I. 이론적 확률분포

II. 표준정규분포

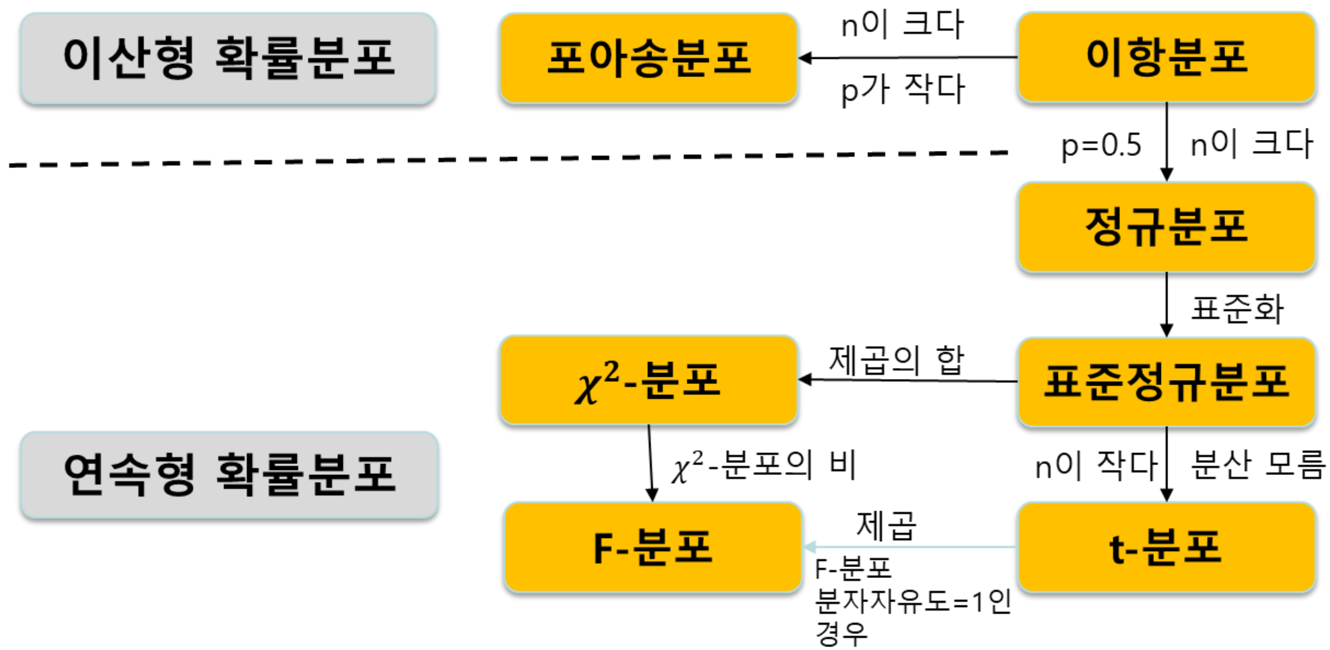
III. χ^2-분포

IV. t-분포

V. F-분포

1. 이론적 확률분포의 관계

- 확률분포는 이산형 확률분포와 연속형 확률분포로 구분
- 이론적 확률분포의 출발점은 정규분포를 표준화한 표준정규분포임
- 확률변수 $Z_1, Z_2, \ldots, Z_n$이 서로 독립적으로 표준정규분포 Z_i~N(0,1)(i=1,2,…,n)를 따를 때, $Z_1, Z_2, \ldots, Z_n$의 제곱합 X=$\sum_{i=1}^{n} Z_i^2$은 자유도가 n인 χ^2 –분포를 따름. 즉, X~χ_n^2
- Z~N(0,1), V~χ^2(ν)이고 Z와 V가 독립이면, T=$\frac{Z}{\sqrt{\frac{V}{v}}}$~t(ν)
- $X_1 \sim \chi_{v1}^2$, $X_2 \sim \chi_{v2}^2$이고 X_1, X_2가 서로 독립이면, F=$\frac{\frac{\chi_1}{v1}}{\frac{\chi_2}{v2}}$~F$(v_1, v_2)$



2. 확률분포 관련 통계함수

- **R에서 확률분포와 관련된 통계함수는 다음과 같음**

분포	R 함수	인수(arguments)
binomial	binom()	size, prob
chi-squared	chisq()	df, ncp
F	f()	df1, df2, ncp
normal	norm()	mean, sd
poison	pois()	lambda
Student's t	t()	df, ncp
uniform	unif()	min, max

- **우리가 원하는 통계량을 얻기 위해서는 함수의 이름 앞에 다음과 같은 접두사를 붙여야 함**

접두사	기능
d	확률밀도함수(PDF)의 확률값, f(x)
p	누적분포함수(CDF)의 확률값, F(x)
q	분위수(quantile) 값, $F^{-1}(x)$
r	무작위 난수 생성

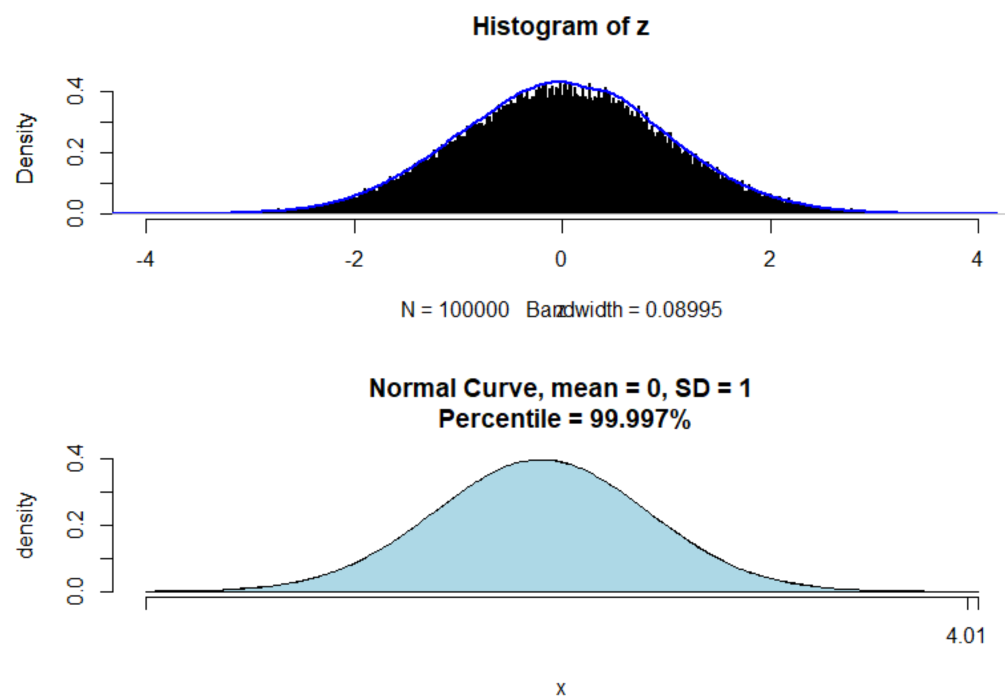
1. 확률함수

- 확률변수 X가 평균 0과 표준편차 1을 갖는 정규분포 즉, 표준정규분포를 따른다고 하면 $X \sim N(0, 1)$
- 확률밀도함수는 다음과 같음

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

b1-ch4-9-new.R

```
set.seed(1234)
n<-100000
z<-rnorm(n,0,1)
(mean(z))
(sd(z))
par(mfrow=c(2,1))
hist(z, freq=F,xlim=c(-4,4),breaks=1000)
par(new=T)
plot(density(z), axes=F, main="", xlim=c(-4,4), lwd=2, co
l="blue")
qnormGC(0.99997,region="below",m=0,s=1, graph=T)
```



2. 확률분포표

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879

b1-ch4-9.R

```
z00<-rep(NA,10);z01<-rep(NA,10);z02<-rep(NA,10);z03<-rep(NA,10);z04<-rep(NA,10)
for(i in 1:10) {z00[i]<-pnorm((i-1)/100, 0, 1)-0.5}
(z00<-round(z00, digits=4))
for(i in 10:19) {z01[i]<-pnorm(i/100, 0, 1)-0.5}
(z01<-round(z01[10:19], digits=4))
for(i in 20:29) {z02[i]<-pnorm(i/100, 0, 1)-0.5}
(z02<-round(z02[20:29], digits=4))
for(i in 30:39) {z03[i]<-pnorm(i/100, 0, 1)-0.5}
(z03<-round(z03[30:39], digits=4))
for(i in 40:49) {z04[i]<-pnorm(i/100, 0, 1)-0.5}
(z04<-round(z04[40:49], digits=4))
zdist<-rbind(z00,z01,z02,z03,z04)
(zdist<-round(zdist, digits=4))
```

```
> (zdist<-round(zdist, digits=4))
      [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]   [,8]   [,9]  [,10]
z00 0.0000 0.0040 0.0080 0.0120 0.0160 0.0199 0.0239 0.0279 0.0319 0.0359
z01 0.0398 0.0438 0.0478 0.0517 0.0557 0.0596 0.0636 0.0675 0.0714 0.0753
z02 0.0793 0.0832 0.0871 0.0910 0.0948 0.0987 0.1026 0.1064 0.1103 0.1141
z03 0.1179 0.1217 0.1255 0.1293 0.1331 0.1368 0.1406 0.1443 0.1480 0.1517
z04 0.1554 0.1591 0.1628 0.1664 0.1700 0.1736 0.1772 0.1808 0.1844 0.1879
```

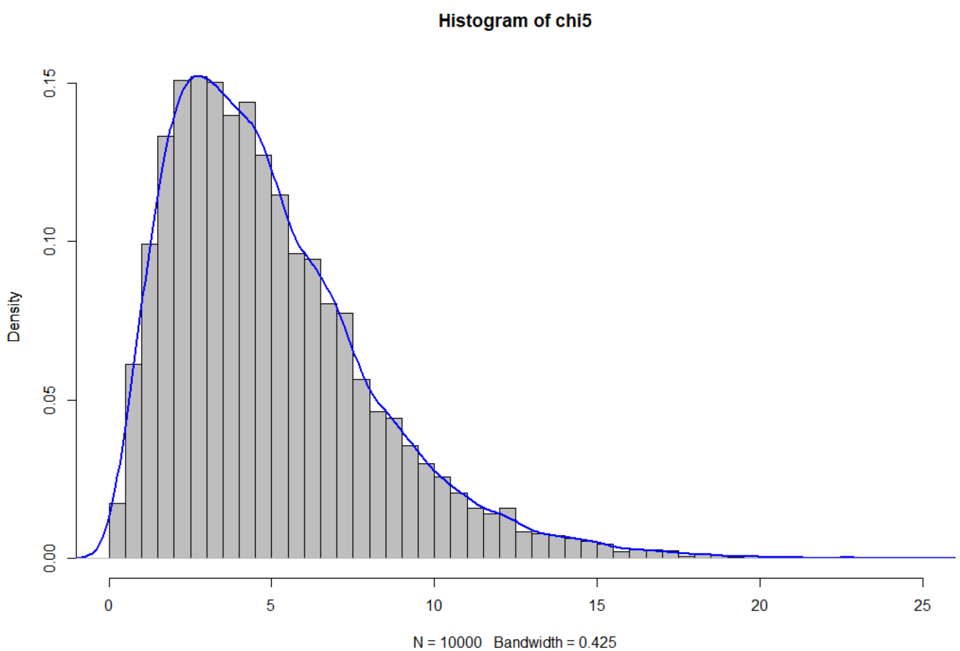
1. 확률분포

- 표준정규분포의 제곱의 합 X=$\sum_{i=1}^{n} Z_i^2$이 자유도가 n인 χ^2-분포에 따름
- X~χ^2(n)
- X가 X~χ^2(n)이라고 할 때, X의 평균은 n, 분산은 2n

b1-ch4-10.R

```
set.seed(12345)

n<-10000;

z1<-rnorm(n,0,1)
z2<-rnorm(n,0,1)
z3<-rnorm(n,0,1)
z4<-rnorm(n,0,1)
z5<-rnorm(n,0,1)

chi5<-z1^2+z2^2+z3^2+z4^2+z5^2
(mean(chi5))
(var(chi5))
hist(chi5,  freq=F, col="grey", xlab="", xlim=c(0, 25),
breaks=100)
par(new=T)
plot(density(chi5), axes=F, main="", xlim=c(0, 25), l
wd=2, col="blue")
```

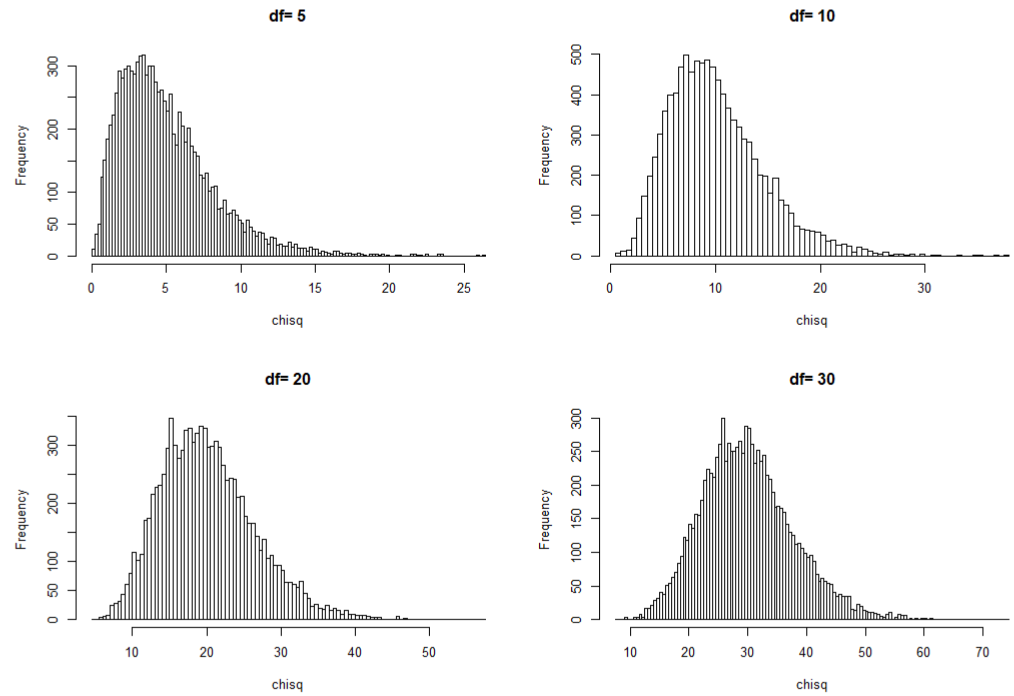
```
> (mean(chi5))
[1] 4.975514
> (var(chi5))
[1] 10.05467
> library(car)
```



- **자유도에 따라 χ^2-분포의 모양이 달라지는데 자유도가 클수록 정규분포와 근사**

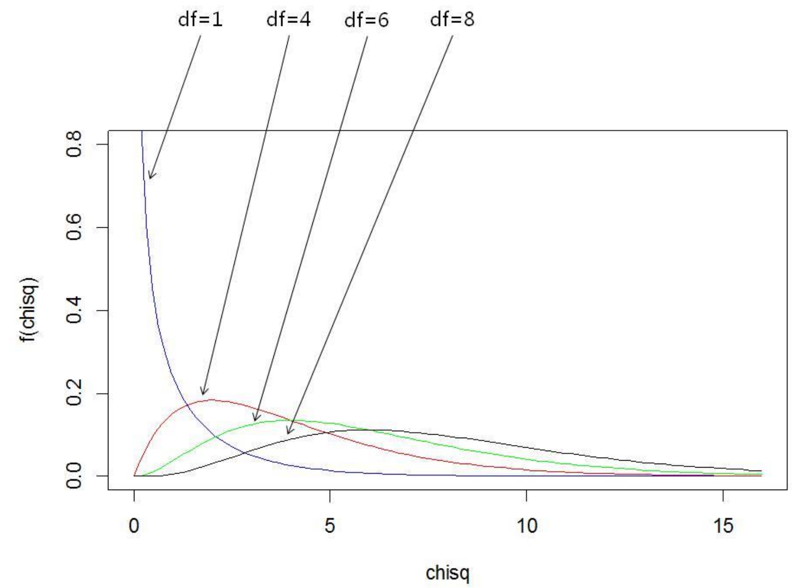
b1-ch4-11.R

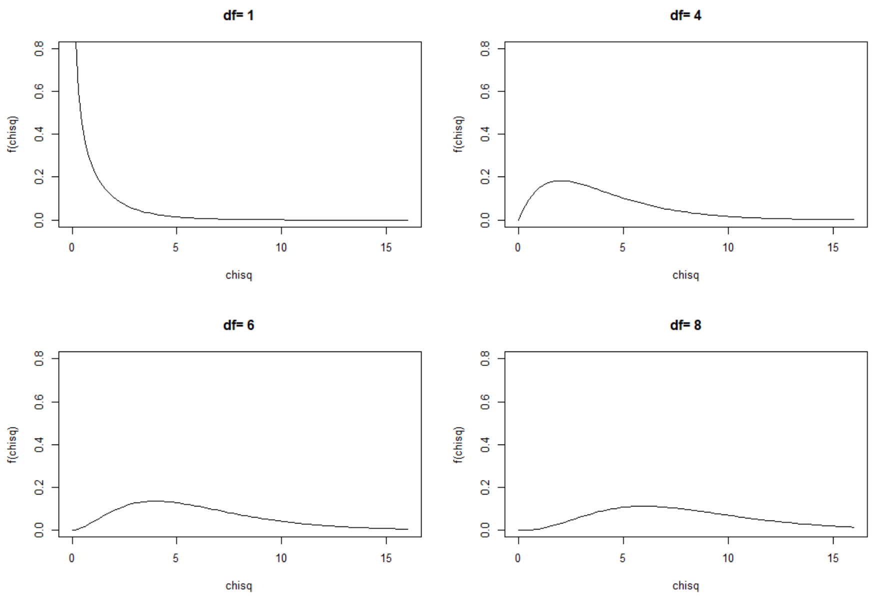
```
set.seed(12345)
n<-10000;
df_list<-c(5,10,20,30)
par(mfrow=c(2,2))
for (i in 1:length(df_list)) {
  hist(rchisq(n, df=df_list[i], ncp=0), breaks=100,
xlab="chisq", main=paste("df=", df_list[i]))
}
```



b1-ch4-12.R

```
n_list<-c(2,5,7,9) # 표본수(n)
df_list<-n_list-1 # 자유도
curve(dchisq(x, 1, ncp=0), col="blue", xlim=c(0, 16
), ylim=c(0, 0.8), xlab="chisq", ylab="f(chisq)")
curve(dchisq(x, 4, ncp=0), add=T, col="red", xlim=
c(0, 16), ylim=c(0, 0.8), xlab="chisq", ylab="f(chisq
)")
curve(dchisq(x, 6, ncp=0), add=T, col="green", xli
m=c(0, 16), ylim=c(0, 0.8), xlab="chisq", ylab="f(ch
isq)")
curve(dchisq(x, 8, ncp=0), add=T, col="black", xlim
=c(0, 16), ylim=c(0, 0.8), xlab="chisq", ylab="f(chis
q)")

par(mfrow=c(2,2))
for (i in 1:length(df_list)) {
  curve(dchisq(x, df_list[i], ncp=0), add=F, xlim=c(
0, 16), ylim=c(0, 0.8), xlab="chisq", ylab="f(chisq)",
main=paste("df=", df_list[i])) }
```





2. 확률분포표

자유도	P=0.99	0.95	0.90	0.10	0.05	0.01
1	0.000157	0.00393	0.0158	2.706	3.841	6.635
2	0.0201	0.103	0.211	4.605	5.991	9.210
3	0.115	0.352	0.584	6.251	7.815	11.341
4	0.297	0.711	1.064	7.779	9.488	13.277
5	0.554	1.145	1.610	9.236	11.070	15.086
6	0.872	1.635	2.204	10.645	12.592	16.812
7	1.239	2.167	2.833	12.017	14.067	18.475
8	1.646	2.733	3.490	13.362	15.507	20.090
9	2.088	3.325	4.168	14.684	16.919	21.666
10	2.558	3.940	4.865	15.987	18.307	23.209

b1-ch4-13.R

```
df<-10
chi1<-numeric(df);chi2<-numeric(df);chi3<-numeric(df)
chi4<-numeric(df);chi5<-numeric(df);chi6<-numeric(df)
 for(j in 1:df) { chi1[j]<-qchisq(0.01,j) }
for(j in 1:df) { chi2[j]<-qchisq(0.05,j) }
for(j in 1:df) { chi3[j]<-qchisq(0.1,j) }
for(j in 1:df) { chi4[j]<-qchisq(0.9,j) }
for(j in 1:df) { chi5[j]<-qchisq(0.95,j) }
for(j in 1:df) { chi6[j]<-qchisq(0.99,j) }
round((chi<-cbind(chi1,chi2,chi3,chi4,chi5,  chi6)),digits
=4)
```

```
> round((chi<-cbind(chi1,chi2,chi3,chi4,chi5, chi6)),digits=4)
        chi1   chi2   chi3    chi4    chi5    chi6
 [1,] 0.0002 0.0039 0.0158  2.7055  3.8415  6.6349
 [2,] 0.0201 0.1026 0.2107  4.6052  5.9915  9.2103
 [3,] 0.1148 0.3518 0.5844  6.2514  7.8147 11.3449
 [4,] 0.2971 0.7107 1.0636  7.7794  9.4877 13.2767
 [5,] 0.5543 1.1455 1.6103  9.2364 11.0705 15.0863
 [6,] 0.8721 1.6354 2.2041 10.6446 12.5916 16.8119
 [7,] 1.2390 2.1673 2.8331 12.0170 14.0671 18.4753
 [8,] 1.6465 2.7326 3.4895 13.3616 15.5073 20.0902
 [9,] 2.0879 3.3251 4.1682 14.6837 16.9190 21.6660
[10,] 2.5582 3.9403 4.8652 15.9872 18.3070 23.2093
```

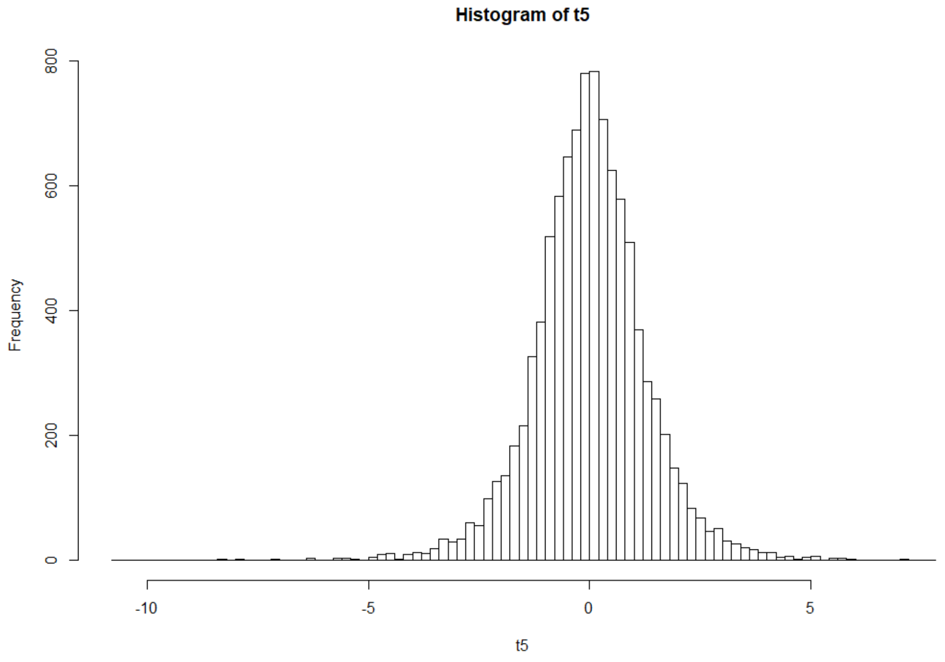
1. 확률분포

- 서로 독립적인 표준정규분포와 χ^2-분포에 의해 t-분포가 도출

  X~t(n-1)

b1-ch4-14.R

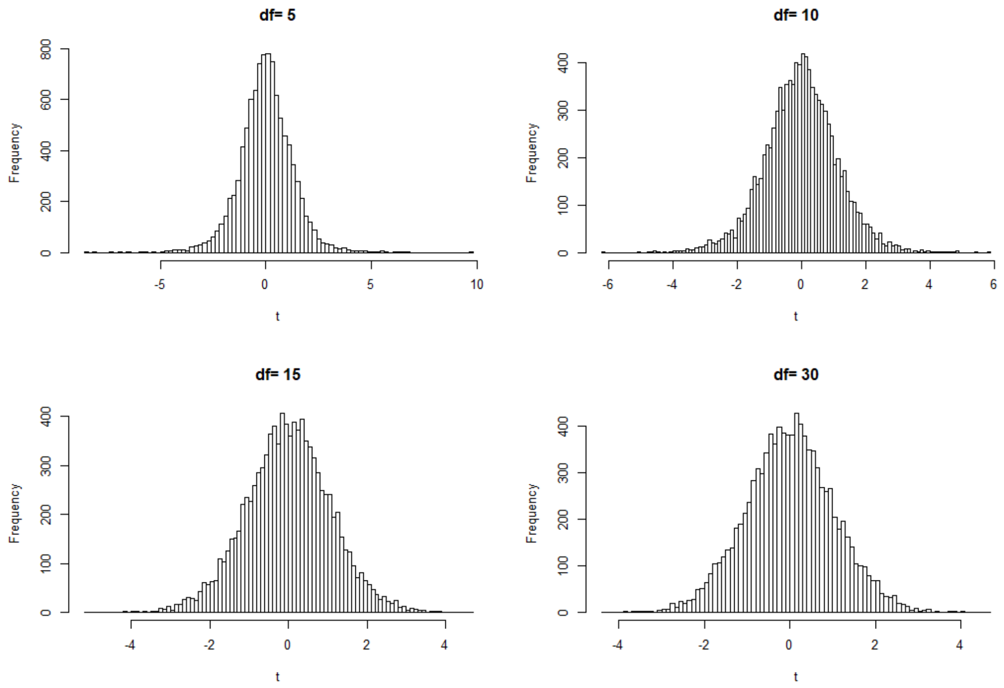
```
set.seed(12345)

n<-10000;

z<-rnorm(n,0,1)
z1<-rnorm(n,0,1)
z2<-rnorm(n,0,1)
z3<-rnorm(n,0,1)
z4<-rnorm(n,0,1)
z5<-rnorm(n,0,1)

chi5<-z1^2+z2^2+z3^2+z4^2+z5^2

sqchi5<-sqrt(chi5/5)

t5<-z/sqchi5

hist(t5, breaks=100)
```



- **자유도에 따라 t-분포의 모양이 달라지는데 자유도가 30 이상이면 표준정규분포와 거의 같아짐**

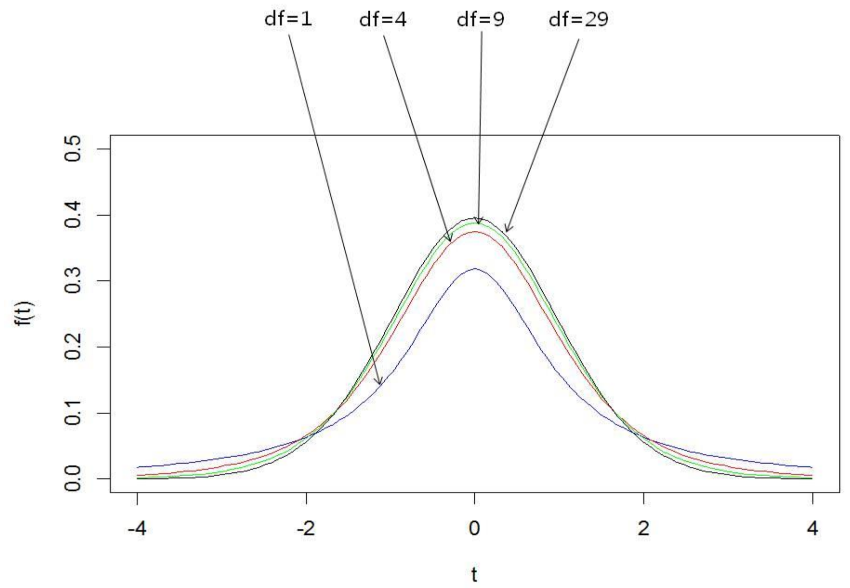
b1-ch4-15.R

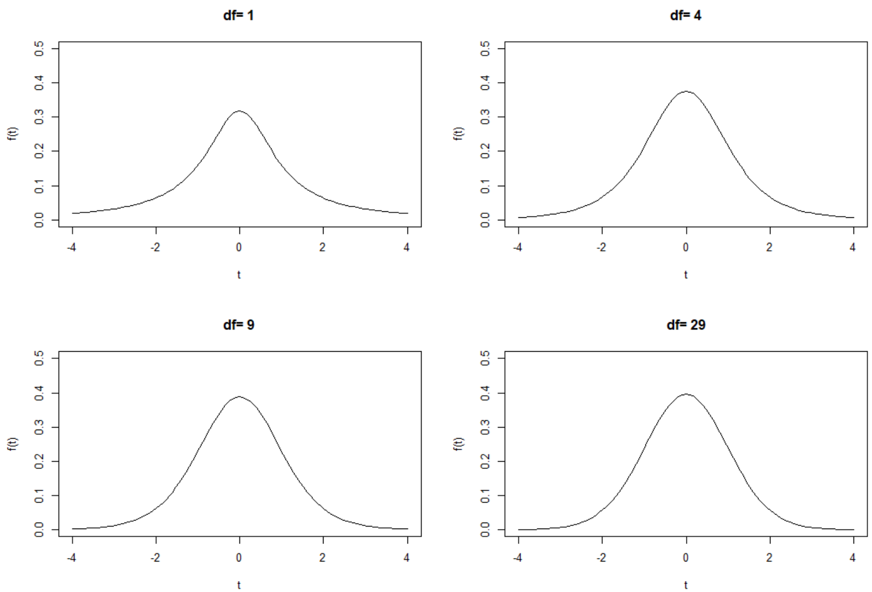
```
set.seed(12345)

n<-10000;
df_list<-c(5,10,15,30)
par(mfrow=c(2,2))

for (i in 1:length(df_list)) {
  hist(rt(n, df=df_list[i], ncp=0), breaks=100, xlab
="t", main=paste("df=", df_list[i]))
 }
```



b1-ch4-16.R

```
n_list<-c(2,5,10,30) # 표본수(n)
df_list<-n_list-1 # 자유도

curve(dt(x, 1, ncp=0), add=T, col="blue", xlim=c(-4
, 4), ylim=c(0, 0.5), xlab="t", ylab="f(t)")
curve(dt(x, 4, ncp=0), add=T, col="red", xlim=c(-4,
4), ylim=c(0, 0.5), xlab="t", ylab="f(t)")
curve(dt(x, 9, ncp=0), add=T, col="green", xlim=c(
-4, 4), ylim=c(0, 0.5), xlab="t", ylab="f(t)")
curve(dt(x, 29, ncp=0), add=T, col="black", xlim=c(
-4, 4), ylim=c(0, 0.5), xlab="t", ylab="f(t)")

par(mfrow=c(2,2))

for (i in 1:length(df_list)) {

  curve(dt(x, df_list[i], ncp=0), xlim=c(-4, 4), ylim=
c(0, 0.5), xlab="t", ylab="f(t)", main=paste("df=", df
_list[i]))

  }
```





2. 확률분포표

v \ p	0.1	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.923
3	0.1638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250

b1-ch4-17.R

```
t11<-rep(NA,9)
t12<-rep(NA,9)
t13<-rep(NA,9)
t14<-rep(NA,9)
t15<-rep(NA,9)

for(i in 1:9) {  t11[i]<-qt(0.9, i) }
for(i in 1:9) {  t12[i]<-qt(0.95, i) }
for(i in 1:9) {  t13[i]<-qt(0.975,i) }
for(i in 1:9) {  t14[i]<-qt(0.99, i) }
for(i in 1:9) {   t15[i]<-qt(0.995, i) }

round((poi<-cbind(t11,t12,t13,t14,t15)), digits=3)
```

```
> round((poi<-cbind(t11,t12,t13,t14,t15)), digits=3)
        t11   t12    t13    t14    t15
 [1,] 3.078 6.314 12.706 31.821 63.657
 [2,] 1.886 2.920  4.303  6.965  9.925
 [3,] 1.638 2.353  3.182  4.541  5.841
 [4,] 1.533 2.132  2.776  3.747  4.604
 [5,] 1.476 2.015  2.571  3.365  4.032
 [6,] 1.440 1.943  2.447  3.143  3.707
 [7,] 1.415 1.895  2.365  2.998  3.499
 [8,] 1.397 1.860  2.306  2.896  3.355
 [9,] 1.383 1.833  2.262  2.821  3.250
```

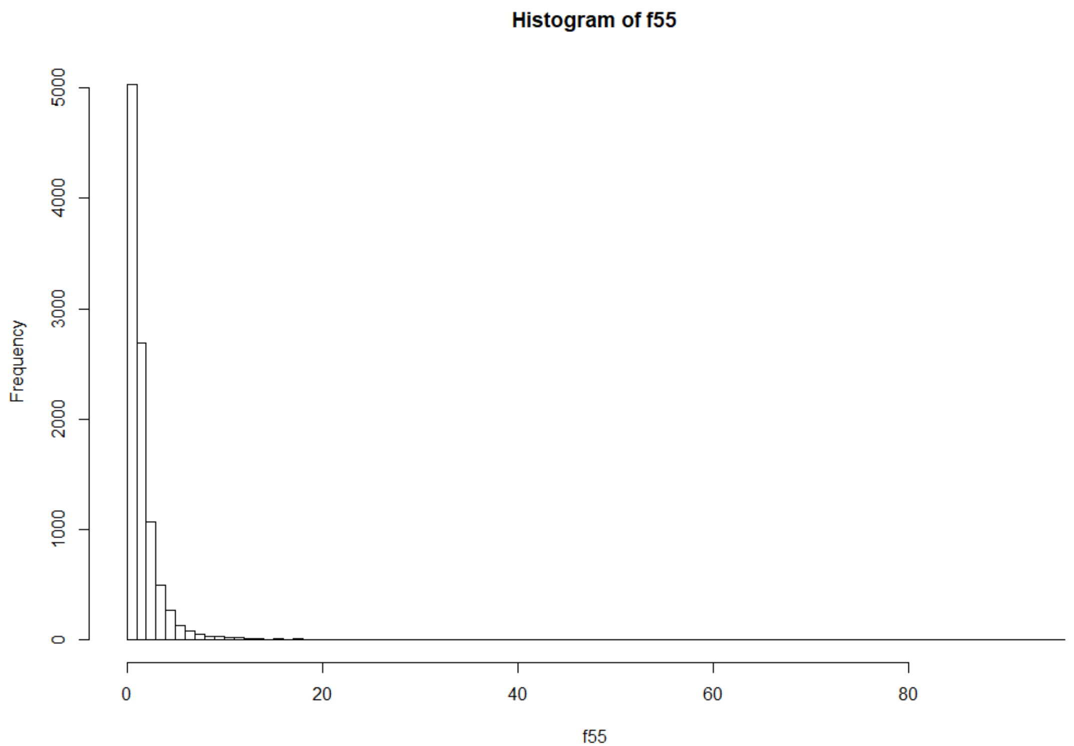
1. 확률분포

- 두 개의 독립적인 χ^2-분포에 의해 F-분포가 도출

  X~F(n_1-1, n_2-1)

b1-ch4-18.R

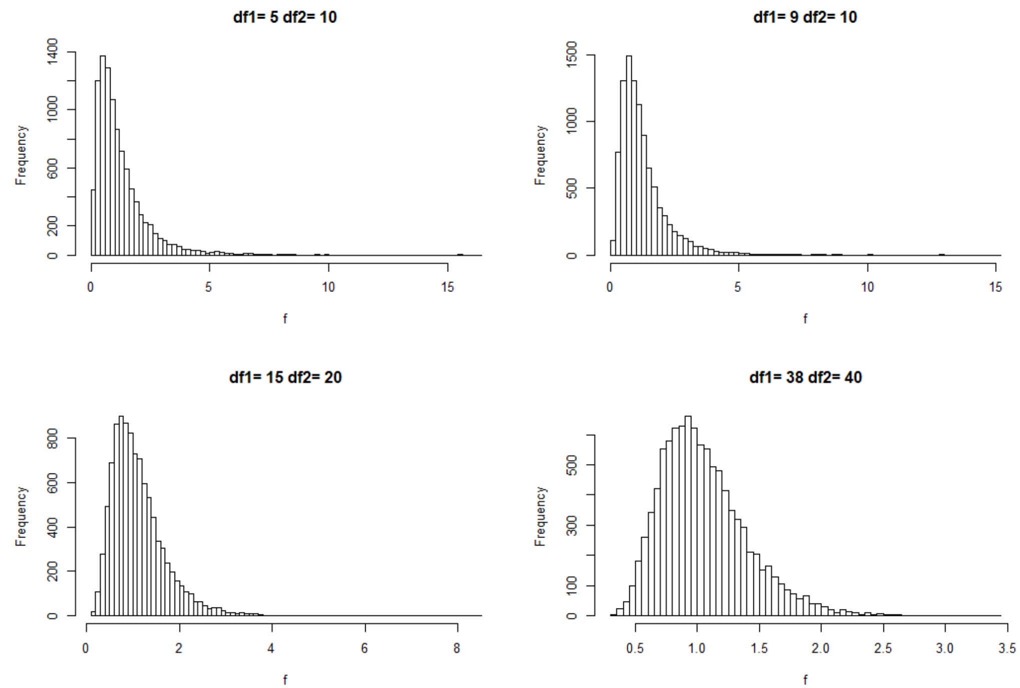
```
set.seed(12345)

n<-10000;

z1<-rnorm(n,0,1)
z2<-rnorm(n,0,1)
z3<-rnorm(n,0,1)
z4<-rnorm(n,0,1)
z5<-rnorm(n,0,1)
z6<-rnorm(n,0,1)
z7<-rnorm(n,0,1)
z8<-rnorm(n,0,1)
z9<-rnorm(n,0,1)
z10<-rnorm(n,0,1)

chi15<-z1^2+z2^2+z3^2+z4^2+z5^2
chi25<-z6^2+z7^2+z8^2+z9^2+z10^2

f55<-(chi15/5)/(chi25/5)

hist(f55, breaks=100)
```



- **분자 및 분모의 자유도에 같아지면서 커질수록 좌우대칭 분포와 비슷하게 됨**

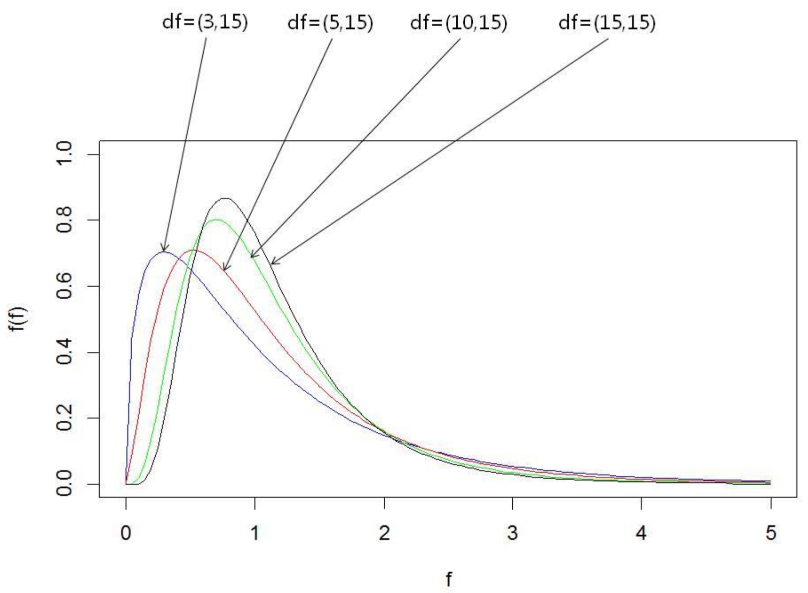
b1-ch4-19.R

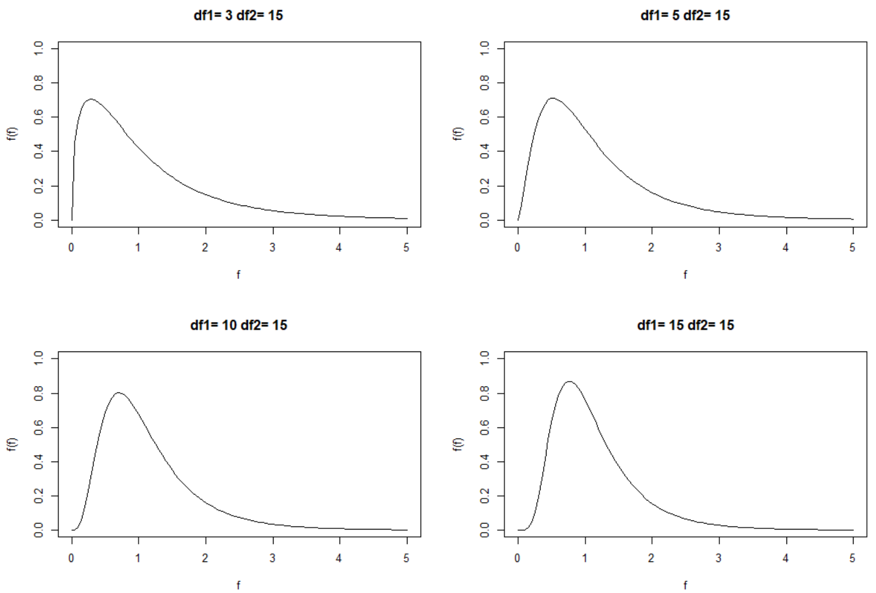
```
set.seed(12345)

n<-10000;
df1_list<-c(5,9,15,38)
df2_list<-c(10,10,20,40)
par(mfrow=c(2,2))

for (i in 1:length(df_list)) {
  hist(rf(n, df1=df1_list[i], df2=df2_list[i], ncp=0),
breaks=100, xlab="f", main=paste("df1=", df1_list[
i], "df2=", df2_list[i]))
}
```



b1-ch4-20.R

```
curve(df(x, 3, 15, ncp=0), col="blue", xlim=c(0,5), y
lim=c(0,1), xlab="f", ylab="f(f)")
curve(df(x, 5, 15, ncp=0), add=T, col="red", xlim=c
(0,5), ylim=c(0,1), xlab="f", ylab="f(f)")
curve(df(x, 10, 15, ncp=0), add=T, col="green", xli
m=c(0,5), ylim=c(0,1), xlab="f", ylab="f(f)")
curve(df(x, 15, 15, ncp=0), add=T, col="black", xli
m=c(0,5), ylim=c(0,1), xlab="f", ylab="f(f)")

df1_list<-c(3,5,10,15)
df2_list<-c(15,15,15,15)
par(mfrow=c(2,2))


for (i in 1:length(df1_list)) {
  curve(df(x, df1=df1_list[i], df2=df2_list[i], ncp=0)
, xlim=c(0,5), ylim=c(0,1), xlab="f", ylab="f(f)", mai
n=paste("df1=", df1_list[i], "df2=", df2_list[i]))
}
```





2. 확률분포표

b1-ch4-21.R

```
11<-rep(NA,10);f12<-rep(NA,10);f13<-rep(NA,1
0);
f14<-rep(NA,10)f15<-rep(NA,10);f16<-rep(NA,1
0);
f17<-rep(NA,10);f18<-rep(NA,10);f19<-rep(NA,1
0);
f110<-rep(NA,10)

for(i in 1:10) {  f11[i]<-qf(0.95, 1, i)}
for(i in 1:10) {  f12[i]<-qf(0.95, 2, i)}
for(i in 1:10) {  f13[i]<-qf(0.95, 3, i)}
for(i in 1:10) {  f14[i]<-qf(0.95, 4, i)}
for(i in 1:10) {  f15[i]<-qf(0.95, 5, i)}
for(i in 1:10) {  f16[i]<-qf(0.95, 6, i)}
for(i in 1:10) {  f17[i]<-qf(0.95, 7, i)}
for(i in 1:10) {  f18[i]<-qf(0.95, 8, i)}
for(i in 1:10) {  f19[i]<-qf(0.95, 9, i)}
for(i in 1:10) {  f110[i]<-qf(0.95, 10, i)}

round((poi<-cbind(f11,f12,f13,f14,f15,f16,f17,f1
8,f19,f110)), digits=2)
```

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.76
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.48	4.82	4.77	4.74
6	5.99	4.74	7.35	4.12	3.94	3.87	3.79	3.73	3.68	3.64
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4346	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98

```
> round((poi<-cbind(f11,f12,f13,f14,f15,f16,f17,f18,f19,f110)), digits=2)
         f11    f12    f13    f14    f15    f16    f17    f18    f19   f110
 [1,] 161.45 199.50 215.71 224.58 230.16 233.99 236.77 238.88 240.54 241.88
 [2,]  18.51  19.00  19.16  19.25  19.30  19.33  19.35  19.37  19.38  19.40
 [3,]  10.13   9.55   9.28   9.12   9.01   8.94   8.89   8.85   8.81   8.79
 [4,]   7.71   6.94   6.59   6.39   6.26   6.16   6.09   6.04   6.00   5.96
 [5,]   6.61   5.79   5.41   5.19   5.05   4.95   4.88   4.82   4.77   4.74
 [6,]   5.99   5.14   4.76   4.53   4.39   4.28   4.21   4.15   4.10   4.06
 [7,]   5.59   4.74   4.35   4.12   3.97   3.87   3.79   3.73   3.68   3.64
 [8,]   5.32   4.46   4.07   3.84   3.69   3.58   3.50   3.44   3.39   3.35
 [9,]   5.12   4.26   3.86   3.63   3.48   3.37   3.29   3.23   3.18   3.14
[10,]   4.96   4.10   3.71   3.48   3.33   3.22   3.14   3.07   3.02   2.98
```