



확률통계 및 프로그래밍

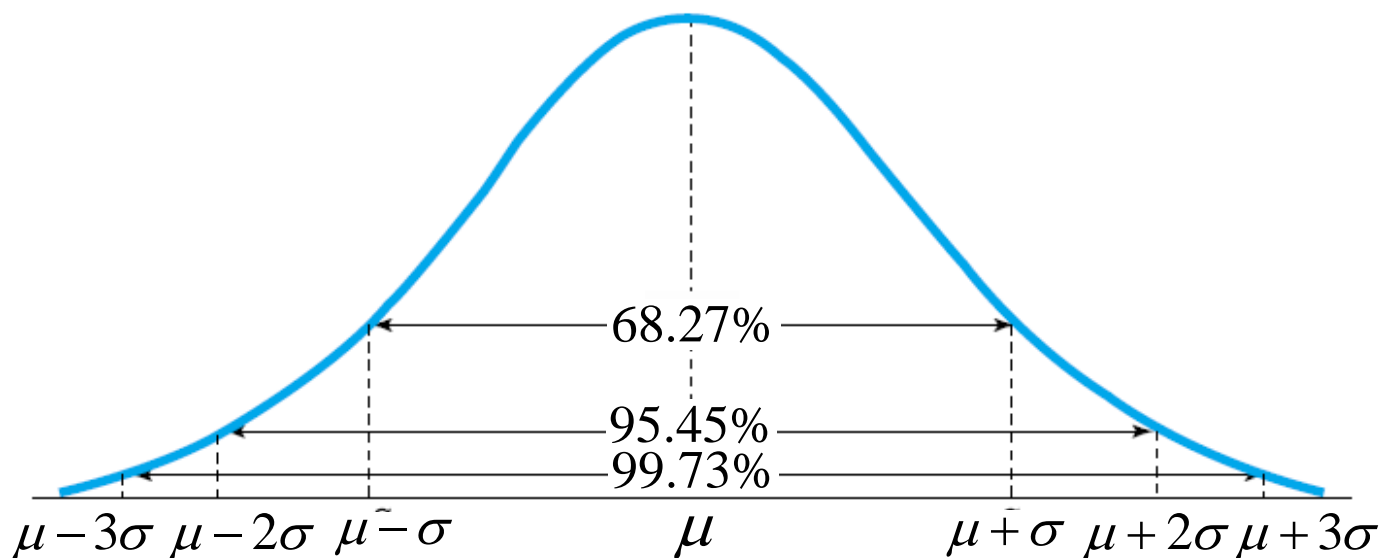
5장 확률분포

: 연속확률분포(정규분포)

[정의] 정규분포(normal distribution) $X \sim N(\mu, \sigma^2)$

기댓값을 중심으로 대칭이며, 중심위치는 기댓값, 산포는 표준편차에 의해 결정되는 얹어 놓은 종 모양의 분포

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



[정의] 표준정규분포(standard normal distribution)

기댓값은 0, 분산(표준편차)는 1인 정규분포

$$Z \sim N(0,1) \quad \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), -\infty < z < \infty$$

- **CDF** $\Phi(y) \equiv P(Z < y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$ **=pnorm(y)**

- **분위수** $z_p \equiv \Phi^{-1}(p) = \{y \mid \Phi(y) = p\}$ **=qnorm(p)**

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

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

정규분포의 R 함수

확률밀도함수 (mean=기댓값, sd=표준편차)

mean, sd를 생략하면 표준정규분포로 계산함 (mean = 0, sd = 1)
dnorm(x, mean = , sd =)

누적분포함수 (q=분위수)

pnorm(q, mean = , sd = , lower.tail = T)

분위수 (p=누적확률)

qnorm(p, mean = , sd = , lower.tail = T)

정규 확률변수 (n=난수의 개수)

rnorm(n, mean = , sd =)

[예제1] 과거 수년간에 걸친 수학과 학생의 영어성적은 $N(75, 9)$ 인 정규분포에 따른다. 이때 어느 한 학생의 영어성적이 **80점 이상일 확률**을 구하여라.

풀이 학생 개개인의 영어성적을 확률변수 X 라 하면 $X \sim N(75, 9)$ 이므로

$$\begin{aligned} P(X \geq 80) &= P\left(Z \geq \frac{80-75}{3}\right) \\ &\simeq P(Z \geq 1.67) = 1 - \Phi(1.67) \\ &= 1 - 0.9525 = 0.0475 \end{aligned}$$

정규분포 그래프 (Normal distribution plot)

x의 범위 정하기

연속인 실수범위는 seq(from=, to=, by=증감도, length = 구간 개수, 연속인 정수는 c(40:120))

`x <- seq(45, 115, length = 300) ; x`

dnorm(구간 x값, mean, sd) : x구간 300개의 각 값들에 대한 정규분포 만들어 주기

`y <- dnorm(x, mean = 75, sd = 3)` # `pnorm(x, mean = 75, sd = 3)` 로 대체하면 누적분포함수

plot(x, y) 그리기

`plot(x, y)` # 동그라미들은 x에 대응하여 300개

`plot(x, y, type = "l")` # 동그라미를 type= "l" (선)

`plot(x, y, type = "l", col = "red")` # 색상

```
x <- seq(-3, 3, length=200) ; x  
plot(x, dnorm(x, mean=0, sd=1), type='l', main="Standard normal distribution")  
  
# 누적정규분포곡선 Cumulative normal distribution plot  
x <- seq(-3, 3, length=200) >  
plot(x, pnorm(x, mean=0, sd=1), type='l', main="Cumulative normal distribution")
```

```
#예제1. x=80보다 클 때의 확률을 정규분포로 구하기
pnorm(80, mean = 75, sd = 3) # 0부터 80까지 누적확률
1 - pnorm(80, mean = 75, sd = 3)
# x=80보다 클 때의 확률을 표준정규분포로 구하기
1 - pnorm(1.67, mean = 0, sd = 1)
# 80부터 오른쪽 끝까지의 면적을 구하고 싶다면 lower.tail = F로 한다)
pnorm(80, mean = 75, sd = 3, lower.tail = F)
pnorm(1.67, mean = 0, sd = 1, lower.tail = F)
```

pnorm(q, mean, sd, lower.tail = TRUE) 이면 분위수 q를 기준으로 왼쪽의 $-\infty$ 부터 q까지의 면적 값을 보여주며,

pnorm(q, mean, sd, lower.tail = FALSE) 이면 분위수 q를 기준으로 q부터 $+\infty$ 까지의 오른쪽으로의 면적 값을 나타낸다.

```
pnorm(q=c(1), mean=0, sd=1, lower.tail = TRUE)
```

```
pnorm(q=c(1), mean=0, sd=1, lower.tail = FALSE)
```