

01  
강

[ 확률의 개념과 응용 ]

# 확률의 기본 개념

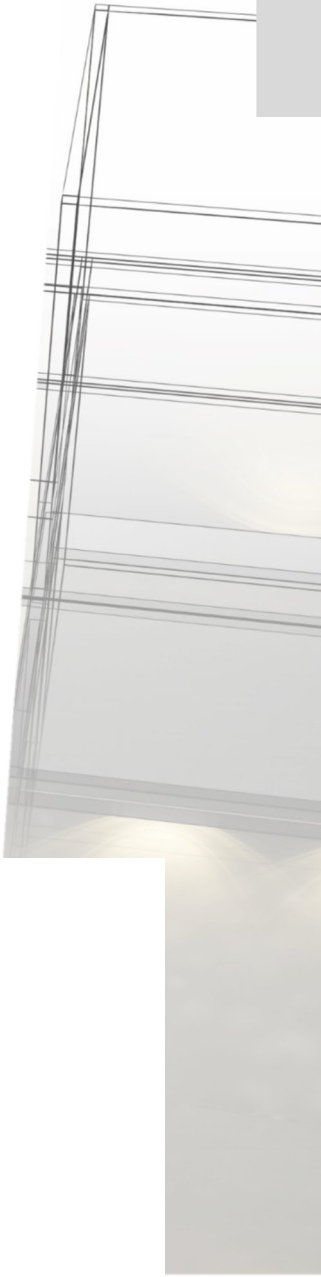


통계·데이터과학과  
이금희 교수

# 학습목표

---

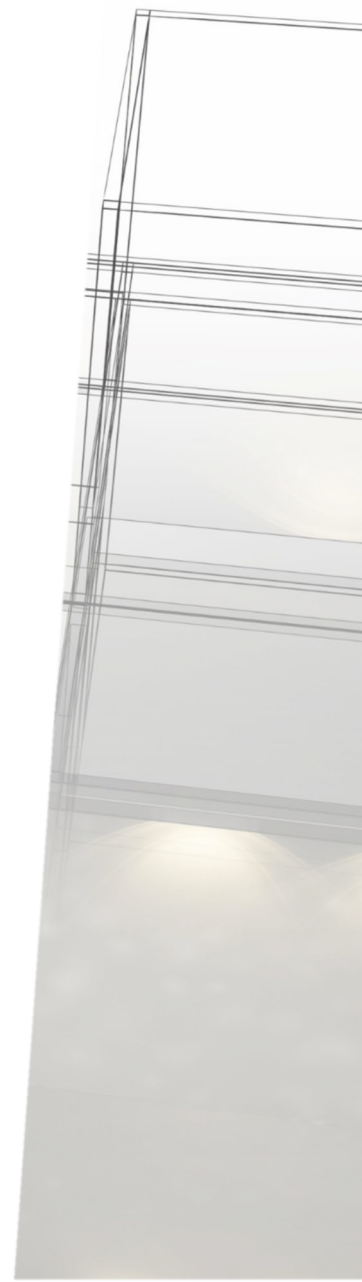
1. 확률의 개념을 이해할 수 있다.
2. 확률의 이용사례를 이해할 수 있다.
3. 확률의 역사를 이해할 수 있다.
4. 빈도론적 확률을 이해할 수 있다.



01

1강 확률의 기본 개념

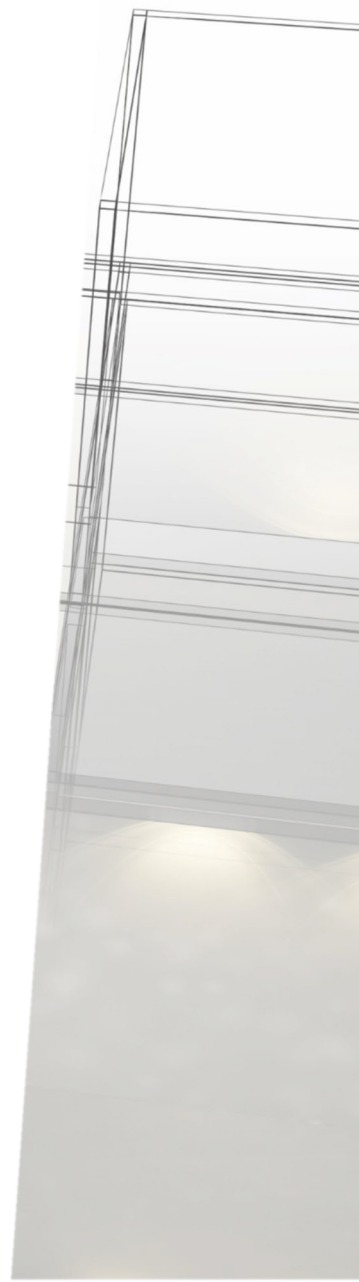
# 확률의 개념



## 운명과 우연

◆ 행운과 불행

◆ 운명, 우연?



# 동전 던지기



◆ 어떤 면이 나올까?

## 우연의 측정

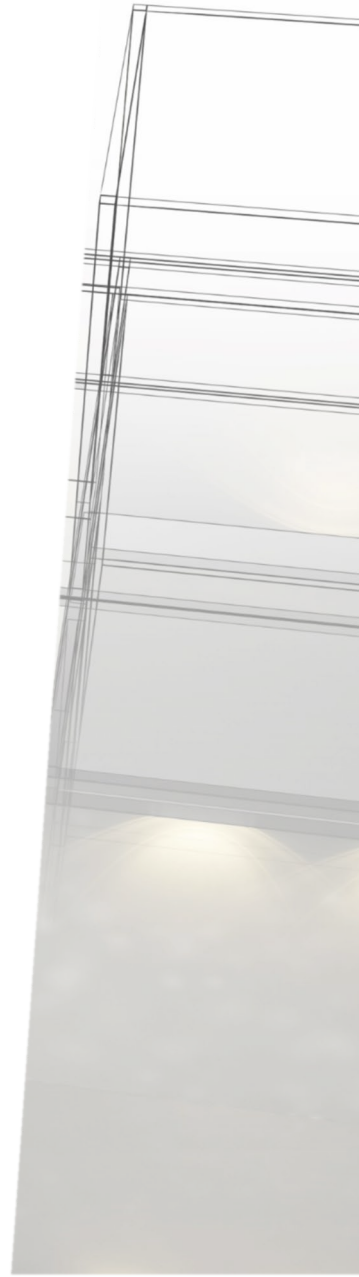
### ◆ 생활주변 비슷한 사건이 반복해서 발생. 실행

- 불확실성 : 사건이 미래에 발생할지 여부는 알 수 없음
- 그러나 사건이 장기적으로 일정 비율로 발생  
→ 확률(probability)



## 불확실한 세상

- ◆ 미래의 불확실성 측정, 이를 감수한 새로운 모험 시도

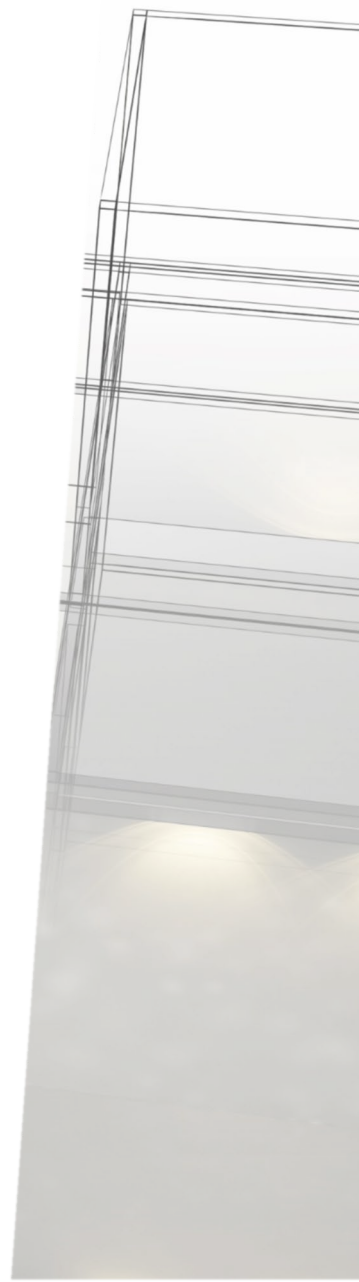


## 확률의 정의

### ◆ 확률 (probability)

어떤 사건이 일어날 가능성을 0 과 1 사이의 실수로서 표시

- 발생 가능성 높음 : 1에 가까운 수
- 발생 가능성 낮음 : 0에 가까운 수



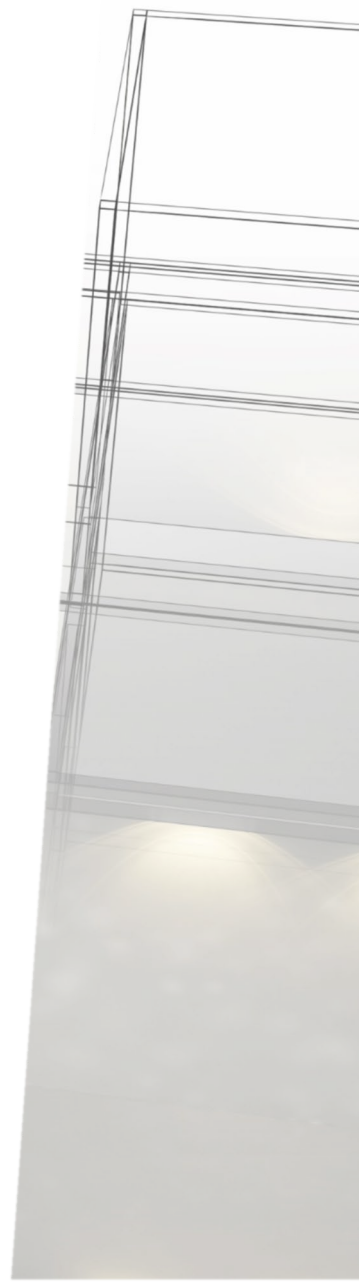


## 확률의 구분

### ◆ 확률 : 객관적 확률과 주관적 확률

- 객관적 확률 : 충분히 많은 시행을 바탕으로 구함  
(빈도론적 확률, 통계적 확률)
- 주관적 확률 : 몇 번의 경험을 바탕으로 구함  
(믿음의 정도)

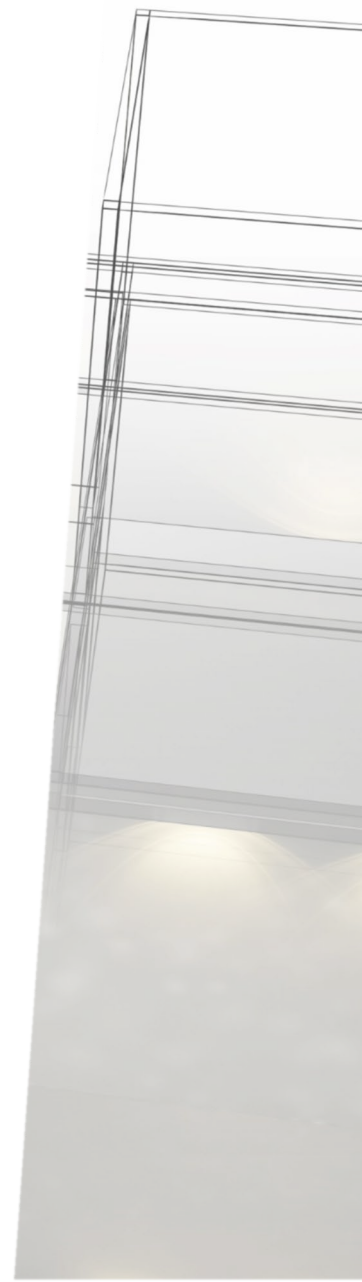
→ 시행횟수가 많아지면 두 확률은 같아짐



# 02

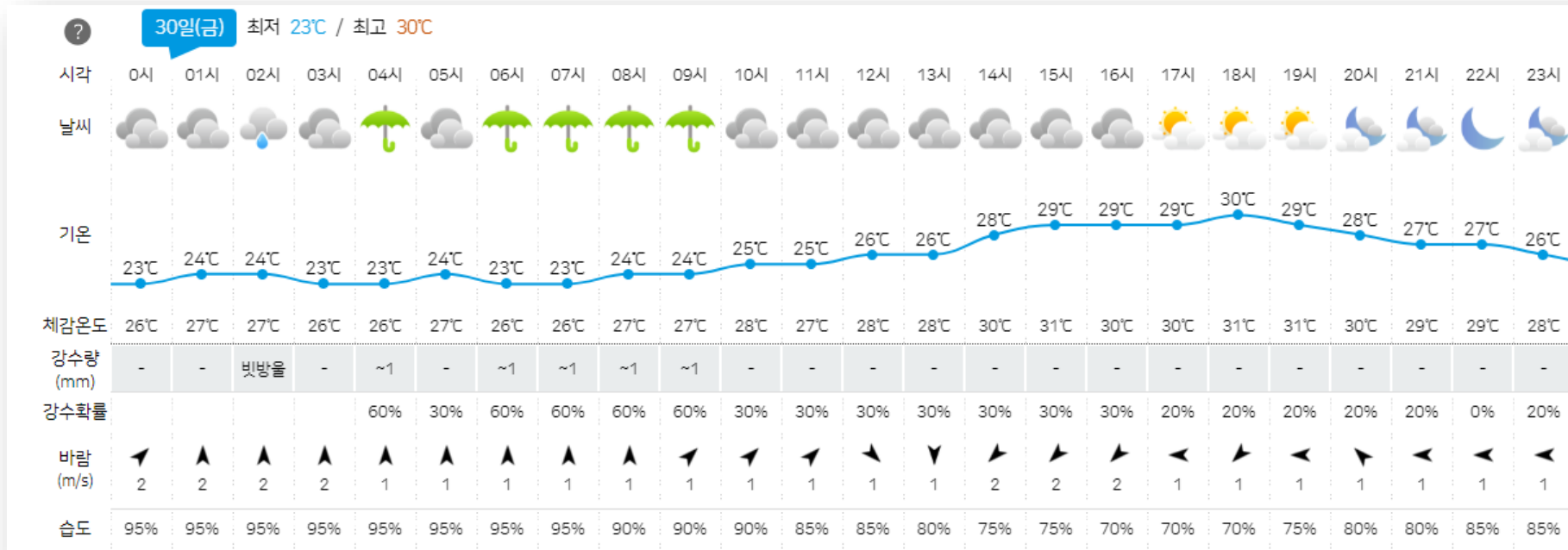
1강 확률의 기본 개념

## 확률의 이용



# 강수 확률

## ◆ 기상청 일기 예보 : 강수확률



# 복권

## ◆ 로또(Lotto) 복권

: 45개의 숫자 중에서 자신이 원하는 6개의 복권번호를  
선택하여 번호를 맞추는 복권



# 복권

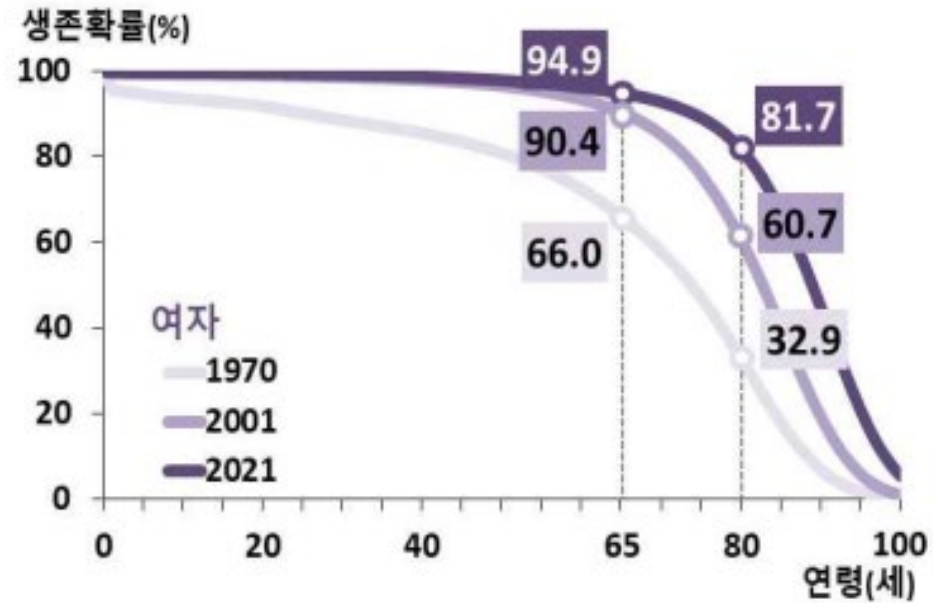
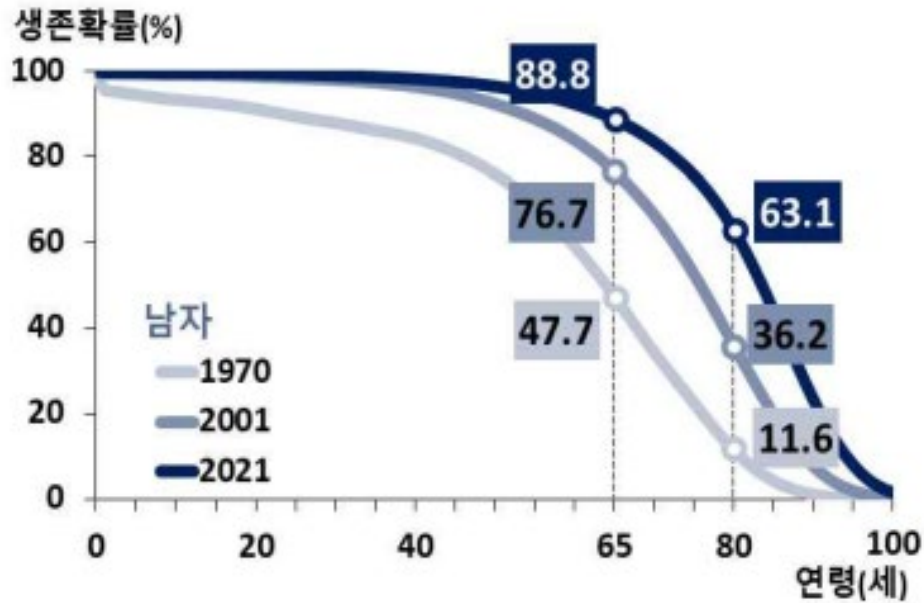
◆ 로또 1등에 당첨될 확률

$${}_{45}C_6 = \frac{45!}{6!39!} = 8,145,060$$

$$\text{1등 당첨확률} = \frac{1}{8,145,060}$$

# 생존확률

## ◆ 2021년 통계청 조사결과



# 생존확률

## ◆ 2021년 통계청 조사결과

| 구분       |    | 남자 생존 도달연령 |      |      |      |     |
|----------|----|------------|------|------|------|-----|
|          |    | 30         | 60   | 80   | 90   | 100 |
| 현재<br>연령 | 0  | 99.0       | 92.4 | 63.1 | 23.3 | 1.4 |
|          | 30 |            | 93.3 | 63.7 | 23.5 | 1.4 |
|          | 60 |            |      | 68.3 | 25.2 | 1.5 |
|          | 80 |            |      |      | 36.9 | 2.2 |
|          | 90 |            |      |      |      | 6.0 |

# 언어모형

### ◆ ChatGPT

김희

언어모형에서 확률은 어떻게 쓰이는지요?



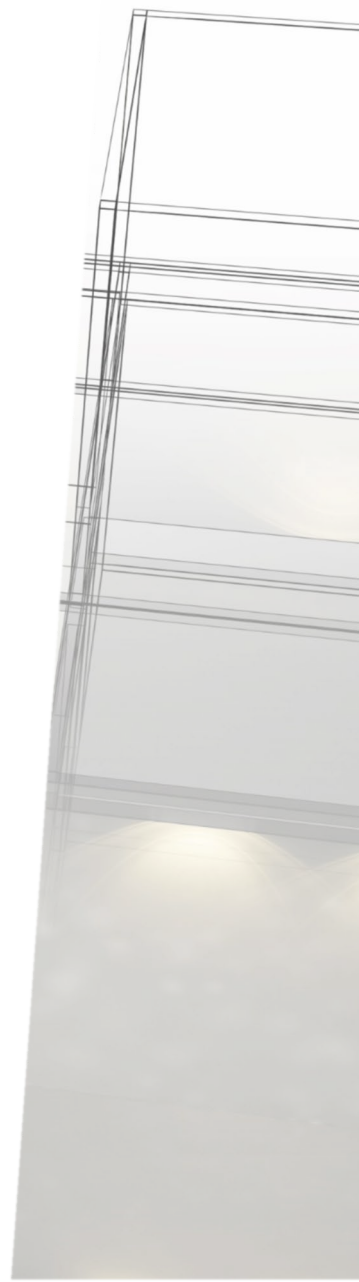
언어 모형에서 확률은 다양한 방법으로 활용됩니다. 주로 언어 생성, 문장 평가, 자연어 이해 등 다양한 자연어 처리(Natural Language Processing, NLP) 작업에 적용됩니다. 몇 가지 예를 살펴보겠습니다.

1. 언어 생성: 확률 모형은 언어 생성에 널리 사용됩니다. 주어진 문맥에서 다음 단어를 예측하거나, 문장을 자연스럽게 생성하기 위해 확률 분포를 모델링합니다. 예를 들어, 주어진 문장의 이전 단어들을 입력으로 받아 다음 단어를 예측하는 언어 모형을 구축할 수 있습니다.
2. 문장 평가: 확률 모형은 주어진 문장의 확률을 평가하는 데 사용될 수 있습니다. 주어진 문맥에서 해당 문장이 얼마나 자연스러운지, 문법적으로 올바른지 등을 확률로 표현하여 문장의 품질을 측정합니다. 이를 통해 문장 생성 과정에서 더 좋은 문장을 선택하거나 평가할 수 있습니다.



# 기적 같은 일과 당연한 일

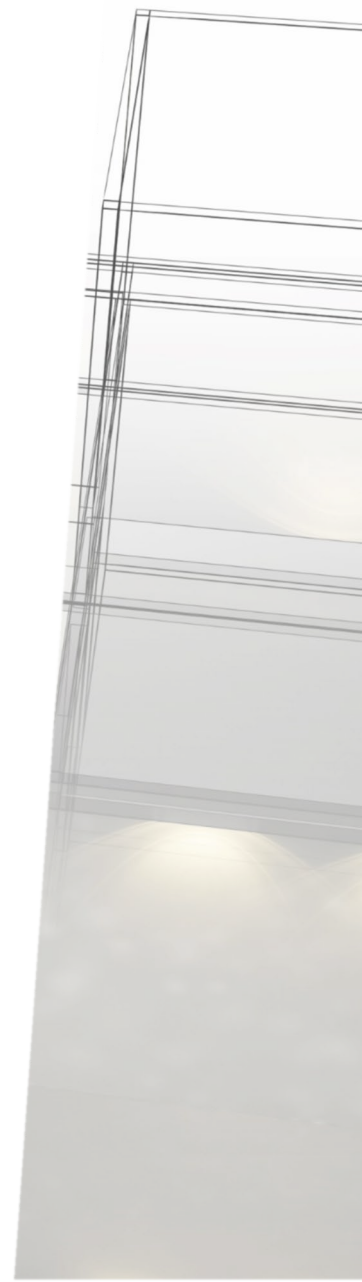
- ◆ 기적 같은 일 : 확률이 거의 0인 사건
- ◆ 당연한 일 : 확률이 거의 1인 사건



# 03

1강 확률의 기본 개념

## 확률의 역사

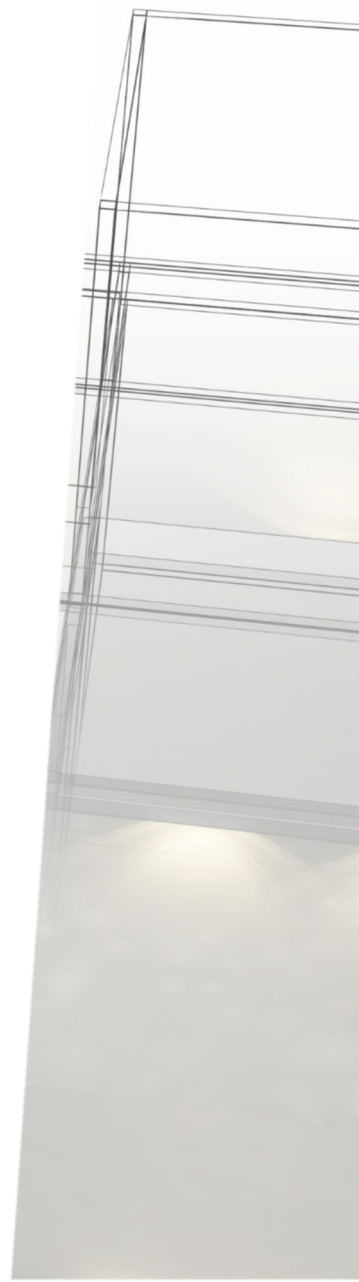


## 자연의 불확실성과 확률

### ◆ 자연의 불확실성을 신의 의지로 설명

- 과학발전으로 자연을 발견된 규칙으로 설명
- 불확실성이 줄었지만 존재

### ◆ 도박, 연금, 보험, 천문학 등에서 확률이 본격적 등장



## 확률의 발견

- ◆ 확률은 역사적으로 도박게임에서 이기기 위한 도구로 활용되면서 발전

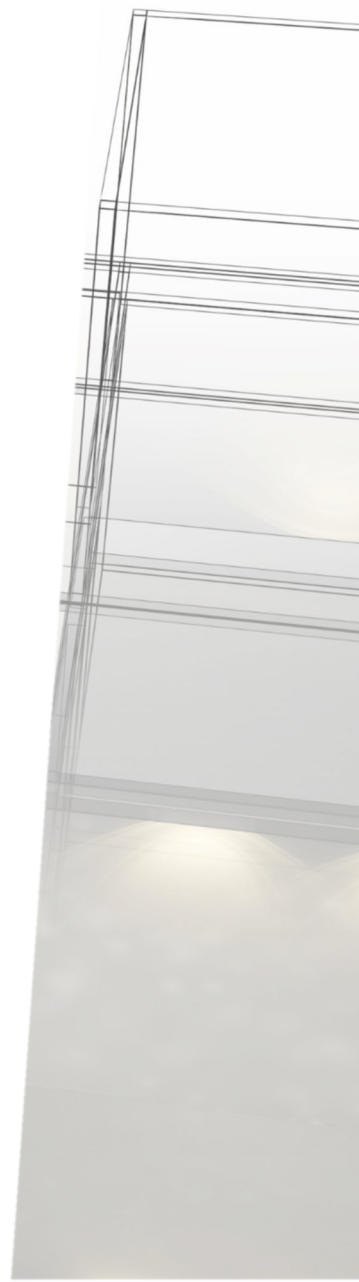
: 1654년 Chevalier de méré는 주사위 도박을 하면서 파스칼과 페르마에게 질문



## 확률의 발견

#### ◆ Chevalier de méré 의 질문

- 주사위 도박의 승산
- 도박판의 배분



## 확률의 발견

#### ◆ 2 종류의 주사위 도박

- ① 주사위 4번 던져서 6의 눈이 적어도 한번 나타남
- ② 2개의 주사위를 24번 던져서 6의 눈이 동시에 나오는 경우가 적어도 한번 나타남

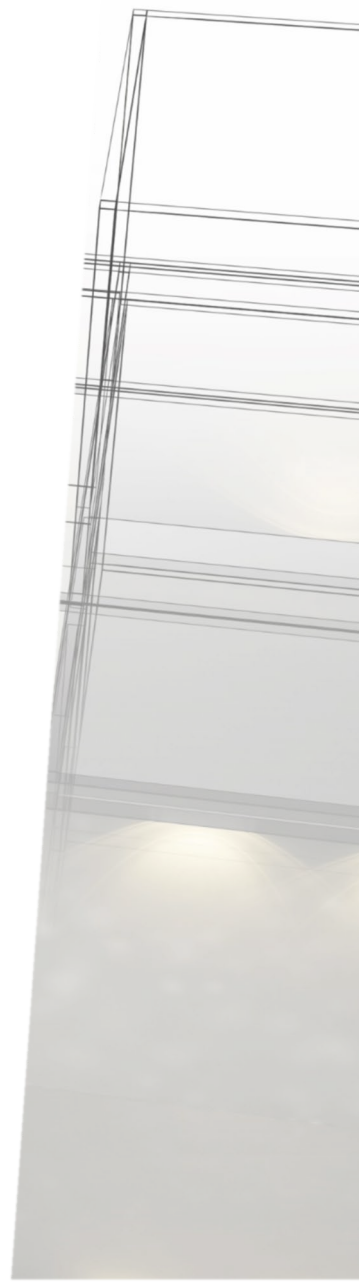


# 확률의 발견

## ◆ 2 종류의 주사위 도박

$$① 1 - \left(\frac{5}{6}\right)^4 = 0.518$$

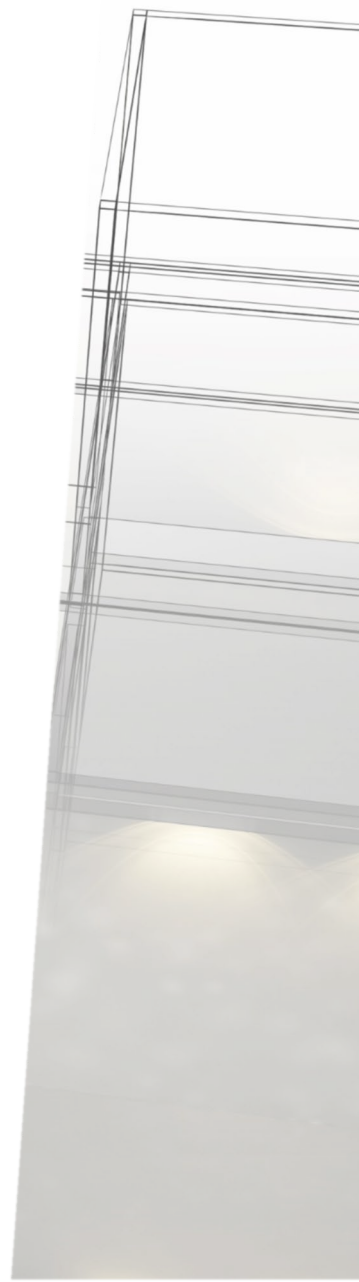
$$② 1 - \left(\frac{35}{36}\right)^{24} = 0.491$$



## 확률의 발견

### ◆ 도박판의 배분

- 32피스톨을 가진 2 명 → 3판 이기면 승리
- A가 2판, B가 1판 이겼을 때 어떻게 배분?

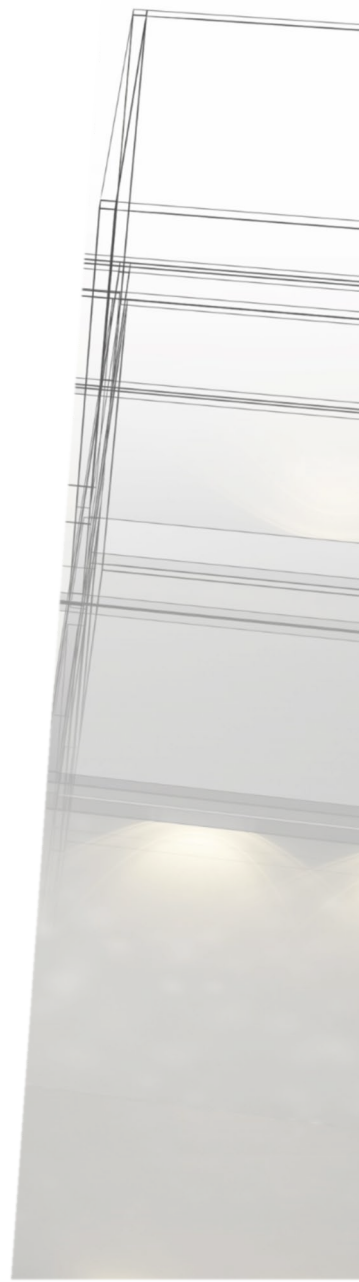




## 확률의 발견

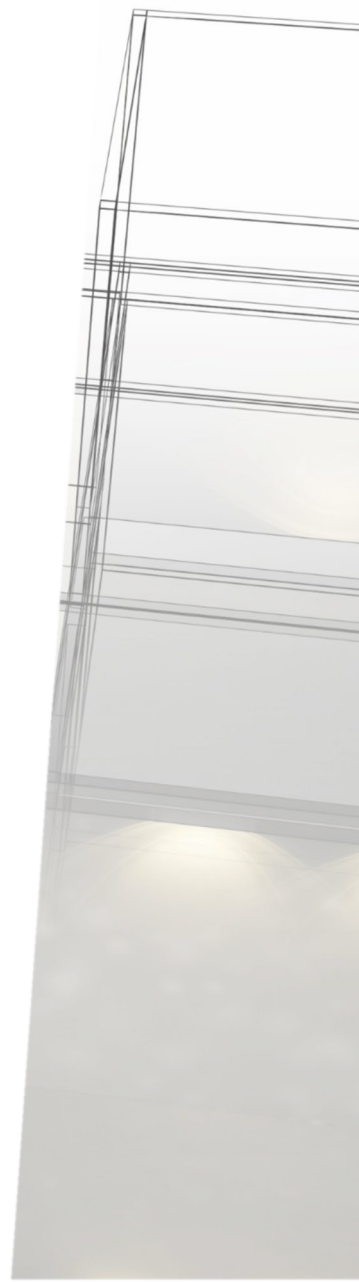
### ◆ 도박판의 배분

- B가 이길 확률 :  $0.5^2 = 0.25$
- A가 이길 확률 :  $1 - 0.25 = 0.75$



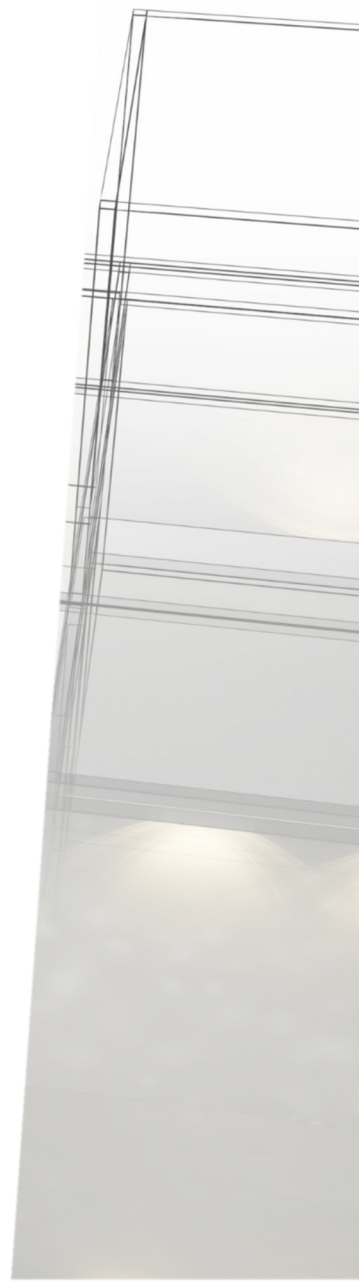
## 확률의 발전

- ◆ 1713년 : 야고프 베르누이(Bernoulli) : 대수의 법칙
- ◆ 1763년 : 베이즈(Bayes) : 베이즈 정리



## 확률의 발전

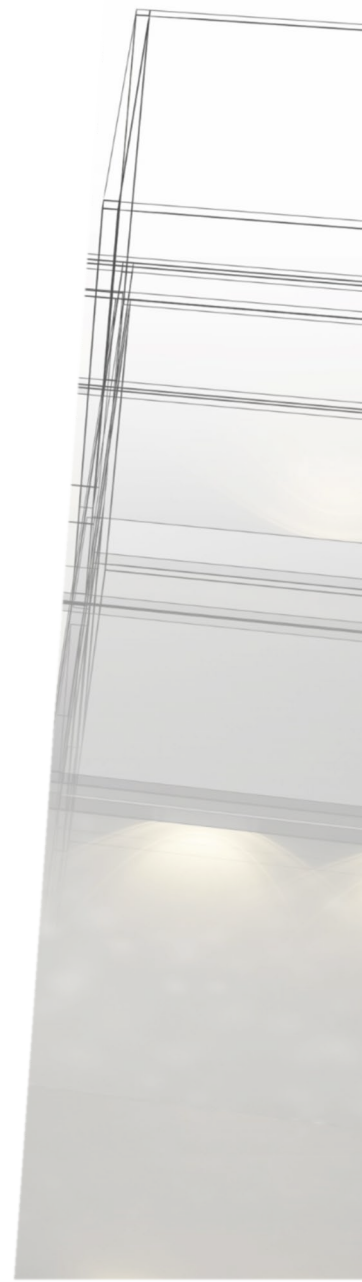
- ◆ 1783년 : 드 무아브르(de Moivre) : 이항분포의 정규근사
- ◆ 1809년 : 가우스(Gauss) : 오차의 분포 → 정규분포
- ◆ 1810년 : 라플라스(Laplace) : 중심극한정리
- ◆ 1933년 : 콜모고로프(Kolmogorov) : 공리적 확률



# 04

1강 확률의 기본 개념

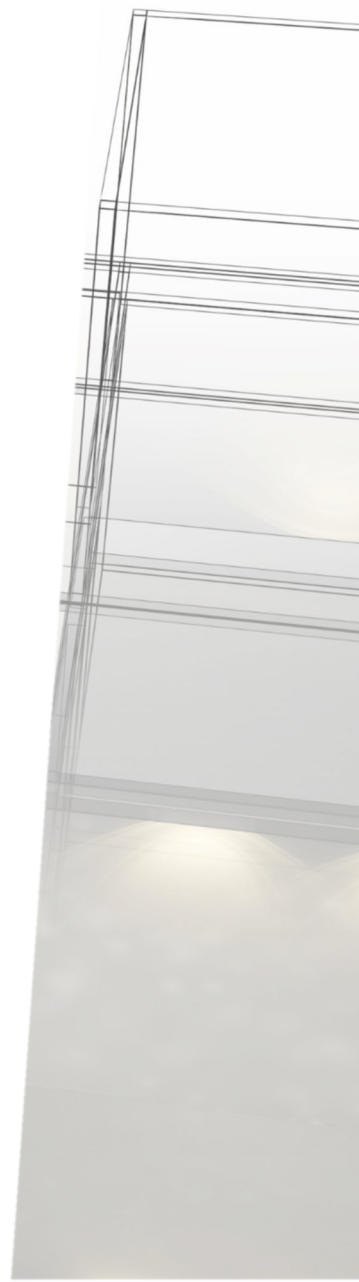
## 확률의 정의



# 확률의 구분

### ◆ 객관적 확률

- 빈도론적 확률
- 고전적 확률
- 공리적 확률



# 동전 앞면이 나올 확률

◆ 동전 앞면이 나올 확률 =  $\frac{\text{앞면이 나온 횟수}}{\text{전체 시행 횟수}}$

- ◆ 통계학자 피어슨(Karl Pearson)  
: 24,000번 던지기

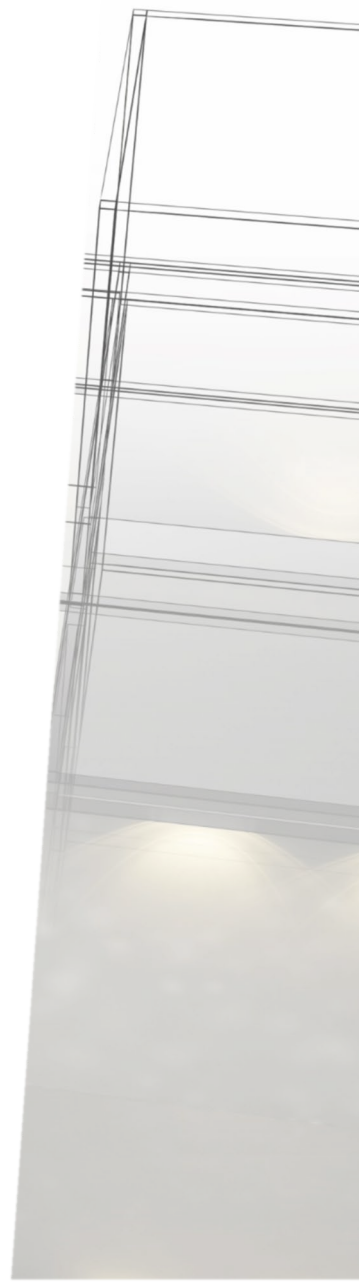


# 상대도수의 예

예

야구 타율

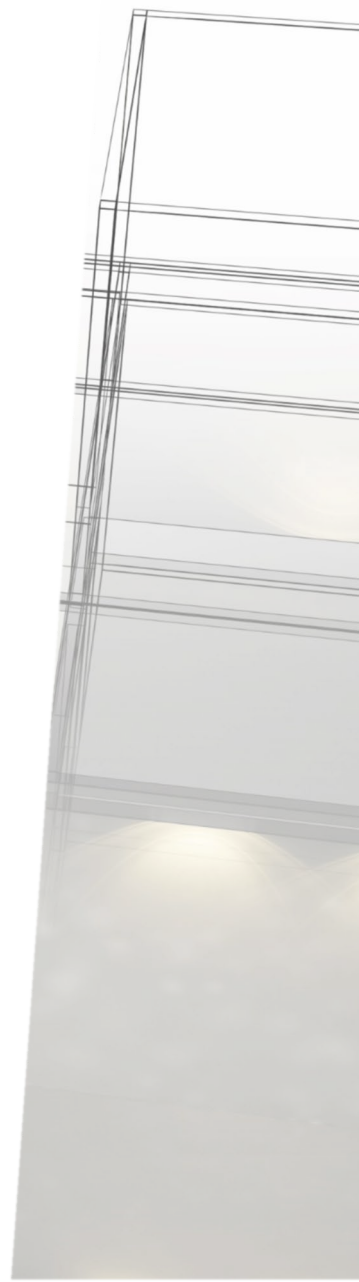
- 타석 수 250, 안타 수 103
- $\frac{103}{250} = 0.412$



# 빈도론적 확률

- ◆ 빈도론적 확률 :  $n$ 번의 시행 중 사건  $A$ 가  $n(A)$ 번 발생했을 때 사건  $A$ 가 일어날 확률  $P(A)$

$$P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}$$

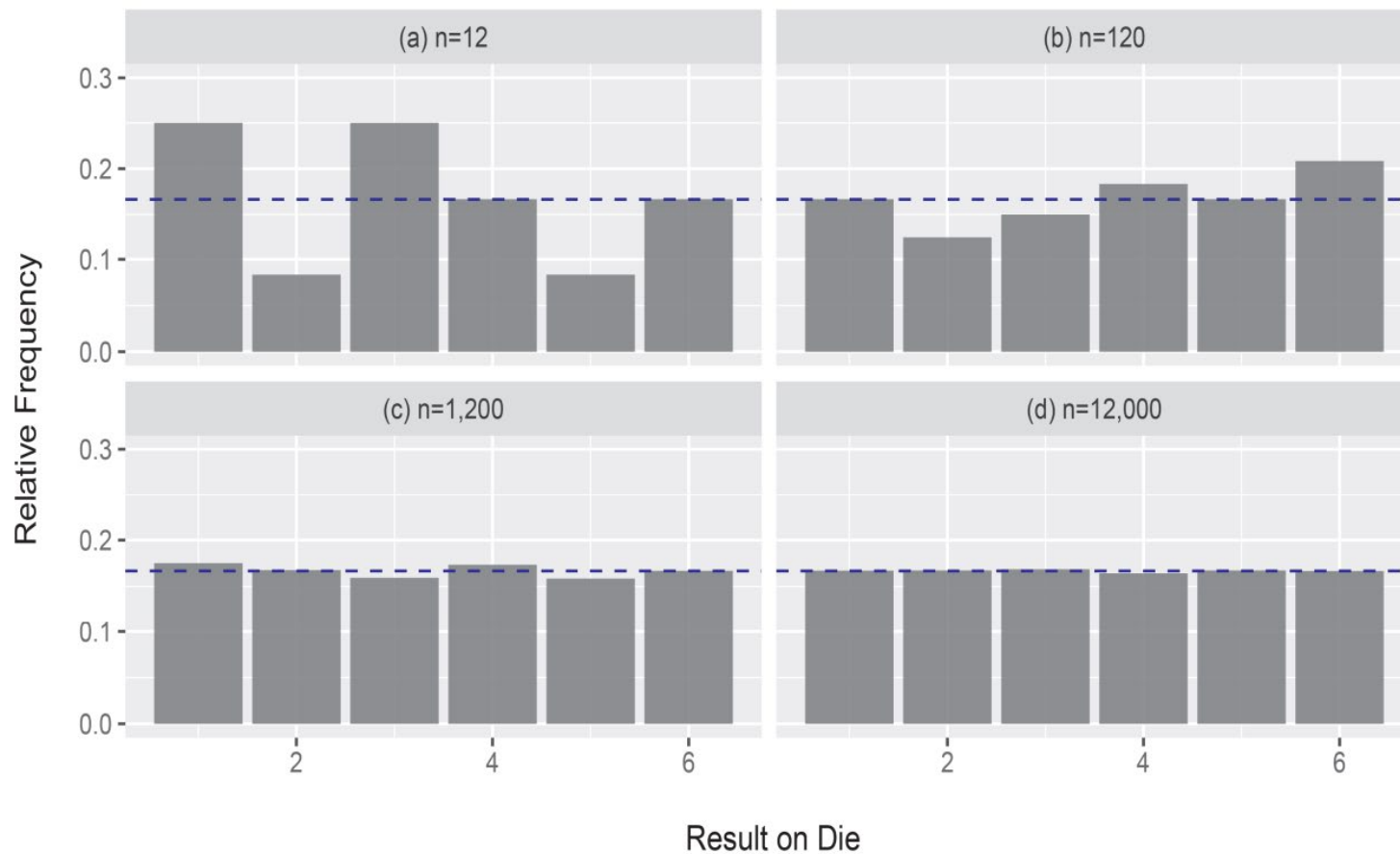




# 주사위 던지기

예

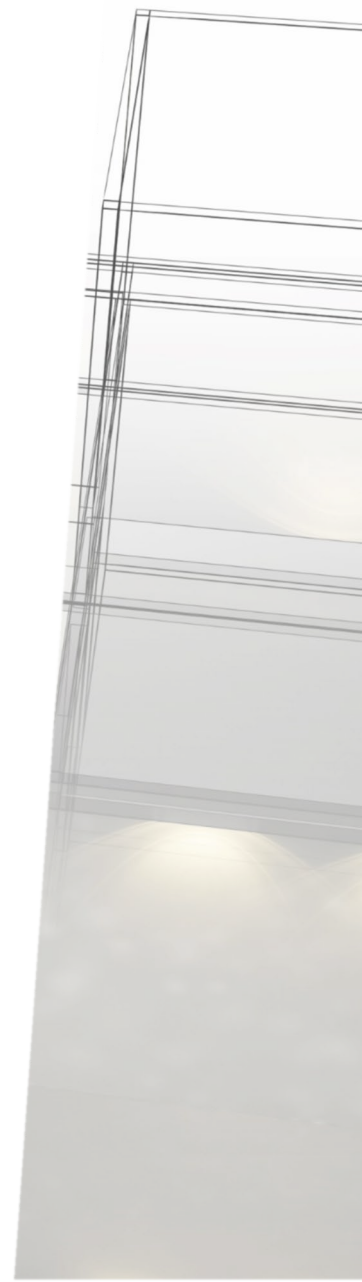
시행 횟수별 주사위 던지기



# 05

1강 확률의 기본 개념

R 실습

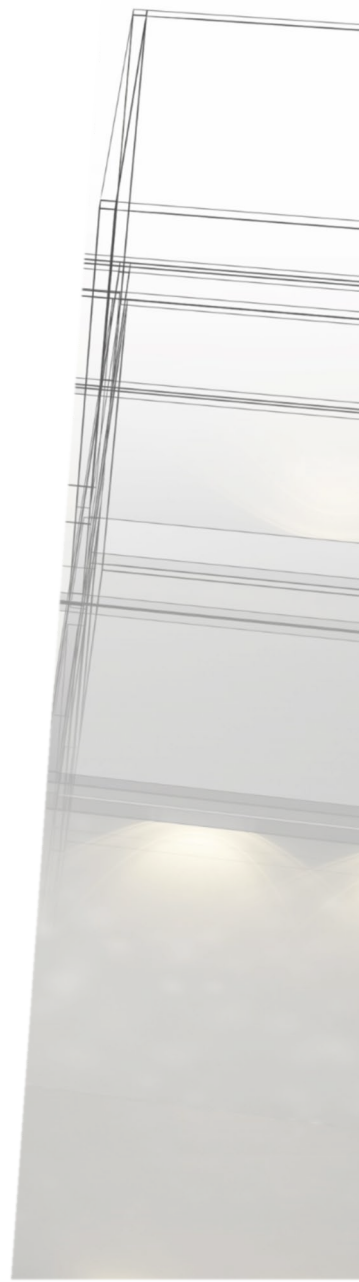


# 주사위 던지기와 상대도수

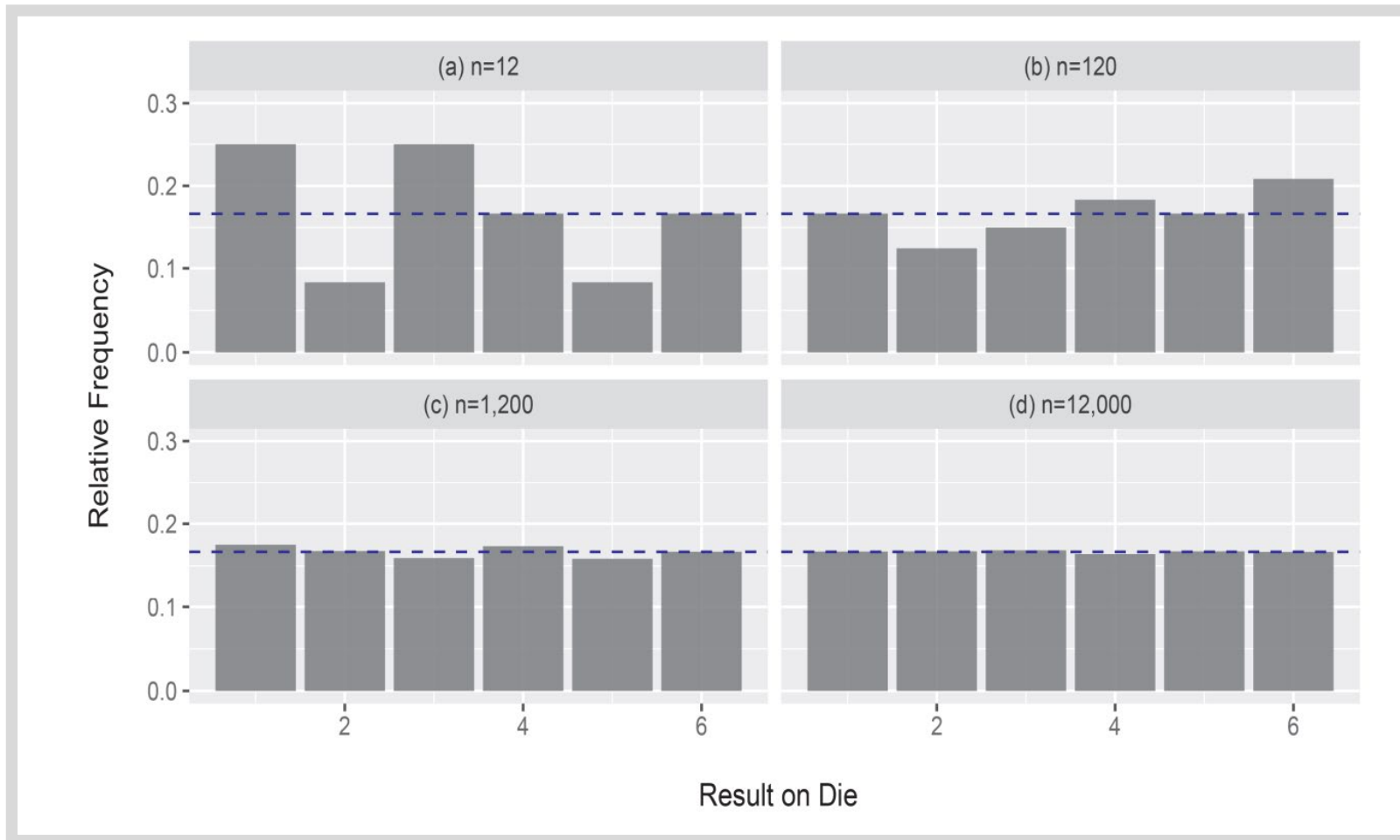
```
> library(ggplot2); library(magrittr)
> RDice=function(nn){sample(1:6, nn, replace=TRUE) %>% table()/nn}
> set.seed(1234567)
> dice_1 = RDice(12)
> dice_2 = RDice(120)
> dice_3 = RDice(1200)
> dice_4 = RDice(12000)
> # data transformation
> dice = c(as.numeric(dice_1), as.numeric(dice_2), as.numeric(dice_3),
           as.numeric(dice_4))
> nn    = c(rep("(a) n=12",6),rep("(b) n=120",6), rep("(c) n=1,200",6),
           rep("(d) n=12,000",6))
```

# 주사위 던지기과 상대도수

```
> dice_result = data.frame(nn, num, dice)
> # ggplot2
> ggplot(data = dice_result, aes(x = num, y = dice)) +
  geom_bar(stat = "identity", alpha = 0.8) +
  xlab("\n Result on Die") + ylab("Relative Frequency\n") +
  geom_hline(yintercept = 1/6, colour=2, lty=2) +
  ylim(0,0.3) + facet_wrap(~nn, ncol = 2)
```



# 주사위 던지기과 상대도수



# 학습정리

---

- 확률은 '어떤 사건이 일어날 가능성을 0과 1사이의 실수로 표시'한 것이다.
- 확률은 객관적 확률과 주관적 확률로 구분된다. 객관적 확률은 고전적·빈도론적·공리적으로 정의된다.

# 학습정리

---

- 빈도론적 확률은 특정한 사건이 발생할 상대도수의 극한값이다.
- 확률은 강수확률, 복권 당첨확률과 생명보험에 이용되는 등 생활과 밀접하다.

# 수고했습니다.

01

강

## 확률의 기본 개념

02

강

## 확률의 정의와 성질

