

기술 통계 (Descriptive Statistics)

수집한 데이터의 전반적인 특성을 요약하고 설명하는 기초적인 방법론입니다.

- 중심 경향 (Central Tendency):** 데이터가 어디에 집중되어 있는지를 나타냅니다.
 - 평균 (Mean):** 모든 값의 합을 개수로 나눈 산술적 평균입니다. 극단적인 값(이상치)에 민감하게 반응합니다.
 - 중앙값 (Median):** 데이터를 크기순으로 나열했을 때 한가운데 위치하는 값입니다. 이상치에 의한 왜곡이 적습니다.
 - 최빈값 (Mode):** 데이터 중에서 가장 빈번하게 나타나는 관측값입니다.
- 산포도 (Dispersion):** 데이터가 얼마나 넓게 퍼져 있는지를 측정합니다.
 - 범위 (Range):** 최댓값과 최솟값의 차이입니다.
 - 분산 (Variance):** 각 관측값이 평균으로부터 떨어진 거리(편차)의 제곱 합을 평균 낸 것입니다.
 - 표준편차 (Standard Deviation):** 분산의 제곱근으로, 데이터의 변동성을 원 데이터와 같은 단위로 직관적으로 보여줍니다.
 - 사분위수 (Quartiles):** 전체 데이터를 4등분하는 지점들(Q1, Q2, Q3)입니다. **IQR(사분위범위) = Q3 - Q1**은 중간 50% 데이터의 확산 정도를 나타냅니다.

추론 통계 (Inferential Statistics)

표본(Sample) 데이터를 분석하여 수집되지 않은 모집단(Population) 전체의 특성을 추측하고 결론을 도출하는 방법입니다.

- 가설 검정 (Hypothesis Testing):**
 - 귀무가설 (Null Hypothesis, H0):** 차이가 없거나 효과가 없다는 '기존의 사실'을 나타내는 가설입니다.
 - 대립가설 (Alternative Hypothesis, H1):** 연구자가 새롭게 입증하고자 하는 '차이가 있다'는 주장입니다.
 - p-value (유의 확률):** 귀무가설이 참이라는 전제하에, 현재 관측된 데이터(또는 그보다 더 극단적인 데이터)가 우연히 나타날 확률입니다.
 - 유의수준 (Significance Level, α):** 귀무가설을 기각할지 결정하는 기준점으로 보통 0.05(5%)를 사용합니다. $p < \alpha$ 일 때 귀무가설을 기각합니다.
- 신뢰 구간 (Confidence Interval):** 모집단의 실제 모수(예: 평균)가 포함될 것으로 기대되는 통계적 범위입니다. (예: 95% 신뢰구간)

주요 확률 분포

- 정규 분포 (Normal Distribution):** 평균을 중심으로 좌우 대칭인 종 모양의 분포로, 자연 및 사회 현상을 설명하는 가장 대표적인 분포입니다.
- 이항 분포 (Binomial Distribution):** 성공/실패와 같이 결과가 두 가지뿐인 베르누이 시행을 독립적으로 여러 번 반복했을 때의 성공 횟수 분포입니다.
- 포아송 분포 (Poisson Distribution):** 정해진 시간이나 공간 단위 내에서 발생하는 희귀한 사건의 횟수에 대한 분포입니다.
- t-분포 (Student's t-Distribution):** 정규 분포와 모양은 비슷하나 양 끝단이 더 두껍습니다. 표본 크기가 작을 때의 평균 추론에 주로 사용됩니다.
- 카이제곱 분포 (Chi-squared Distribution):** 독립성 검정이나 적합도 검정 등 범주형 데이터 분석에 널리 쓰입니다.

주요 통계 검정 방법

- t-검정 (t-test):** 두 집단의 평균 차이가 통계적으로 유의미한지 확인합니다.
 - 독립 표본 t-검정:** 서로 연관이 없는 두 집단의 평균을 비교합니다. (예: 남녀 성적 비교)
 - 대응 표본 t-검정:** 동일 집단의 사전-사후 변화를 비교합니다. (예: 다이어트 전후 체중 비교)
- 분산 분석 (ANOVA):** 세 집단 이상의 평균값 차이를 한꺼번에 비교할 때 사용합니다.
- 카이제곱 검정 (Chi-squared Test):** 두 범주형 변수 간의 연관성이나 독립성을 검정합니다.
- 상관 분석 (Correlation Analysis):** 두 변수 간의 선형적인 관계가 얼마나 강한지 확인합니다.
 - 피어슨 상관 계수 (r):** -1(완벽한 음의 상관)에서 1(완벽한 양의 상관) 사이의 값을 가집니다.

회귀 분석 (Regression Analysis)

하나 이상의 독립 변수(X)를 통해 종속 변수(Y)의 값을 예측하는 함수적 모델을 만드는 방법입니다.

- 단순 선형 회귀:** 하나의 독립 변수로 종속 변수와의 직선 관계를 모델링합니다.
- 다중 선형 회귀:** 여러 개의 독립 변수를 사용하여 종속 변수를 더 정밀하게 예측합니다.
- 로지스틱 회귀:** 종속 변수가 범주형(예: 합격/불합격)일 때, 특정 사건이 발생할 확률을 모델링합니다.

통계적 오류

- 1종 오류 (Type I Error, α):** 실제로는 참인 귀무가설을 실수로 기각하는 오류입니다. (위양성)
- 2종 오류 (Type II Error, β):** 실제로는 거짓인 귀무가설을 기각하지 못하고 채택하는 오류입니다. (위음성)