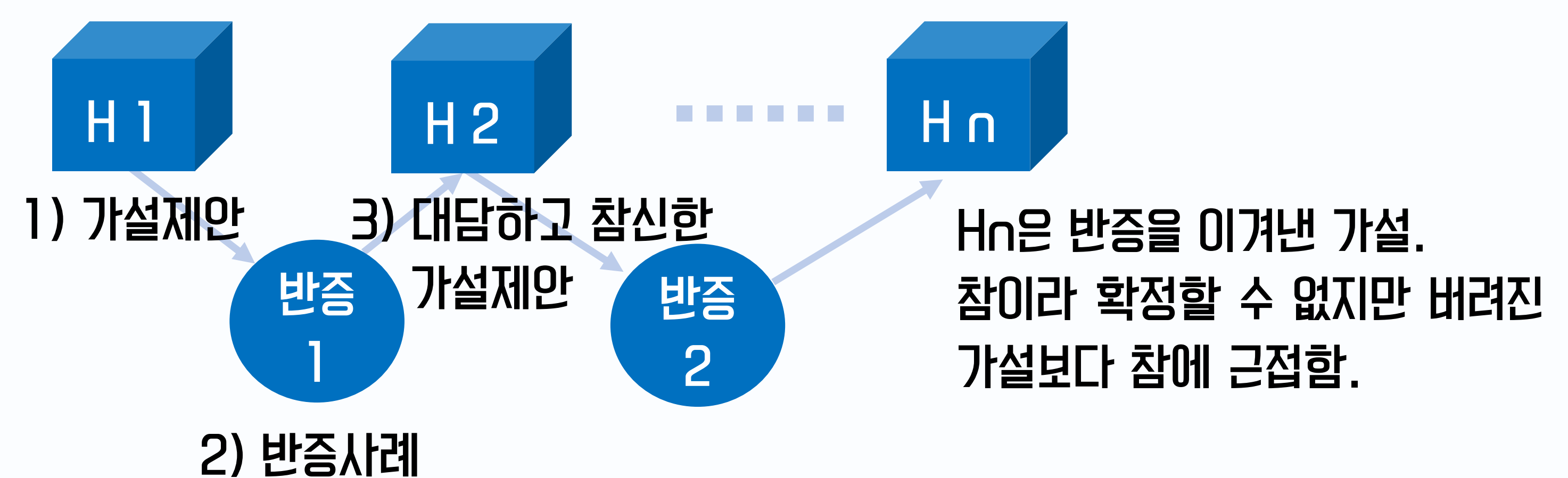
귀납주의 과학관은 관찰을 과학적 법칙을 위한 기초로서
연역주의 과학관은 기존 가설에 문제를 제시하는 사례로서 이해.

관찰은 객관적임을 암암리에 전제함.

[귀납주의 과학관]

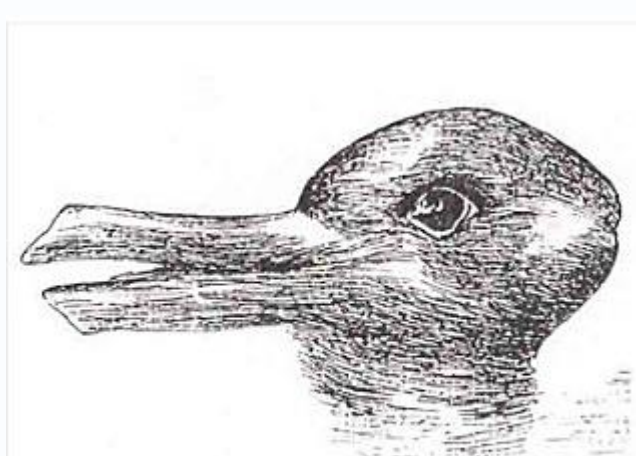


[연역주의 과학관] 20C 초 칼 포퍼의 반증주의



현대적인 과학관

[게슈탈트 전환]



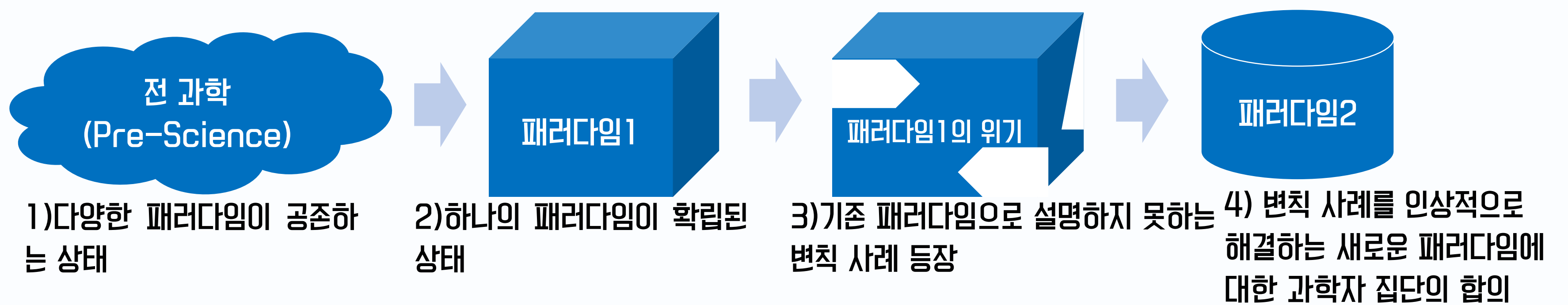
오리? 토끼?

관찰은 객관적이지 않음.
인간의 감각은 불완전함.

관찰이 객관적이라 가정하는 전통적인 과학관에 대한 비판 및 회의가 제기됨.

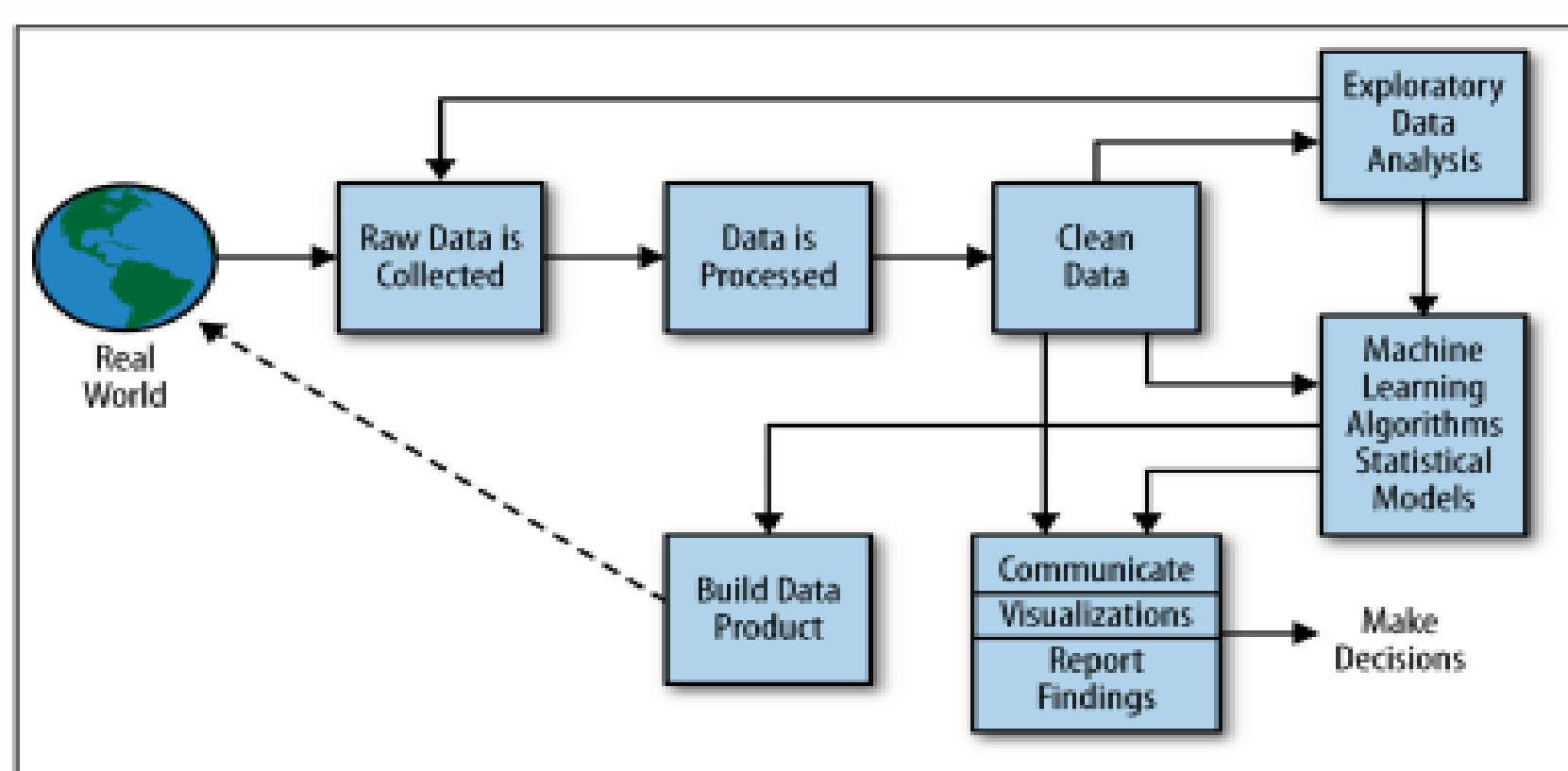
[전체론적 과학관]

1962년 토마스 쿤의 [과학혁명의 구조]에서 제시된 과학관



데이터 과학은 어떤 과학관일까?

[데이터 과학의 분석과정]



데이터 과학자들은 우선 현실 세계를 반영하는 raw data를 수집한다. 데이터 과학자들은 데이터를 정제된 뒤 탐색적 데이터 분석을 통해 데이터의 특징(Feature)을 파악한다. 파악된 데이터의 특징을 바탕으로 머신러닝 알고리즘에 기반한 확률 모델을 만든다. 이렇게 만들어진 확률 모델은 현실세계에 발생하는 현상을 예측한다. 만약 예측 오차가 발생할 경우 오차와 관련된 데이터를 통해 확률 모델을 다듬는 과정을 거친다. 이를 통해 현실 세계를 더 잘 설명하는 확률 모델을 만든다.

[귀납주의 과학관과 비교]

데이터 과학이 기본적으로 과학적 법칙을 파악하는 과정은 귀납주의 과학관과 유사함. 하지만 데이터 과학을 통해 얻은 법칙은 귀납주의 과학관과 달리 법칙이 **내재적 (Implicit)**이라는 점에서 가장 큰 차이가 존재한다. 이러한 특징 때문에 데이터 과학에서 만들어진 확률 모델은 주로 블랙박스 모형으로도 불린다. 데이터 과학을 통해 얻은 모형을 통해 현상을 예측할 수 있지만 왜 그렇게 예측되는지 설명하기 어렵다. 대표적인 예시로 구글 딥마인드의 알파고가 있다. 알파고는 가장 효과적인 수를 예측할 수 있지만 왜 그 수가 가장 효과적인지 설명하기 어렵다. 알파고는 딥 러닝 알고리즘에 기반하여 확률 모형이라는 내재적으로 만들어진 과학적 법칙에 의해 행동하기 때문이다.

데이터 과학 또한 귀납주의 과학관과 같이 과학적 법칙의 보편화에 대한 어려움과 데이터 과학을 통해 얻은 법칙 간의 인과성을 파악하기 어렵다는 한계가 있다. 하지만 전자는 최근 IT 기술을 활용해 인간의 인지 능력의 범위를 비약적인 확장으로 극복하려는 노력을 하고 있다. 후자의 경우 데이터 과학자들은 크게 문제 삼지 않는데 이는 데이터 과학자는 급변하는 사회에서 인과관계보다 상관관계를 파악하는 것이 더 효율적이라 생각하기 때문이다.

결론

데이터 과학은 귀납주의적 과학관, 연역주의적 과학관 전체론적 과학관 중 귀납주의적 과학관과 가장 유사한 성격을 갖는다. 때문에 일각에서는 데이터 과학이 그저 유행에 불과하고 실체가 없다고 비판하기도 한다. 하지만 4차 산업혁명과 함께 변화의 속도가 점점 가속화 되면서 기존의 과학관으로는 관찰을 과학적 법칙으로 만들고 현실을 예측하는데 많은 어려움이 있다. 이러한 상황에서 데이터 과학은 하나의 대안이 될 수도 있을 것이다.

참고자료

Cathy O'Neil, Rachel Schutt, "Doing Data Science.", (O'REILLY, 2013)
Harlan Harris, Merck Veisman, Sean Murphy, "Analyzing the Analyzers.", (O'REILLY, 2013)
한국데이터베이스진흥원, "데이터 분석 전문가 가이드(ADP, APsP).", (한국데이터베이스진흥원, 2014)
John Tukey, "The Future of Data Analysis.", (Annals of Mathematical Statistics, Volume 33, 1962)

Andrew Gelman, American Statistical Association (2004), "Exploratory Data Analysis for Complex Models"
네이버 캐스트, "데이터과학", http://nevercast.never.com/contents.nhn?rid=122&contents_id=104547, (2016. 11. 23)
네이버 캐스트, "데이터 과학자", <http://terms.never.com/entry.nhn?docId=1691563&cid=42171&categoryId=42183>, (2016. 11. 23)