

I. Excel 확률분포 시뮬레이션

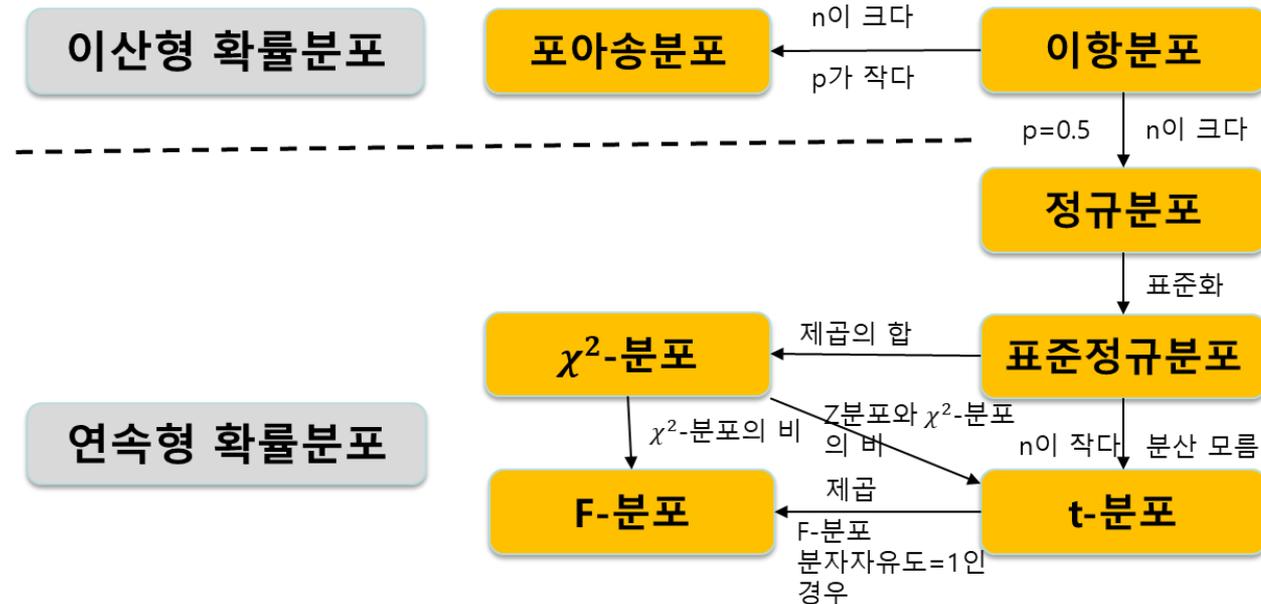
II. R 확률분포 시뮬레이션



## 0. 이론적 확률분포의 관계

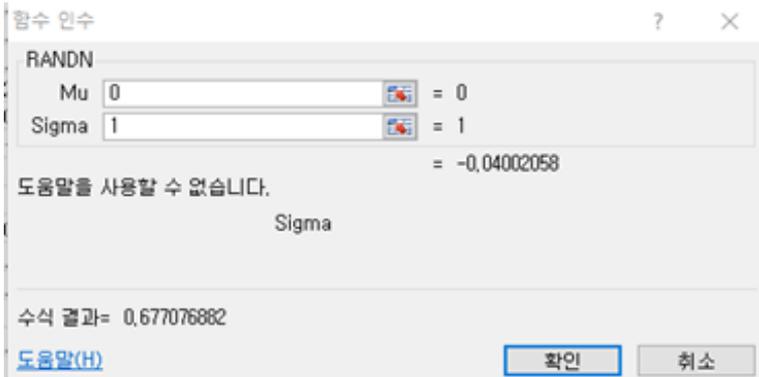
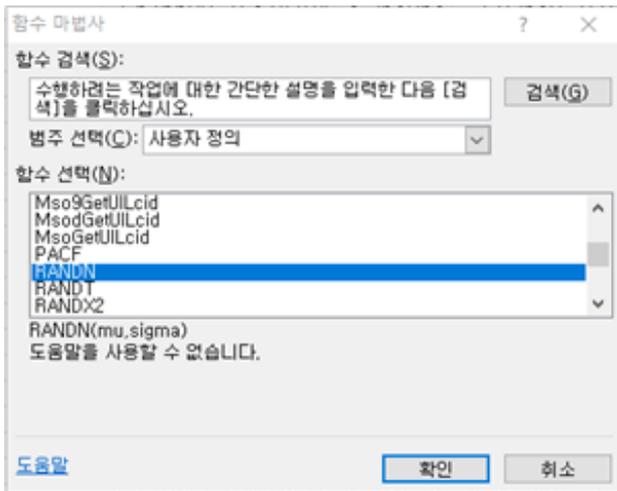
- 확률분포는 이산형 확률분포와 연속형 확률분포로 구분
- 이론적 확률분포의 출발점은 정규분포를 표준화한 표준정규분포임
- 확률변수  $Z_1, Z_2, \dots, Z_n$ 이 서로 독립적으로 표준정규분포  $Z_i \sim N(0,1) (i=1,2,\dots,n)$ 를 따를 때,  $Z_1, Z_2, \dots, Z_n$ 의 제곱합  $X = \sum_{i=1}^n Z_i^2$ 은 자유도가  $n$ 인  $\chi^2$ -분포를 따름. 즉,  $X \sim \chi_n^2$
- $Z \sim N(0,1), V \sim \chi^2(v)$ 이고  $Z$ 와  $V$ 가 독립이면,  $T = \frac{Z}{\sqrt{\frac{V}{v}}} \sim t(v)$
- $X_1 \sim \chi_{v_1}^2, X_2 \sim \chi_{v_2}^2$ 이고  $X_1, X_2$ 가 서로 독립이면,  $F = \frac{\frac{X_1}{v_1}}{\frac{X_2}{v_2}} \sim F(v_1, v_2)$

### 이론적 확률분포의 관계



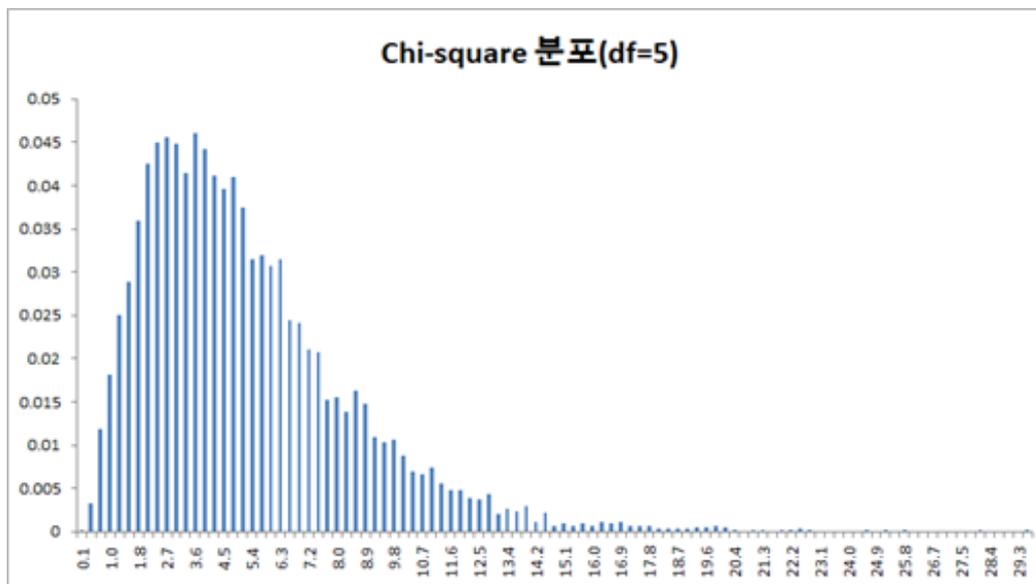
# 1. 표준정규분포 및 시뮬레이션

- 평균이 0이고, 표준편차가 1인 표준정규분포는 모든 확률분포의 출발점
- A1셀에서 함수삽입 아이콘을 클릭하면 나타나는 함수마법사에서 표준정규분포에 따르는 데이터를 생성해 주는 RANDN을 선택
- RANDN의 함수 인수 대화상자에서 그림과 같이 입력하면 표준정규분포에 따르는 임의 수 1개가 생성
- A10000까지 복사하면 표준정규분포에 따르는  $Z_1$ 의 임의 수 10000개가 생성되고,  $Z_1 \sim N(0,1)$
- 데이터-분석-데이터 분석을 실행하면 나타나는 통계 데이터분석 대화상자에서 히스토그램을 선택하고, 히스토그램대화상자에서 입력범위에 \$A\$1:\$A\$10000을 입력하고 차트출력에 체크한 후 확인 클릭



## 2. $\chi^2$ -분포 및 시뮬레이션

- 앞에서  $Z_1$ 을 생성했던 것과 동일한 방법으로 B1셀부터 E1셀까지 각각  $Z_2, Z_3, Z_4, Z_5$ 를 생성한 후 각 셀을 10000번째 행까지 복사
- G1셀부터 K1셀까지 이미 생성된  $Z_1, Z_2, Z_3, Z_4, Z_5$ 를 각각 제곱한  $Z_1^2, Z_2^2, Z_3^2, Z_4^2, Z_5^2$ 을 만들고 각 셀을 10000번째 행까지 복사
- M1셀에  $Z_1^2, Z_2^2, Z_3^2, Z_4^2, Z_5^2$ 을 모두 합하고 10000번째 행까지 복사하면 이 값은 자유도가 5인  $\chi^2$ -분포를 따르게 됨
- 즉,  $X = \sum_{i=1}^n Z_i^2 \sim \chi_5^2$
- 데이터-분석-데이터 분석을 실행하면 나타나는 통계 데이터분석 대화상자에서 히스토그램을 선택하고, 히스토그램대화상자에서 입력범위에  $\$M\$1:\$M\$10000$ 을 입력하고 차트출력에 체크한 후 확인 클릭



## 1. 표준정규분포 및 시뮬레이션

- `set.seed` : 난수 생성기의 상태를 통제하는 것으로 동일한 숫자(seed)는 동일한 난수를 생성
- `z<-rnorm(n,0,1)` : 표준정규분포에서 임의로 10000개의 데이터를 생성
- `par(mfrow=c(1,2))` : 그래픽 매개변수를 설정하는데 `mfrow=c(1,2)`는 그림을 하나의 행, 2개의 열로 배열
- `par(new=T)` : `new=T`는 히스토그램 위에 다른 그래프를 그림
- `plot(density(z), ...)` : `plot`은 R 객체를 그리는 함수이고, `density(z)`는 데이터 `z`의 밀도
- `curve(dnorm(x,0,1), ...)` : `curve`는 함수에 해당하는 곡선을 그리는데 `dnorm(x,0,1)`는 표준정규분포의 확률밀도함수

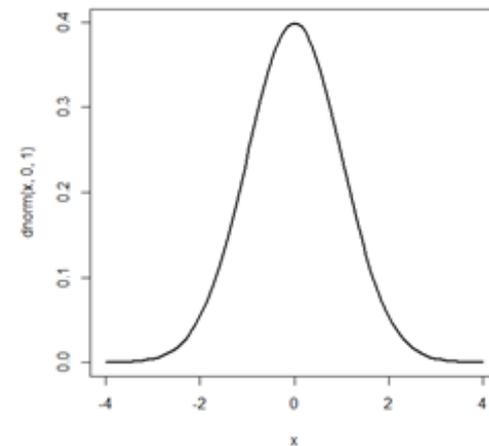
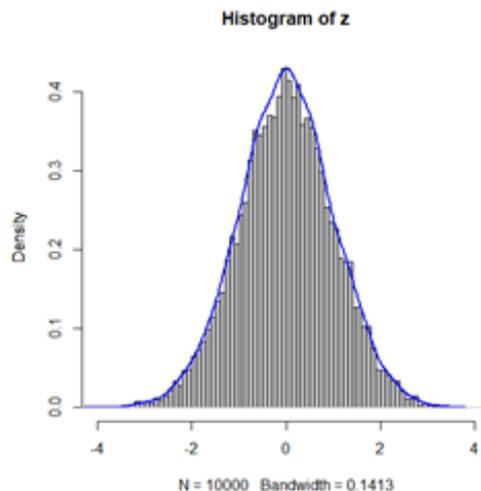
```

b3-ch3-7.R

set.seed(12345)

n<-10000;
z<-rnorm(n,0,1)
par(mfrow=c(1,2))

hist(z, freq=F, col="grey", xlab="", xlim=c(-4, 4), breaks
=100)
par(new=T)
plot(density(z), axes=F, main="", xlim=c(-4, 4), lwd=2, c
ol="blue")
curve(dnorm(x,0,1),xlim=c(-4, 4), lwd=2)
    
```



## 2. $\chi^2$ -분포 및 시뮬레이션

- `plot(density(chi5), ....)` : `plot`은 R 객체를 그리는 함수이고, `density(chi5)`는 데이터 `chi5`의 밀도
- `curve(dchisq(x,5,ncp=0), ....)` : `curve`는 함수에 해당하는 곡선을 그리는데 `dchisq(x,5, ncp=0)`는 자유도가 5인  $\chi^2$ -분포의 확률밀도함수

```

b3-ch3-8.R

set.seed(12345)

n<-10000;

par(mfrow=c(1,2))

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

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), lwd=2, col="blue")
curve(dchisq(x, 5, ncp=0), col="black", xlim=c(0, 25), ylim=c(0, 0.15), xlab="chisq", ylab="f(chisq)")
    
```

