통계학

모집단 : 연구의 대상이 되는 모든 개체들을 모은 집합 ( 모든 대상을 분석하는 것 불가능 )

표본 : 모집단의 일부분의 관측값들

모수(모집단의 요약값) : 수치로 표현되는 모집단의 특성

통계량 : 표본의 관측값들에 의해서 결정되는 양

* 통계량을 통해 모수를 알아내는 것이 통계학의 목표
* 자료가 달라지면 행하는 분석이 달라진다

수치형( 연속형 / 이산형 )자료 | 범주형( 순위형 / 명목형 )자료

중심 경향값(대표값)

* 평균(Mean)
* 중앙값(Median) : 크기순으로 정렬시켜 중앙에 위치한 값
* 최빈값(Mode) : 가장 자주 나오는 값

산포도 (퍼진정도)

* 분산 : 분산이 작으면 위로 솟은 모양, 분산이 크면 낮아지는 모양
* 사분위수 범위

정규분포

* 위치는 평균에 의해, 모양은 분산에 의해 결정

분포도

* 왜도(Skewness)

분포의 비대칭 정도 ( 꼬리가 긴쪽의 skewness ) 왼쪽이 길면 left-skewed

* 첨도

분포의 꼬리 부분의 비중에 대한 측도

K = 0 뾰족한 정도가 정규분포와 동일

표본평균, 표본분산 ( 모집단의 파라미터를 알기위해 표본에서부터 얻은 추정량 )

표본분산을 구할 때 표본평균을 이용한다. 모집단의 평균을 모르기 때문에. ?? 분모가 n-1인 이유는?

< 확률의 기초 >

주사위를 던진다.

표본공간(sample space) : S = {1,2,3,4,5,6}

근원사건(sample outcom) 표본공간의 원소 : 1,2,3,4,5,6

사건(event) : 짝수가 나오는 사건 A = {2,4,6} p(A) = 1/2

조건부확률

* P(A|B) = P(A N B)/P(B)

독립

* P(A|B) = P(A), P(B|A) = P(B) => P(A N B) = P(A)P(B)