NASA Exoplanet Archive 기반 발견시설별 외계행성 크기 분포와 표본 축적 편향 연구

장연희(*), 정은비(**), 황세웅(***)†

(*, **, ***) 선문대학교 AI소프트웨어학과, {j_astronaut, jeung9232}@naver.com, indimoa@gmail.com

† 지도교수(Supervisor) : 황세웅

1. 연구 배경

외계행성 탐사는 Kepler, K2, TESS 등 우주 기반 발견시설과 다양한 지상 관측 시설을 통해 빠르게 확대되어 왔다[3]. 그러나 NASA Exoplanet Archive에 기록된 자료는 실제 우주의 완전한 행성 분포라기보다, 각 발견시설과 발견방법의 관측 특성이 반영된 확인 행성 표본으로 이해해야 한다[1][2].

NASA Exoplanet Archive는 외계행성 연구를 지원하는 공식 데이터베이스이며, Planetary Systems Composite Parameters(PSCompPars)는 외계행성의 행성·항성·시스템 파라미터를 종합한 표이다[1][2]. 본 연구는 데이터 내 발견시설 및 행성 반지름 정보를 활용하여 Kepler, K2, TESS, 기타 발견시설의 행성 크기 분포를 비교한다. 이를 통해 NASA Archive에 축적된 표본에서 발견 시설별 크기 분포 차이를 분석하고, Transit 발견 행성만을 별도로 검토하여 발견방법을 제한한 조건에서도 그 차이가 유지되는지 확인하고자 한다.

2. 연구 내용

본 연구는 NASA Exoplanet Archive의 PSCompPars 자료를 사용하였다[2]. 분석 대상은 1992년부터 2025년까지 기록된 외계행성 표본이며, 행성 반지름 변수인 pl_rade 값이 결측되지 않은 6,035개 표본을 중심으로 분석하였다.

disc_facility는 행성 발견에 사용된 시설 또는 미션을, pl_rade는 지구 반지름 단위의 행성 반지름을 의미한다[2]. 본 연구에서는 pl_rade를 기준으로 Earth−size($0.5 \le R \le 1.5$), Super−Earth($1.5 < R \le 2.0$), Neptune−like($2.0 < R \le 6.0$), Jupiter−like($R > 6.0$)의 네 구간을 정의하였다. Sub−Earth or Other 구간은 8개 표본에 불과하여 카이제곱 검정에서는 제외하였고, 최종적으로 6,027개 표본을 검정 대상으로 사용하였다. 발견시설 그룹은 disc_facility 값을 기준으로 Kepler, K2, TESS, Other Facilities로 구분하였다. 분석 결과는 표 1과 그림 1에서 제시하였다.

표 1. 발견 시설별 행성 크기 구간 비율

발견시설 그룹	N (표본 수)	Earth -size	Super -Earth	Neptune -like	Jupiter -like
Kepler	2,772	25.6%	19.8%	48.4%	6.2%
K2	549	21.3%	18.4%	50.6%	9.7%
TESS	739	11.9%	9.1%	40.5%	38.6%
Other Facilities	1,967	3.9%	2.7%	16.2%	77.3%

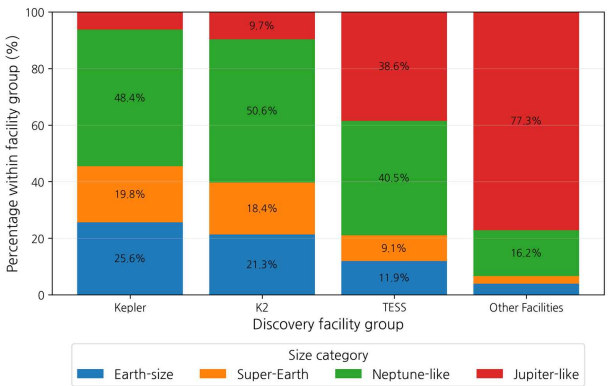


그림 1. 발견시설 그룹별 외계행성 크기 구간 구성 비율

표 1과 그림 1에서 Kepler와 K2는 소형 및 Neptune−like 행성 비율이 높고, Other Facilities는 Jupiter−like에 편중되었다. TESS는 Neptune−like와

Jupiter-like가 함께 높은 중간적 분포를 보였다. 이러한 분포 차이가 통계적으로 유의한지 확인하기 위해 전체 표본과 Transit-only 표본에 대해 카이제곱 독립성 검정을 수행하였다(표 2).

표 2. 표본별 카이제곱 독립성 검정 결과

분석 대상	N	χ^2	df	p-value	Cramér's V
전체 표본	6,027	2820.688	9	<.001	0.395
Transit-only	4,486	2087.623	9	<.001	0.394

카이제곱 독립성 검정 결과, 전체 표본과 Transit-only 표본 모두에서 발견시설 그룹과 행성 크기 구간 사이의 유의한 관련성이 확인되었다. Cramér's V도 각각 0.395, 0.394로 유사하게 나타나, 발견방법을 Transit으로 제한해도 시설별 크기 분포 차이가 유지됨을 확인하였다.

3. 현실적 제한 조건에 대한 검토

본 연구의 대상은 NASA Exoplanet Archive에 기록된 확인 외계행성 표본에 한정되므로, 실제 우주의 전체 행성 크기 분포로 일반화하기는 어렵다. 또한, Other Facilities는 다양한 발견시설이 포함된 이질적 그룹이므로, 결과는 NASA Exoplanet Archive에 축적된 발견 표본의 분포 특성으로 한정해 해석해야 한다.

4. 기대효과 및 활용

본 결과는 외계행성 데이터 분석에서 표본의 발견 수뿐 아니라, 어떤 발견시설을 통해 축적되었는지도 함께 고려해야 함을 보여준다. 특히 Kepler, K2, TESS가 모두 Transit 기반 발견시설임에도[3] 크기 분포가 동일하지 않았다는 점은 동일한 발견방법 안에서도 발견 시설별 표본 축적 특성이 다르게 나타날 수 있음을 시사한다. 따라서 본 연구는 외계행성 발견 데이터 해석에서 관측 편향과 표본 축적 특성을 고려하는 자료로 활용될 수 있다.

5. 결론

본 연구는 NASA Exoplanet Archive의 PSCompPars 자료를 기반으로 Kepler, K2, TESS 및 기타 발견시설의 외계행성 크기 분포를 비교하였다. 그 결과 Kepler와 K2는 소형 및 Neptune-like 행성 비율이 높고, Other Facilities는 Jupiter-like에 편중되었으며 TESS는 중간적 분포를 보였다. 또한, Transit-only 분석에서도 유사한 결과가 유지되어, NASA Archive 자료에는 발견방법뿐 아니라 발견 시설별 표본 축적 특성이 반영되어 있음을 확인하였다. 이는 Transit 기반 발견시설의 관측 방식과 후속 확인 과정의 차이가 표본 구성에 일부 반영된 결과로 해석된다[3]. 이에 따라 향후 연구에서는 공전주기, 모항성 특성, 거리, 관측 기간 및 후속 검증 여부를 함께 고려해 발견 시설별 표본 구성 차이의 원인을 분석할 필요가 있다.

감사의 글

본 연구는 2026년도 과학기술정통신부 및 정보통신기획평가원의 'SW중심대학지원사업'의 연구 결과로 수행되었습니다. (2024-0-00023)

참고 문헌 (참고자료)

[1] R. L. Akeson et al., "The NASA Exoplanet Archive: Data and Tools for Exoplanet Research," PASP, Vol.125, pp.989-999, 2013, doi: 10.1086/672273.

[2] NASA Exoplanet Archive, "Planetary Systems Composite Parameters Table Documentation." [Online]. Available: https://exoplanetarchive.ipac.caltech.edu/docs/pscp_about.html. Accessed: May 18, 2026.

[3] NASA Science Editorial Team, "How do you find and confirm – a planet?," NASA Science, Sep. 23, 2018. [Online]. Available: https://science.nasa.gov/universe/exoplanets/how-do-you-find-and-confirm-a-planet-10-things-about-the-search-for-exoplanets/. Accessed: May 18, 2026.