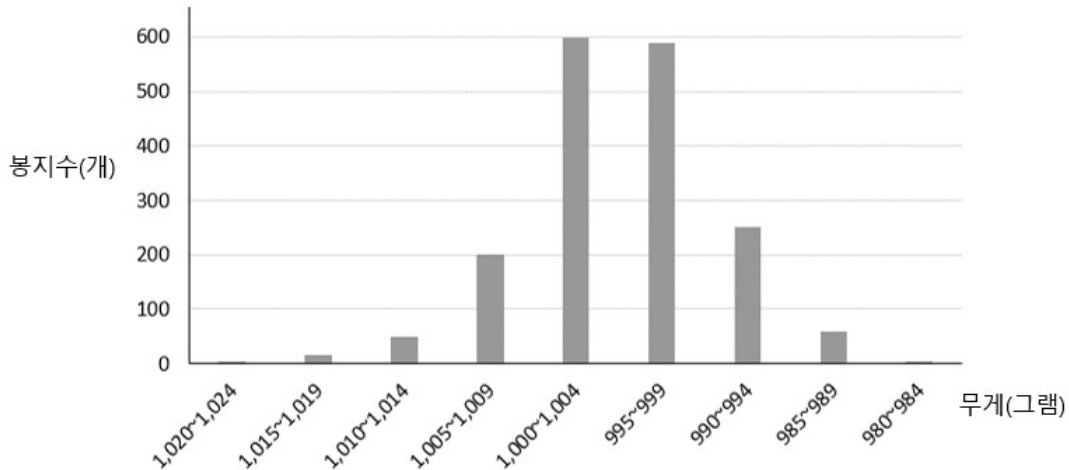


정규분포

사람의 키나 혈압, 제품의 치수나 중량 등과 같이 자연이나 사회에서 관측되는 데이터는 일정한 형태의 분포를 나타낸다. 일 예로, 1Kg 용량의 봉지설탕을 생산하는 공장에서 하루 생산된 모든 봉지설탕의 무게를 측정하면 봉지설탕의 무게는 아래의 히스토그램(Histogram)과 유사한 분포를 보이게 될 것이다.



이처럼 대부분의 데이터가 평균에 가까이 있고 그 평균값을 기준으로 좌우 대칭을 이루면서 평균값에서 멀어질수록 데이터가 존재할 가능성이 적어지는 경우, 그 데이터가 정규분포(Normal Distribution) 또는 가우스분포(Gaussian Distribution)를 보인다고 표현한다.

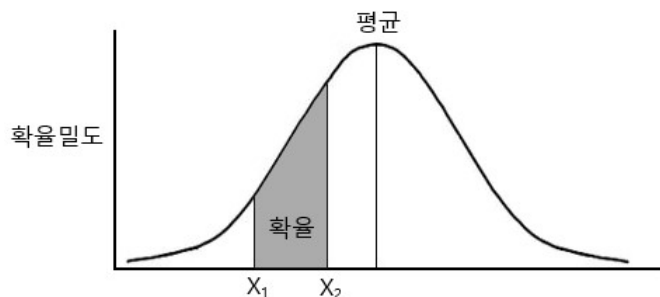
주사위를 던져 6이 나올 확률은 1/6이다. 통계학에서 주사위의 숫자 또는 동전의 면처럼 개별적으로 확률이 정해지는 변수를 이산확률변수(Discrete Random Variable)라 한다.

위의 봉지설탕의 예에서 봉지설탕의 무게가 정확히 1Kg일 확률은 영(Zero)이다. 사람의 키나 어느 지역의 연간 강수량과 같은 변수도 이처럼 연속된 일정 구간이 가지는 확률을 확인할 수 밖에 없는 데 이러한 변수를 연속확률변수(Continuous Random Variables)라 한다. 따라서 연속확률변수의 확률분포 즉, 연속확률분포(Continuous Probability Distribution)는 확률밀도함수(Probability Density Function)를 이용하여 정의한다.

정규분포는 아래의 확률밀도함수로 정의되는 연속확률분포를 의미한다.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{여기서 } x \text{ 는 변수, } \mu \text{ 는 평균, } \sigma \text{ 는 표준편차})$$

즉, 정규분포를 나타내는 어떤 변수의 평균과 표준편차를 위의 공식에 대입하면 아래 그림과 같은 확률밀도함수그래프를 얻을 수 있다. 정규분포는 항상 종 모양을 하고 있어 Bell Curve라 칭하기도 한다.



정규분포의 경우, 확률분포는 평균을 중심으로 좌우가 대칭이며, 변수 X1, X2 범위가 가지는 확률은 위의 그림에서 음영으로 표시된 구간의 면적이 된다.