**주가 조작 가능성 분석 보고서**

202401833\_신해솔

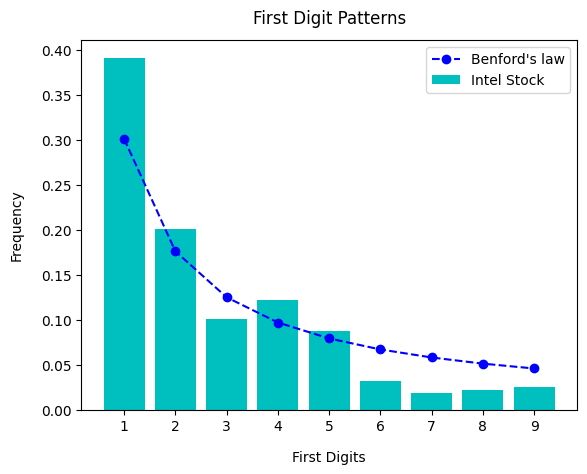
1. **서론**

벤포드 법칙이란, 자연적으로 발생하는 수치 데이터의 첫째 자리 숫자의 빈도가 특정 비율을 따른다는 법칙이다. 이 법칙에 따르면, 자연적으로 발생하는 수치 데이터에서 첫째 자리 숫자는 1로 시작하는 경우가 가장 많고, 9로 시작하는 경우는 가장 적다. 주가 데이터 역시 자연적으로 발생하는 수치 데이터로 간주될 수 있기에, 조작되지 않았다면 벤포드 법칙에 따른 분포를 보일 것으로 예상된다.

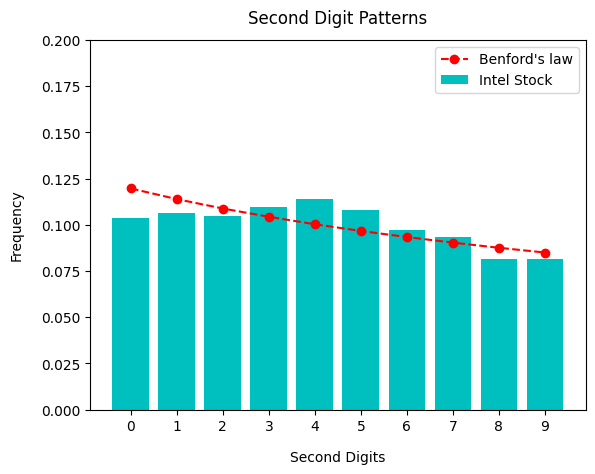
Benford(1938)와 Nigrini(1999)에 따르면, 표본 데이터집합들이 최소한 4 또는 5 자리 이상의 숫자이여야 하며, 만약 숫자가 4자릿수 보다 작다면, 둘째 숫자의 빈도를 이용해야 한다고 한다. 분석의 사용될 주가 데이터의 정수 부분의 자릿수의 평균은 2.3이므로, 벤포드 법칙에 따른 빈도와 실제 데이터의 첫째 자리 숫자 빈도와 둘째 자리 숫자의 빈도를 각각 비교하여 벤포드 법칙 위반 여부를 판단할 것이다.

주가 데이터가 벤포드 법칙에 따른 빈도를 보이는지는 평균절대분산(Mean Absolute Deviation, MAD)를 이용해 확인하고자 한다. MAD는 관측된 첫째 자리 숫자 또는 둘째 자리 숫자의 빈도와 벤포드 법칙에 따른 기대 빈도 간의 평균 절대 편차를 측정하는 지표이다. 이를 통해 데이터의 분포가 벤포드 법칙과 얼마나 일치하는지를 수치적으로 평가할 수 있다. MAD 값이 0에 가까울수록 데이터가 벤포드 법칙을 잘 따른다는 의미이다.

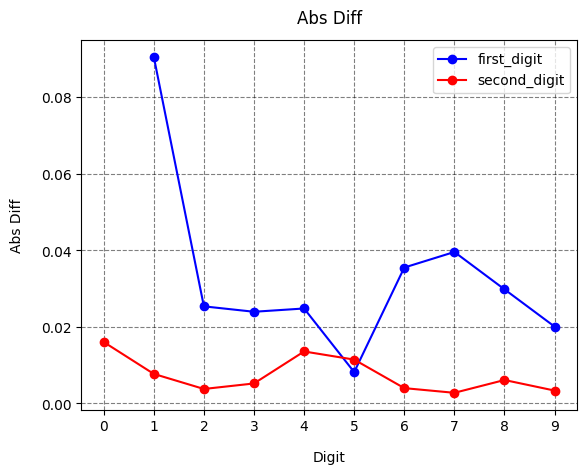
1. **본론**



먼저, 주가 데이터와 벤포드 법칙에 따른 첫째 자리 빈도를 비교했다. 첫째 자리가 1인 부분에서 약 10%p의 차이가 나타났고, 6부터 9의 범위에서도 주가데이터의 첫째 자리의 빈도가 벤포드 법칙에 따른 첫째 자리 빈도보다 다소 적게 나타났다. 첫째 자리의 빈도로만 비교했을 때에는 주가 데이터가 조작되었을 가능성이 높다고 판단된다.

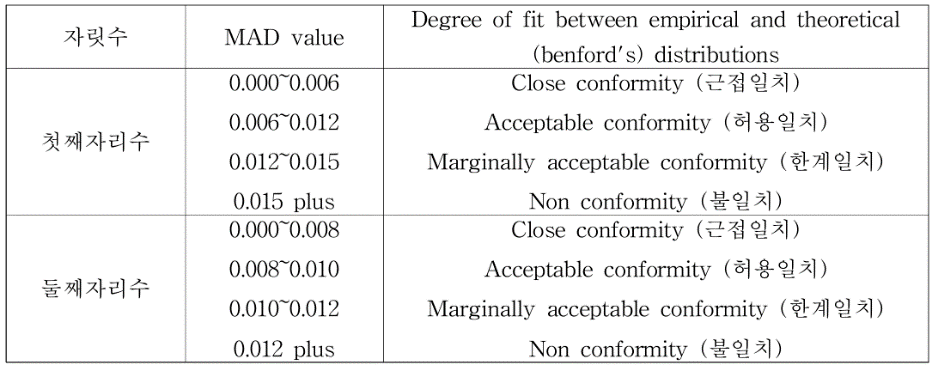


다음으로는 주가 데이터와 벤포드 법칙에 따른 둘째 자리 빈도를 비교했다. 첫째 자리와 달리 둘째 자리의 빈도는 최대 약 2.5%p의 차이를 보였다. 주가 데이터의 정수 부분의 평균 자리수가 2.3이므로 둘째 자리를 비교한 데이터가 첫째 자리를 비교한 데이터보다 주가 조작 여부 판단 시 더 높은 신뢰도를 가질 것이다. 분석할 수치 데이터가 4자릿수 보다 작다면 둘째자리를 이용해 벤포드 법칙을 적용해야 하기 때문이다.



보다 구체적인 차이를 확인하기 위해 각 자릿수에 대해 벤포드 법칙에 따른 빈도와 주가 데이터의 첫째, 둘째 자리 빈도의 절대 차이(Abs Diff)를 구하고 시각화했다. 앞에서 보았던 것과 같이 첫째 자리 빈도의 절대 차이는 1, 6, 7, 8, 9에서 크게 나타났으며, 둘째 자리 빈도의 절대 차이는 비교적 적게 나타났다. 벤포드 법칙과 주가 데이터의 첫째, 둘째 자리의 빈도가 얼마나 일치하는지 산술적으로 평가하기 위해 MAD값을 각각 구해보았다. 첫째 자리의 MAD는 0.03311이, 둘째 자리의 MAD는 0.00742로 나타났다.

<MAD 값에 따른 신뢰 수준>



첫째 자리 빈도는 불일치 구간에 있음을, 둘째 자리 빈도는 근접일치 구간에 있음을 확인할 수 있다.

1. **결론**

벤포드의 법칙과 주가 데이터에 따른 첫째 자리의 빈도는 높은 유사성을 보이지 않았고 MAD 값 또한 불일치 구간에 있었다. 따라서, 주가 데이터의 첫째 자리만 고려했을때는 주가 데이터가 조작되었을 가능성이 높다고 판단된다.

하지만, 주가 데이터 정수 부분의 평균 자릿수가 2.3임을 고려했을때는 첫째 자리 빈도가 아닌 둘째 자리의 빈도를 통해 조작 여부를 판단하는 것이 타당한 듯하다. 벤포드의 법칙과 주가 데이터에 따른 둘째 자리의 빈도는 높은 유사성을 보였으며, MAD 값 또한 근접일치 구간에 있었으므로 주가 데이터가 조작되었을 가능성은 낮다고 판단된다.

그러나 벤포드 법칙을 적용하는 범위를 임의로 정수 부분으로 한정하고 그 자릿수에 따라 둘째 자리 빈도로 주가 데이터의 조작여부를 판단한 것이므로, 특정 소수점 아래까지 고려한다면, 주가 데이터가 조작 가능성이 높다고 판단할 수 있을 것이다. 또한, 벤포드 법칙과 주가 데이터의 유사성을 MAD로만 판단하였기에 결과를 신뢰할 수 있는 정도가 그리 높지 않다고 볼 수 있다. 다른 검정 방식을 사용한다면 더 높은 신뢰도를 갖는 결론을 도출할 수 있을 것이다.

**참고 문헌**

서원율. (2016, August). 벤포드 법칙을 이용한 신용등급별 보고이익 수치의 신뢰성 검증.

Benford. (1938). *The Law of Anomalous Numbers*.

Nigiri. (1999). *The Law of Anomalous Numbers*.