제5장 정규분포

5.1 연속확률변수의 확률모형

(1) 상대도수 히스토그램의 근사적 확률밀도곡선

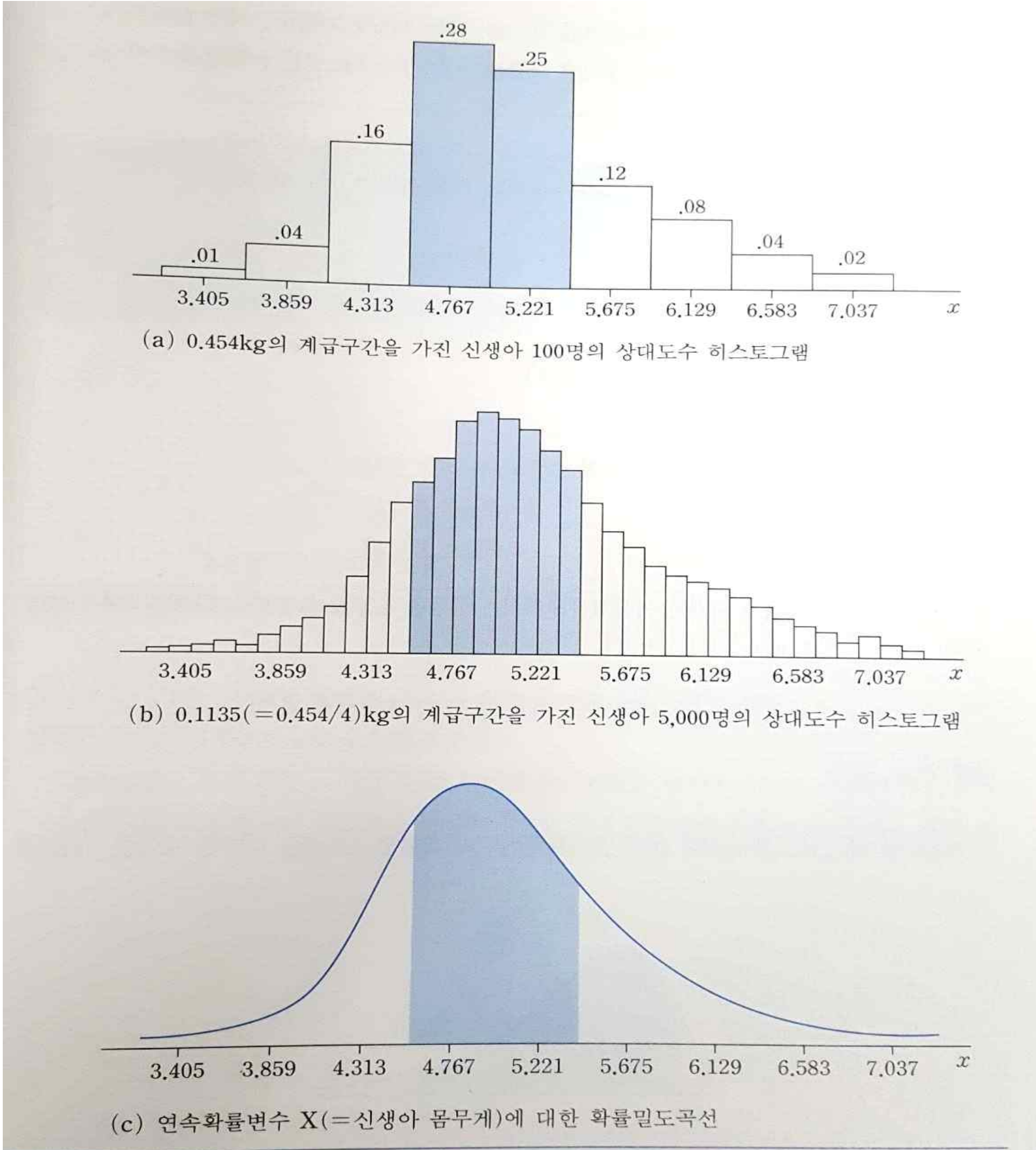


그림 5-1 | 상대도수 히스토그램의 근사적 확률밀도곡선

(2) 확률밀도함수 $f(x)$의 성질

① 확률밀도곡선 아래의 총면적은 1이다.

② $P[a \le X \le b]$

③ $f(x) \ge 0 \quad \forall x$

(3) $P[X=x]=0$ 여기서 x: 연속적 확률변수

$P[X=a]=P[X=b]=0$

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

(4) $P[a < X < b]$ = (b의 왼쪽 면적) - (a의 왼쪽 면적)

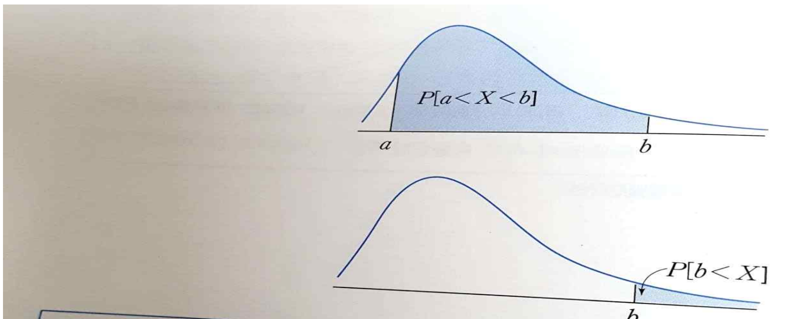


그림 5-2 | $P[b<X]=1-$(b의 왼편 면적)

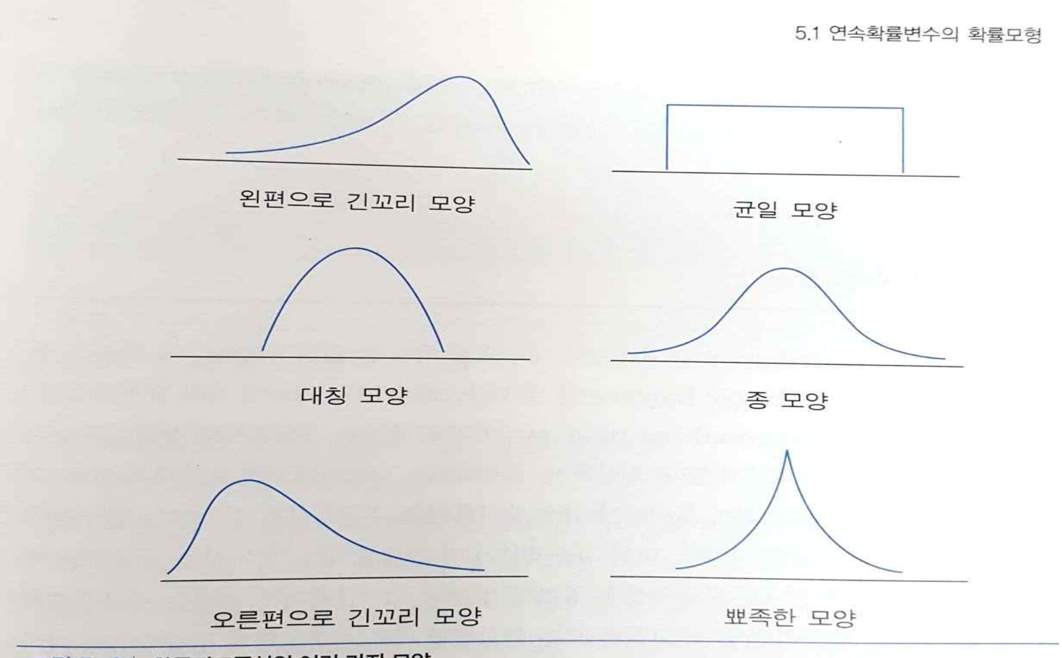


그림 5-3 | 확률밀도곡선의 여러 가지 모양

(5) 확률밀도함수의 평균과 중앙값

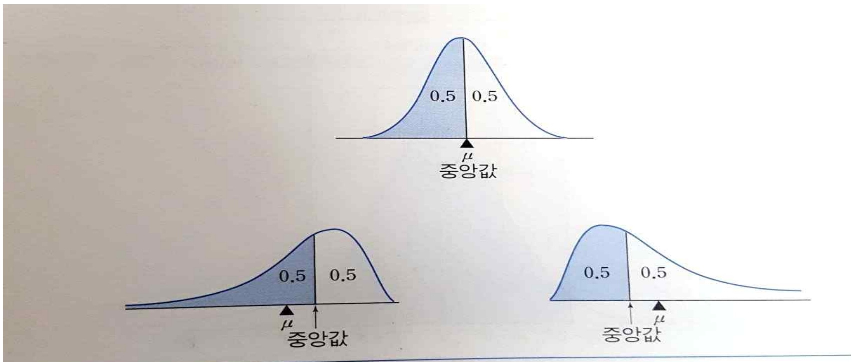


그림 5-4 | 세 가지 확률분포에서 평균과 중앙값

5.2 정규분포

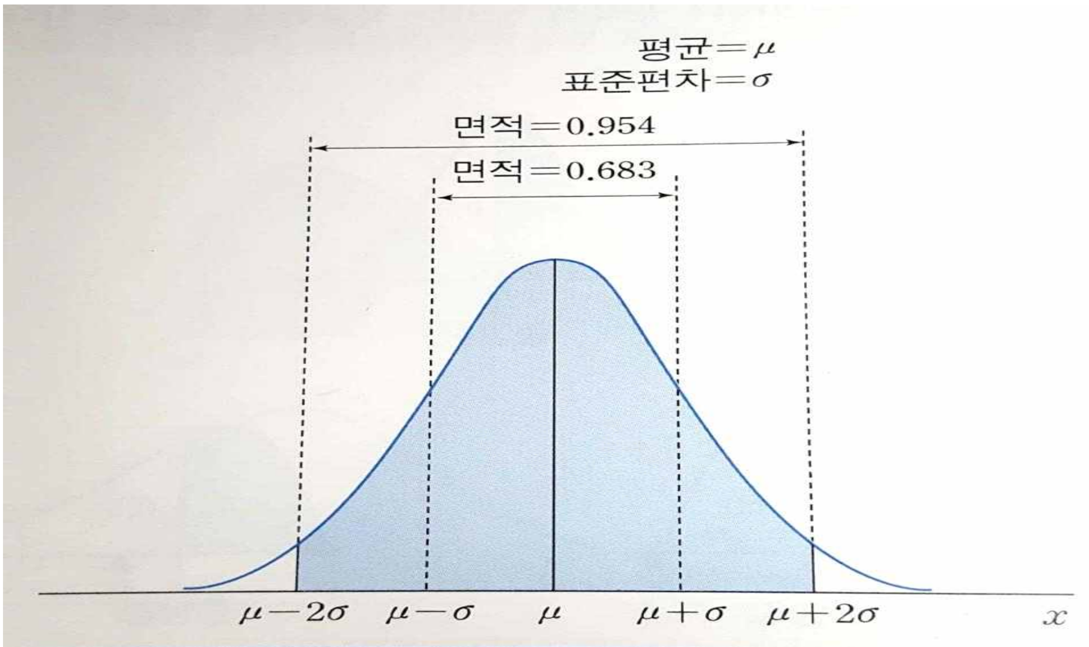
(1) 정규분포는 Pierre Laplace와 Carl Gauss에 의해 발전되었다.

(2) 정규분포의 확률밀도함수

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

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, \ \pi = 3.1416, \ e = 2.7183$$

(3) 정규분포의 그림



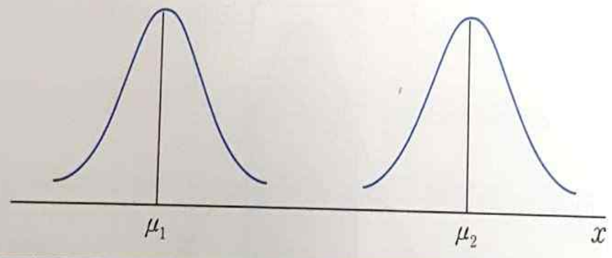


그림 5-6 | 평균($\mu_1 < \mu_2$)이 다르고 표준편차가 같은 두 정규분포

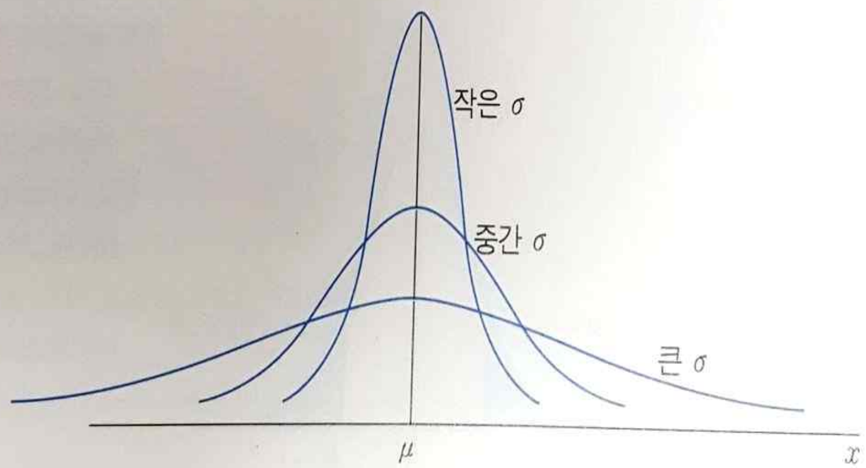


그림 5-7 | 평균이 같고 표준편차가 다른 정규분포들

5.3 표준정규분포

(1) 표준정규분포의 정의

$$X \sim N(\mu, \sigma^2) \qquad Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

(2) 표준정규분포의 그림 (p.5)

(3) 공식

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

$$P[a \le X \le b] = P\left[\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right] \quad \text{where } \ Z \sim N(0,1)$$

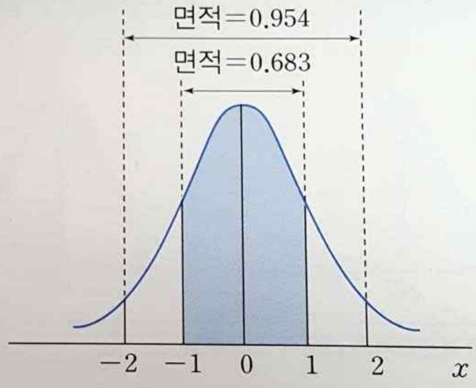


그림 5-8 | 표준정규곡선

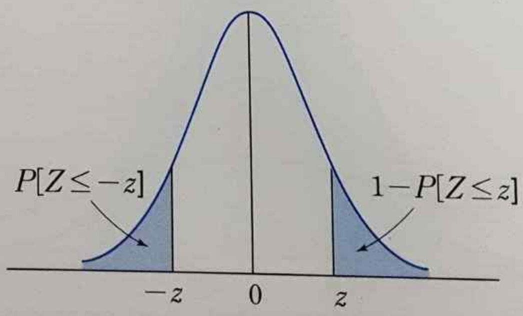


그림 5-9 | 표준정규분포

[예제 5.1]

$P[Z \le 1.37]$과 $P[Z > 1.37]$은 얼마인가?

【풀이】

$P[Z \le 1.37] = 0.9147$

$P[Z > 1.37] = 1 - P[Z \le 1.37] = 1 - 0.9147 = 0.0853$

[예제 5.4]

$P[Z > z] = 0.025$를 만족하는 z를 찾아라.

【풀이】

$P[Z < z] = 1 - 0.025 = 0.975$

표준정규분포표에서 0.975에 해당하는 주변값을 찾으면 $z = 1.96$이다.

5.4 정규분포의 확률계산

> X가 $N(\mu, \sigma^2)$를 따르면 표준화된 변수
> $$Z = \frac{X-\mu}{\sigma}$$
> 는 표준정규분포 $N(0,1)$을 따른다.

[예제 5.6]

X가 정규분포 $N(60, 4^2)$을 따를 때 $P[55 \le X \le 63]$을 구하라.

【풀이】

표준화된 변수: $Z = \dfrac{X-\mu}{\sigma} = \dfrac{X-60}{4}$

$X = 55$인 경우: $Z = \dfrac{X-\mu}{\sigma} = \dfrac{55-60}{4} = -1.25$

$X = 63$인 경우: $Z = \dfrac{X-\mu}{\sigma} = \dfrac{63-60}{4} = 0.75$

$P[55 \le X \le 63] = P[-1.25 \le Z \le 0.75]$
$= P[Z \le 0.75] - P[Z \le -1.25] = 0.7734 - 0.1056 = 0.6678$

> X가 $N(\mu, \sigma^2)$를 따르면
> $$P[a \le X \le b] = P\left[\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right]$$
> 여기서 Z는 표준정규분포이다.

[예제 5.7]

점심식사로 곁들여 먹는 상추의 칼로리가 평균 200이고 표준편차가 5인 정규분포를 한다고 하자. 다음의 확률은 얼마인가?

(1) 208 칼로리가 넘는다.

【풀이】

X: 상추의 칼로리, $Z = \dfrac{X-\mu}{\sigma} = \dfrac{X-200}{5}$

$X = 208$인 경우: $Z = \dfrac{X-\mu}{\sigma} = \dfrac{208-200}{5} = 1.6$

$P[X > 208] = P[Z > 1.6] = 1 - P[Z \le 1.6] = 1 - 0.9452 = 0.0548$

(2) 190과 200 칼로리 사이에 있다.

【풀이】

$X=190$인 경우: $Z=\dfrac{X-\mu}{\sigma}=\dfrac{190-200}{5}=-2.0$

$X=200$인 경우: $Z=\dfrac{X-\mu}{\sigma}=\dfrac{200-200}{5}=0.0$

$P[190 \leq X \leq 200]=P[-2.0 \leq Z \leq 0.0]=0.5-P[Z \leq -2.0]=0.5-0.0228=0.4772$

5.5 이항분포의 정규근사

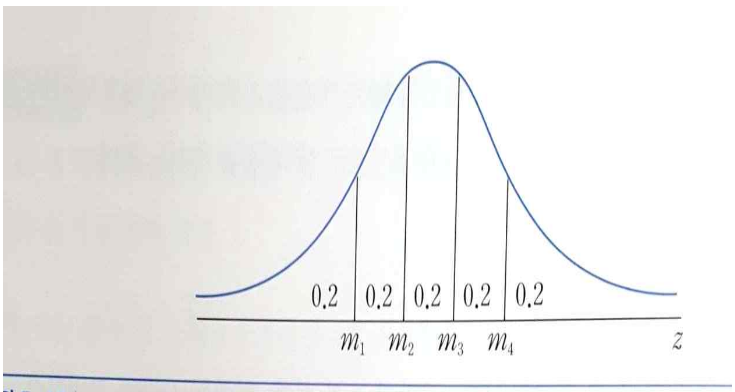
np와 $n(1-p)$가 모두 클 때, 즉 15보다 클 때, 이항분포는 평균=np와 표준편차=$\sqrt{np(1-p)}$를 갖는 정규분포에 의해 잘 근사된다. 즉,

$$Z=\frac{X-np}{\sqrt{np(1-p)}}$$

는 근사적으로 $N(0,1)$을 따른다.

5.6 정규모형가정의 점검

❑ 정규점수그림



림 5-18 | n=4일 때 $N(0,1)$ 분포와 정규점수

표 5-2 |

정규점수	순서화된 표본
$m_1=-0.84$	44
$m_2=-0.25$	68
$m_3=0.25$	75
$m_4=0.84$	82

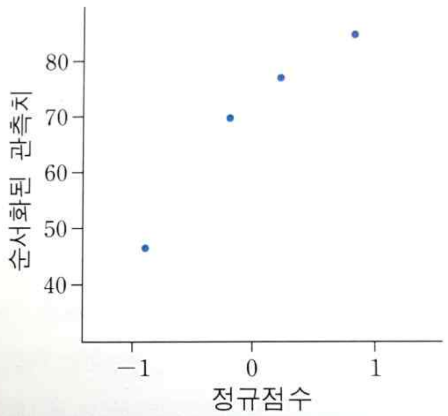
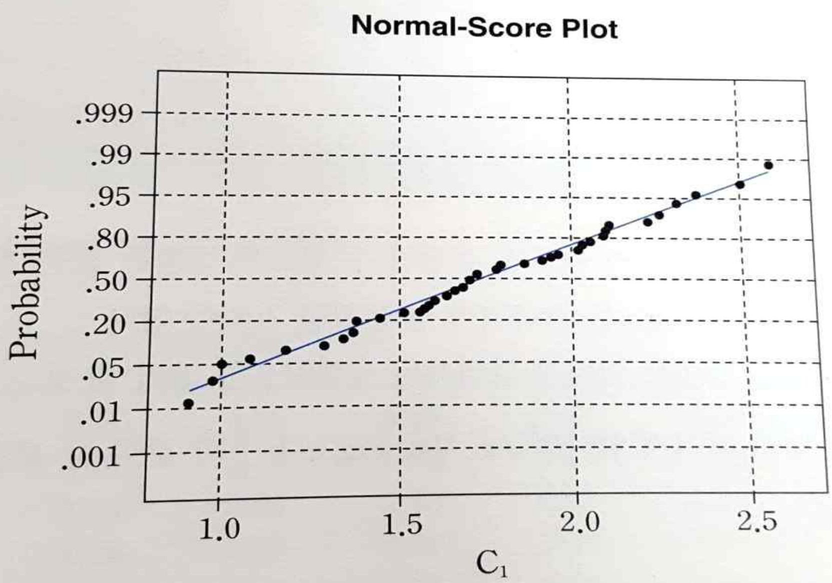


그림 5-19 | [표 5.2]의 정규점수 그림



5.7 근사적 정규성을 위한 변환

❑ 1보다 큰 값을 크게 만들기: x^2, x^3

❑ 1보다 큰 값을 작게 만들기: $\sqrt{x}$, $\sqrt[4]{x}$, $\log_e x$, $\dfrac{1}{x}$

[예제 5.12]

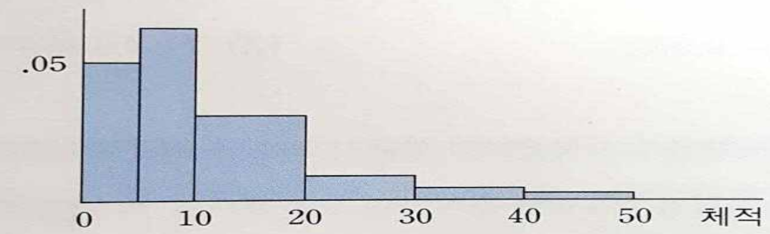
[표 5-3]은 어느 산림에서 수집된 49 지역에 대한 목재의 체적을 조사한 자료다. 자료를 정규성을 거의 만족하도록 변환하여라.

| 표 5-3 | 목재의 체적(단위 : 코드)

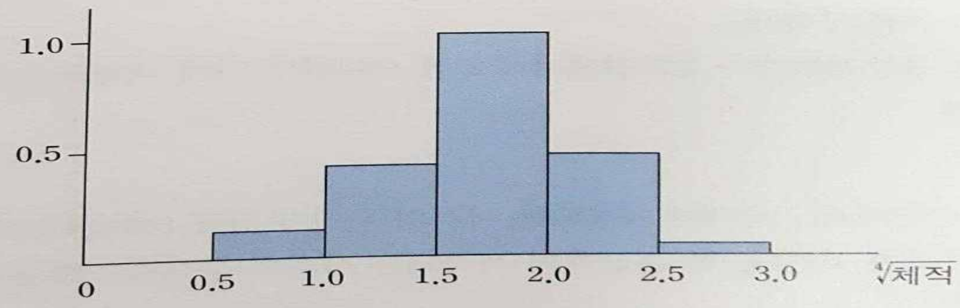
39.3	14.8	6.3	0.9	6.5
3.5	8.3	10.0	1.3	7.15
6.0	17.1	16.8	0.7	7.9
2.7	26.2	24.3	17.7	3.2
7.4	6.6	5.2	8.3	5.9
3.5	8.3	44.8	8.3	13.4
19.4	19.0	14.1	1.9	12.0
19.7	10.3	3.4	16.7	4.3
1.0	7.6	28.3	26.2	31.7
8.7	18.9	3.4	10.0	

| 표 5-4 | 변환된 자료 : $\sqrt[4]{체적}$

2.50	1.96	1.58	0.97	1.60
1.37	1.70	1.78	1.07	1.63
1.57	2.03	2.02	0.91	1.68
1.28	2.26	2.22	2.05	1.34
1.64	1.60	1.51	1.70	1.56
1.37	1.70	2.59	1.70	1.91
2.10	2.09	1.94	1.17	1.86
2.11	1.79	1.36	2.02	1.44
1.00	1.66	2.31	2.26	2.37
1.72	2.09	1.36	1.78	



(a) 목재 체적의 히스토그램



(b) $\sqrt[4]{체적}$의 히스토그램