<논문1>

**요약 및 해석: 데이터 분석에서의 로그 변환과 그 의미**

**요약:**

로그 변환은 생물 의학 및 심리 사회 연구에서 왜도 데이터(skewed data)를 처리하기 위해 널리 사용됩니다. 이 논문은 왜도 데이터를 처리하는 데 있어 이 고전적 접근 방식의 심각한 문제점을 강조합니다. 일반적으로 로그 변환이 데이터의 변동성을 줄이고 데이터가 정규 분포에 더 가깝게 만들 수 있다고 믿어지지만, 이는 실제로 그렇지 않은 경우가 많습니다. 또한, 로그 변환된 데이터에 대해 수행된 표준 통계 테스트의 결과는 원래의 비변환 데이터에 대해 관련성이 없을 수 있습니다. 우리는 시뮬레이션된 데이터를 사용한 예를 통해 이러한 문제를 설명합니다. 데이터 변환을 사용할 경우 매우 신중하게 적용해야 하며, 대부분의 경우 연구자들은 왜도 데이터를 처리하는 이러한 전통적인 방법을 버리고 일반화 추정 방정식(GEE)과 같은 최신 분석 방법을 사용할 것을 권장합니다.

**키워드:** 가설 검정, 이상치, 로그 정규 분포, 정규 분포, 왜도

**해석:**

**1. 서론**

로그 변환은 왜도 데이터를 처리하기 위해 널리 사용되는 방법입니다. 사용의 용이성과 인기 덕분에, SAS, Splus, SPSS와 같은 주요 통계 소프트웨어 패키지에 포함되어 있습니다. 그러나 그 인기로 인해 오용의 가능성도 큽니다. 특히 통계학자들조차도 로그 변환을 잘못 해석하거나 오용하여 실험 결과를 잘못 해석할 수 있습니다. 이 논문은 로그 변환의 문제점과 그 사용에 대한 한계점을 논의합니다.

**2. 로그 정규 변환**

**2.1 로그 변환을 사용하여 데이터의 정규성 확보**

정규 분포는 연속적인 결과를 모델링하기 위해 기초 및 임상 연구에서 널리 사용됩니다. 그러나 대칭적인 종 모양의 분포가 연구 프로젝트에서 관찰된 데이터를 적절히 설명하지 못하는 경우가 많습니다. 이러한 경우 데이터 변환을 통해 데이터를 ‘정규’에 가깝게 만듭니다. 하지만 많은 연구 데이터는 로그 정규 분포를 따르지 않기 때문에 로그 변환이 분포의 왜도를 줄이지 못할 뿐만 아니라, 오히려 더 왜곡시킬 수 있습니다.

**2.2 로그 변환을 사용하여 데이터 변동성 감소**

로그 변환의 또 다른 인기 있는 용도는 데이터의 변동성을 줄이는 것입니다. 그러나 이 일반적인 믿음과는 달리, 로그 변환은 데이터 변동성을 줄이는 대신 오히려 증가시킬 수 있습니다. 로그 변환 후 데이터의 해석 또한 어렵기 때문에, 변환된 데이터와 원래 데이터 간의 차이를 이해하는 것이 중요합니다.

**3. 로그 변환된 데이터의 모델 추정의 해석**

로그 변환된 데이터에 대해 많은 통계적 방법을 적용할 수 있습니다. 그러나 로그 변환된 데이터의 평균을 원래 데이터의 평균으로 추정하는 것은 부정확한 결과를 초래할 수 있습니다. 또한, 가설 검정에 로그 변환된 데이터를 사용할 때, 원래 데이터에 대한 가설을 테스트하는 것이 아닐 수 있습니다.

**4. 결론**

로그 변환 및 기타 데이터 변환은 문제가 많습니다. 데이터 변환을 사용할 경우, 그 한계점을 명심하고 원래 데이터에 대한 가설과의 관련성을 신중하게 해석해야 합니다. 많은 경우, 데이터의 분포를 변환하여 모델링하는 것보다는 분포에 의존하지 않는 최신 방법을 사용하는 것이 더 나을 수 있습니다. 예를 들어, 일반화 추정 방정식(GEE) 접근법은 데이터의 분포와 상관없이 유효한 추론을 제공합니다.

**결론:**

이 논문은 로그 변환의 한계와 문제점을 강조하며, 데이터 분석 시 분포에 의존하지 않는 현대적인 방법을 사용할 것을 제안합니다.

<논문2>

**요약 및 해석: 통계에서 로그 변환의 최선의 사용 방법**

**요약:**

로그 변환은 측정 변수의 왜도를 줄이기 위해 자주 사용됩니다. 변환 후 분포가 대칭적이라면 Welch t-검정을 사용해 그룹을 비교할 수 있습니다. 또한 분포가 정규에 가까워진다면 참조 구간을 결정할 수도 있습니다.

**키워드:** 통계, 로그 변환, 참조 구간, t-검정

**해석:**

**1. 서론**

이 논문의 목적은 왜도 데이터에 적합한 변환을 사용하는 최선의 방법을 설명하는 것입니다. 예제로는 National Health and Nutrition Examination Study (NHANES) 코호트의 데이터를 사용합니다. 2017-2018년 데이터 중 18-29세 사이의 건강한 피험자들의 요알부민/크레아티닌 비율을 분석합니다.

**2. 변환 선택**

알부민 데이터는 일반적으로 양수이거나 0으로, 높은 값으로 갈수록 긴 꼬리를 가집니다. 적합한 변환으로는 제곱근 변환이나 로그 변환이 있습니다. 변환의 목표는 분포를 대칭적으로 만들어 추가적인 통계 기법의 기초를 제공하는 것입니다. 변환된 분포가 완전히 정규 분포일 필요는 없지만, 그렇게 되면 더 작은 샘플 기반의 테스트에 더 많은 신뢰를 가질 수 있습니다.

**3. 자연 로그 또는 일반 로그**

자연 로그 (기저 e) 또는 일반 로그 (기저 10)를 사용할 수 있습니다. 일반 로그의 장점은 해석이 더 쉽다는 점입니다. 예를 들어, log10 값이 ‘2.xxx’이면 100과 1000 사이에 해당합니다.

**4. 역변환**

원래의 단위(μg/mL)로 생각하는 것이 더 도움이 되므로 역변환이 필요합니다. 예를 들어 log10 값이 0.933이면 역변환 후 8.6 μg/mL가 됩니다. 이는 요알부민 측정의 산술 평균이 아니라 기하 평균입니다.

**5. 참조 구간**

참조 구간을 얻기 위해서는 (a) 측정을 변환하여 정규 분포에 가깝게 만들고, (b) 변환된 척도에서 참조 구간을 계산한 후, (c) 다시 변환합니다. 변환된 로그10 변수의 참조 구간을 계산하면 남성의 경우 1.2–62.2 μg/mL, 여성의 경우 1.2–92.2 μg/mL가 됩니다.

**6. 요약**

이 예시는 실제 샘플 데이터를 사용하였으며, 로그 변환의 사용과 오류 가능성을 설명합니다. 분포 비교에는 Wilcox 순위 합 검정이 더 간단하며, 참조 구간의 경우 로그 변환이 이상치로 인해 오차를 유발할 수 있습니다.

**결론:**

로그 변환은 데이터의 왜도를 줄이고 분포를 대칭적으로 만드는 데 사용될 수 있지만, 실제 데이터를 다룰 때에는 주의가 필요합니다. 참조 구간을 설정하거나 그룹 간 비교를 할 때는 변환 방법의 한계와 오류 가능성을 염두에 두어야 합니다.

**Acknowledgements:**

이 논문 작성에 도움을 준 ACB 편집팀에게 감사드립니다.

**Declaration of Conflicting Interests:**

저자는 이 연구, 저작 및 출판과 관련하여 잠재적인 이해 충돌이 없음을 선언합니다.

**Funding:**

저자는 이 연구, 저작 및 출판에 대한 재정적 지원을 받지 않았습니다.

**Ethical Approval:**

사용된 데이터는 공공 데이터로 별도의 윤리적 승인 필요 없음.

**Guarantor:**

RMW.

**Contributorship:**

RMW 단독 기여.

이 논문은 로그 변환의 사용 및 한계를 실제 데이터를 통해 설명하며, 변환 후 통계적 분석의 정확성을 유지하는 방법을 제시합니다.