

## 2.3 정규분포로의 근사

- 1) 통계자료의 단위변환과 그 사례
- 2) 정규분포곡선의 특징
- 3) 분위수의 의미와 활용

# 1. 통계자료의 단위변환과 그 사례

## 단위 변환

- 상수를 더하거나 곱하는 변환.
- 예: 섭씨 온도( $y$ )= $5/9$ (화씨 온도( $x$ )-32). 즉  $y=-160/9+(5/9)x$
- 표준화(standardization): 평균을 빼주고 표준편차로 나누어 주는 변환.
- 확률변수가  $X \sim N(\mu, \sigma^2)$  일 때

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

# 1. 통계자료의 단위변환과 그 사례

## 수능 표준 점수

- 사회탐구 과목당 50점 (2과목)  
과학탐구 과목당 50점 (2과목)
- 언어 100점  
수리 100점  
외국어 100점
- 사회탐구, 과학탐구 과목당 표준점수

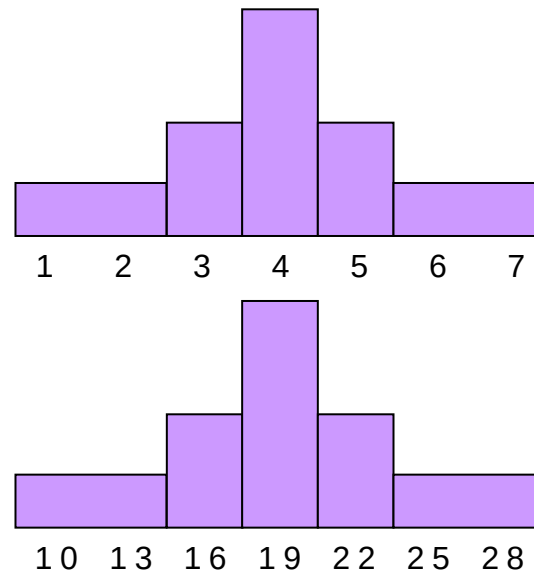
$$(\text{표준점수}) = 50 + 10z$$

- 언어, 수리, 외국어 표준점수

$$(\text{표준점수}) = 2(50 + 10z)$$

# 1. 통계자료의 단위변환과 그 사례

단위 변환



분포의 전반적 형태는  $y=7+3x$  의 단위변환 이후에도 불변임.

## 2. 정규분포곡선의 특징

### 정규분포곡선

- 하나의 이상적인 히스토그램. 하나의 수학적 모형. 개념상 모집단의 분포.
- 정규분포의 확률밀도함수(probability density function)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty, \quad e = 2.71828 \dots$$

- $\mu$  를 모평균,  $\sigma$  를 모표준편차라고 부름.
- 모집단: 모평균과 모표준편차
- 표본: 표본평균과 표본표준편차

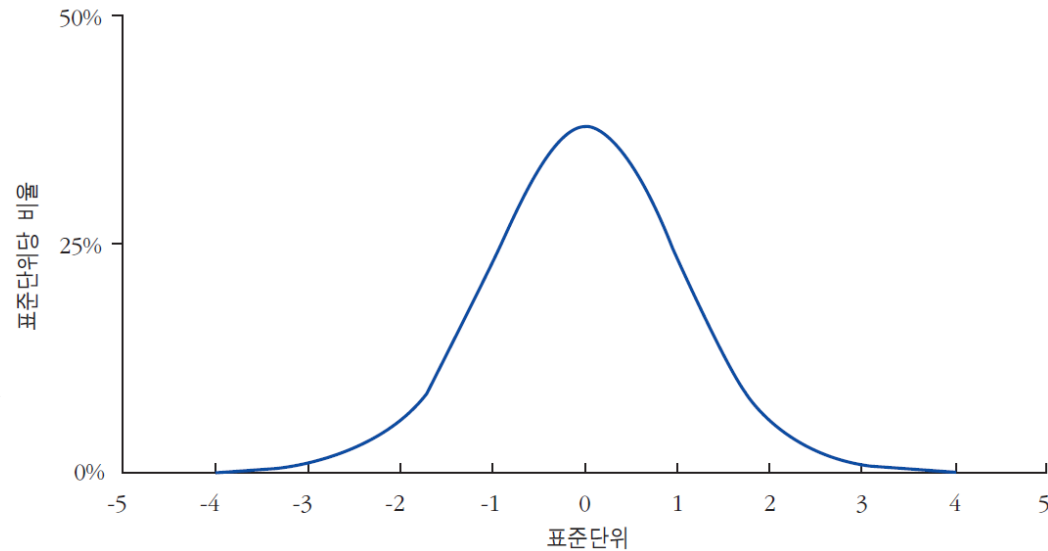
## 2. 정규분포곡선의 특징

### 표준정규분포 (standard normal distribution)

평균이 0이고 표준편차가 1인

정규분포:  $Z \sim N(0, 1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty$$



류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 76

정규분포곡선의 68-95-99.7 규칙

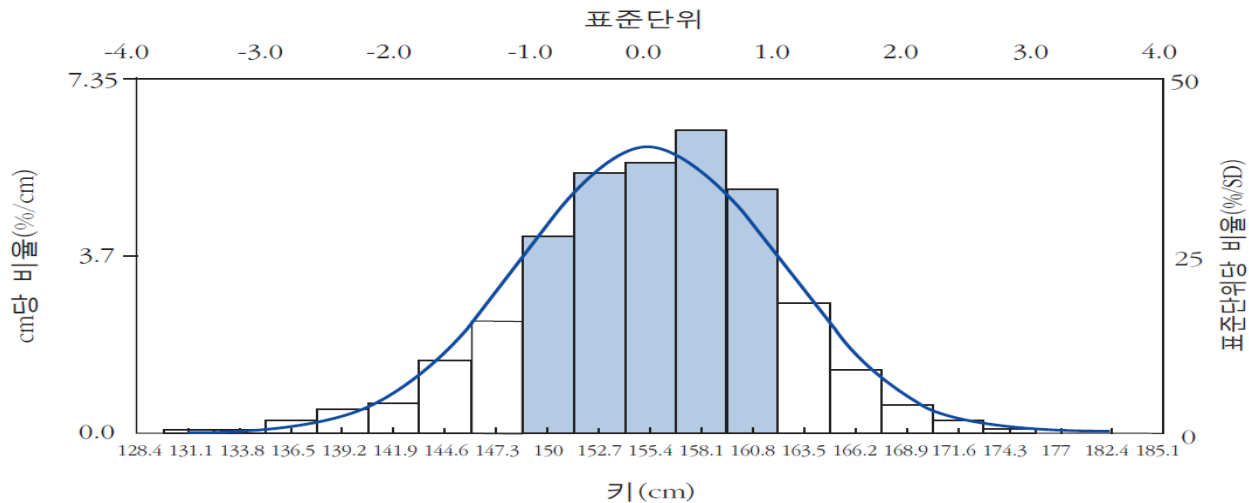
- 표준단위로 -1부터 1까지 영역의 넓이 : 약 68%
- 표준단위로 -2부터 2까지 영역의 넓이 : 약 95%
- 표준단위로 -3부터 3까지 영역의 넓이 : 약 99.7%

## 2. 정규분포곡선의 특징

### 정규분포곡선의 모양

- 평균을 중심으로 좌우 대칭(symmetric)
- 종 모양(bell-shaped)
- 봉우리가 하나(single-peaked)

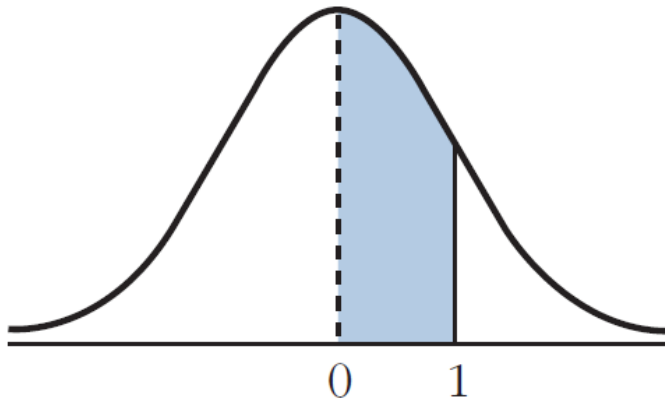
그림 4-2 여성의 키 분포를 나타내는 히스토그램과 표준정규분포곡선



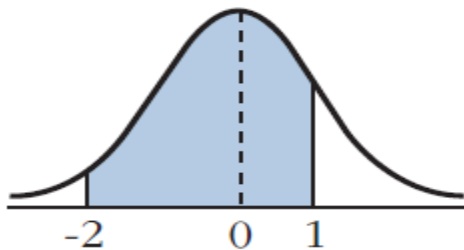
## 2. 정규분포곡선의 특징

표준정규분포곡선 아래의 영역 찾기

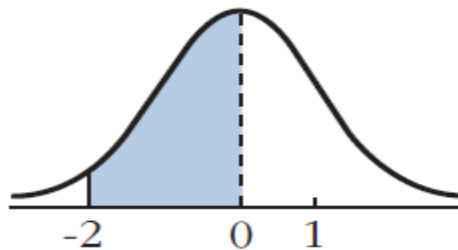
0과 1 사이 구간의 면적은 전체의 34.13%



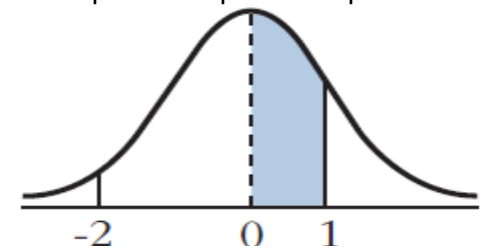
-2와 1 사이 구간의 면적은 48% + 34% = 82% 이다.



=



+



z	.00	.01	.02
0.0	.0000	.0040	.0080
0.1	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.1179	.1217	.1255
0.4	.1554	.1591	.1628
0.5	.1915	.1950	.1985
0.6	.2257	.2291	.2324
0.7	.2580	.2611	.2642
0.8	.2881	.2910	.2939
0.9	.3159	.3186	.3212
1.0	.3413	.3438	.3461
1.1	.3643	.3665	.3686



## 2. 정규분포곡선의 특징

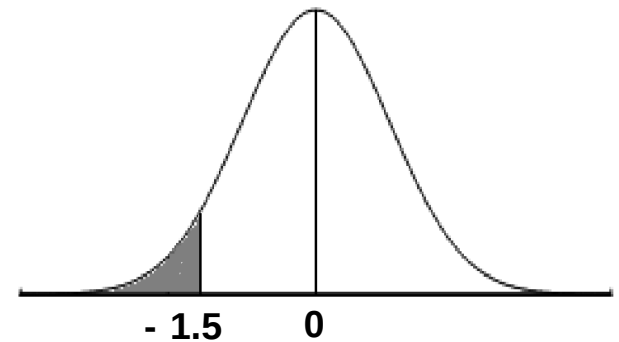
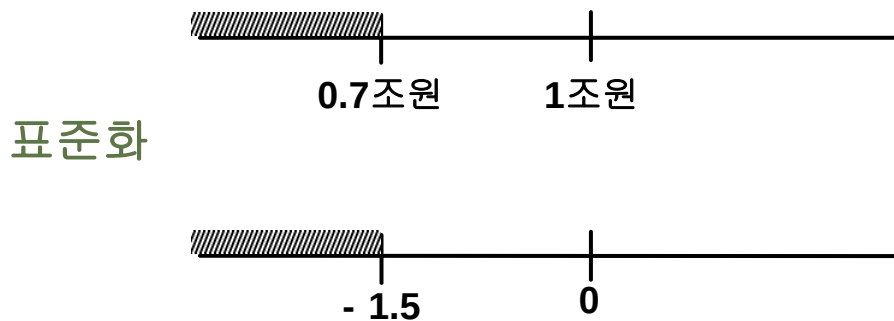
### 평균과 표준편차

- 정규분포곡선은 평균과 표준편차에 의해 그 모양이 완벽하게 묘사된다.
- 즉, 정규분포를 따르는 히스토그램은 중심과 중심 주위로 퍼진 정도 등 두 정보만으로 100% 묘사된다.

## 2. 정규분포곡선의 특징

### 정규 근사

예: 한 은행이 특정 영업일에 지급준비금 부족을 겪을 가능성은? (답: 6.68%)  
(매일 영업이 끝난 뒤 이 은행에 남아있는 잔고는 평균이 1조원이고 표준편차가 0.2조원인 정규분포에 의해 잘 근사된다고 가정. 지급준비금은 0.7조원 이상이어야 한다고 가정.)



표준정규분포곡선에  
해당 구간을 표시

# 3. 분위수의 의미와 활용

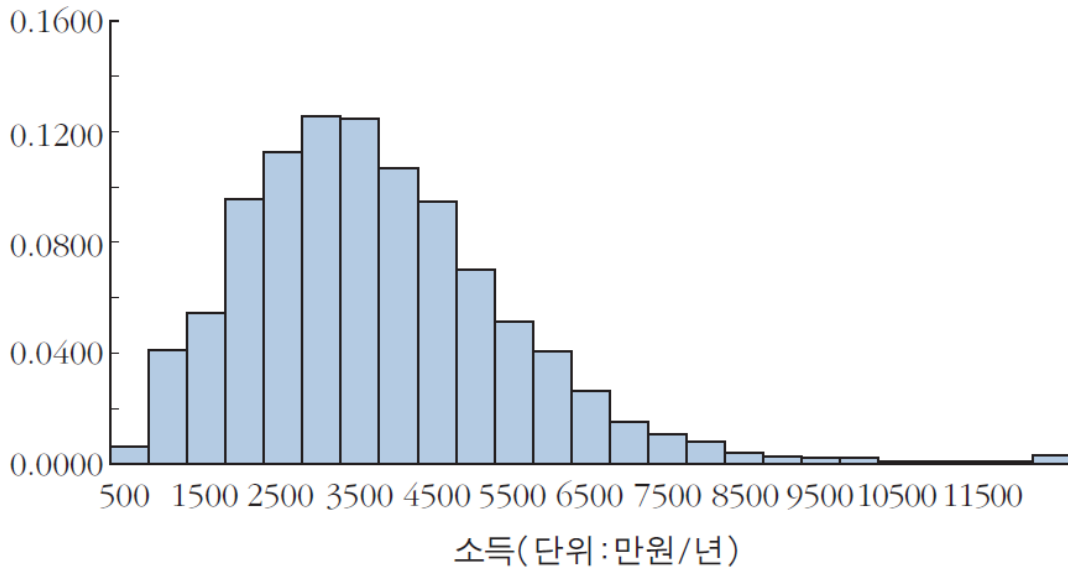
## 백분위수의 정의

- 백분위수(percentile)는 하나의 히스토그램을 100개의 균등한 영역으로 나누는 99개의 경계점 값들.
- 제  $p$  백분위수는 그 값보다 작은 값이  $p\%$ , 큰 값이  $(100-p)\%$ 가 되는 경계값.
- 많은 히스토그램은 정규분포곡선과 다름.(평균과 표준편차만으로는 부족)
- 이러한 히스토그램을 요약할 때는 백분위수 개념이 유용.

# 3. 분위수의 의미와 활용

## 연간 가구소득 분포

연간 가구소득 분포



출처: 통계청 마이크로데이터

류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 82

연간 가구소득 분포 : 2008년, 한국

백분위	백분위수(만원/년)
1	589.5
10	1484.7
25	2242.8
50	3254.1
75	4438.5
90	5687.3
99	8887.9

류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 83

# 3. 분위수의 의미와 활용

## 사분위수

- 백분위수 가운데 25번째, 50번째, 75번째 백분위수를 특별히 제1사분위수 (first quartile), 제2사분위수(second quartile), 제3사분위수(third quartile)라 부름.
- 50번째 백분위수는 제2사분위수이면서 중앙값(median)임.
- 사분위수 범위(interquartile range)

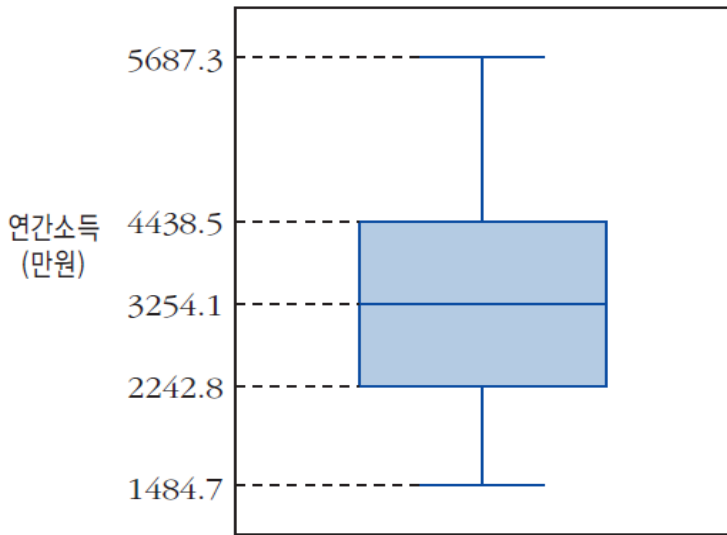
$$(\text{사분위수 범위}) = (\text{제3사분위수}) - (\text{제1사분위수})$$

- 다섯 숫자 요약(five number summary): 최소값, 제1사분위수, 제2사분위수, 제3사분위수, 최대값
  - \*(최소값, 최대값) 쌍 대신 (제5백분위수, 제95 백분위수) 쌍 또는 (제1백분위수, 제99 백분위수) 쌍을 사용하기도 함

# 3. 분위수의 의미와 활용

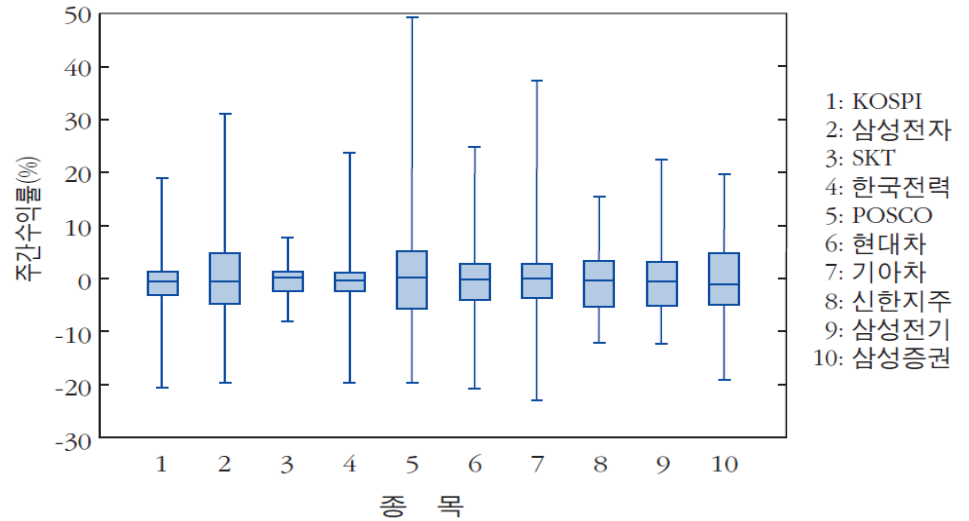
## 상자그림 (box plot)

연간 가구소득 분포



류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 84

상자그림 : KOSPI 지수와 거래소 개별종목의 2008년 주간수익률

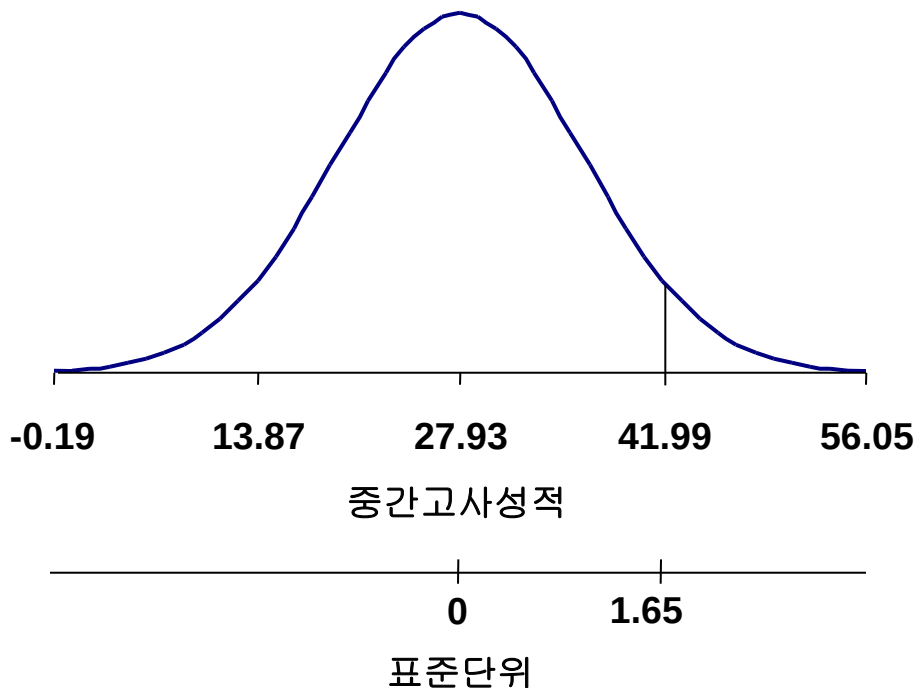


류근관. (2013). 통계학, 제 3 판. 서울: 법문사. P. 85

# 3. 분위수의 의미와 활용

## 백분위수 찾기

예: 2009년도 1학기 통계학 중간고사에서 상위 5%에 해당하는 학생의 점수를 추정하라. 단 평균점수는 27.93점이고 표준편차는 8.52점이다.



$z = 1.65$ 일 때  $[0, 1.65]$  구간의 면적이 45% 이므로, 상위 5% 학생의  $z$  값은 1.65이다. 이 학생은 평균보다  $1.65 * 8.52 = 14.06$ 점 높을 것으로 추정된다. 즉, 이 학생은  $27.93 + 14.06 = 41.99$ 점을 받았을 것으로 추정된다.