

베이지데이터분석 / 이재용 교수

03 강

R의 소개와 실습

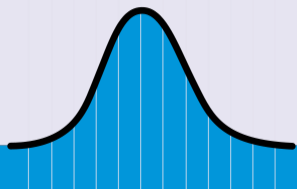




목차

➤ R, RStudio, R 마크다운 사용법

➤ 이항모형 실습

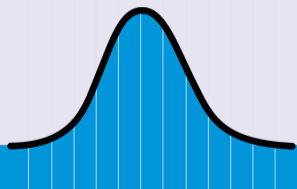




목차

➤ R, RStudio, R 마크다운 사용법

➤ 이항모형 실습



- R이란? R은 자료 분석, 계산, 시각화를 위한 통합 통계 소프트웨어다.
- (R의 역사)
 - S 언어는 벨랩의 Rick Becker, John Chambers, Allan Wilks에 의해 개발되었다.
 - S 언어는 상업용 소프트웨어인 S-Plus와 공짜 소프트웨어인 R로 발전했다.
 - R은 현재 통계학의 핵심 언어로 성장하였다.

➤ R의 장단점

- 많은 패키지. R은 base R environment 와 많은 패키지로 구성되어 있다. 많은 통계학자들은 자신이 개발한 통계 방법론을 패키지로 만들어 공개하고 있다.
- 공짜라는 장점.
- R은 유닉스, 맥, 윈도우에서 다 돌아간다.
- 단점. 속도가 느리다.

➤ CRAN 웹사이트

- (<http://cran.r-project.org>)

- RStudio 수행

- 워킹디렉토리(working directory) 만들고 세팅하기.

워킹디렉토리로 만들 폴더를 컴퓨터에 만들고,

Session > Set Working Directory > Choose Directory, 를 이용해 세팅을 한다.

- R 스크립트 파일 작성 및 수행.

File > New file > R Script, 를 선택하면 R script 파일이 만들어진다.

➤ 다음의 코드로 R 스크립트를 작성한다.

```
x <- seq(-pi, pi, len=50)
```

```
y <- x
```

```
f <- outer(x, y, function(x, y) cos(y)/(1 + x^2))
```

```
contour(x, y, f)
```

```
contour(x, y, f, nlevels=15, add=TRUE)
```

R 마크다운 소개

- 마크다운은 코드와 텍스트를 같이 쓸 수 있어 분석 결과를 보관하고 타인에게 보고하는데 편리하다.
- (세 가지 구분)
 - YAML 헤더: — 로 둘러 싸여 있다.
 - R의 청크 코드: ““{r, echo = FALSE} ... ““
 - 텍스트: 텍스트와 수식을 쓸 수 있다. 수식은 라텍스의 형식을 따른다.
- (Rstudio에서 코드의 실행)
 - 청크 코드를 실행할 때는 Cmd/Ctrl-Shift-Enter
 - 라인 하나만 실행할 때는 Cmd/Ctrl-Enter
 - 문서 전체를 실행할 때는 Knit 버튼이나 Cmd/Ctrl-Shift-K

R 마크다운 소개

```
1 ---
2 title: "R, Rstudio, R 마크다운 실습"
3 author: "이재용"
4 date:
5 output:
6   html_document :
7     toc : true
8 ---
9
10
11 ## 등고선 그림 그리기
12
13 ```{r}
14 x <- seq(-pi, pi, len=50)
15 y <- x
16 f <- outer(x, y, function(x, y) cos(y)/(1 + x^2))
17 contour(x, y, f)
18 #contour(x, y, f, nlevels=15, add=TRUE)
19 ```
20
21
```

R 마크다운 소개

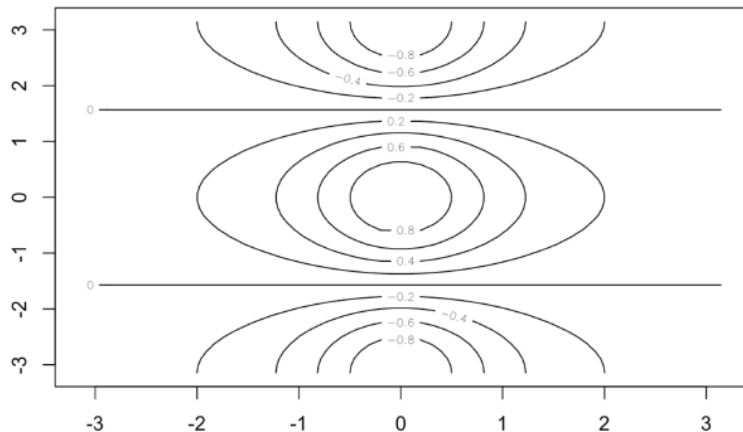
R, Rstudio, R 마크다운 실습

이재용

- 등고선 그림 그리기
- 이항모형 실습

등고선 그림 그리기

```
x <- seq(-pi, pi, len=50)
y <- x
f <- outer(x, y, function(x, y) cos(y)/(1 + x^2))
contour(x, y, f)
```

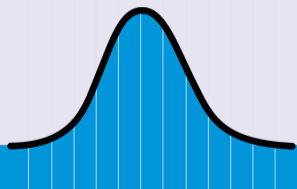




목차

➤ R, RStudio, R 마크다운 사용법

➤ 이항모형 실습



문제

1. 모형 : $X \sim \text{Bin}(n, \theta), \theta \in (0, 1)$
2. 사전분포 : $\theta \sim \text{Beta}(\alpha, \beta), \alpha, \beta > 0$
3. 사후분포 : $\theta|x \sim \text{Beta}(\alpha + x, \beta + n - x)$

위의 모형에서 $n = 10$ 일 때, $X = 3$ 을 관측하였다. 이 때, 사전분포, 사후분포, 가능도를 한 그림에 그리고 레전드를 넣으시오.

이항모형 실습. 풀이.

사전분포와 모형의 파라미터와 관측치

$\alpha = 1$

$\beta = 1$

$n = 10$

$x = 3$

사전분포, 사후분포, 가능도를 벡터로 계산

$\theta = \text{seq}(\text{from}=0, \text{to}=1, \text{length}=100)$

$\text{prior.den} = \text{dbeta}(\theta, \alpha, \beta)$

$\text{post.den} = \text{dbeta}(\theta, \alpha + x, \beta + n - x)$

$\text{likelihood} = \text{dbinom}(x, \text{size}=n, \text{prob}=\theta)$

$\text{likelihood} = 100 * \text{likelihood} / \text{sum}(\text{likelihood})$

문제

1) 모형 : $X \sim \text{Bin}(n, \theta), \theta \in (0, 1)$

2) 사전분포 : $\theta \sim \text{Beta}(\alpha, \beta), \alpha, \beta > 0$

3) 사후분포 : $\theta | x \sim \text{Beta}(\alpha + x, \beta + n - x)$

위의 모형에서 $n = 10$ 일 때, $X = 3$ 을 관측하였다. 이 때, 사전분포, 사후분포, 가능도를 한 그림에 그리고 레전드를 넣으시오.

세 개의 그림을 따로 따로 그려본다

```
par(mfrow=c(2,2))
```

```
plot(theta, prior.den, type="l", ylab="prior density")
```

```
plot(theta, likelihood, type="l", ylab="likelihood")
```

```
plot(theta, post.den, type="l", ylab="posterior density")
```

```
par(mfrow=c(1,1))
```

세 개의 그림을 한 윈도우에 그린다

```
ymax = max(prior.den, likelihood, post.den)
```

```
plot(theta, post.den, type="l", ylab="density", ylim=c(0,ymax))
```

```
lines(theta, prior.den, col="magenta")
```

```
lines(theta, likelihood, col="cyan")
```

```
legend(0.6, 2.8, c("posterior", "prior", "likelihood"), lty=c(1,1,1),
```

```
lwd=c(2.5, 2.5), col=c("black", "magenta", "cyan"))
```

점추정 구간추정 실습

- 1) 위에 주어진 모형에 대하여, 사후분포의 평균, 중앙값, 최빈값, 표준편차, 2.5%와 97.5% 백분위수를 구하시오.
- 2) 사후분포의 밀도함수를 그리고, 위의 값들을 사후분포 밀도함수 위에 그리시오.

1번 문제.

```
post.summary = numeric(6)
```

```
names(post.summary) = c("post mean", "post median", "MAP", "post sd", "2.5%", "97.5%")
```

```
post.summary["post mean"] = (alpha+x)/(alpha + beta + n)
```

```
post.summary["post median"] = qbeta(0.5, alpha+x, beta + n-x)
```

```
post.summary["MAP"] = (alpha+x-1)/(alpha + beta + n-2)
```

```
post.summary["post sd"] = sqrt( (alpha+x)*(beta+n-x)/((alpha+beta+n)^2 *(alpha+beta+n+1)))
```

```
post.summary["2.5%"] = qbeta(0.025, alpha+x, beta + n-x)
```

```
post.summary["97.5%"] = qbeta(0.975, alpha+x, beta + n-x)
```

```
post.summary
```

2번 문제.

```
plot(theta, post.den, type="l", ylab="density", ylim=c(0, ymax))
xpts1 = seq(f=post.summary["2.5%"], t=post.summary["97.5%"], len=100)
xpts = c(xpts1, xpts1[100:1])
ypts1 = dbeta(xpts1, alpha+x, beta+n-x)
ypts = c(rep(0,100), ypts1[100:1])
polygon(x = xpts, y = ypts, col = "cyan4", border = NA, fillOddEven=TRUE)
lines(x=rep(post.summary["post mean"], 2),
y=c(0,dbeta(post.summary["post mean"],alpha+x, beta+n-x)), col="magenta")
lines(x=rep(post.summary["post median"], 2),
y=c(0,dbeta(post.summary["post median"],alpha+x, beta+n-x)), col="magenta4")
lines(x=rep(post.summary["MAP"], 2),
y=c(0,dbeta(post.summary["MAP"],alpha+x, beta+n-x)), col="maroon")
legend(0.5, 2.8, c("posterior mean","posterior median", "MAP", "95% credible interval"),
lty=c(1,1,1, 1), lwd=c(2.5,2.5, 2.5, 5),
col=c("magenta","magenta4","maroon", "cyan4"))s
```


다음시간

04 강 _____

베이지스 가설 검증

