

수문학적 데이터의 자릿수 빈도 분석 적용가능성 연구

A study on applicability of the digit frequency analysis to Hydrological Data

박정은*, 맹승진**, 임광섭***

Jung Eun Park, Seung Jin Maeng, Kwang Suop Lim

요 지

벤포드 법칙(Benford's Law)은 실생활에서 관찰되는 수치 데이터를 첫 자리 숫자에 따라 분류할 때 첫 자리의 숫자가 커질수록 그 분포가 점차 감소되는 현상을 말한다. 이러한 벤포드 법칙은 일반식으로 도출하여 다양한 자릿수로 확장하여 적용할 수 있는 연구결과가 제시되었으며, 회계학, 사회과학, 물리학, 컴퓨터과학, 생물학 등 다방면의 수치 자료에서 그 유효성이 확인되고 있다. 자릿수의 관찰빈도를 분석하는 것만으로 많은 양의 실생활 데이터에서 빠르고 쉽게 데이터 조작 여부를 탐지하거나 1차적인 데이터 품질검사에 효과적으로 활용되고 있다.

본 연구에서는 다학제적 연구의 측면에서 수학·물리적 법칙인 벤포드 법칙을 일유량 등 다양한 수문학 측정자료에 적용하여 그 적용가능성을 확인하고 자료의 불균질성과 신뢰성을 빠르게 탐지할 수 있는 방법론을 제시하고자 한다. 수문자료는 공인심의를 통해 자료의 신뢰도를 확보하고 있으나 확정·배포까지 약 2년이 소요되어 활용기간 단축에 대한 사용자 요구가 지속되고 있는 실정이다. 따라서 본 연구에서는 분석대상 데이터의 자릿수 관찰빈도가 벤포드 법칙에 의한 예상 자릿수 빈도를 따르는지 여부에 대한 가설을 설정하고 카이제곱 검정 또는 Kolmogorov-Smirnov(K-S) 검정 등을 통해 적합도에 대한 통계적 유의미함을 분석함으로써 대략적으로나마 빠르고 쉽게 측정자료의 신뢰성을 판단할 수 있다. 본 연구는 다양한 학문과의 결합을 통한 새로운 접근을 시도함으로써 빅데이터 시대에 효과적으로 수자원의 개발, 관리 및 운영의 의사결정을 하는데 도움이 될 수 있을 것으로 판단된다.

핵심용어 : 벤포드 법칙, 자릿수 빈도 분석, 품질 검사, 수문자료

* 정회원 · 행정안전부 국립재난안전연구원 시설연구사 · E-mail : jungeunpark@korea.kr

** 정회원 · 충북대학교 농업생명환경대학 지역건설공학과 교수 · E-mail : maeng@chungbuk.ac.kr

*** 정회원 · K-water연구원 책임연구원 · E-mail : oklim@kwater.or.kr