

7장 확률변수와 상관관계

6장에서는 확률을 어떻게 정의하는지 살펴보았다. 이 장에서는 실제 데이터값이 어떻게 나올지 묘사하는 데 확률을 사용하는 방법을 알아본다. 확률변수는 특정 데이터의 값을 대표하는 확률모형이다. 먼저 확률변수의 기댓값, 분산, 표준편차 등의 수학적 특성을 공부한다. 다음으로 여러 종류의 데이터가 존재할 때 어떤 변수의 값이 변함에 따라 이와 관련된 다른 변수의 특성이 어떻게 달라지는지 알아보고 예측 문제에 어떻게 활용되는지 살펴본다.

학습 목표

- 확률변수의 정의와 데이터 분석에 확률변수가 어떻게 사용되는지 알아본다.
- 기댓값, 분산, 표준편차의 수학적 정의와 특성을 공부한다.
- 다변수 확률변수간의 공분산과 독립의 의미를 살펴보고 상관계수로 상관관계를 측정하는 법을 알아본다.
- 조건부 기댓값의 정의와 특성을 공부하고 예측 문제에 어떻게 쓰이는지 살펴본다.