

Contents

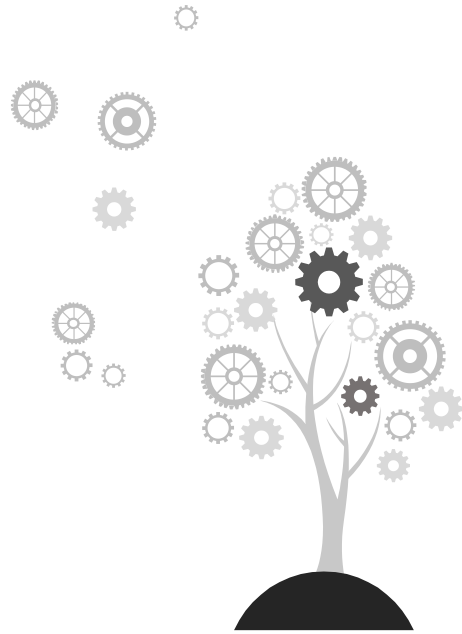
Chapter 8 정규분포

8.2 연속확률분포

8.3 정규분포의 일반적인 성질 및 확률계산

8.4 이항분포의 정규분포근사

8.5 정규분포가정의 조사



Chapter 8 정규분포

01 연속확률분포

- 연속확률변수
: 구간의 모든 값을 가질 수 있음
주어진 구간에서 확률이 어떻게 분포하는지 함수를 이용해 표현
- 확률밀도함수 (probability density function, pdf)
 - 1) 모든 x 값에 대해 $f(x) \geq 0$
 - 2) $P(a \leq X \leq b) = \int_a^b f_X(x)dx$
 - 3) $P(-\infty \leq X \leq \infty) = 1$

를 만족하는 $f(x)$ 를 X 의 확률밀도함수라고 한다.
- 예제 1, 예제 2

Chapter 8 정규분포

02 정규분포의 일반적인 성질 및 확률계산

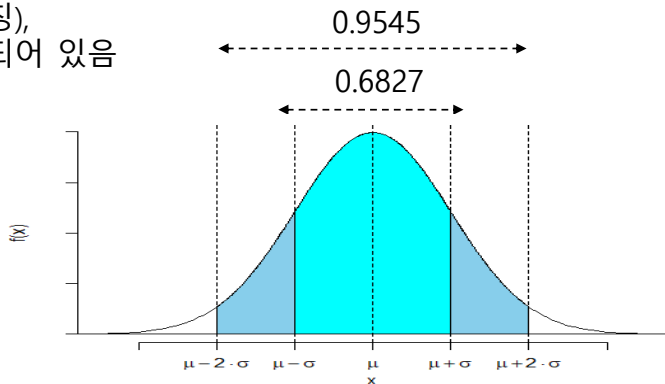
- 정규분포 (normal distribution)

- ① 연속확률변수 X 의 확률밀도함수가 $f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty < x < \infty$ 인 확률분포
- ② 평균 μ 와 분산 σ^2 에 의해 분포가 확정됨

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

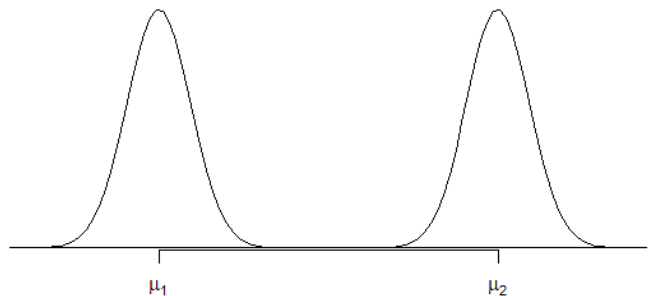
③ 특징

- 1) 평균 = 최빈값 = 중앙값 (평균을 중심으로 좌우대칭),
- 2) μ 를 중심으로 $\pm 3\sigma$ 안에 확률이 거의 (0.9973) 집중되어 있음
(평균으로부터 멀어지면 함수값이 급격히 작아짐)

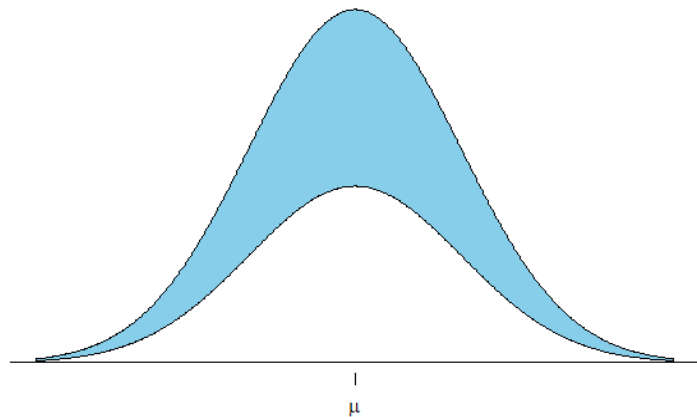


Chapter 8 정규분포

분산은 같고
평균이 다른 두 정규분포

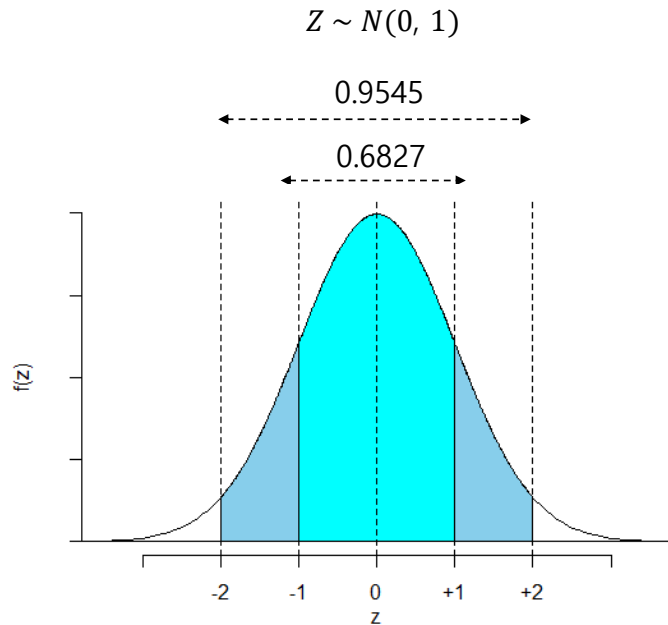


평균은 같고
분산이 다른 두 정규분포



Chapter 8 정규분포

- 표준정규분포(standard normal distribution)
: 평균이 0이고 분산이 1인 정규분포



Chapter 8

정규분포

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

- 정규분포의 확률계산

$$E\left(\frac{X-\mu}{\sigma}\right) = E\left(\frac{1}{\sigma} \cdot X - \frac{\mu}{\sigma}\right) = \frac{E(X) - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$

$$P(Z \leq -z) = P(Z \geq z) = P(Z > z) = 1 - P(Z \leq z)$$

- Z가 구간 $[a, b]$ 에 있을 확률

$$P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$$

- 예제 3, 예제 4, 예제 5

$$\cdot \text{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} \cdot \text{Var}(X) = 1$$

- 표준정규확률변수

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

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

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

↑ "표준화"

- 예제 6, 예제 7

Chapter 8

정규분포

03 이항분포의 정규분포 근사

- $X \sim \text{Bin}(n, p)$ 이고 np 나 $n(1-p)$ 가 모두 클 경우에 (보통 10이상) X 는 근사적으로 $N(np, np(1-p))$ 를 따른다.

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

$$X \sim N(np, np(1-p))$$

이 때 ('근사적일')을 붙여놓지
이항분포 핵심. 정규분포

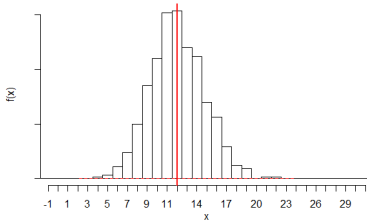
연속성 수정 (continuity correction) : $P[a \leq X \leq b] = P(a - \frac{1}{2} \leq X \leq b + \frac{1}{2})$

- 예제 8, 예제 9 $P(X=x) \div P(x - \frac{1}{2} \leq X \leq x + \frac{1}{2}) \therefore X \sim N(np, np(1-p))$

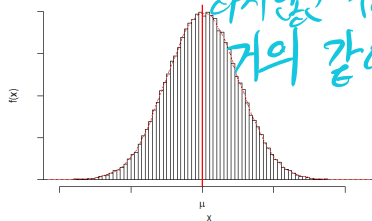
연속성 수정을 해야
정확한 정규분포에
가까운 확률분포를
얻을 수 있다.

$n = 5, p = 0.4$

$np = 2$



$n = 30, p = 0.4$



$n = \infty, p = 0.4$

np 와 $n(1-p)$ 가
10이 되면
연속성 수정
사용을 의의할 것

$np \geq 10$ 보통 $p \approx 0.5$
 $n(1-p) \geq 10$

<"A" $np \geq 10$ ">

N 이 충분히 크다!

N 이 엄청 크면,
연속성 수정을 하지
확률과
하차일 것 확률은
거의 같다.

$$n(1-p) = ?$$

$$\begin{cases} np = 12 \\ n(1-p) = 18 \end{cases}$$

$$\therefore X \sim N(12, (2.6833)^2)$$

04 정규분포가정의 조사

- 정규분포 가정의 조사

: 정규분포가정이 맞는지를 확인하는 작업, "정규확률 그림"을 이용

- 정규확률그림(정규점수그림) (= Q-Q plot)

: 표본이 이상적인 정규분포와 얼마나 유사한지를 보여주는 그림



정규확률그림 그리는 순서

"정규점수"

① 자료를 작은 것부터 크기 순으로 나열한다

② 각 자료에 해당하는 점수를 계산한다

③ i 번째 순서의 자료와 i 번째 순서의 정규점수를 하나의 쌍으로 2차원 공간상에 나타낸다.

- 정규확률그림을 이용한 정규성 판정

: 정규확률그림이 직선식을 나타내면 정규분포의 가정이 타당하고 "곡선형태" 등 직선식을 벗어나면 정규분포의 가정이 의문시된다고 할 수 있다

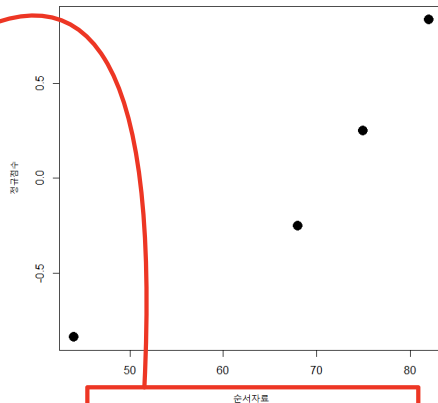
Chapter 8

정규분포

표본내 원소 갯수만큼
등화물로 쪼갬다.
(여러 개
→ 쪼개는 건 4개)

이들을 매칭하여,
두들 생성.

정규점수 (m_i)	* 크기순으로 나열된 관측값 x	이상적인 x 값
$m_1 = -0.84$	$x_{(1)} = 44$	$\mu + \sigma m_1$
$m_2 = -0.25$	$x_{(2)} = 68$	$\mu + \sigma m_2$
$m_3 = 0.25$	$x_{(3)} = 75$	$\mu + \sigma m_3$
$m_4 = 0.84$	$x_{(4)} = 82$	$\mu + \sigma m_4$



$$P(z \leq -0.84) = 0.2 \quad P(z \leq -0.25) = 0.4$$

- 자료의 변환

: 자료가 정규분포를 벗어 났을 때, 자료를 변환하여 정규분포를 따르도록 시도

Ex) $x^2, x^3, x^{\frac{1}{4}}, \sqrt{x}, \log x, \frac{1}{x}, \dots$

↑ 로그정규분포.

" $y = z \cdot m + \ln$ "
↑ 계수가 'z'이고,
절편이 'ln'인 것