データの分析と知識発見

Introduction to Data Analysis

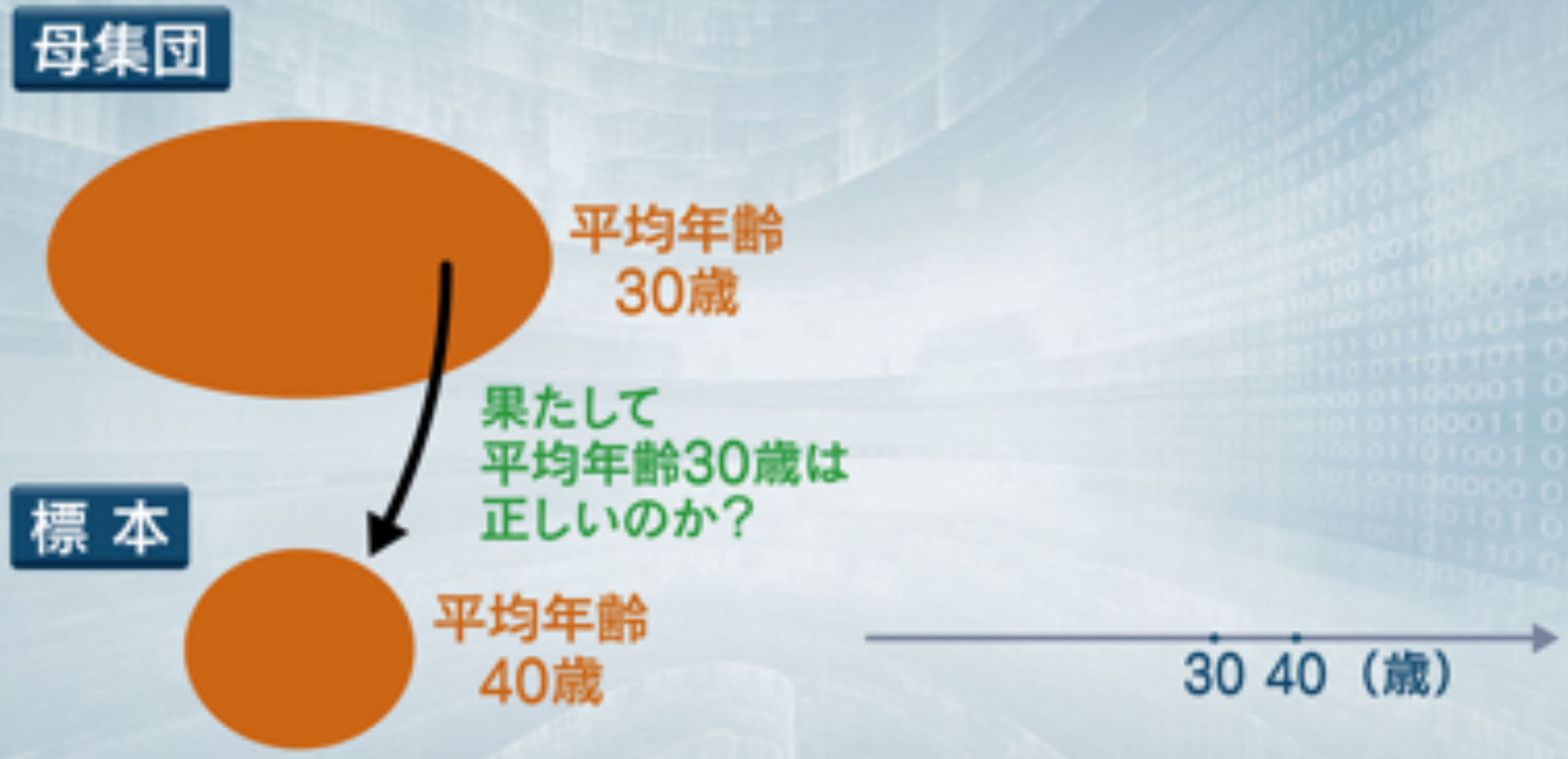
今回の構成

統計的検定の考え方について学ぶ

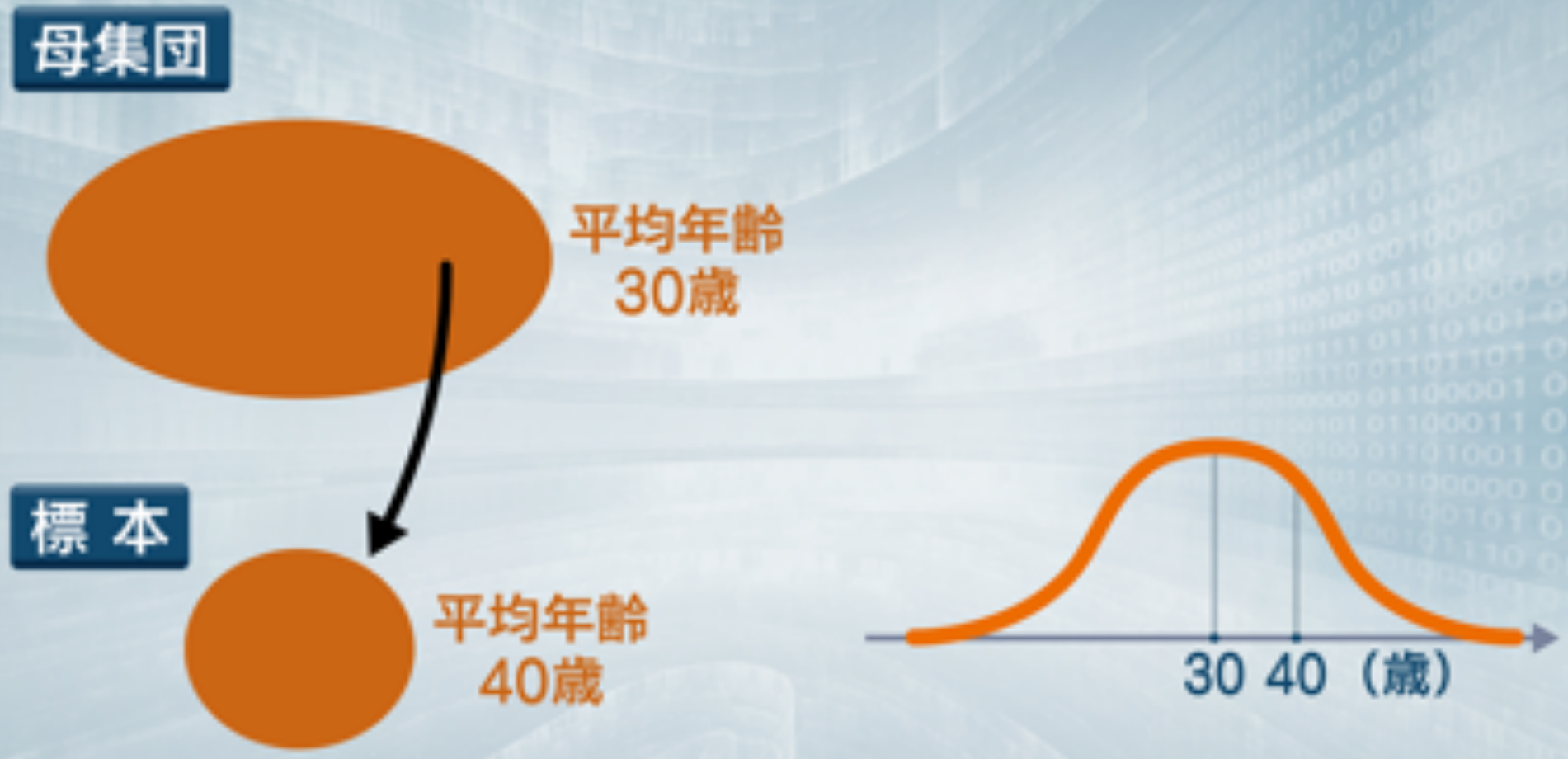
離散的確率分布、連続的確率分布
について理解する

Rを用いてカイ２乗検定を行う

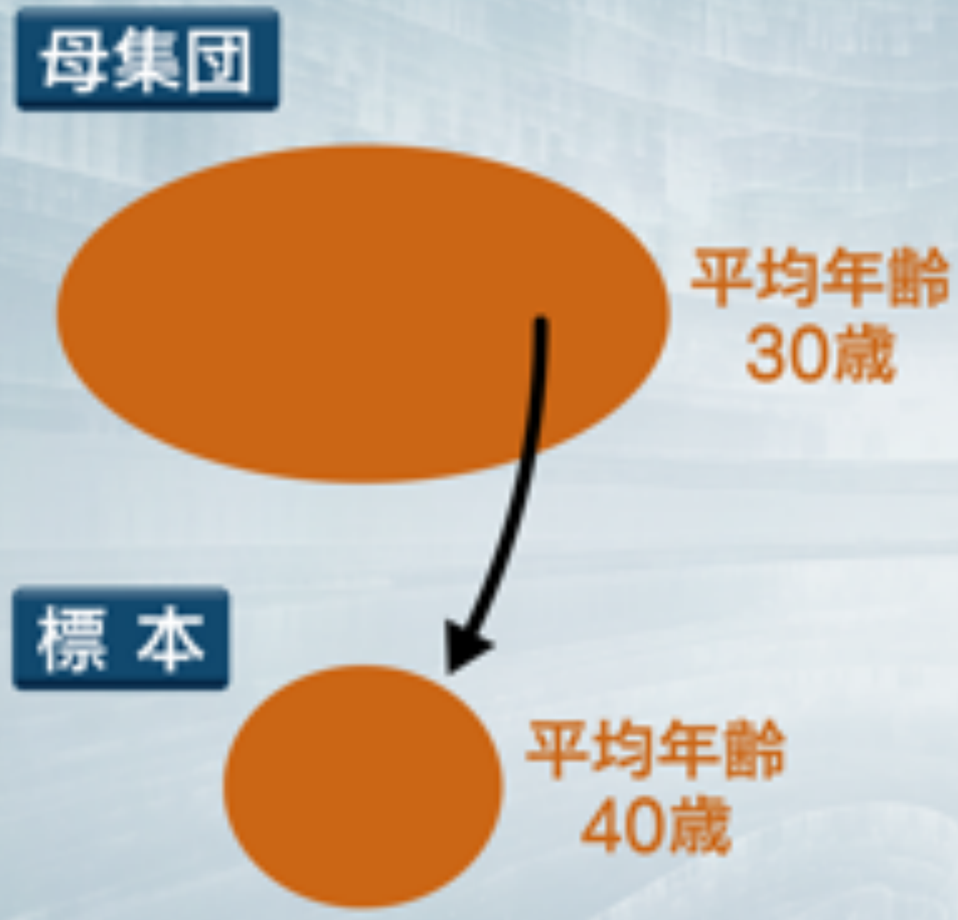
統計的仮説検定

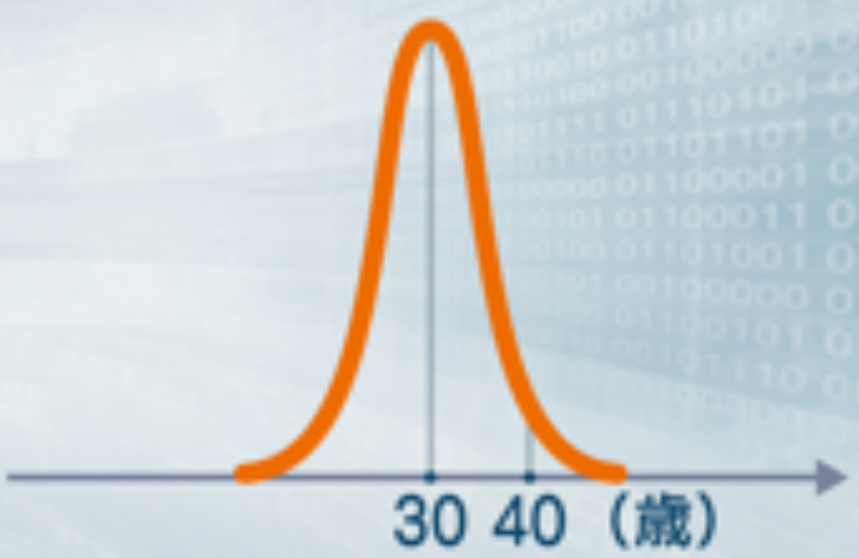


統計的仮説検定

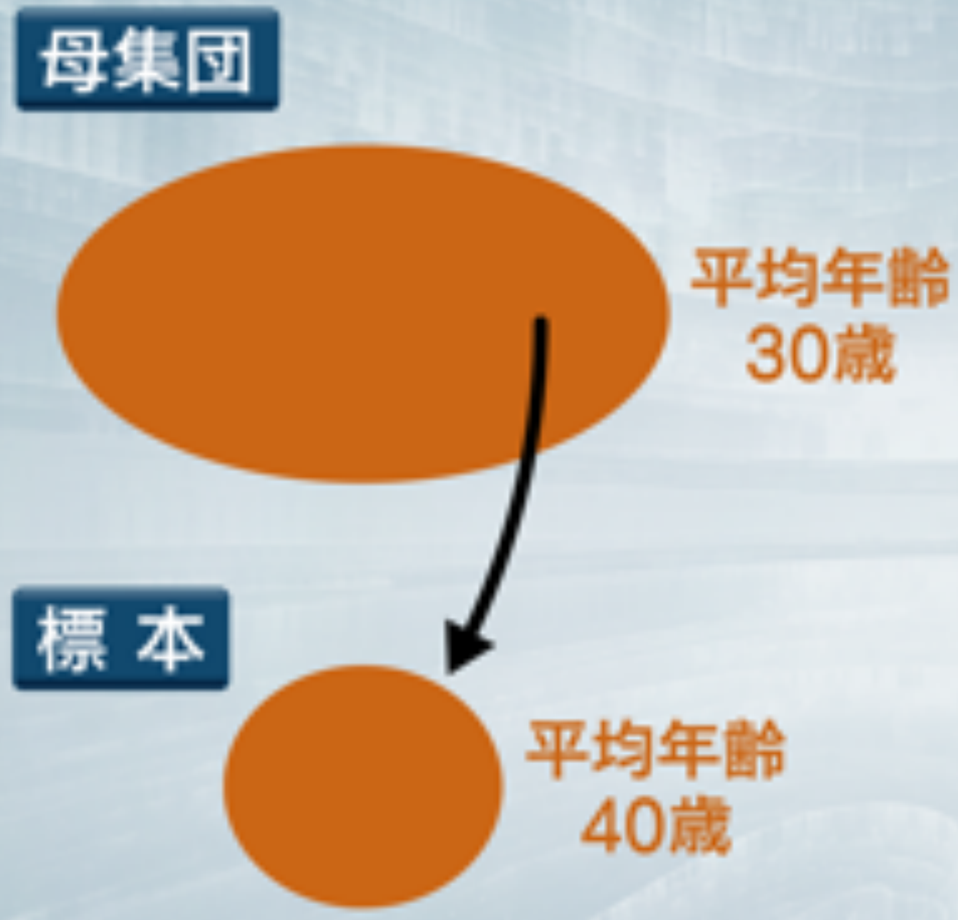


統計的仮説検定

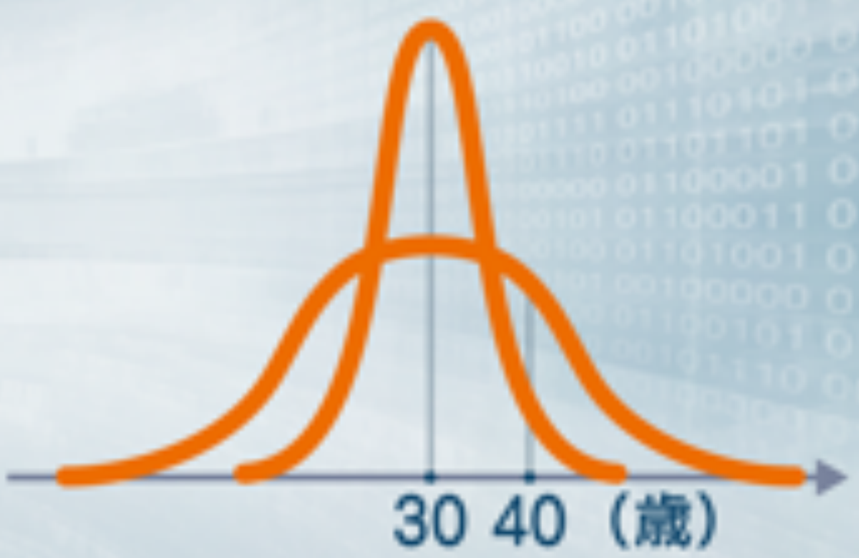




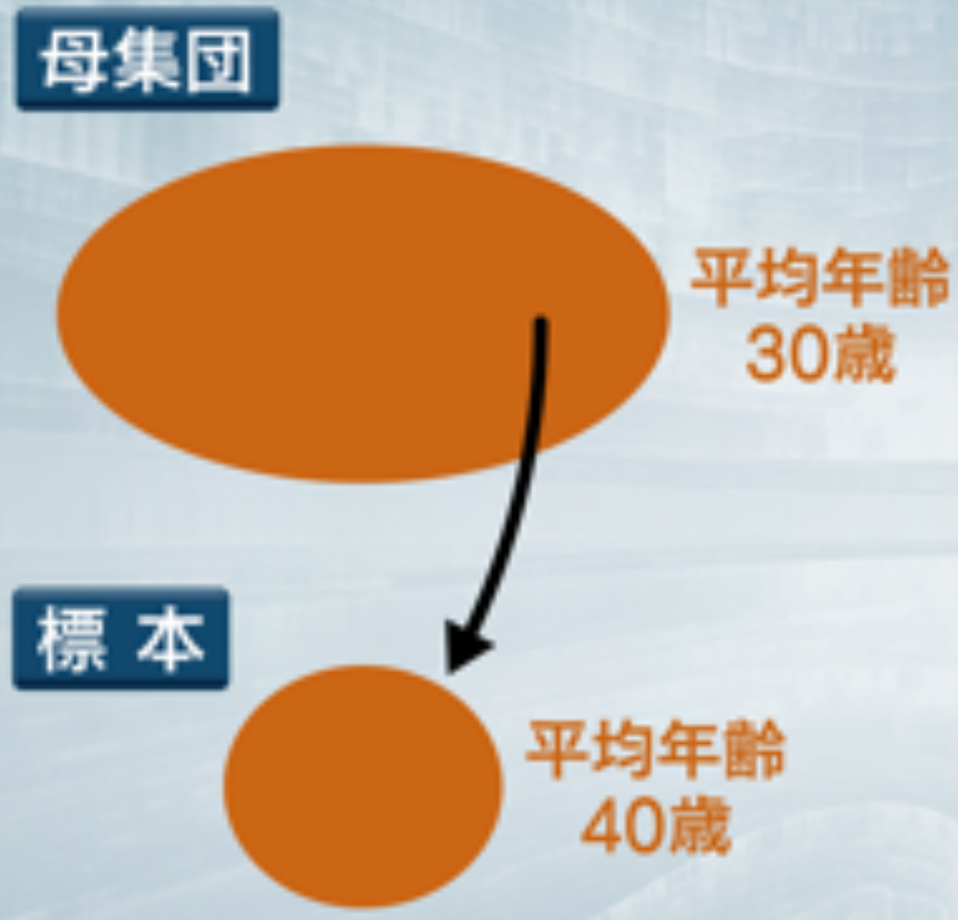
統計的仮説検定



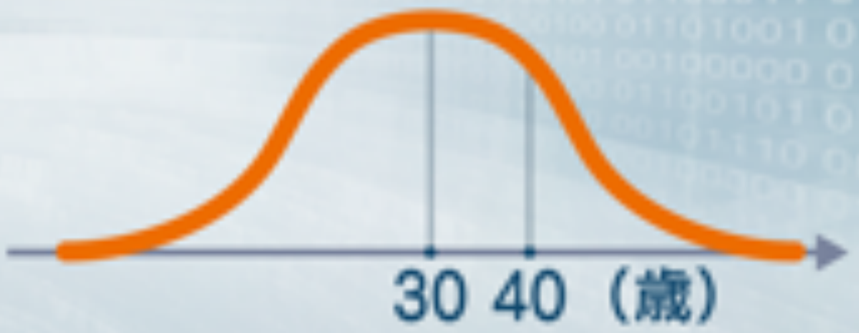
分布（ヒストグラム）が
得られるわけではない



統計的仮説検定



- 確率分布を仮定
- 確率を計算
- ある基準(有意確率)で判断



2種類の過誤

	帰無仮説が正しい	帰無仮説が正しくない
帰無仮説を棄却	第1種の過誤	正しい判定
帰無仮説を棄却しない	正しい判定	第2種の過誤

主な確率分布

離散型分布

- 離散型一様分布
- 二項分布
- 幾何分布
- 多項分布
- ポアソン分布

⋮

連続型分布

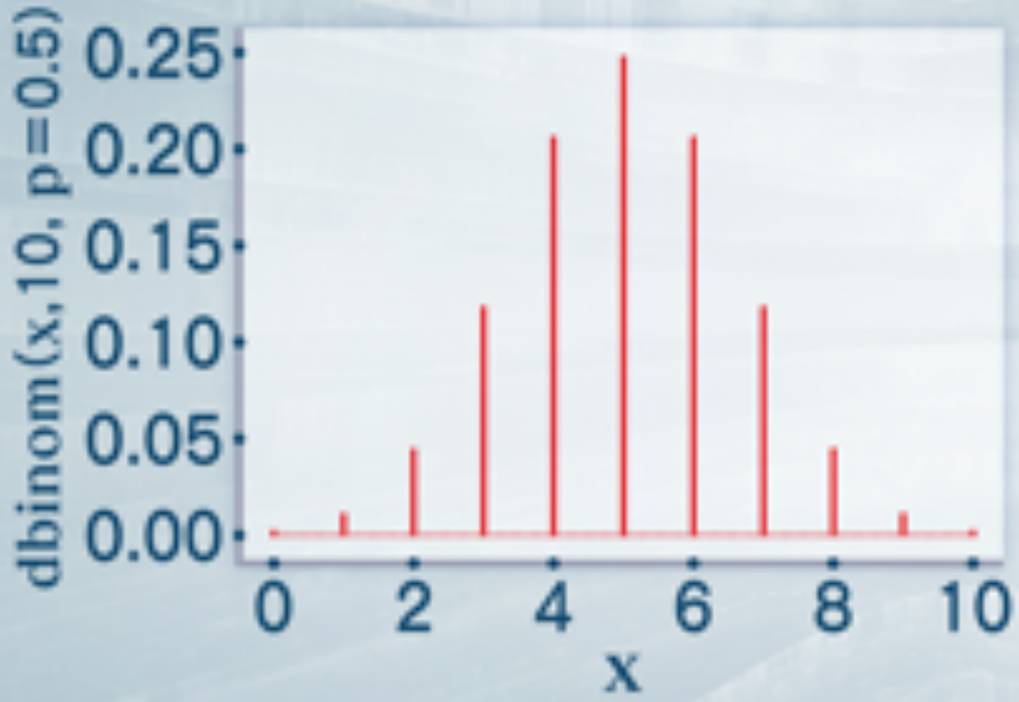
- 指数分布
- 一様分布
- 正規分布
- t 分布
- F分布
- カイ2乗分布

⋮

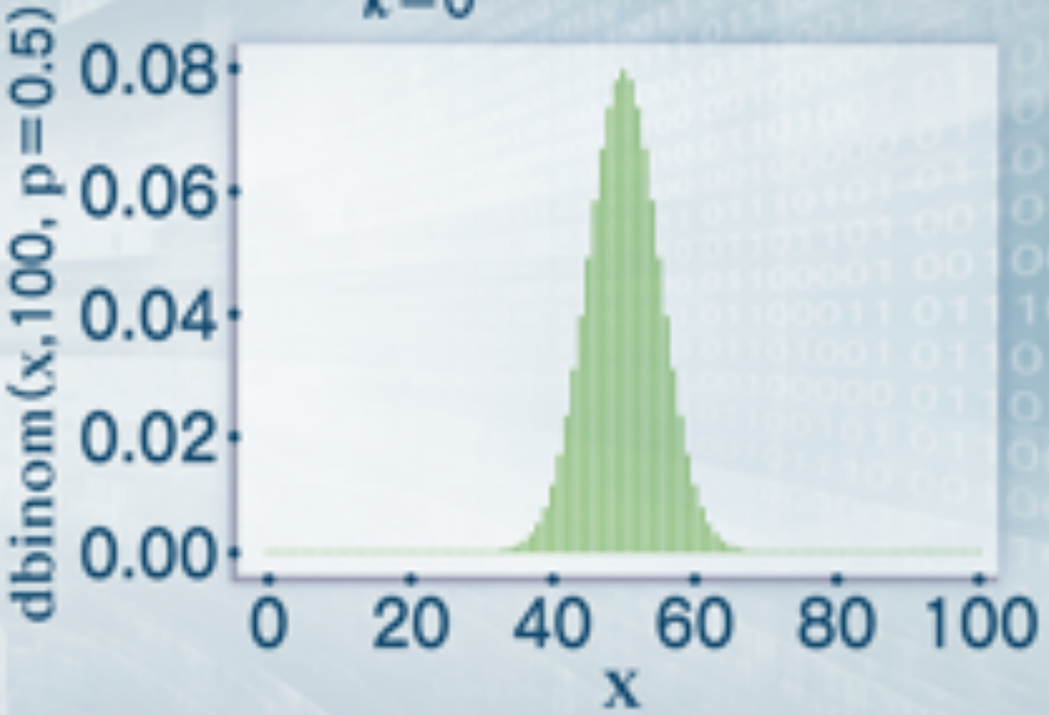
二項分布

二項分布

$$P(X=k)={}_nC_k p^k(1-p)^{n-k}$$



$$\sum_{k=0}^{n} P(X=k)=1$$

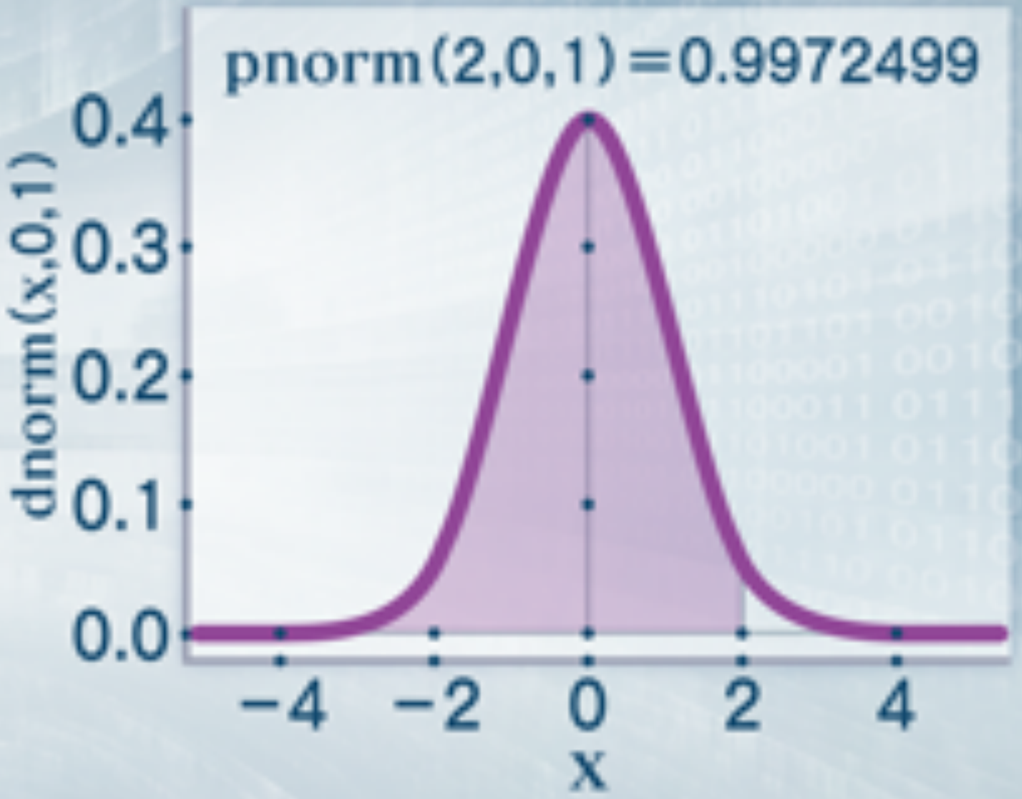


正規分布

正規分布

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

$$P(X\leq x)=\int_{-\infty}^{x}f(t)dt$$



Rと確率分布

２項分布	ポアソン分布	一様分布	正規分布	F分布	t分布	カイ２乗分布	
dbinom	dpois	dunif	dnorm	df	dt	dchisq	密度関数
pbinom	ppois	punif	pnorm	pf	pt	pchisq	確率
qbinom	qpois	qunif	qnorm	qf	qt	qchisq	確率点
rbinom	rpois	runif	rnorm	rf	rt	rchisq	乱数
				var.test	t.test	chisq.test	検定

カイ2乗検定のシミュレーション

```
sim.chi1<-function(ite){
  sc<-numeric(ite)
  for(i in 1:ite){
    x<-runif(200)
    y<-ifelse(x<0.4,1,ifelse(x<0.7,2,ifelse(x<0.9,3,4)))
    z1<-table(y)
    z2<-c(40,30,20,10)*2
    z3<-(z1-z2)^2/z2
    sc[i]<-sum(z3)
  }
  return(sc)
}
```

カイ2乗検定のシミュレーション

乱数の発生

```
> source("simchi.R")
> sim1 <- sim.chi1(1000)
> hist(sim1,ylim=c(0,0.5), breaks="Scott", freq=F)
```

乱数の発生

```
> curve(dchisq(x,3), add=T)
```