**확률분포와 파라미터**

확률분포는 통계학에서 자료를 설명하는 기본적인 도구이다. 이론적인 확률분포는 수식으로 표현되며, 분포의 모양을 결정하는 파라미터(모수)를 가진다. 파라미터를 알면 분포의 성질을 이해할 수 있다.

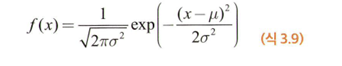
데이터 분석의 핵심 목표는 모집단의 성질을 파악하는 것이다. 예를 들어 “평균이 ○○, 분산이 □□인 정규분포를 따른다”라고 말할 수 있다면, 모집단을 이해한 셈이다.

**정규분포**

가장 대표적인 확률분포는 정규분포(가우스 분포)이다.

정규분포는 평균과 표준편차 두 개의 파라미터로 정의된다.

수식:



평균 μ는 분포의 중심을, 표준편차 σ는 분포의 퍼짐 정도를 결정한다.

특징:

1. 평균을 중심으로 좌우 대칭이다. 2. 평균에서 가까울수록 빈도가 높고, 멀어질수록 빈도가 낮다.

3. 키, 몸무게 등 다양한 자연현상에서 관찰된다.

성질:

μ ± 1σ 범위: 전체 데이터의 약 68% 포함

μ ± 2σ 범위: 약 95% 포함

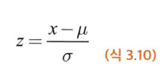
μ ± 3σ 범위: 약 99.7% 포함

또한 “어떤 값이 평균에서 몇 σ 떨어져 있다”는 표현을 통해 해당 값이 얼마나 드문지 설명할 수 있다. 예를 들어 “평균에서 3σ 떨어져 있다”면 약 상위 0.15%에 속한다는 의미이다.

**표준화**

정규분포에서 어떤 값 x가 평균에서 얼마나 떨어져 있는지를 표현하기 위해 표준화를 사용한다.

공식:



=> 평균을 0, 표준편차를 1로 변환한다. z값은 해당 값이 평균에서 몇 표준편차만큼 떨어져 있는지를 나타낸다.

**다양한 확률분포**

정규분포 외에도 다양한 확률분포가 있다.

균등분포: 모든 값이 같은 확률로 발생

이항분포: 두 가지 결과(성공/실패)만 있는 경우

포아송분포: 일정한 시간·공간 내 사건 발생 횟수

지수분포: 사건 사이의 간격을 나타냄

가우스 분포: 연속형, 대표적인 정규분포

또한 통계적 검정에 사용되는 t분포, F분포, χ²분포 등도 존재한다.