

제13주: 분포함수

1. 분포와 관련된 함수

R에서 통계 분포 함수의 이름은 “접두어 + 분포명”입니다.

- 접두어:

접두어	의미	기능
d	density	확률밀도함수
p	probability	(누적)확률분포함수
q	quantile	(누적)확률분포함수의 y값에 해당하는 x값
r	random	난수생성

- 분포명:

연속형분포		이산형분포	
분포명	분포	분포명	분포
beta	Beta	binom	Binomial
cauchy	Cauchy	geom	Geometric
chisq	Chisquare	hyper	Hypergeometric
exp	Exponential	nbinom	음이항분포
f	F	pois	Poisson
gamma	Gamma	wilcox	Wilcoxon
logis	Logistic		
lnorm	Lognormal		
norm	Normal		
t	Student t		
unif	Uniform		
tukey	Tukey		
weibull	Weibull		

- 사용예1: 정규분포

- `dnorm(x, mean = 0, sd = 1, log = FALSE)`
- `pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)`
- `qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)`
- `rnorm(n, mean = 0, sd = 1, log = FALSE)`

```
# dnorm()
# 정규분포의 확률밀도함수(PDF)

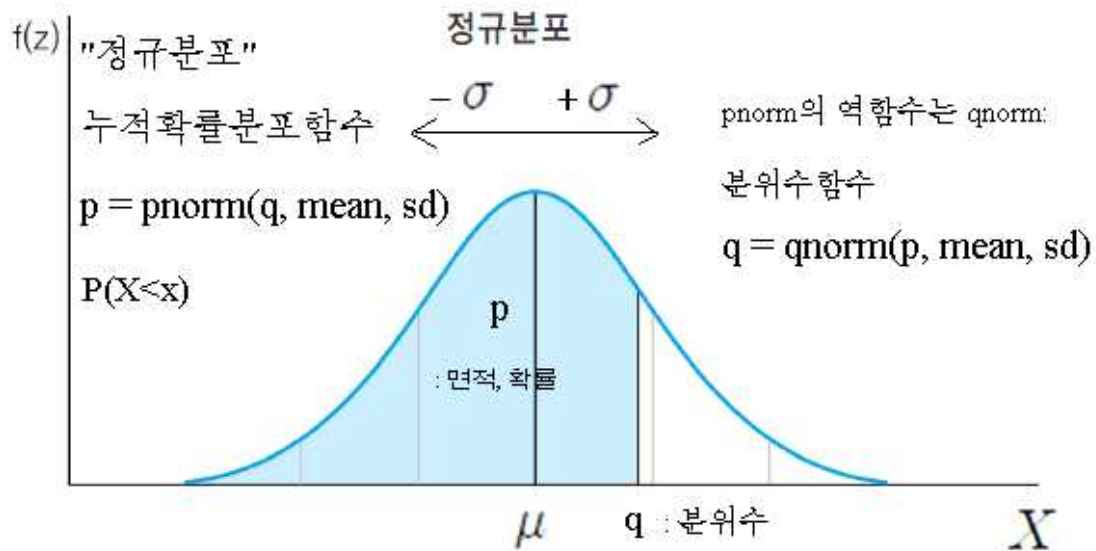
# 
$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

dnorm(0, mean = 0, sd = 1)
## [1] 0.3989423

# 평균과 표준편차에 기본값을 이용:
dnorm(0)
## [1] 0.3989423

# 평균과 표준편차에 다른 값을 이용:
dnorm(2, mean = 5, sd = 3)
## [1] 0.08065691

# 표준정규분포의 PDF(확률밀도함수) 그림 그리기:
z_scores <- seq(-3, 3, by = .1)    #z-값들의 벡터 생성
z_scores
dvalues <- dnorm(z_scores)         #z-값들에 대한 함수값 계산
dvalues
plot(z_scores, dvalues,          #x=z_scores와 y=dvalues의 그림
     type = "l",                #선(line)그림
     main = "표준정규분포의 확률밀도함수",
     xlab = "Z-score")
```



```
# pnorm()
pnorm(0)
## [1] 0.5
pnorm(2)
## [1] 0.9772499
pnorm(2, mean = 5, sd = 3)
## [1] 0.1586553
# lower.tail = FALSE를 설정하면 q부터 ∞까지의 영역 면적 계산
pnorm(2, mean = 5, sd = 3, lower.tail = FALSE)
## [1] 0.8413447
# pnorm(q)는 1-pnorm(q, lower.tail = FALSE)와 동일하다.
1-pnorm(2, mean = 5, sd = 3, lower.tail = FALSE)
## [1] 0.1586553
# 표준정규분포의 CDF(누적확률분포함수) 그리기
pvalues <- pnorm(z_scores)
plot(z_scores, pvalues,
     type = "l", # Make it a line plot
     main = "표준정규분포의 누적확률분포함수",
     xlab = "Quantiles",
     ylab = "Probability Density")
```

```

# qnorm()
qnorm(.5)
## [1] 0
# 제96분위수 구하기
qnorm(0.96)
# 제99분위수 구하기
qnorm(0.99)
# 역함수관계
pnorm(qnorm(0))
qnorm(pnorm(0))

```

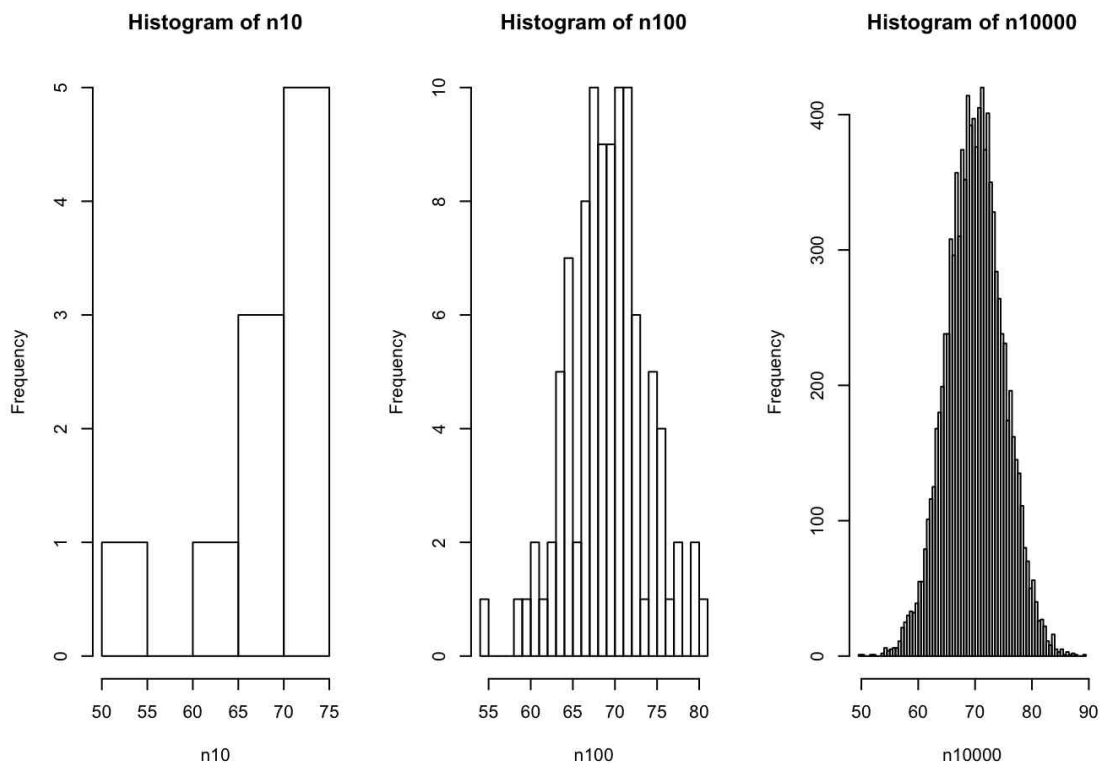
정규분포를 갖는 난수값들을 생성하고자 한다면 `rnorm(n, mean, sd)` 함수를 사용한다. 첫 번째 인자 `n`은 생성하고자 하는 난수의 개수이다. `rnorm()`을 이용하여 약한 대수의 법칙(weak law of large numbers)을 보이는 예제이다:

```

# set.seed() 함수는 난수생성을 위한 시드를 설정한다.
set.seed(12-5-2018)
rnorm(5)
## [1] 0.3130206 0.7406870 -0.4864897 -1.8161197
1.4202653
set.seed(12-5-2018)
rnorm(5)
## [1] 0.3130206 0.7406870 -0.4864897 -1.8161197
1.4202653
# 길이가 다른 벡터를 세 개 만들기
n10 <- rnorm(10, mean = 70, sd = 5)
n100 <- rnorm(100, mean = 70, sd = 5)
n10000 <- rnorm(10000, mean = 70, sd = 5)
n10

```

```
# 세 개의 히스토그램을 하나의 창에 나란히 그리기
# hist()의 전달인자 breaks=는 막대의 개수를 설정한다.
oldpar <- par() # 현재 그래픽 설정을 저장
par(mfrow = c(1,3))
hist(n10, breaks = 5)
hist(n100, breaks = 20)
hist(n10000, breaks = 100)
par(oldpar) # 이전 그래픽 설정을 회복
```



심화 예제: 중심극한정리에 대한 시뮬레이션

모집단의 분포가 평균 μ , 표준편차 σ 를 갖는다고 가정한다. 이 경우, 중심극한정리는 모집단의 분포에 상관없이 표본의 수가 크면 표본평균 \bar{X} 의 분포가 근사적으로 평균이 μ 이고 표준편차가 σ/\sqrt{n} 인 정규분포를 따른다는 것이다.

본 예에서는 한 쪽으로 치우쳐진 포아송분포로부터 추출한 표본평균이 표본의 개수가 증가함에 따라 정규분포를 따르는 것을 R을 이용하여 시각적으로 나타내보려 한다.

```
# 1. 포아송분포의 확률밀도함수 그리기
#   포아송분포의 모수는 λ
#    $f(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, x=0,1,2,\dots$ 
x <- 0:10
y1 <- dpois(x, lambda = 1)
y4 <- dpois(x, lambda = 4)
plot(x, y1, type = "b", col = "blue", lwd = 2,
      ylab = "Density", xlab = "x")
lines(x, y4, type = "b", col = "red", lwd = 2, lty = 2)

# 2. 표본추출1회 시행 및 상대도수 히스토그램 그리기
#   이산형자료이므로 막대그래프를 이용해 그린다.
n <- 100
lam <- 1
y <- rpois(n = n, lambda = lam)
z <- table(y)
z
barplot(z/sum(z), ylab = "상대도수")

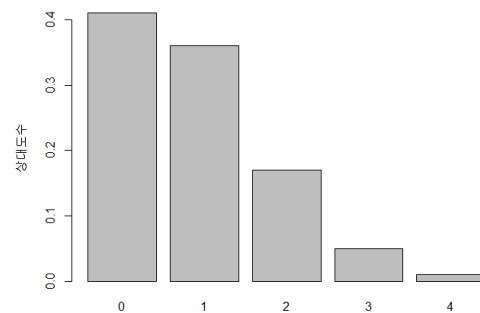
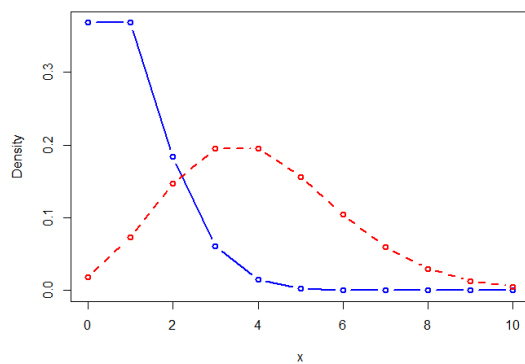
# 3. 표본추출 및 평균 계산을 반복하는 함수 작성하기
#   함수이름: sample_mean_pois
#   입력: 표본크기 n,
#         포아송분포 모수 lambda,
#         반복횟수 n.repeat
#   출력: 길이 n.repeat 인 표본평균값의 벡터
sample_mean_pois <- function(n, lambda, n.repeat) {
  ret <- NULL
```

```

for (i in 1:n.repeat) {
  x <- rpois(n, lambda)
  ret[i] <- mean(x)
}
return(ret)
}
xbar <- sample_mean_pois(10, 1, 100)
hist(xbar, freq = FALSE, main = "표본평균의 표집분포", xlab =
expression(bar(x)))

```

표본크기를 바꿔가면서 표본평균의 표집분포가 변해가는 모양을 관찰하는 R코드를 작성하시오.



표본평균의 표집분포

