6 . 정규분포

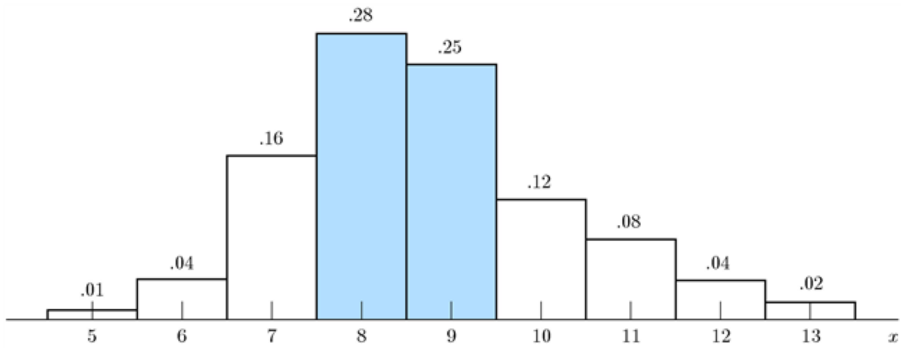
1. 연속확률변수의 확률모형

- **연속확률변수**
  - 어떤 구간에 있는 임의의 값을 연속적으로 취하는 확률변수
  - 몸무게, 힘, 수명, 온도 등
- **오랜 시행에서의 상대도수로의 확률**

  :연속확률분포는 많은 측정값들의 상대도수 히스토그램

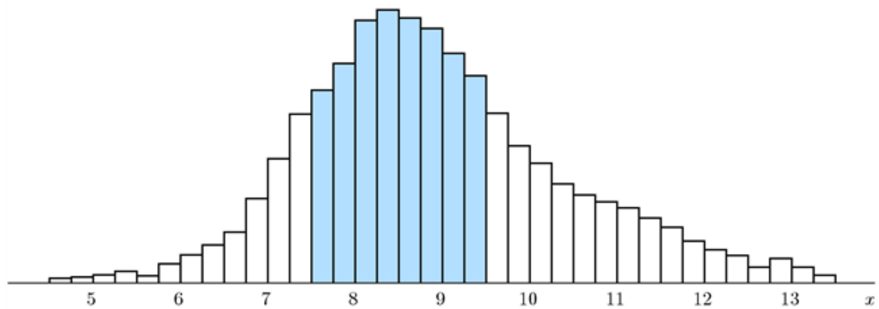
<상대도수 히스토그램 성질>

1. 히스토그램으로 나타나는 전체의 면적은 1이다.
2. 어느 구간에 포함되는 측정값들의 상대도수는 이 구간에서의 사각형의 면적 이다.



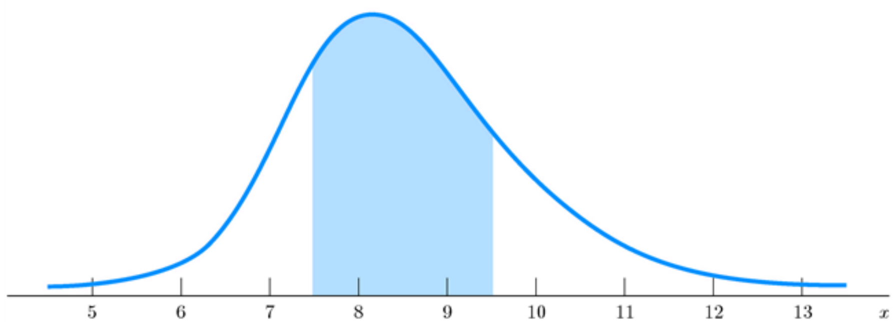
(a)

100명의 측정값



(b)

5000명의 측정값



(c)

상대도수 히스토그램의
극한 형태로 얻어진 곡선

확률밀도함수
(Probability density function: pdf)

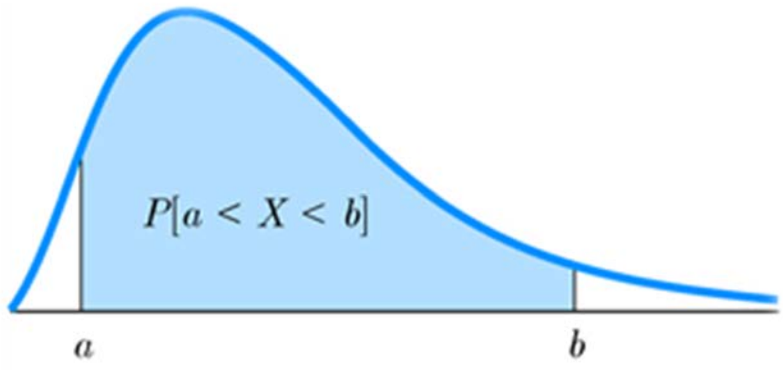
확률밀도함수 $f(x)$ 는 다음 성질을 만족한다.

1. 확률밀도곡선 아래의 전체 면적은 1이다.
2. $P[a \le X \le b]$ 는 a와 b사이의 구간에서 확률밀도곡선 아래의 면적이다.
3. 모든 x 에 대해 $f(x) \ge 0$ 이다.

연속확률변수에서 $P[X = x] = 0$

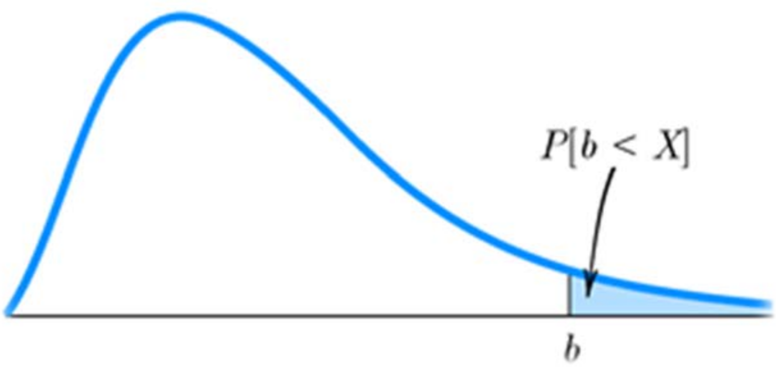
$$P[a \le X \le b] = P[a < X \le b] = P[a \le X < b] = P[a < X < b]$$

$$P[a < X < b]$$



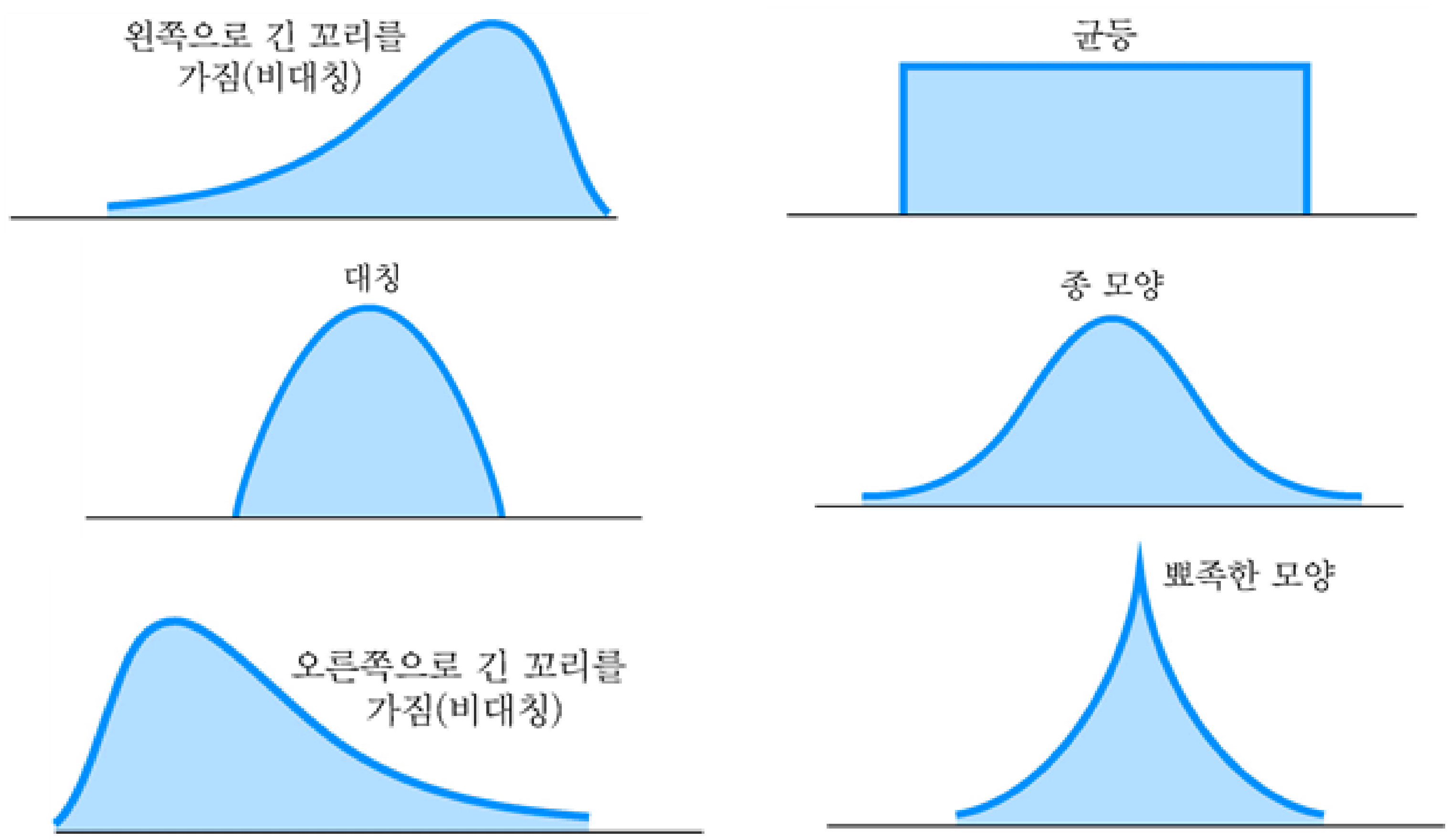
$$P[b < X]$$

$$= 1 - P[X \le b]$$

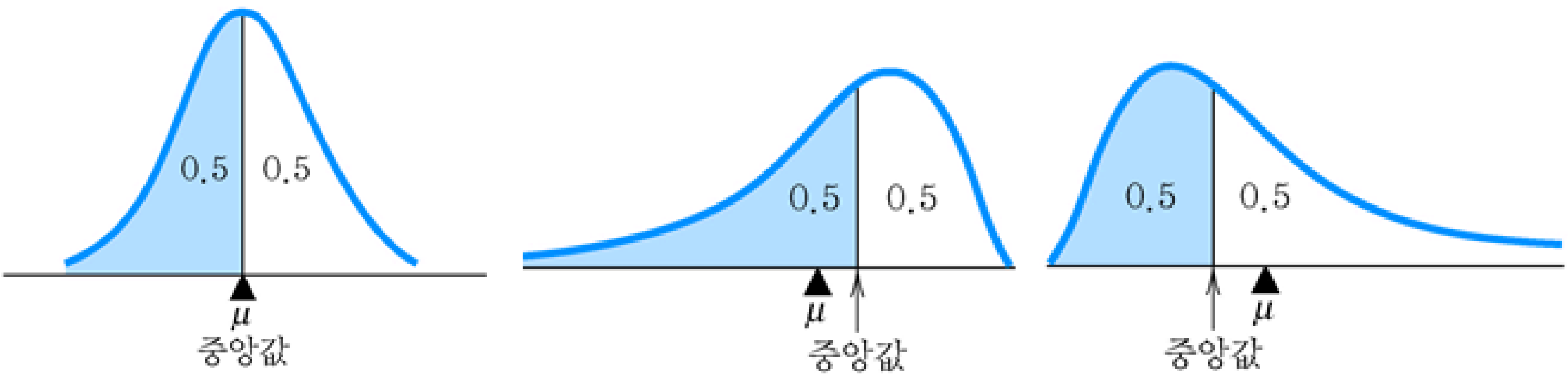


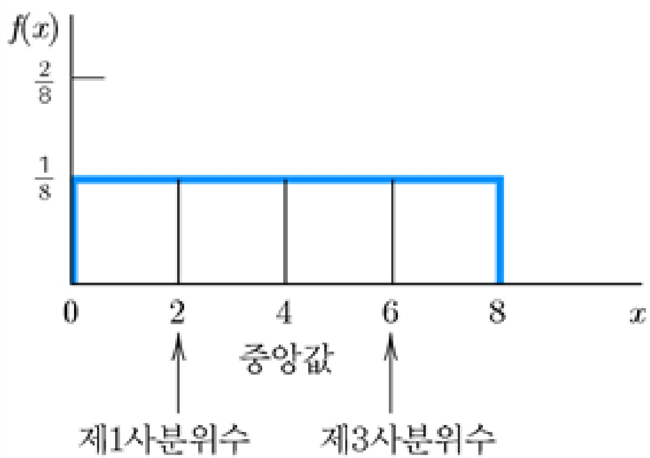
연속분포의 특징

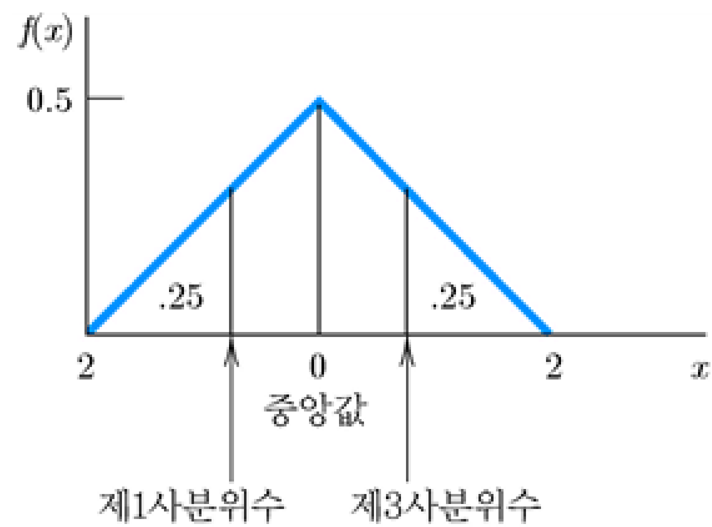
- 연속확률변수의 확률밀도곡선은 다양한 모양을 가진다.



평균, 중앙값, 사분위수, 백분위수







표준화된 변수(standardized variable)

$$Z = \frac{X - \mu}{\sigma} = \frac{X - \text{평균}}{\text{표준편차}}$$

Z는

평균이 **0**

표준편차 **1**

2. **정**규분포-일반적인 특징

정규분포(normal distribution) 는 추론과정에서 다양하고 폭넓게 사용되고 있고 통계 분석의 중추적 역할을 하고 있다.

정규분포는 종 모양의 확률밀도함수를 가지며 다음을 만족한다.

평균 = μ 표준편차 = σ

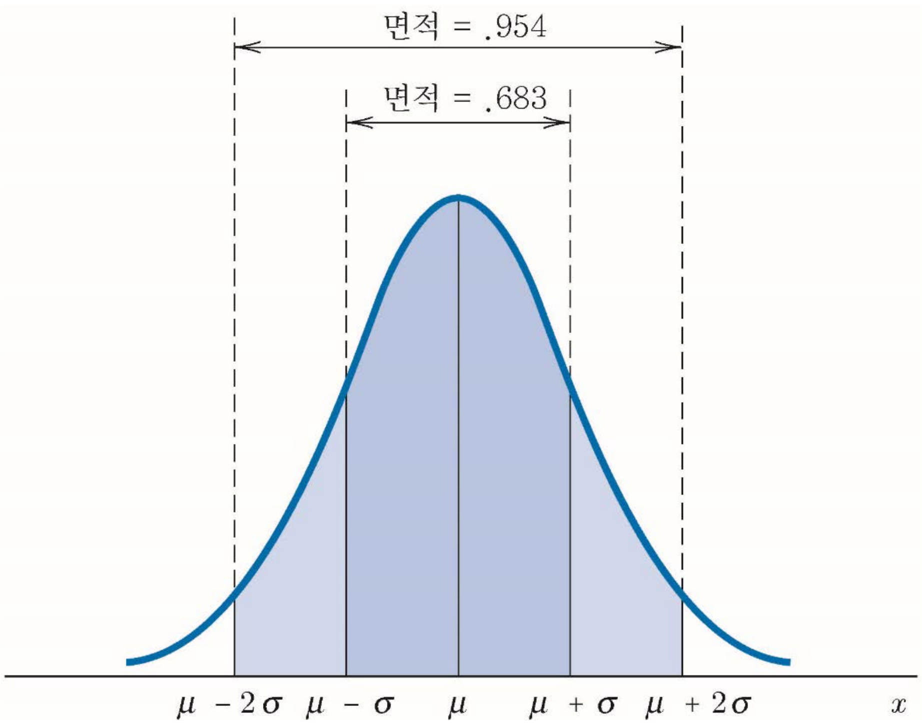
p49 좌우대칭인 종모양분포의 경험적 규칙

평균의 양 방향으로 한 배의 표준편차인 구간: $P[\mu - \sigma < X < \mu + \sigma] = 0.683$

평균의 양 방향으로 두 배의 표준편차인 구간: $P[\mu - 2\sigma < X < \mu + 2\sigma] = 0.954$

평균의 양 방향으로 세 배의 표준편차인 구간: $P[\mu - 3\sigma < X < \mu + 3\sigma] = 0.998$

2. **정**규분포-일반적인 특징



$X \sim N(\mu\,,\sigma^2)$

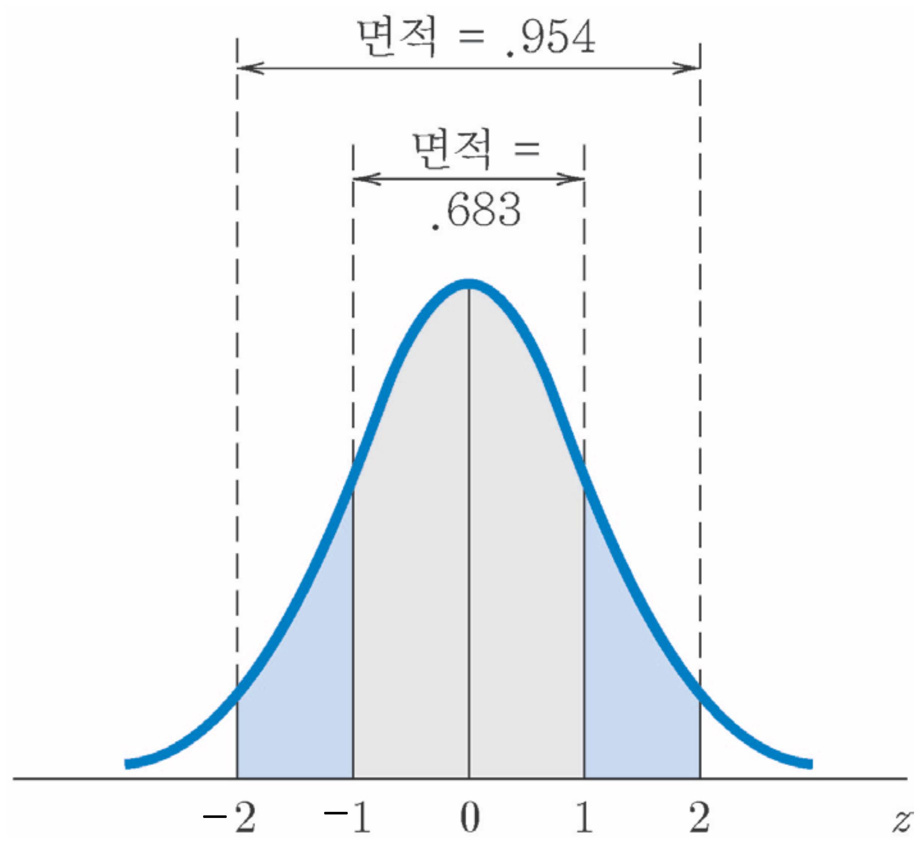
X의 확률밀도함수

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

평균이 μ 이고 표준편차 σ 인 정규분포

$N(\mu,\ \sigma^2)$

3. 표준정규분포 (standard normal distribution)



평균이 $\mu = 0$

이고 $\sigma = 1$

표준편차

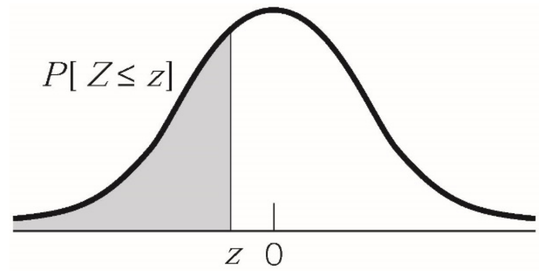
인 표준정규분포

$N(0,\ 1)$

표준정규분포표 이용법

표준정규분포표는 지정된 값 Z의 왼쪽부분의 면적을 나타낸다.

$$P[Z \le z]$$



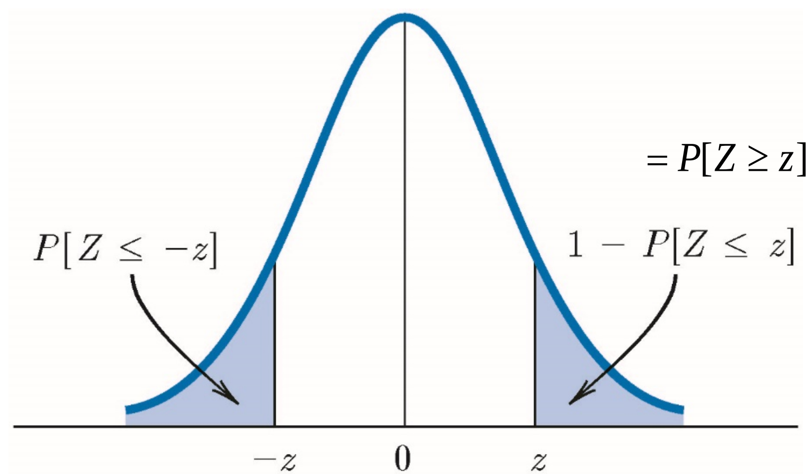
$$P[a \le Z \le b]$$

= b의 왼쪽 부분의 면적 – a의 왼쪽 부분의 면적 $= P[Z \le b] - P[Z < a]$

성질

$$P[Z \le 0] = 0.5$$

$$P[Z \le -z] = 1 - P[Z \le z] = P[Z \ge z]$$



예제 1

$P[Z \leq 1.37]$ 과 $P[Z > 1.37]$ 을 구하라.

예제 3

$P[Z < -1.9 \text{ 또는 } Z > 2.1]$ 을 구하라.

4. 정규분포에 대한 확률 계산

만일 X가 $N(\mu,\ \sigma^2)$ 의 분포를 따른다면, 표준화된 확률변수

$$Z = \frac{X - \mu}{\sigma}$$

는 표준정규분포를 따른다.

$$P[a \le X \le b] = P[\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}]$$

X를 Z로 변환하여 표준정규분포를 이용하여 확률을 계산

예제 6

확률변수 X가 정규분포 $N(60, 4^2)$를 따를 때,

$P[55 \le X \le 63]$ 을 구하라.

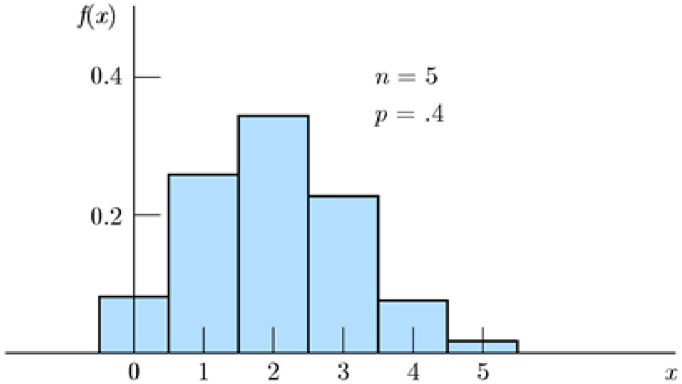
예제 7

점심 메뉴의 샐러드에 포함되어 있는 칼로리(X)는 평균이 200이고 표준편차가 5인 정규분포를 따른다고 한다.

(a) 208을 초과할 확률

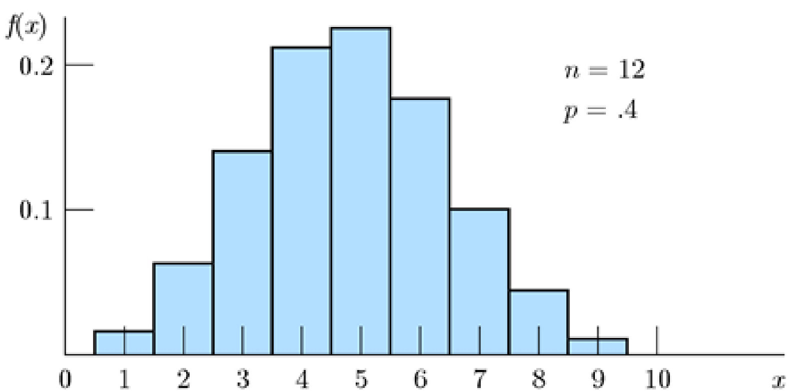
(b) 190과 200 사이일 확률

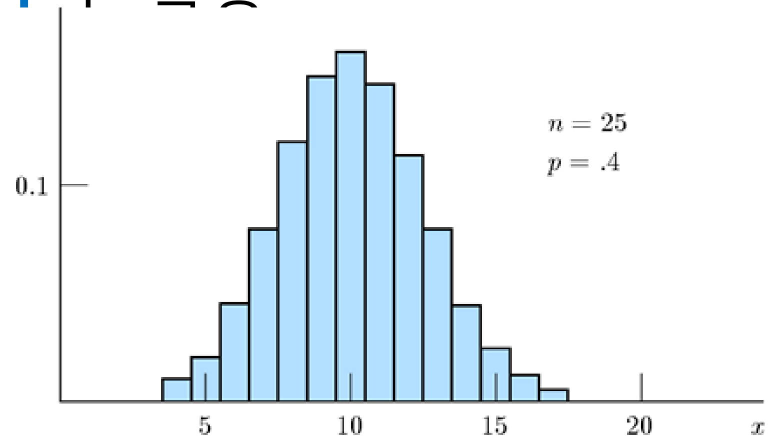
5. 이항분포의 정규근사



성공할 확률 p가 0이나 1에 너무 가깝지 않고 시행횟수가 크면

정규분포를 **이항확률의 좋은 근사분포**로 활용





이항분포에 대한 정규근사

- np 와 $n(1-p)$ 가 둘 다 클 때 (15보다 클 때)

이항분포는 평균이 np 이고, 표준편차 $\sqrt{np(1-p)}$ 인 정규분포에 잘 근사한다.

$$X \sim B(n, p), \qquad np > 15, n(1-p) > 15$$

$$\Rightarrow X \approx N(np, np(1-p))$$

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \approx N(0,\ 1)$$

예제 9

- 어린이 TV 시청에 관한 대규모 조사에 의해 생후 몇 개월 된 아기 중 약 40%가 TV를 규칙적으로 보는 것으로 파악. 이 연령대의 아기 150명의 표본, X : 규칙적으로 TV를 보는 아기의 수.

(a) X 가 52 이상 71 이하일 확률

(b) X가 67보다 더 클 확률