

추론통계 개요

#01. 데이터 분석의 구분

| | 탐색적 데이터 분석(EDA) | 확증적 데이터 분석(CDA) |
|----|-----------------------------------|---|
| 활동 | 기술통계 | 추론통계 |
| 정의 | 데이터를 요약해 설명하는 기법 | 단순히 숫자를 요약하는 것을 넘어 어떤 값이 발생할 확률을 계산하는 통계 기법. 모집단에서 샘플링한 표본을 가지고 모집단의 특성을 추론하고 그 결과가 신뢰성이 있는지 검정하는 것이다. |
| 예시 | 사람들이 받는 월급을 집계해 전체 월급 평균을 구한다. | 수집된 데이터에서 성별에 따라 월급에 차이가 있는 것으로 나타났을 때 이런 차이가 우연히 발생할 확률을 계산한다. |
| 과정 | 데이터 수집 > 시각화 탐색 > 패턴 도출 > 인사이트 발견 | 가설 설정 > 데이터 수집 > 탐색적 데이터 분석 > 추론통계 > 가설검증 |

#02. 추론통계

정의

모집단으로부터 추출된 표본의 표본통계량으로 부터 모집단의 특성인 모수에 관해 통계적으로 추론하는 절차

실질적인 데이터 분석(통계)를 의미하는 과정

- 자료의 정보를 이용해 집단에 관한 추측, 결론을 이끌어내는 과정
- 수집된 자료를 이용해 대상 집단(모집단)에 대한 의사결정을 하는 것으로 Sample을 통해 모집 단을 추정하는 것을 의미
- 제한된 표본을 바탕으로 모집단에 대한 일반적인 결론을 유도하려는 시도이므로 본질적으로 불확실성을 수반함.

추론 통계의 결론

- 성별에 따른 월급의 차이가 **우연히 나타날 확률이 작다면** 통계적으로 ****유의하다(statistically significant)****라고 결론 내린다.
- 성별에 따른 월급의 차이가 **우연히 나타날 확률이 크다면** 통계적으로 ****유의하지 않다(not statistically significant)****고 결론 내린다.

일반적으로 통계 분석을 수행했다는 것은 추론 통계를 이용해 가설 검정을 했다는 의미.