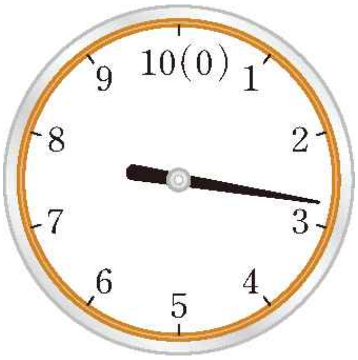
Ⅲ_2. 연속확률변수의 확률분포

[12확통03-01] 확률변수와 확률분포의 뜻을 안다.
[12확통03-04] 정규분포의 뜻을 알고, 그 성질을 이해한다.

1 연속확률변수

확률변수 X가 어떤 구간에 속하는 모든 실수의 값을 가질 때, X를 '연속확률변수'라고 한다.

☑ 길이, 무게, 온도, 시간 등의 값을 확률변수 X라 하면 X는 어떤 범위에 속하는 모든 실수의 값을 갖는다.

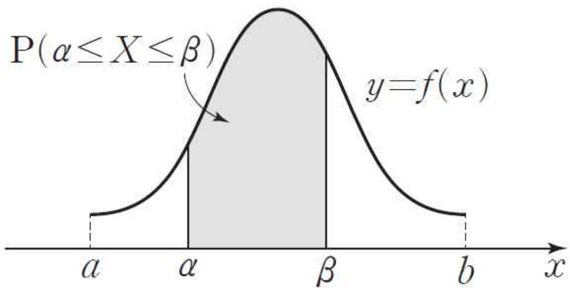


2 확률밀도함수 ❶

일반적으로 $a \le X \le b$의 모든 실수의 값을 가지는 연속확률변수 X에 대하여 $a \le x \le b$에서 정의된 함수 $f(x)$가 다음 세 가지를 모두 만족시킬 때, 함수 $f(x)$를 '연속확률변수 X의 확률밀도함수'라고 한다.

(1) $f(x) \ge 0$ (단, $a \le x \le b$)

(2) 함수 $y=f(x)$의 그래프와 x축 및 두 직선 $x=a$, $x=b$로 둘러싸인 부분의 넓이는 1이다.

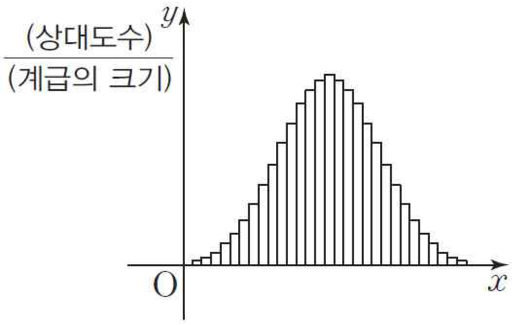


(3) $\mathrm{P}(\alpha \le X \le \beta)$는 함수 $y=f(x)$의 그래프와 x축 및 두 직선 $x=\alpha$, $x=\beta$로 둘러싸인 부분의 넓이와 같다. (단, $a \le \alpha \le x \le \beta \le b$)

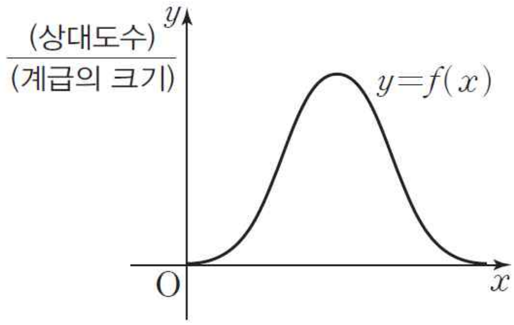
2 확률밀도함수 ❷

☑ 연속확률변수 X의 $\dfrac{(\text{상대도수})}{(\text{계급의 크기})}$를 히스토그램으로 나타내면 히스토그램의 각 구간에 세워진 직사각형의 넓이는 각 구간의 상대도수를 나타내고, 상대도수의 합이 1이므로 모든 직사각형들의 넓이의 합은 항상 1이다.
이때 조사 대상의 수를 한없이 늘리고, 계급의 크기를 한없이 0에 가깝게 하여 히스토그램을 그리면 [그림 1]과 같이 어떤 곡선 모양에 가까워지고, 이 과정을 계속하면 [그림 2]와 같이 매끄러운 곡선이 된다.

2 확률밀도함수 ❸



[그림 1]



[그림 2]

☑ 연속확률변수 X가 하나의 값을 가질 확률은 0이다. 즉,
$\mathrm{P}(X=\alpha)=\mathrm{P}(X=\beta)=0$이므로

$$\begin{aligned}\mathrm{P}(\alpha < X < \beta) &= \mathrm{P}(\alpha \le X < \beta)\\ &= \mathrm{P}(\alpha < X \le \beta)\\ &= \mathrm{P}(\alpha \le X \le \beta)\end{aligned}$$

3 정규분포

연속확률변수 X가 모든 실수의 값을 갖고, 그 확률밀도함수 $f(x)$가 두 상수 m, σ $(\sigma > 0)$에 대하여

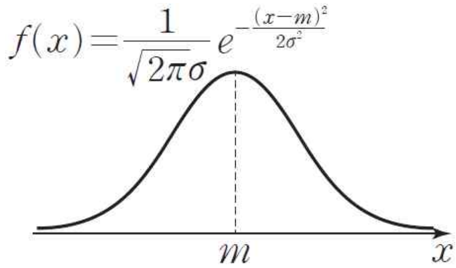
$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (e\text{는 } 2.718281\cdots \text{ 인 무리수})$$

일 때, X의 확률분포를 '정규분포'라고 한다.
이때 확률변수 X의 평균과 표준편차는 각각 m, σ임이 알려져 있다. 또한 평균이 m, 표준편차가 σ인 정규분포를 $\mathrm{N}(m, \sigma^2)$과 같이 나타내고, 확률변수 X는 '정규분포 $\mathrm{N}(m, \sigma^2)$을 따른다'고 한다.

☑ N은 정규분포를 뜻하는 Normal distribution의 첫 글자.

④ 정규분포를 따르는 확률변수의 확률밀도함수의 그래프

정규분포 $\mathrm{N}\left(m,\ \sigma^2\right)$을 따르는 연속확률변수 X의 확률밀도함수



$$f(x)=\frac{1}{\sqrt{2\pi}\ \sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$$ 의 그래프는

오른쪽 그림과 같은 모양이고, 다음과 같은 성질을 가지고 있음이 알려져 있다.

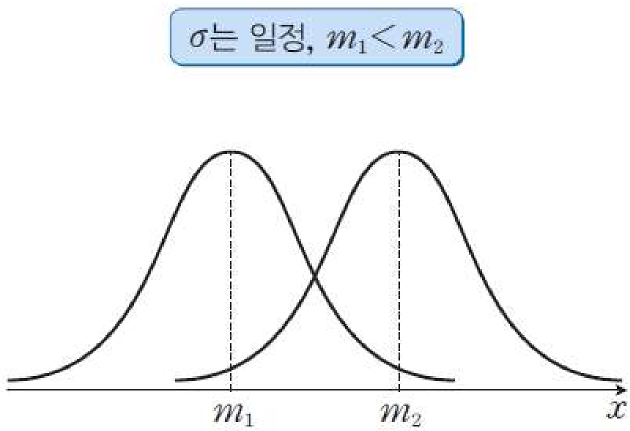
(1) 직선 $x=m$에 대하여 대칭인 종 모양의 곡선이다.

(2) x축을 점근선으로 하며, $x=m$일 때 최댓값 $\frac{1}{\sqrt{2\pi}\ \sigma}$을 갖는다.

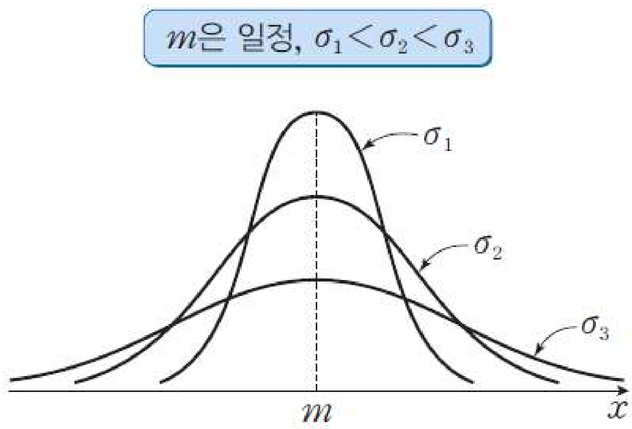
(3) 곡선과 x축 사이의 넓이는 1이다.

(4) 표준편차 σ의 값이 일정할 때, [그림 1]와 같이 m의 값이 변하면 대칭축의 위치는 바뀌지만 곡선의 모양은 같다.

(5) 평균 m의 값이 일정할 때, [그림 2]과 같이 σ의 값이 작아지면 곡선의 중앙 부분이 높아지면서 좁아지지만 대칭축의 위치는 같다.

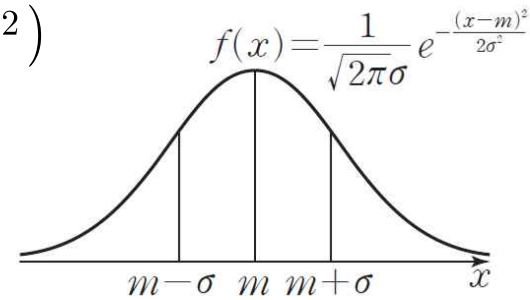


[그림 1]



[그림 2]

☑ 확률변수 X가 정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따를 때, 다음을 만족시킨다.



① $\mathrm{P}(X \le m) = \mathrm{P}(X \ge m)$

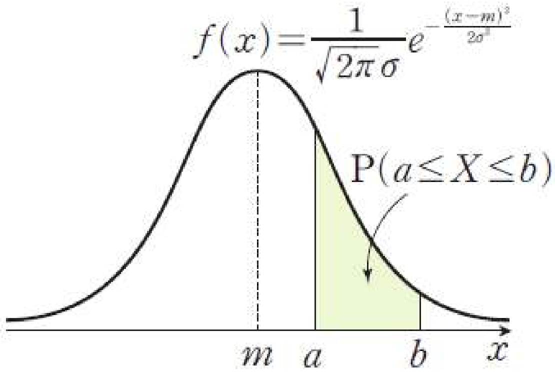
$= 0.5$

② $\mathrm{P}(m-\sigma \le X \le m) = \mathrm{P}(m \le X \le m+\sigma)$

③ $\mathrm{P}(m-k\sigma \le X \le m) = \mathrm{P}(m \le X \le m+k\sigma)$

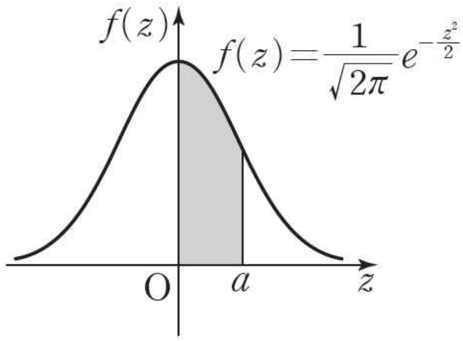
(단, k는 양의 상수)

☑ $\mathrm{P}(a \le X \le b) = \int_a^b f(x)\,dx$



5 표준정규분포 ❶

(1) 정규분포 중에서 평균이 0, 표준편차가 1인 정규분포 $\mathrm{N}(0,\ 1)$을 '표준정규분포'라고 한다.



(2) 확률변수 Z가 표준정규분포 $\mathrm{N}(0,\ 1)$을 따를 때, Z의 확률밀도함수 $f(z)$는

$$f(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \quad (e\text{는 } 2.718281\cdots \text{ 인 무리수})$$

이다. 이때 임의의 양수 a에 대하여 $\mathrm{P}(0 \le Z \le a)$는 오른쪽 그림에서 색칠된 부분의 넓이와 같다.

5 표준정규분포 ❷

☑ 확률변수 Z가 표준정규분포 $N(0, 1)$을 따를 때, $P(0 \le Z \le a)$의 값은 표준정규분포표를 이용하여 구할 수 있다.

z	0.00	0.01	$\cdots$	0.06	$\cdots$
0.0	.0000	.0040	$\cdots$	.0239	$\cdots$
0.1	.0398	.0438	$\cdots$	.0636	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1.9	.4713	.4719	$\cdots$	.4750	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

예를 들어 $P(0 \le Z \le 1.96)$의 값은 표준정규분포표의 왼쪽에 있는 수 중에서 1.9를 찾고, 표의 위쪽에 있는 수 중에서 0.06을 찾아 1.9의 가로줄과 0.06의 세로줄이 만나는 곳의 수를 찾으면 된다. $\therefore\ P(0 \le Z \le 1.96) = 0.4750$

6 정규분포와 표준정규분포의 관계 ❶

확률변수 X가 정규분포 $N(m, \sigma^2)(\sigma > 0)$을 따를 때, 확률변수

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

은 표준정규분포 $N(0, 1)$을 따른다는 사실이 알려져 있다.

이때 $P(a \le X \le b)$는 $Z = \dfrac{X-m}{\sigma}$을 이용하여 다음과 같이 표준정규분포 $N(0, 1)$을 따르는 확률변수 Z로 바꾸어 구한다.

$$P(a \le X \le b) = P\left(\frac{a-m}{\sigma} \le \frac{X-m}{\sigma} \le \frac{b-m}{\sigma}\right)$$
$$= P\left(\frac{a-m}{\sigma} \le Z \le \frac{b-m}{\sigma}\right)$$

$\boxed{6}$ 정규분포와 표준정규분포의 관계 ❷

$\boxed{\text{예}}$ 확률변수 X가 정규분포 $\mathrm{N}(10,\ 2^2)$을 따를 때,

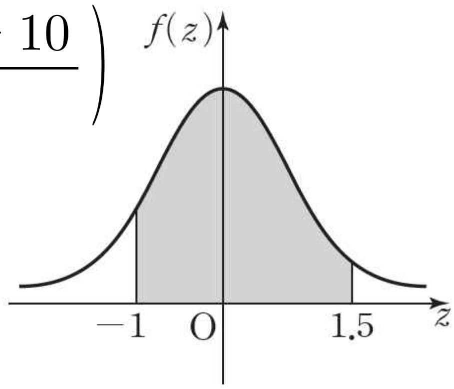
$\mathrm{P}(8 \le X \le 13)$의 값을 구해 보자.

$Z = \dfrac{X-10}{2}$이라 하면 확률변수 Z는 표준정규분포

$\mathrm{N}(0,\ 1)$을 따르고, 표준정규분포표에서

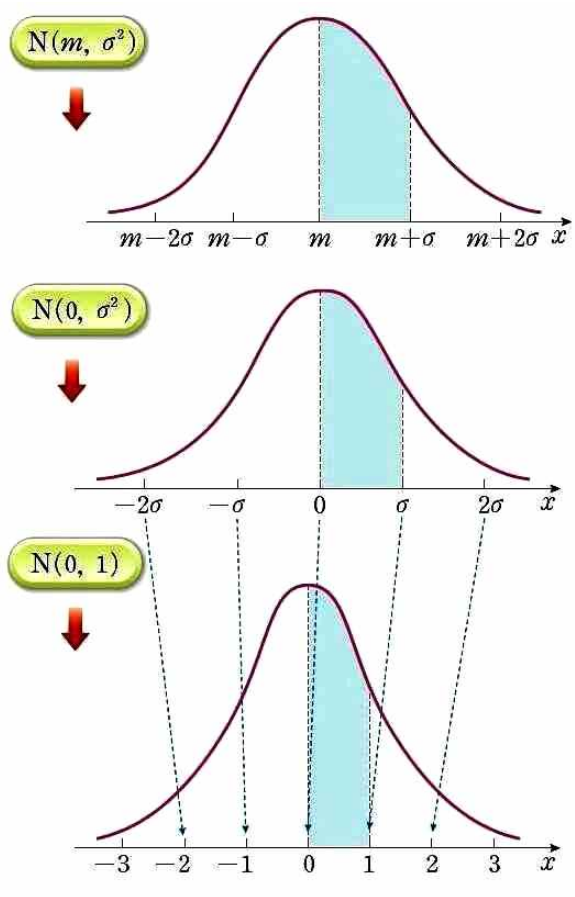
$\mathrm{P}(0 \le Z \le 1) = 0.3413$, $\mathrm{P}(0 \le Z \le 1.5) = 0.4332$이므로

$$\mathrm{P}(8 \le X \le 13) = \mathrm{P}\left(8-10 \le Z \le \frac{13-10}{2}\right)$$



$$= \mathrm{P}(-1 \le Z \le 1.5)$$
$$= \mathrm{P}(0 \le Z \le 1) + \mathrm{P}(0 \le Z \le 1.5)$$
$$= 0.3413 + 0.4332 = 0.7745$$

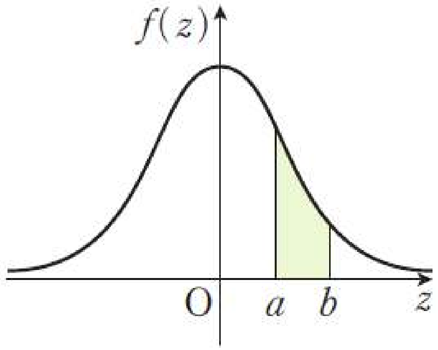
✡ $\mathrm{P}(Z \le 0) = \mathrm{P}(Z \ge 0) = 0.5$

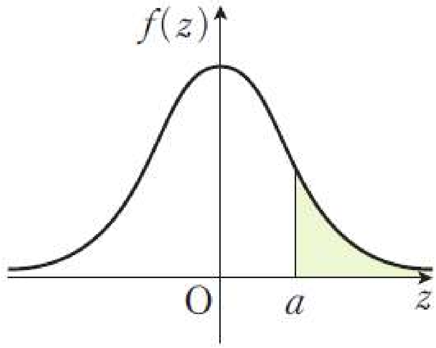


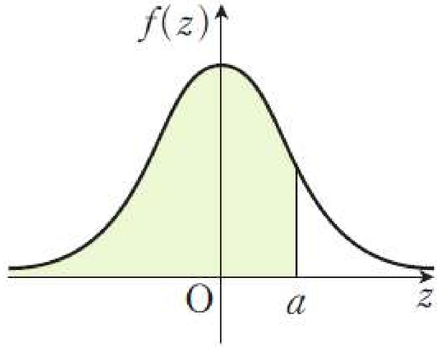
✡ 표준정규분포의 값

(1) $\mathrm{P}(-a \le Z \le 0) = \mathrm{P}(0 \le Z \le a)$

(2) $\mathrm{P}(a \le Z \le b) = \mathrm{P}(0 \le Z \le b) - \mathrm{P}(0 \le Z \le a)$



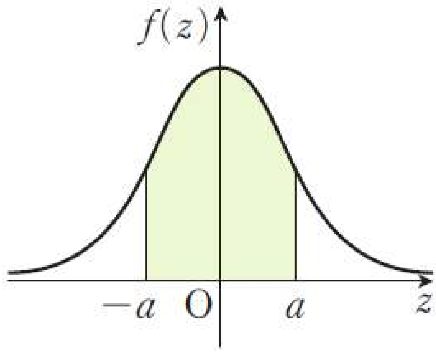




(3) $\mathrm{P}(Z \ge a) = 0.5 - \mathrm{P}(0 \le Z \le a)$

(4) $\mathrm{P}(Z \le a) = 0.5 + \mathrm{P}(0 \le Z \le a)$

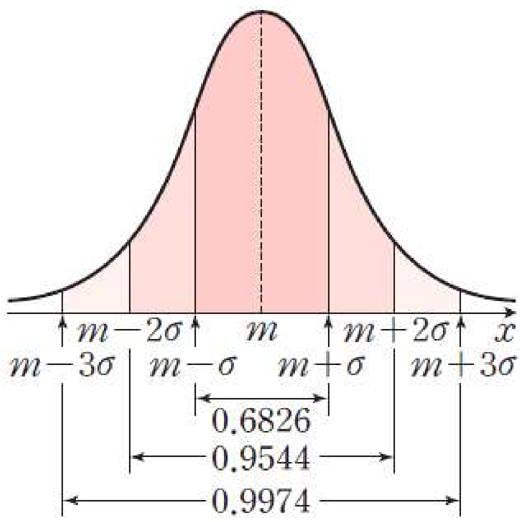
(5) $\mathrm{P}(|Z| \le a) = 2 \times \mathrm{P}(0 \le Z \le a)$



✡ 중요 확률

정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따르는

확률변수 X에 대하여



(1) $\mathrm{P}(|X-m| \le \sigma)$

$= \mathrm{P}(-1 \le Z \le 1)$

$= 2 \times 0.3413 = 0.6826$

(2) $\mathrm{P}(|X-m| \le 2\sigma)$

$= \mathrm{P}(-2 \le Z \le 2) = 2 \times 0.4772 = 0.9544$

(3) $\mathrm{P}(|X-m| \le 3\sigma) = \mathrm{P}(-3 \le Z \le 3)$

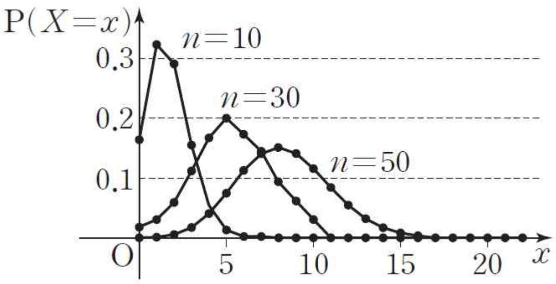
$= 2 \times 0.4987 = 0.9974$

∴ X와 평균의 차가 σ, 2σ, 3σ이내에 있을 확률이 각각 0.6826, 0.9544, 0.9974이다.

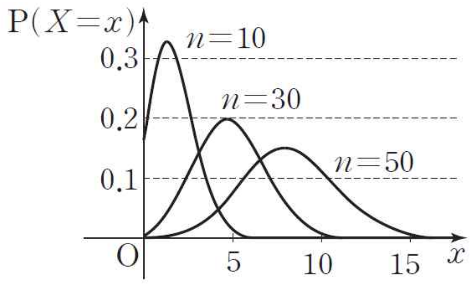
$\boxed{7}$ 이항분포와 정규분포의 관계 ❶

(1) 이항분포와 정규분포의 관계를 나타내는 그래프

한 개의 주사위를 n번 던질 때, 1의 눈이 나오는 횟수를 X라 하면 확률변수 X는 이항분포 $\mathrm{B}\left(n,\ \frac{1}{6}\right)$을 따른다. [그림 1]은 주사위를 던지는 횟수가 $n=10$, $n=30$, $n=50$일 때의 이항분포를 그래프로 나타낸 것이고, 점들을 부드럽게 연결하면 [그림 2]를 얻을 수 있다.



[그림 1]



[그림 2]

$\boxed{7}$ 이항분포와 정규분포의 관계 ❷

일반적으로 이항분포 $\mathrm{B}(n,\ p)$의 그래프는 n의 값이 커지면 정규분포의 확률밀도함수의 그래프와 가까워짐이 알려져 있다.

(2) 이항분포와 정규분포의 관계

확률변수 X가 이항분포 $\mathrm{B}(n,\ p)$를 따를 때, n이 충분히 크면 X는 근사적으로 정규분포 $\mathrm{N}(np,\ npq)$를 따른다.

(단, $q=1-p$) 이때 확률변수 $Z=\dfrac{X-np}{\sqrt{npq}}$는

표준정규분포 $\mathrm{N}(0,\ 1)$을 따른다.

☑ 일반적으로 $np \geq 5$, $nq \geq 5$이면 n이 충분히 큰 것으로 생각한다.

7 이항분포와 정규분포의 관계 ❸

예 확률변수 X가 이항분포 $\mathrm{B}\left(450,\ \frac{1}{3}\right)$을 따를 때, $\mathrm{P}(160 \le X \le 170)$의 값을 구해 보자.

$$\mathrm{E}(X) = 450 \times \frac{1}{3} = 150$$

$$\mathrm{V}(X) = 450 \times \frac{1}{3} \times \frac{2}{3} = 100$$

이때 450은 충분히 큰 수이므로 확률변수 X는 근사적으로 정규분포 $\mathrm{N}\left(150,\ 10^2\right)$을 따르고, $Z = \frac{X-150}{10}$이라 하면 확률변수 Z는 표준정규분포 $\mathrm{N}(0,\ 1)$을 따른다.

7 이항분포와 정규분포의 관계 ❹

따라서

$$\begin{aligned}\mathrm{P}(160 \le X \le 170) &= \mathrm{P}\left(\frac{160-150}{10} \le Z \le \frac{170-150}{10}\right)\\ &= \mathrm{P}(1 \le Z \le 2)\\ &= \mathrm{P}(0 \le Z \le 2) - \mathrm{P}(0 \le Z \le 1)\\ &= 0.4772 - 0.3413 = 0.1359\end{aligned}$$