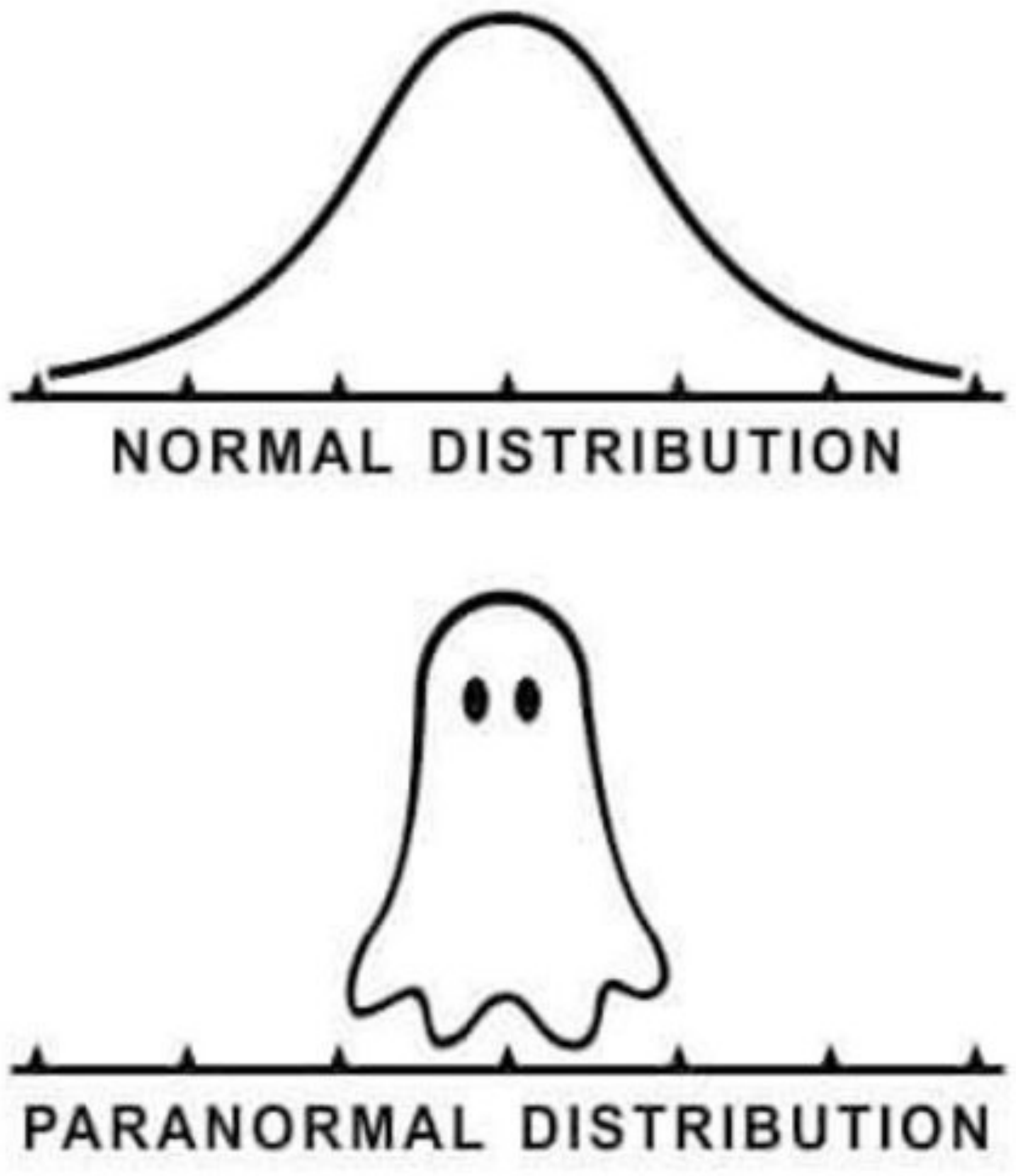
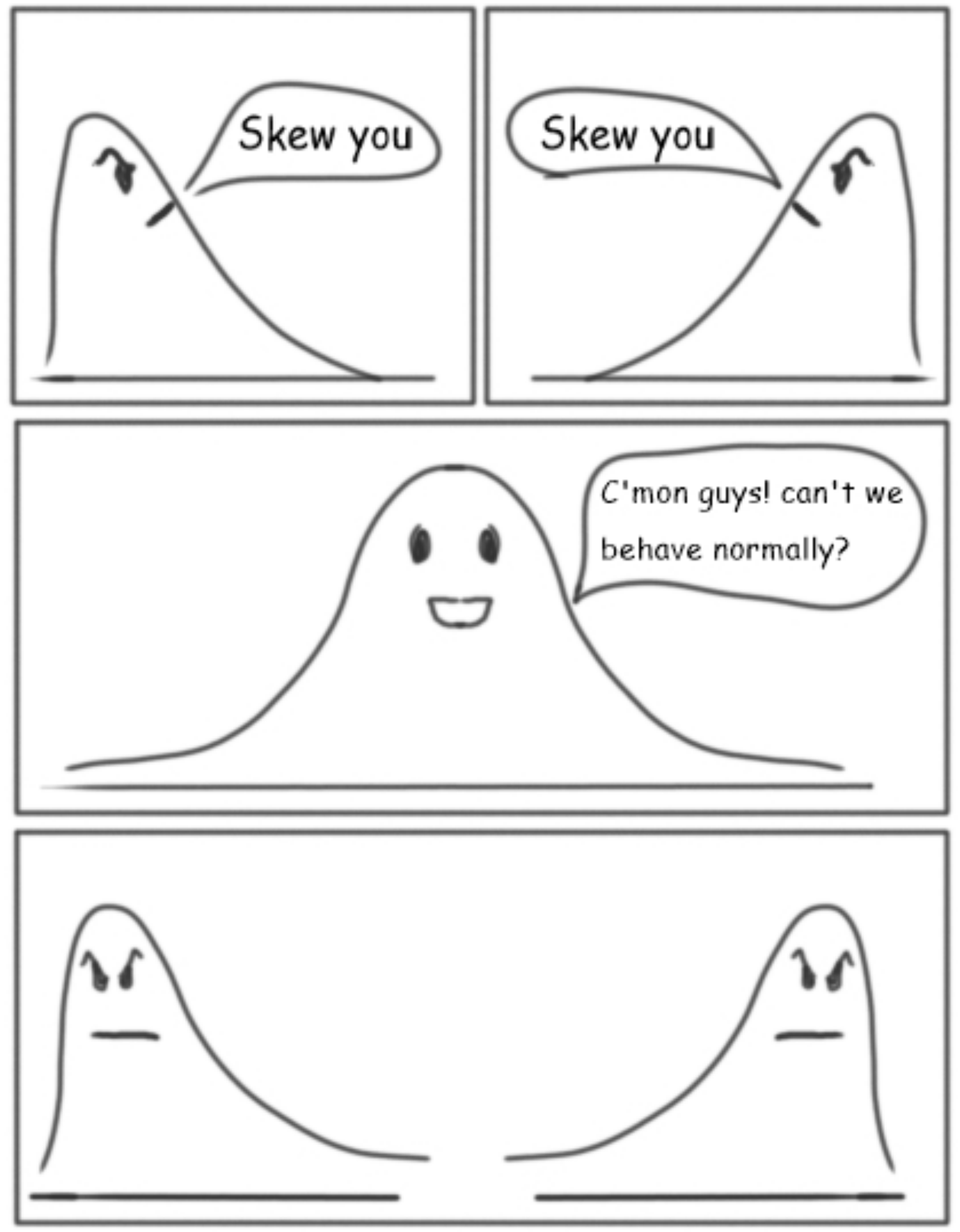




실험심리언어학도를 위한 기초 확률과 통계

3. 추정과 가설검정

2022. 07. 15.

박기효

Cornell University

KONKUK UNIVERSITY

Brain & Cognition Research Center

수업 내용 - 7월 15일(금)

- 추정
- 가설검정

➡ (여러분께) 바라는 것: 추정과 가설검정이랑 친해지기

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

추정

- 우리는 표본으로부터 얻은 정보를 활용해 모집단에 대한 정보를 추측한다고 했다.
- 이러한 과정은 통계적 추론(statistical inference)이라 하며, 이 추론 두 가지로 나누어진다.
- **추정**

  ✓ (표본)통계량을 이용해 모집단에 대한 특성치인 모수에 대해 추론
- **검정**

  ✓ 모집단에 대한 주장(=가설)을 표본을 통해 어느 정도 믿을 수 있는가를 확인하는 추론

점 추정과 구간 추정

- **점 추정(point estimation)**
  - ✓ 표본을 바탕으로 모수를 하나의 값으로 구해서 추정하는 것
  - ✓ "비문법적인 문장의 평균 읽기 시간은 800ms다."
- **구간 추정(interval estimation)**
  - ✓ 표본을 바탕으로 모수가 들어가있을만한 범위를 추정하는 것
  - ✓ "비문법적인 문장의 평균 읽기 시간은 일정한 신뢰수준(예: 95%) 하에서 750ms ~ 850ms 사이이다."

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

점 추정

- **추정량(estimator)**

  ✓ 모수를 추론하기 위해 표본을 갖고 추정하는 방법/함수(=통계량)

- **추정치(estimate)**

  ✓ 추정량에 표본 값을 집어넣어서 실제로 추정한 값

점 추정

- 추정량이 함수(혹은 방법)라고 중고등학교 때 나온 $f(x)$ 등을 생각하지 않아도 된다.
- 아래 공식 또한 함수다!

✓ $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

✓ 왜? 함수는 기본적으로 어떤 값이 입력되면 다른 값으로 출력되는 작은 기계와 같은 것이니깐!

Cornell University KONKUK UNIVERSITY Brain & Cognition Research Center

점 추정

- $f(X) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

  ✓ 비문법적인 문장 10개에 대한 읽기 시간을 점 추정해보자.

  ✓ 읽기 시간 = (750, 760, 770, 780, 790, 800, 810, 820, 830, 840)

  ✓ 추정량 = $\bar{X}$

  ✓ 추정치 = $\bar{x}$ = 읽기 시간의 평균 = 795ms

Cornell University

점 추정과 표준오차

- 한편, 추정량은 표본에 따라 달라질 수 있고, 그에 따라 추정치도 변동.
- 이러한 변동을 추정량의 표준편차로 생각해볼 수 있다.

➡ 표준오차(Standard error)

✓ 표준오차는 추정량의 정확도를 의미하며, 오차가 적을수록 정확도가 높다고 간주.

✓ 표준 오차의 공식은 기존 표준 편차와 동일!

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

구간 추정

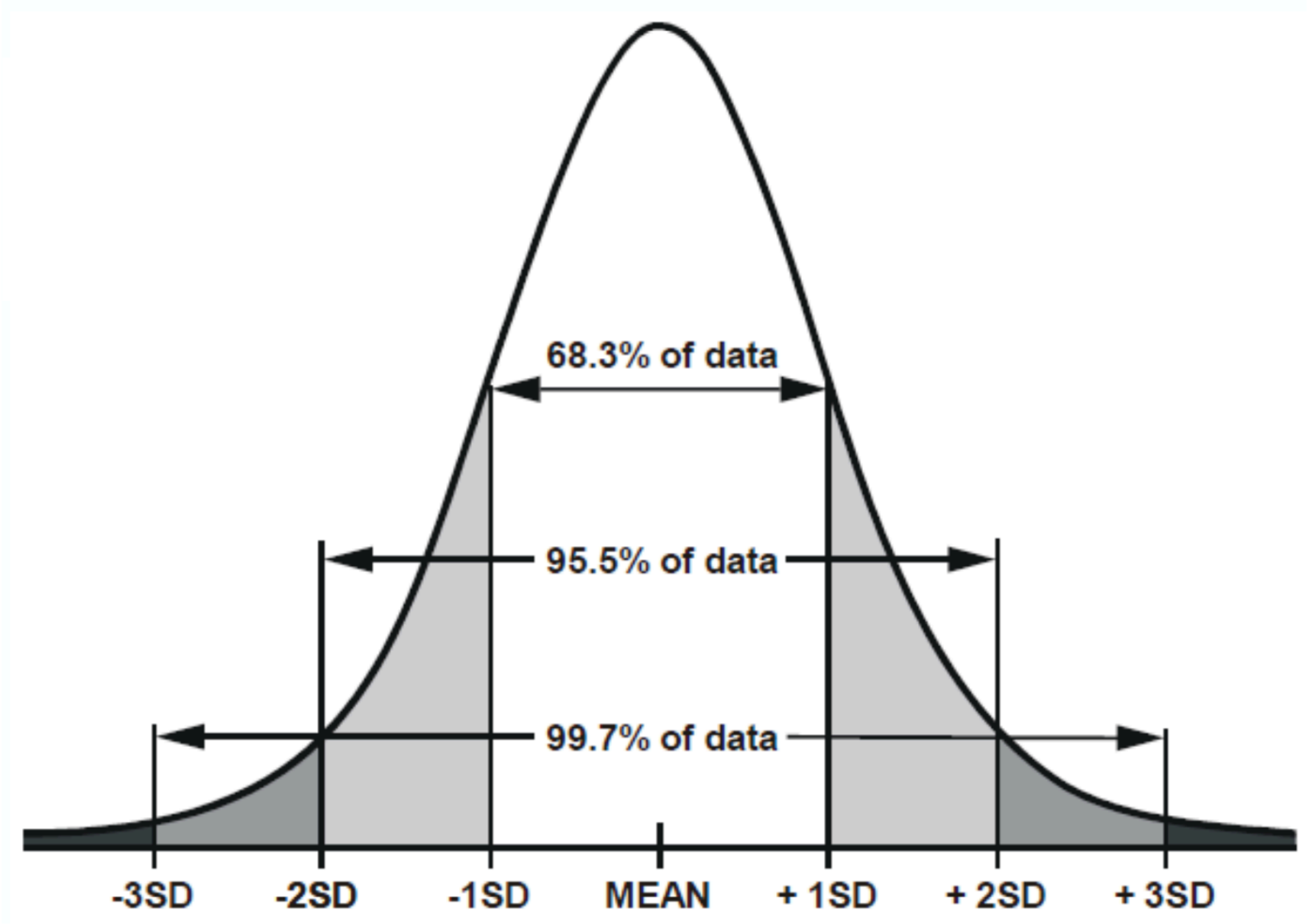
- 점 추정은 모집단의 특성치로써 모수를 추정.
- 그러나 표본이 서로 제각기, 즉, 독립이라는 점을 고려하면 정확한 하나의 값을 추정하는 것보다 구간으로 추정하는 것이 더 풍부한 얘기를 가져다줄 수 있음.
- 대표적인 구간 추정이 바로 <u>신뢰구간!</u>

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

구간 추정



신뢰 구간

- 다음과 같은 것들이 주어졌다고 해보자.

  ✓ 표본 데이터: y

  ✓ 표본 평균: $\bar{y}$

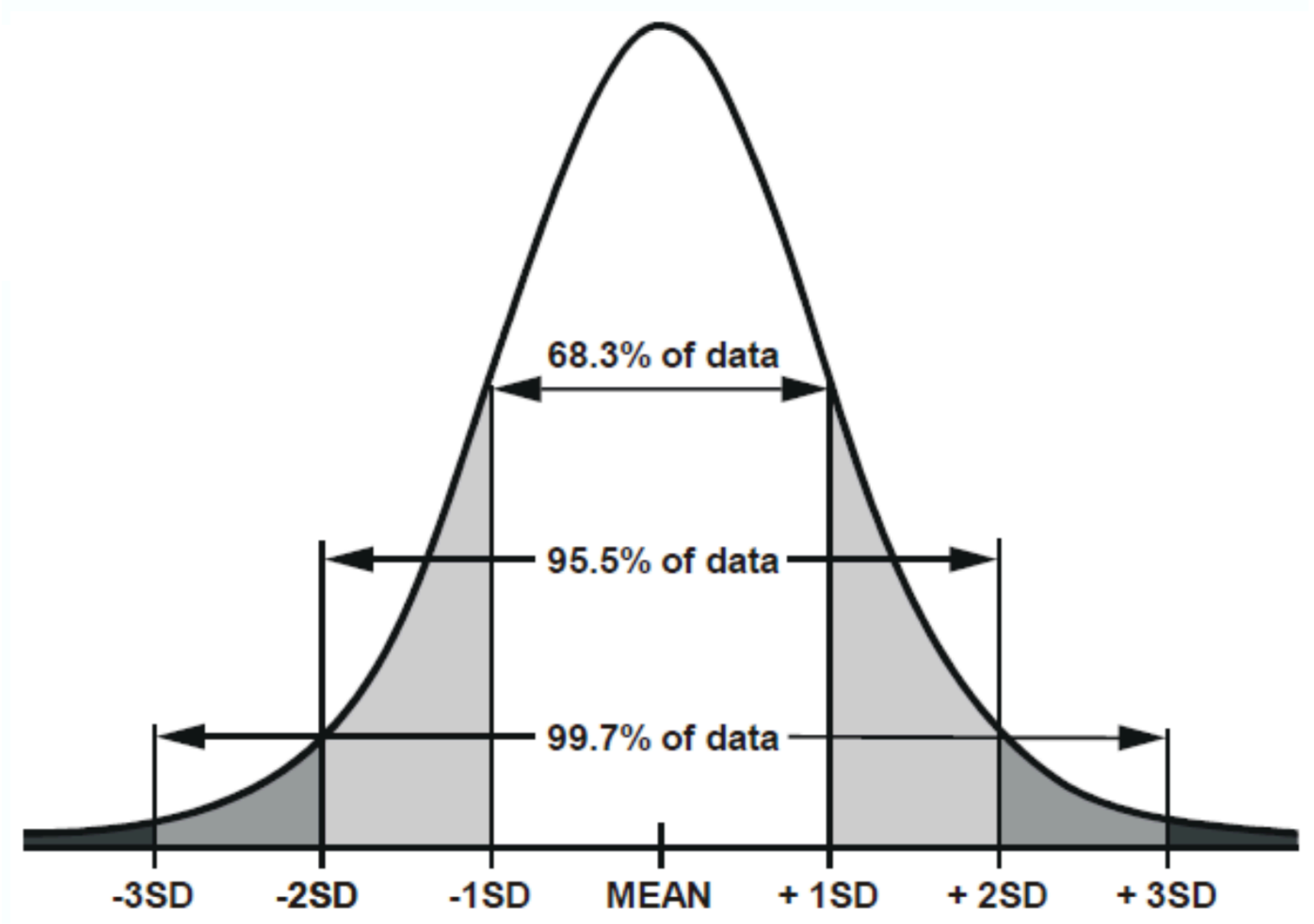
  ✓ 표준 오차: $SE = \dfrac{s}{\sqrt{n}}$

  ✓ 이 추정량을 갖고서 흔히들 말하는 95% 신뢰 구간을 다음과 같이 정의함.

    ➡ **신뢰 구간(Confidence Interval)**

      ✓ $\bar{y} \pm 2SE$

신뢰 구간

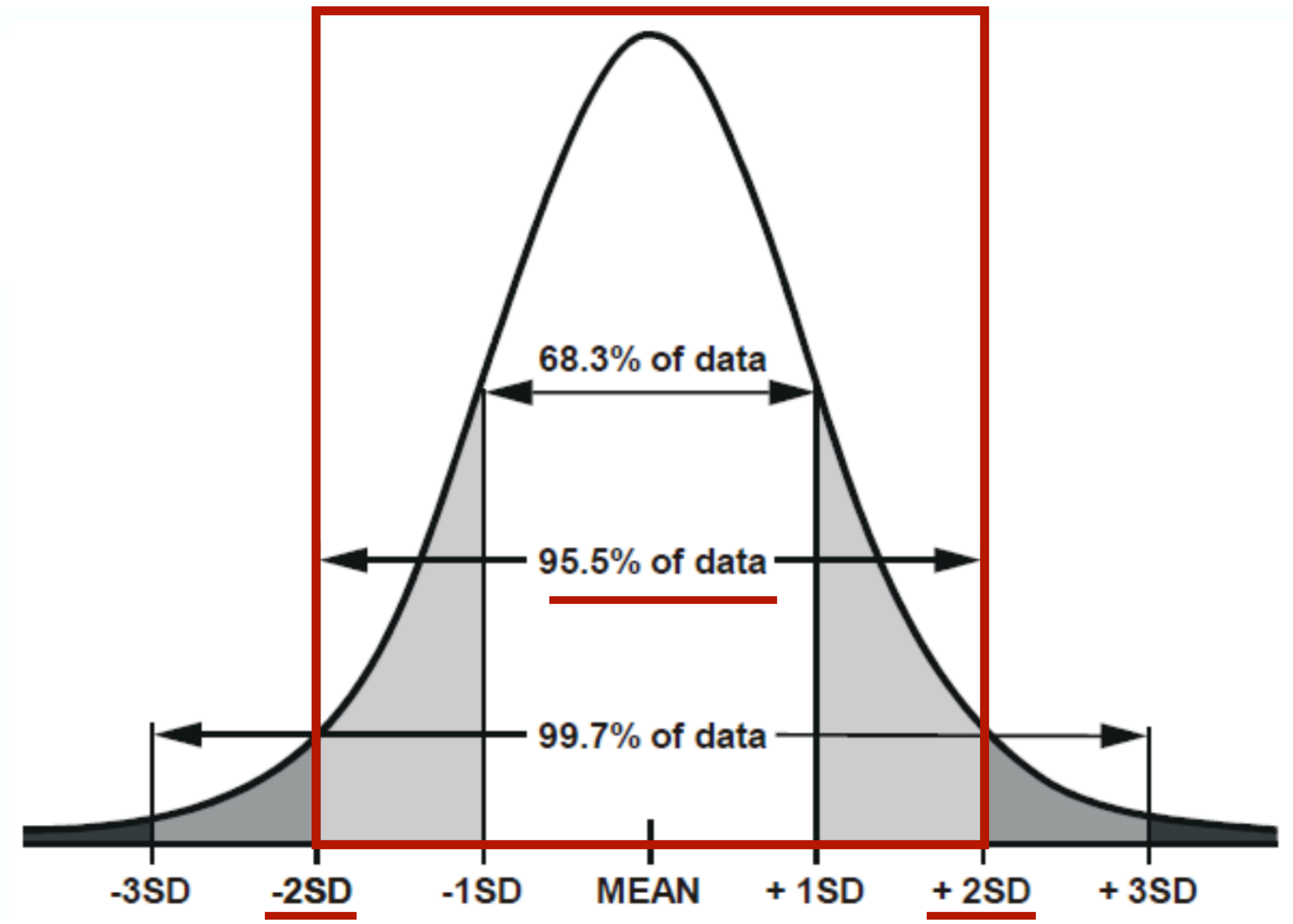


Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

신뢰 구간



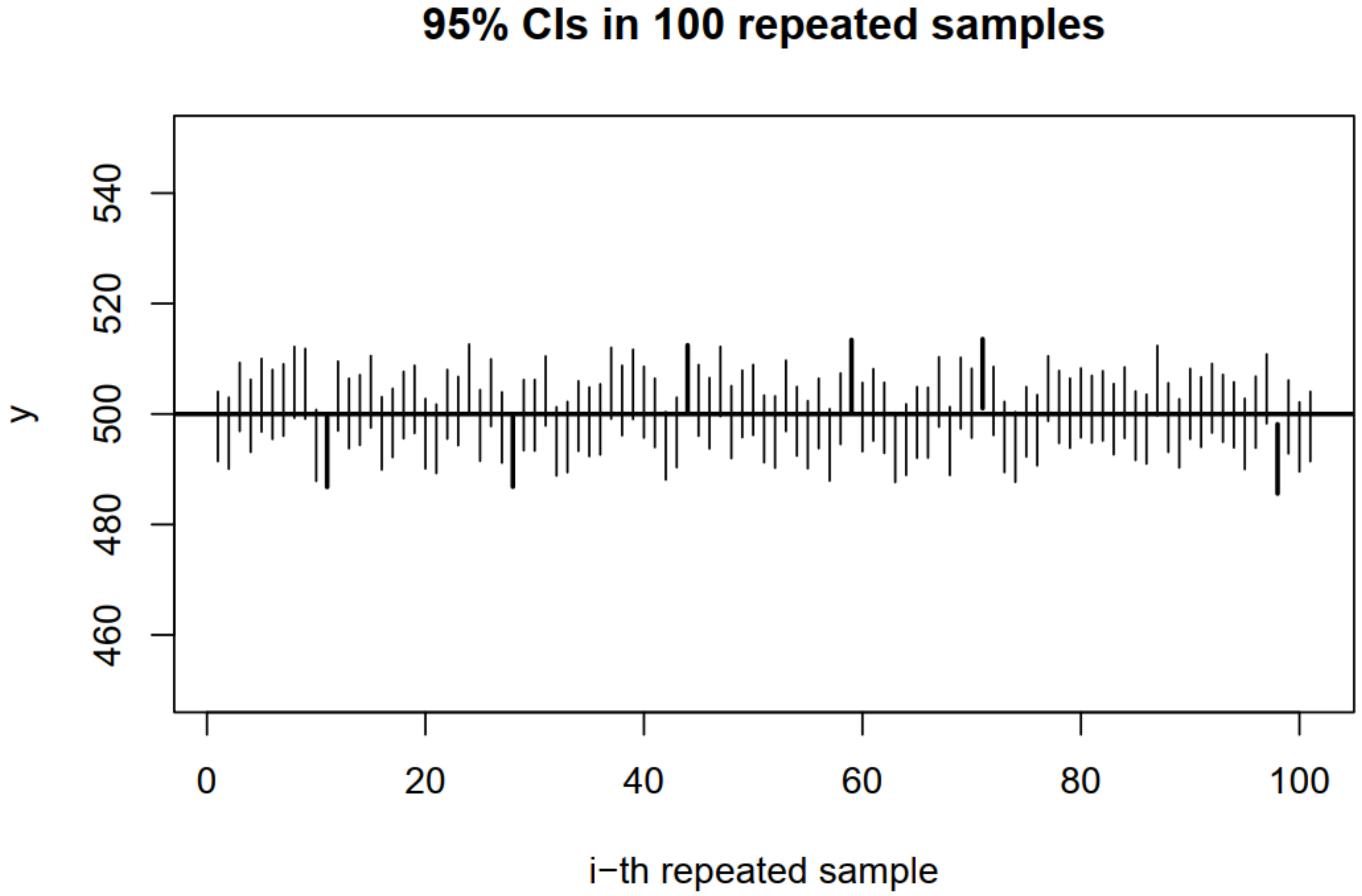
Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

신뢰 구간

- 신뢰 구간의 의미는 다음과 같다.
  - ✓ 표본을 반복적으로 수집 한 뒤 각 표본이 지닐 수 있는 신뢰 구간을 매번 계산할 때 마다, 해당 신뢰 구간의 95% 비율은 모평균을 갖게 된다.



Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

신뢰 구간

- 신뢰 구간 해석에 있어 종종 하는 실수

  ✓ “신뢰 구간은 모평균이 될 수 있는 값들을 구간으로 표현한 것이다!”

- 땡! 모평균은 점 추정을 통해 추정할 수 있는 하나의 ‘단일값’이지, 어떤 확률 분포를 가질 수 있는 ‘구간’이 아니다.

- 무엇보다, 신뢰구간은 표본 분포를 ‘반복적으로’ 수집해 얻은 <u>표본 분포를 기준으로 정의되는 것</u>이지, 모평균을 포함하고 있는 어떤 확률 분포에 의해 정의되는 것이 아니다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

신뢰 구간

- 그렇다면 신뢰 구간을 어디에다가 써먹는가?
- 신뢰 구간은 평균의 표본 분포의 넓이를 대략적으로 보여주는 것
- 한편, 넓이가 넓을수록 어떠한 변이(variability)들을 갖고 있다.
- 즉, 신뢰 구간이 넓을수록 우리가 수집 표본이 기존 실험 데이터 수집 과정이나 설계 과정에서 미처 고려하지 못했던 것들이 담겨 있다는 것을 내포한다.
- 따라서 신뢰 구간은 표본평균의 추정치에 대해서 얼마나 '불확실한가'를 보여주는 것이라 할 수 있다. (신뢰구간 넓으면 넓을수록 다양한 변이점들이 많다는 건 알지만 구체적으로 어떤 변이점들이 있는지는 확실하게 모르므로)
- 그러므로 신뢰 구간은 좁으면 좁을수록 좋다!

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

모수, 추정량, 추정치

- **모수(parameter)**
  - ✓ 일반적으로 알 수 없는 모집단의 진짜 parameter
  - ✓ 보통 그리스 문자로 표시: μ = 모평균, σ^2 = 모분산 등
- **추정량(estimator)**
  - ✓ 표본으로부터 모수를 추정하는 방법/함수
  - ✓ 보통 그리스 문자에 모자를 씌움: $\hat{\mu}, \widehat{\sigma^2}$ 등
- **추정치(estimate)**
  - ✓ 추정량에 표본값을 대입해서 실제로 추정한 값
  - ✓ 보통 로마 문자로 표시: m, s^2 등

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

좋은 추정량의 속성

- **점근적 정규성(Asymptotic Normality)**

  ✓ "추정량과 모수의 차이는 어떤 정규분포에 수렴한다."

- **비편향성/불편성(Unbiasedness)**

  ✓ "추정량의 평균/기댓값은 모수여야 한다."

- **일치성(Consistency)**

  ✓ "표본 크기가 무한히 커질수록, 추정량이 모수에 가까워져야 한다."

- **효율성(Efficiency)**

  ✓ "같은 조건이라면 모수에 가능한 한 더 가까워야 한다."

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

좋은 추정량의 속성

- 이러한 속성을 만족하는 추정량이 있을까?

  ✓ 최소 분산 비편향 추정량(Minimum Variance Unbiased Estimator; MVUE)

  ✓ 하지만 존재하지 않는 경우가 많고, 존재하더라도 구하기가 어려움.

  ➡ 따라서 결국 자주 쓰는 건 <u>최대 가능도 추정(Maximum Likelihood Estimation)</u>

  ✓ MVUE보단 이론적으로 '살짝' 모자라나, 범용성 덕분에 제일 자주 사용.

실험심리언어학에서 사용되는 실험설계

- **현실과 이상**

  ✓ 엄밀하게 말하면 우리는 무작위 단위로 실험참여자들을 모집해야함.

실험심리언어학에서 사용되는 실험설계

- **현실과 이상**

  ✓ 엄밀하게 말하면 우리는 무작위 단위로 실험참여자들을 모집해야함.

  ✓ 왜냐하면 우리가 연구하고자 하는 모집단에 대해서 일반화할 여건을 마련해야 하기 때문.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- **현실과 이상**
  - ✓ 엄밀하게 말하면 우리는 무작위 단위로 실험참여자들을 모집해야함.
  - ✓ 왜냐하면 우리가 연구하고자 하는 모집단에 대해서 일반화할 여건을 마련해야 하기 때문.
  - ✓ 하지만 현실은 녹록치 않으므로 타협을 한다.

Cornell University KONKUK UNIVERSITY Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- **현실과 이상**
  - ✓ 엄밀하게 말하면 우리는 무작위 단위로 실험참여자들을 모집해야함.
  - ✓ 왜냐하면 우리가 연구하고자 하는 모집단에 대해서 일반화할 여건을 마련해야 하기 때문.
  - ✓ 하지만 현실은 녹록치 않으므로 타협을 한다.
  - ✓ 예를 들면 심리학/언어학 수업을 듣는 대학생들 모아서 하기!

실험심리언어학에서 사용되는 실험설계

- **현실과 이상**

  ✓ 엄밀하게 말하면 우리는 무작위 단위로 실험참여자들을 모집해야함.

  ✓ 왜냐하면 우리가 연구하고자 하는 모집단에 대해서 일반화할 여건을 마련해야 하기 때문.

  ✓ 하지만 현실은 녹록치 않으므로 타협을 한다.

  ✓ 예를 들면 심리학/언어학/어문학 수업을 듣는 대학생들 모아서 하기!

  ➡ 하지만 본 세미나에서 우리는 모든 데이터가 무작위 단위로 모아졌다, 즉, 무작위 표본추출을 했다라고 **가정**할 것임.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- **독립과 의존**

  ✓ 우리 실험심리언어학자는 개개인의 실험참여자에게 각 조건에 대해서 단 한번의 응답만 수집할지, 아니면 동일한 실험참여자에게서 반복적으로 응답을 얻을지 결정해야함.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- **독립과 의존**

  ✓ 우리 실험심리언어학자는 개개인의 실험참여자에게 각 조건에 대해서 단 한번의 응답만 수집할지, 아니면 동일한 실험참여자에게서 반복적으로 응답을 얻을지 결정해야함.

  ✓ 전자라면 각각의 응답들은 서로 '독립'.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- **독립과 의존**
  - ✓ 우리 실험심리언어학자는 개개인의 실험참여자에게 각 조건에 대해서 단 한번의 응답만 수집할지, 아니면 동일한 실험참여자에게서 반복적으로 응답을 얻을지 결정해야함.
  - ✓ 전자라면 각각의 응답들은 서로 '독립'.
  - ✓ 후자라면 해당 응답들은 서로 '의존'이며, 다른 말로 하면 '반복 측정(repeated measures)'.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- **독립과 의존**

  ✓ 우리 실험심리언어학자는 개개인의 실험참여자에게 각 조건에 대해서 단 한번의 응답만 수집할지, 아니면 동일한 실험참여자에게서 반복적으로 응답을 얻을지 결정해야함.

  ✓ 전자라면 각각의 응답들은 서로 ‘<u>독립</u>’.

  ✓ 후자라면 해당 응답들은 서로 ‘<u>의존</u>'이며, 다른 말로 하면 '<u>반복 측정(repeated measures)’.</u>

  ➡ 실험심리언어학의 경우 **반복 측정**이 사실상 표준.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- **라틴 방진 설계 (Latin square design)**

  ✓ 실험심리언어학에서 라틴 방진 설계 사용은 굉장히 흔함.

1번 그룹	2번 그룹
조건 A	조건 B
조건 B	조건 A

< 표1 > 2개 조건 (첫번째 행을 제외한 2 x 2 방진 설계)

Cf. '라틴'은 알파벳을 뜻하며, 흔히들 수학자 레온하르트 오일러가 쓴 것에 기원함. 하지만 사실 그보다 더 일찍 조선 후기 수학자 최석정이 먼저 기록하였음.

실험심리언어학에서 사용되는 실험설계

- **라틴 방진 설계의 특징**

✓ 각 조건(혹은 알파벳)이 행과 열마다 단 한번만 나타남.

1번 그룹	2번 그룹	3번 그룹	4번 그룹
조건 A	조건 B	조건 C	조건 D
조건 B	조건 C	조건 D	조건 A
조건 C	조건 D	조건 A	조건 B
조건 D	조건 A	조건 B	조건 C

< 표2 > 4개 조건 (첫번째 행을 제외한 4 x 4 방진 설계)

실험심리언어학에서 사용되는 실험설계

- **라틴 방진의 장점**

✓ 쉽게 확장시킬 수 있으며, 이론상 무한 확장 가능!

1번 그룹	2번 그룹	3번 그룹	4번 그룹	5번 그룹	6번 그룹	7번 그룹	8번 그룹
조건 A	조건 B	조건 C	조건 D	조건 E	조건 F	조건 G	조건 H
조건 B	조건 C	조건 D	조건 E	조건 F	조건 G	조건 H	조건 A
조건 C	조건 D	조건 E	조건 F	조건 G	조건 H	조건 A	조건 B
조건 D	조건 E	조건 F	조건 G	조건 H	조건 A	조건 B	조건 C
조건 E	조건 F	조건 G	조건 H	조건 A	조건 B	조건 C	조건 D
조건 F	조건 G	조건 H	조건 A	조건 B	조건 C	조건 D	조건 E
조건 G	조건 H	조건 A	조건 B	조건 C	조건 D	조건 E	조건 F
조건 H	조건 A	조건 B	조건 C	조건 D	조건 E	조건 F	조건 G

< 표3 > 8개 조건 (첫번째 행을 제외한 8 x 8 방진 설계)

실험심리언어학에서 사용되는 실험설계

- **실험심리언어학에서 라틴 방진 설계의 유용성**

  ✓ 특히 앞서 말한 각 행과 열에서 각각의 조건들이 단 한 번만 나타나는 경우가 매우 유용.

  ✓ 이는 각 실험차여자들로 하여금 특정 조건에 속하는 실험재료들을 한 번씩만 노출시킴으로써 효율적인 반복측정을 진행할 수 있는 여건 마련!

실험재료 번호	1번 그룹	2번 그룹
1번	조건 A	조건 B
2번	조건 B	조건 A
3번	조건 A	조건 B
4번	조건 B	조건 A

< 표3 > 2개 조건 기준 총 4개의 실험재료를 포함하는 라틴 방진

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- **실험심리언어학에서 라틴 방진 설계의 유용성**

  ✓ 이는 결국 각 조건에 대한 개개의 실험참여자들에 대한 추정값을 좀 더 정확하게 이끌어내고 일반화할 수 있다는 점에서 큰 장점!

  ✓ 물론 라틴 방진 설계가 아닌 다른 설계 또한 있으나, 이는 실험 설계(Experimental design)라 하여 통계학에서 아예 따로 분과가 있으므로 더 구체적으로 가진 않겠음.

  ✓ 시뮬레이션을 통해서 라틴 방진 설계를 살펴보자.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계 - 같이 해보기

```
> condition <- c(rep(letters[1:2], 2), rep(letters[2:1], 2))
> item <- rep(1:4, 2)
> subj <- rep(1:2, each = 4)
> df_example <- data.frame(subj, item, condition)
> df_example
  subj item condition
1    1    1         a
2    1    2         b
3    1    3         a
4    1    4         b
5    2    1         b
6    2    2         a
7    2    3         b
8    2    4         a
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

```
> condition <- c(rep(letters[1:2], 2), rep(letters[2:1], 2))
> item <- rep(1:4, 2)
> subj <- rep(1:2, each = 4)
> df_example <- data.frame(subj, item, condition)
> df_example
  subj item condition
1    1    1         a
2    1    2         b
3    1    3         a
4    1    4         b
5    2    1         b
6    2    2         a
7    2    3         b
8    2    4         a
```

✓ 이 경우가 바로 각 실험참여자가 한번씩 실험재료에 노출된 것을 지칭!

➡ 이를 <u>완전 교차(fully crossed) 실험참여자 및 실험재료 설계</u>라고함.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

실험심리언어학에서 사용되는 실험설계

- 좀 더 확실하게 확인해보자.

```
> xtabs(~ subj + item, df_example)
     item
subj. 1 2 3 4
   1 1 1 1 1
   2 1 1 1 1
> xtabs(~ subj + condition, df_example)
     condition
subj a b
   1 2 2
   2 2 2
> xtabs(~ item + condition, df_example)
     condition
item a b
   1 1 1
   2 1 1
   3 1 1
   4 1 1
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

가설검정 - t검정 (t-test)

- 이제 가장 간단한 가설검정 중 하나인 t검정에 대해서 알아보자.
- t검정은 사실상 실험심리언어학 연구에 사용된 다양한 가설검정들 중 뼈대가 되는 것들 중 하나다.

  ✓ 분산분석, 선형 회귀 등등…

- 특히나 오늘날 실험심리언어학에서 주로 사용되고 사실상 표준으로 여겨지는 선형 혼합 효과 모형(Linear mixed-effects model)은 t검정과 밀접한 연관성을 갖고 있다.
- 따라서 오늘은 t검정을 굉장히 구체적으로 살펴봄으로써 가설검정이 어떠한 (수리)통계학적 메커니즘으로 돌아가는지 살펴보고자 한다.

단일표본 t검정 (one sample t-test)

- 다음과 같은 데이터가 주어졌다고 가정해보자.

  ✓ n만큼 크기의 무작위 표본 y이 있고,

  ✓ 이 데이터는 $N(\mu, \sigma)$ 분포를 따른다.

- t-검정의 주요 가정 중 하나는 각각의 데이터 값, 즉 관측값이 각각 독립이라는 것이다.
- 이를 바탕으로 우리는 반복적 (가설) 표본 수집(=시뮬레이션)을 통해 평균의 표본 분포를 추정할 수 있다.
- 즉, 하나의 가설적 통계 모형을 세워볼 수 있다.

➡ $N(\hat{\mu}, \frac{\hat{\sigma}}{\sqrt{n}})$

단일표본 t검정 (one sample t-test)

- 한 가지 중요한 점은 이와 같은 표본 분포는 모평균과 모표준편차의 추정치를 실제와 가깝게 설정할 수 있게끔 해야지만이 (합리적인) 통계적 모형이 된다는 것이다.
- 예컨대 우리가 추정하고자 하는 모평균과 모표준편차는 비문법적인 문장에 대한 읽기 시간에 대한 것이라면, 문법적인 문장에 대한 읽기 시간에 대해서 추정하는 건 옳지 않다.
- 우리가 시뮬레이션할 통계적 모형은 다음과 같이 R에서 구현해볼 수 있다.

```
> n <- 1000
> mu <- 500
> sigma <- 100
> ## generate simulated data:
> y <- rnorm(n, mean = 500, sd = 100)
> ## compute summary statistics:
> y_bar <- mean(y)
> SE <- sd(y) / sqrt(n)
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- 귀무가설 유의성검정(Null Hypothesis Significance Testing; NHST)

  ✓ 어떤 평균값 μ를 갖고 있는 귀무가설을 세우는 것에서 시작!

  ✓ 예컨대 우리는 다음과 같이 귀무가설을 세워볼 수 있다.

  ✓ $H_0 : \mu = 450$

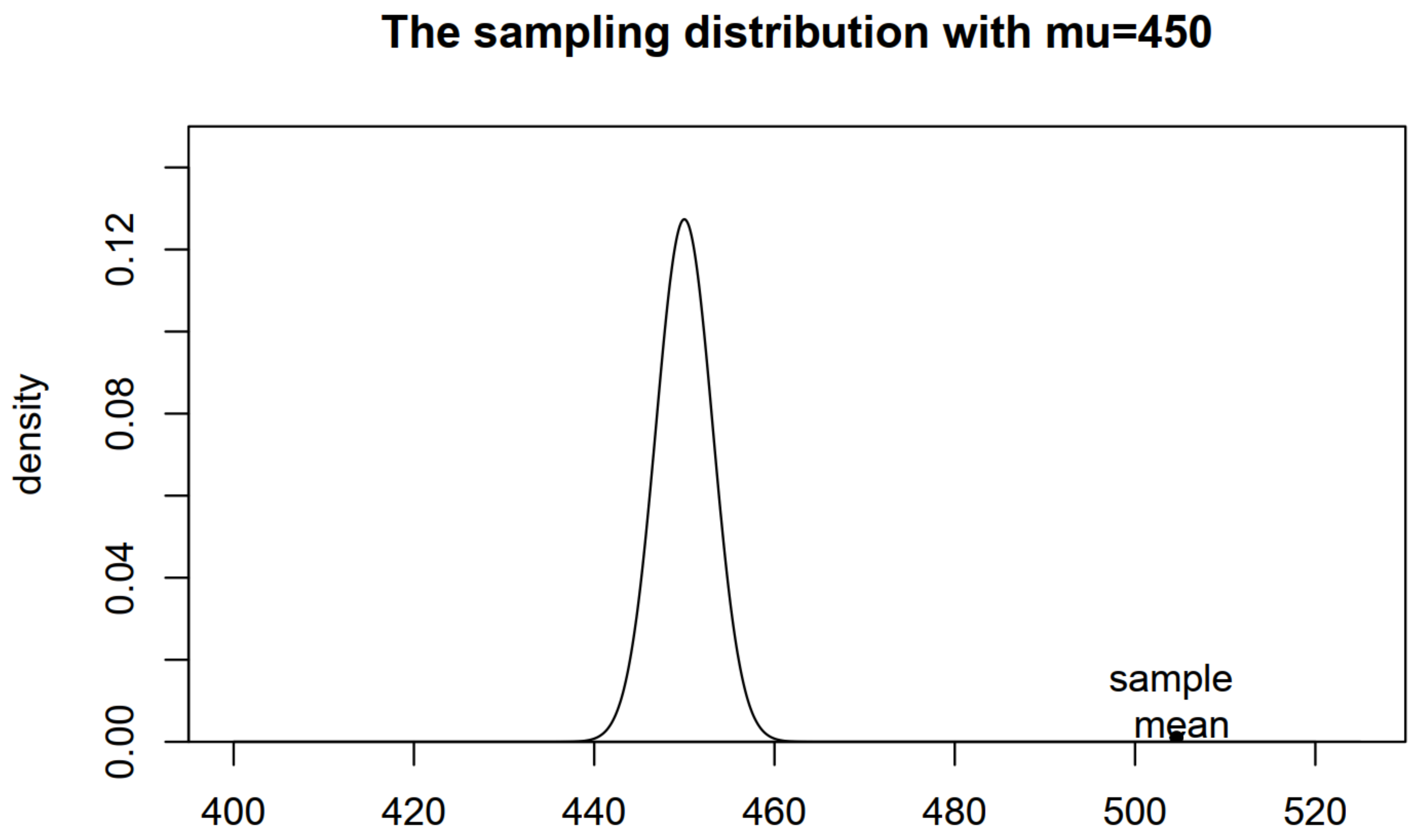
  ➡ 이를 그림으로 표현하면..

Cornell University

KU KONKUK UNIVERSITY

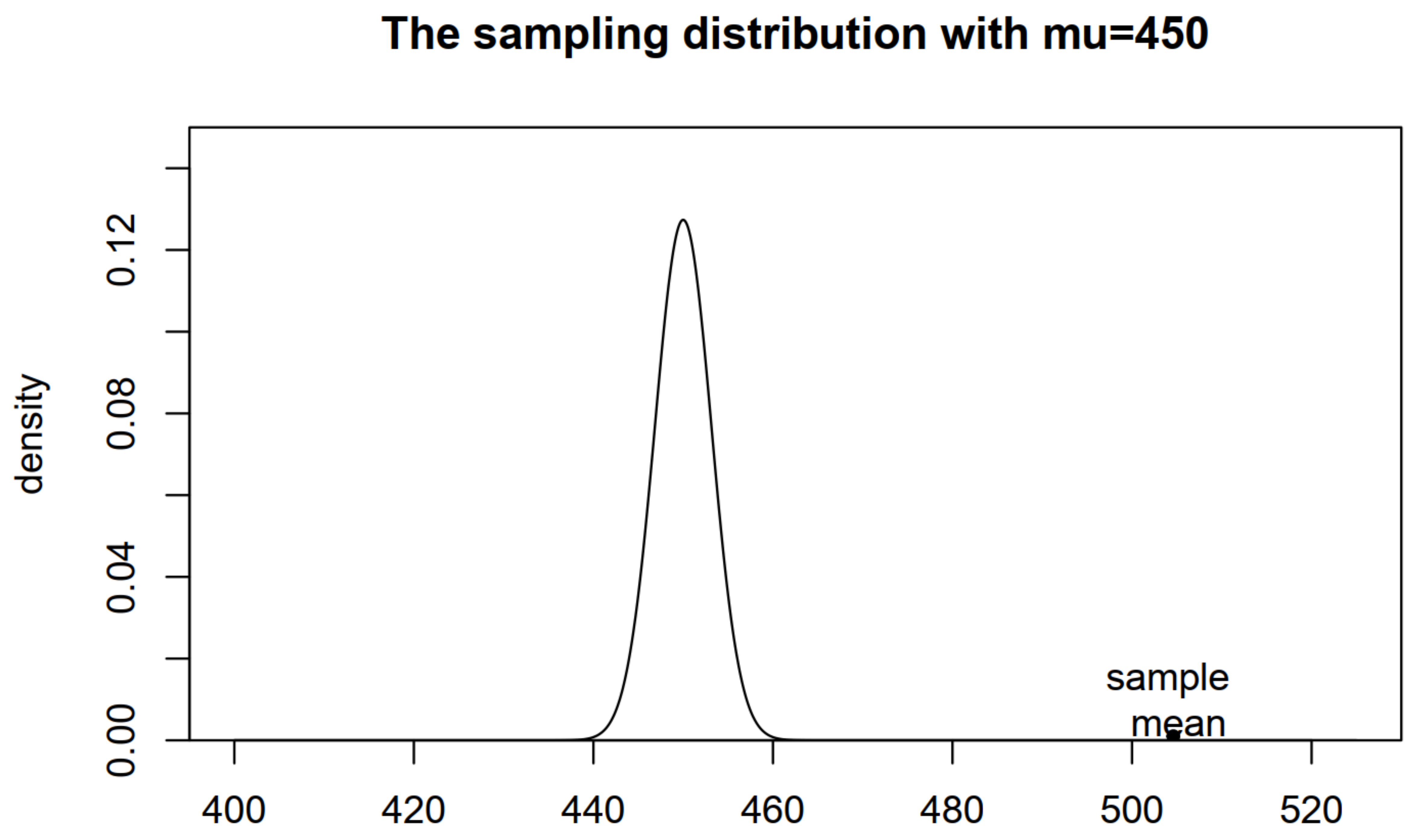
Brain & Cognition Research Center

단표본 t검정 (one sample t-test)



- 여기서 sample mean, 즉, 표본 평균은 실제 수집한 데이터의 평균을 의미한다.

단일표본 t검정 (one sample t-test)



- 위 그림에서 우리가 생각해볼 수 있는 것 두 개가 있다.

✓ 표본 평균이 가설 평균 '근처'에 있다면, 우리는 실제 데이터와 귀무가설에 부합한다.

✓ 반대로 '멀리' 있다면, 우리는 실제 데이터가 귀무가설에 부합하지 않다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- 그렇담 이 ‘근처’와 ‘멀리’를 어떻게 통계학적으로 얘기할 수 있을까?

  ✓ 즉, 귀무가설이 제시하는 평균과 실제 데이터를 통해 얻은 평균 사이의 거리는 어떻게 <u>정량화</u> 할 수 있을까?

- 생각해보자. 귀무가설을 보여주고 있는 통계적 모형은 하나의 ‘분포’다.
- 그리고 이러한 분포는 ‘추정’을 한 것이며, 이는 항상 (표준) 오차가 있기 마련이다.
- 그렇다면 이 오차를 기준값으로 해서 얼만큼 가깝게/멀리 떨어져있는지를 정량화할 수 있다!

  ➡ 즉, 표본 평균이 가설 평균으로부터 <u>t 표준오차만큼 멀어진 것</u>이라고 할 수 있다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- $t \cdot SE = \bar{x} - \mu$
- 양변을 표준 오차로 나누어주면…

  ✓ $t_{obs} = \dfrac{\bar{x} - \mu}{SE}$

  ➡ 이 관측된 t값이 바로 얼마만큼 가까운지/멀어져있는지를 측정해주는 값!
- 무엇보다 이 t값은 사실 어떠한 값을 출력해주는 하나의 함수다!

  ✓ 즉, 확률 변수라는 것을 의미하며 이는 하나의 확률 분포를 생성한다는 것을 의미한다.

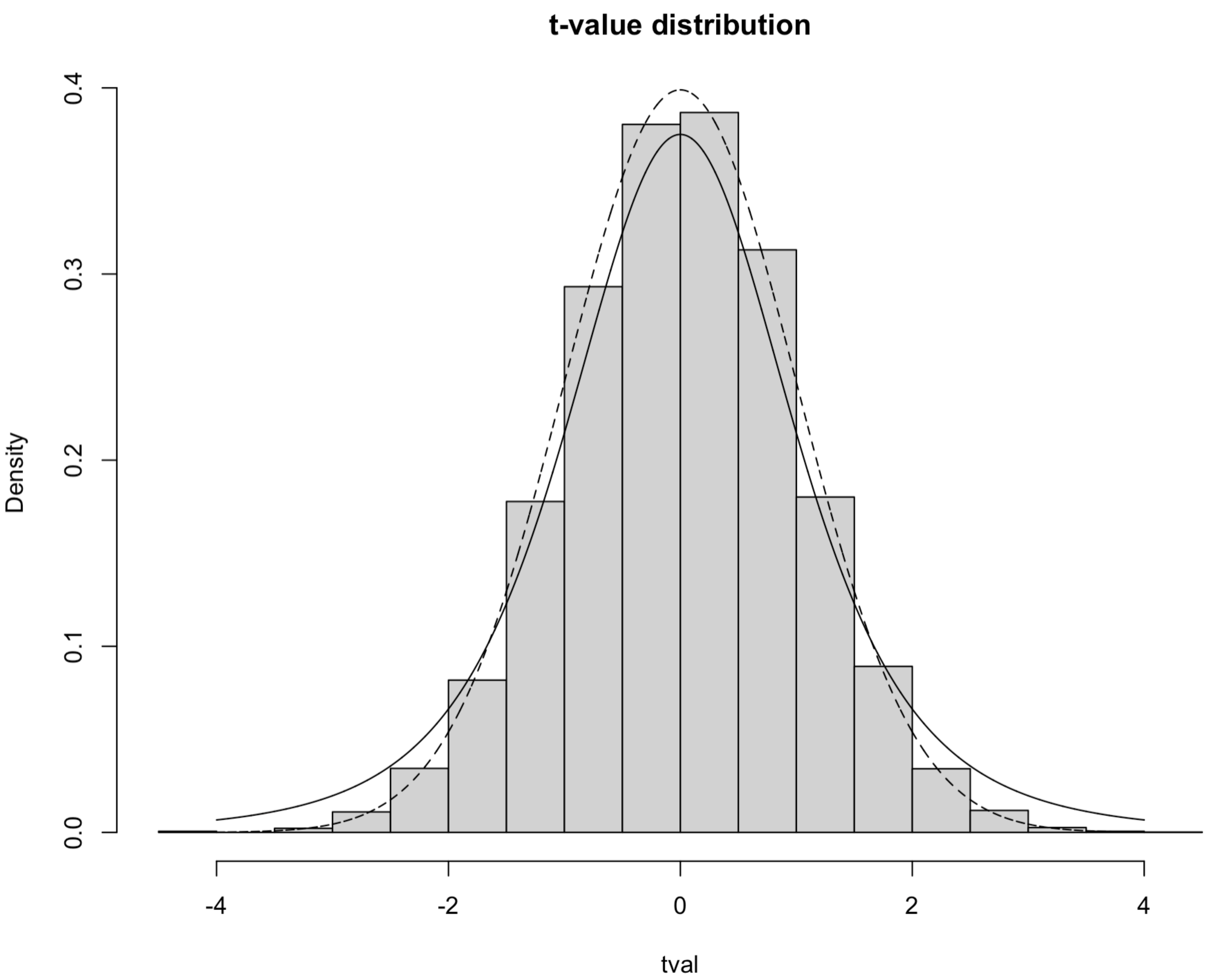
  ➡ 그리고 이 분포가 바로 t분포다!

단일표본 t검정 (one sample t-test)

```
nsim <- 10000
n <- 5
mu <- 450
## null hypothesis mean:
mu0 <- 450
sigma <- 100
tval <- rep(NA, nsim)
se <- sigma / sqrt(n)
for (i in 1:nsim) {
  y <- rnorm(n, mean = mu, sd = sigma)
  xbar <- mean(y)
  tval[i] <- (xbar - mu0) / se
}

hist(tval, freq = FALSE, main = "t-value distribution")
x <- seq(-4, 4, by = 0.001)
lines(x, dt(x, df = n-1)) # t-distribution
lines(x, dnorm(x), lty = 2) # standard normal distribution
```

단일표본 t검정 (one sample t-test)



Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

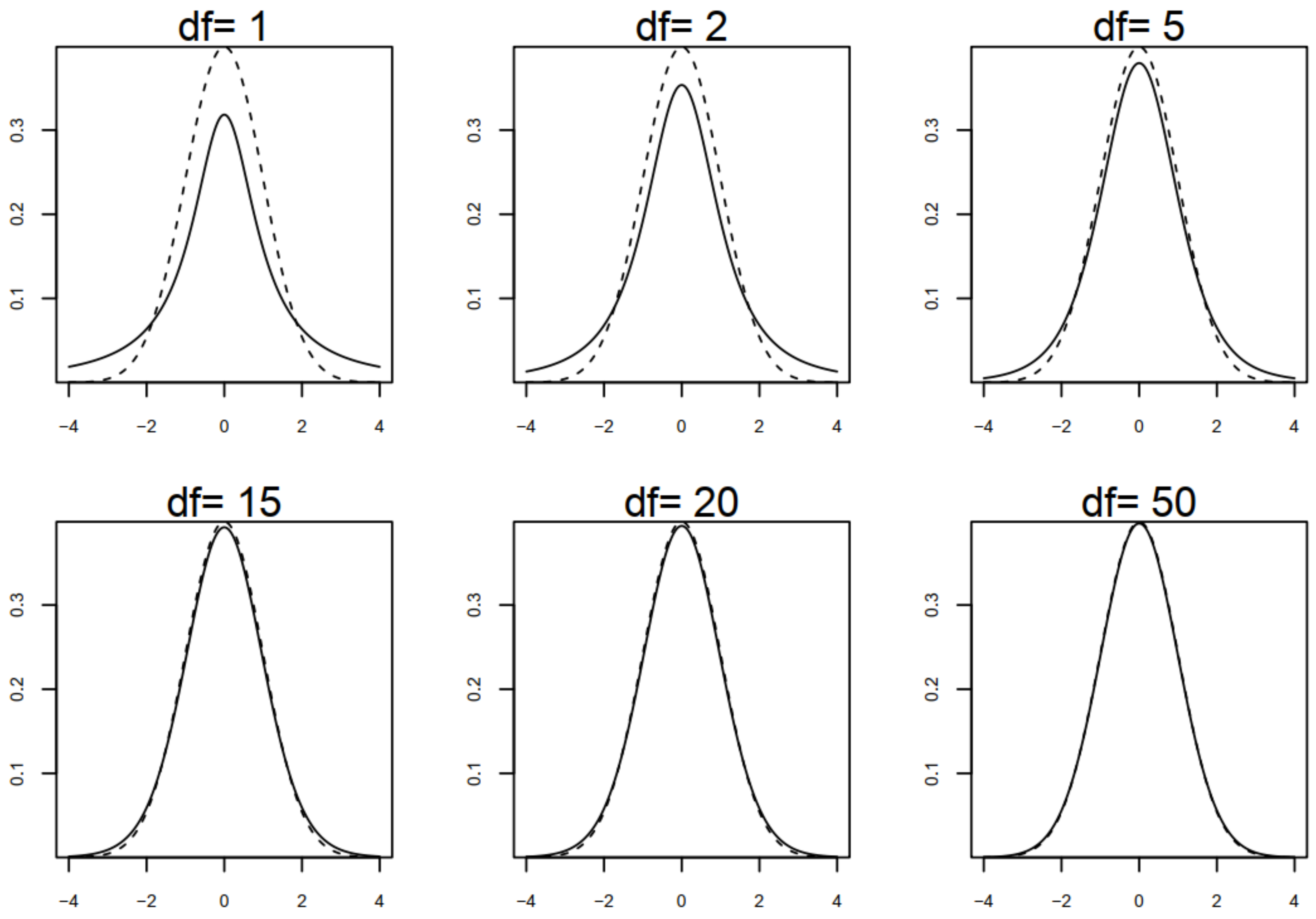
단일표본 t검정 (one sample t-test)

- 따라서 우리는 “관측된 t값은 (표본이 반복 수집된) $n-1$의 자유도를 지닌 t분포에서 생성된 것으로 가정한다.”라는 통계적 모형을 만들 수가 있으며, 이는 아래와 같이 표현할 수 있다.

✓ $T \sim t(n-1)$

✓ 그리고 어제 확률론 시간에 살펴보았듯, t분포는 자유도가 커지면 커질수록 표준정규분포 $N(0,1)$에 가까워진다.

단일표본 t검정 (one sample t-test)



Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- 결국 t분포를 활용한 귀무가설검정은 다음과 같은 단계를 통해 진행되는 것으로 볼 수 있다.

1. 귀무가설 설정

2. 추정값 계산

3. 실제 데이터를 통해 관측된 t값 계산: $t = \dfrac{\bar{y} - \mu}{s/\sqrt{n}}$

4-1. 만약 t값이 '크면' 귀무가설 '기각(reject)'

4-2. 만약 t값이 '작으면' 귀무가설 '기각 실패(fail to reject)'

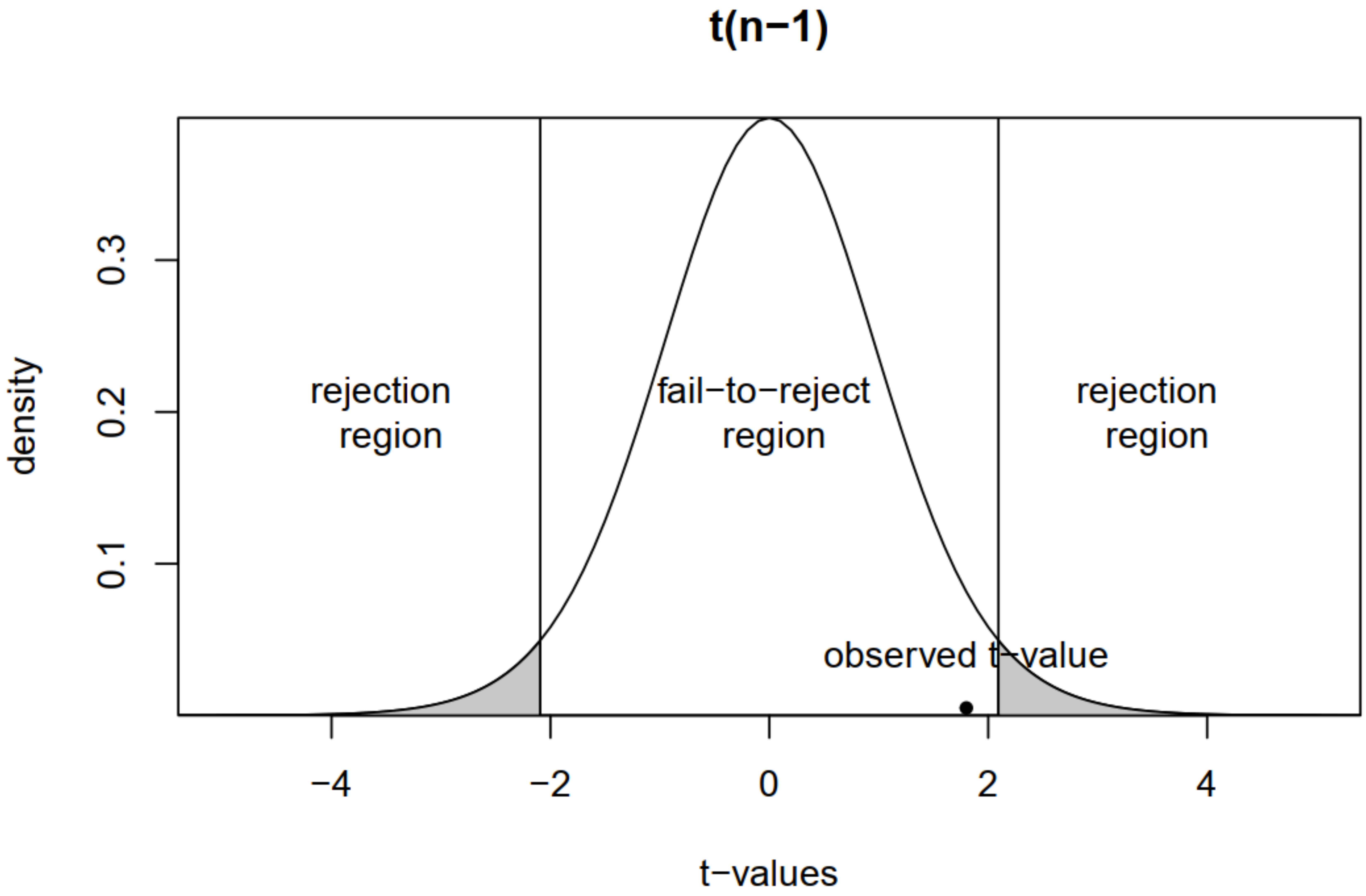
➡ 그렇다면 t값이 얼마여야지 귀무가설 기각 여부를 결정하는가?

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)



Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- **절댓값 t값이 2보다 크면 유의하다고 간주하고, 이는 귀무가설 기각을 의미한다.**

✓ 확률론 시간 때 배운 t 분포의 자유도와 관련된 특징을 떠올려보자 (뭘까요?)

Cornell University

KU KONKUK UNIVERSITY

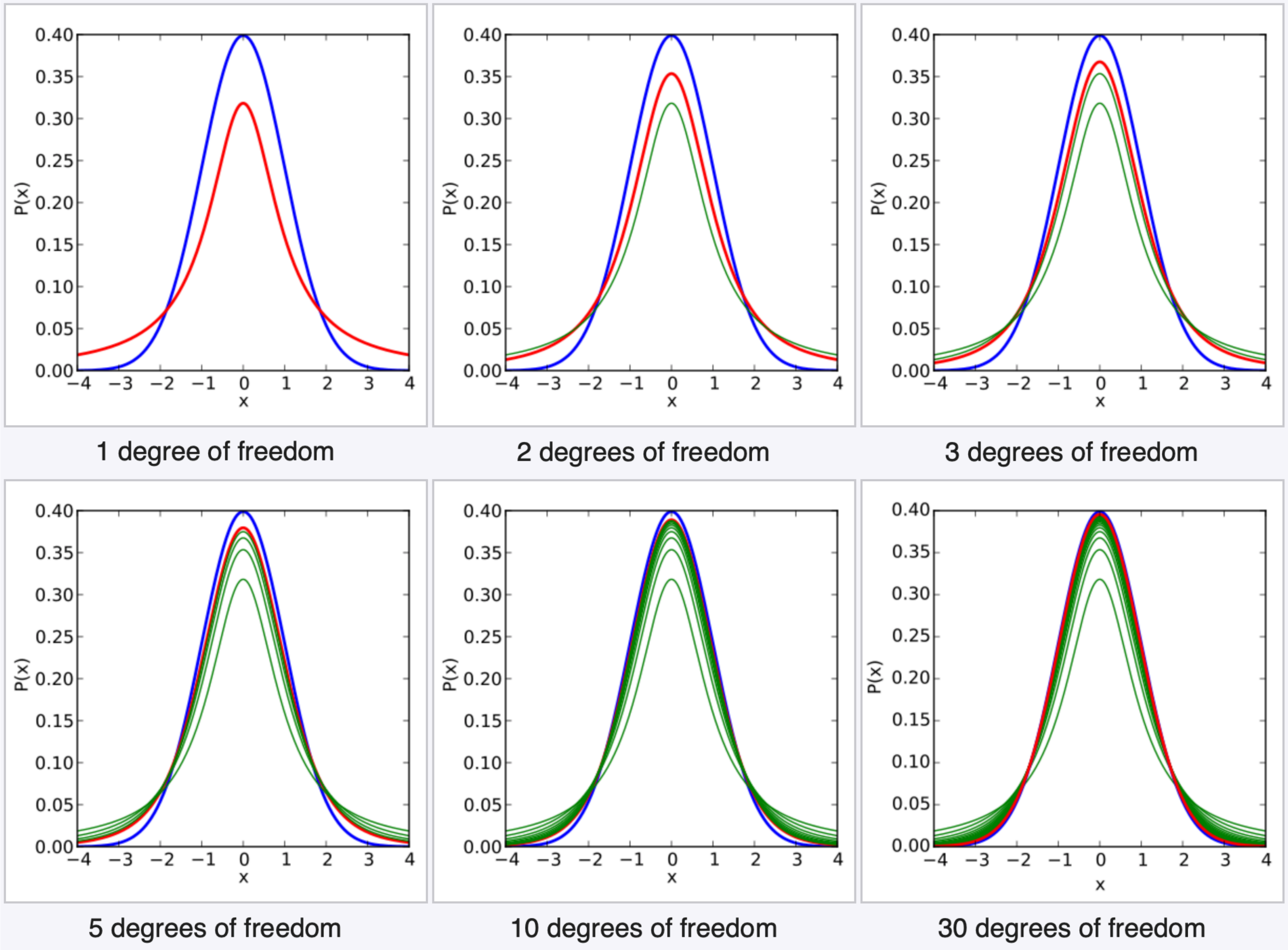
Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- **절댓값 t값이 2보다 크면 유의하다고 간주하고, 이는 귀무가설 기각을 의미한다.**

  ✓ 확률론 시간 때 배운 t 분포의 자유도와 관련된 특징을 떠올려보자 (뭘까요?)

  ✓ 우리는 **t 분포도의 자유도(=표본수)가 높으면 높을수록 표준 정규 분포와 사실상 동일**해진다는 걸 이미 안다.



각 자유도에 따른 t 분포의 변화 (파란색: 표준 정규 분포)

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

p값에서 절대값 t값으로

- **실제 실험심리언어학 현장에서는…**

✓ 4개 조건을 기준으로 하여 실험 문장이 최소 8개 ~ 16개 이상을 만듦.

✓ 실험참여자들은 최소 20명대 후반 ~ 30명대 이상을 모집.

➡ t 분포도 또한 자연스럽게 표준 정규 분포에 가까워진다.

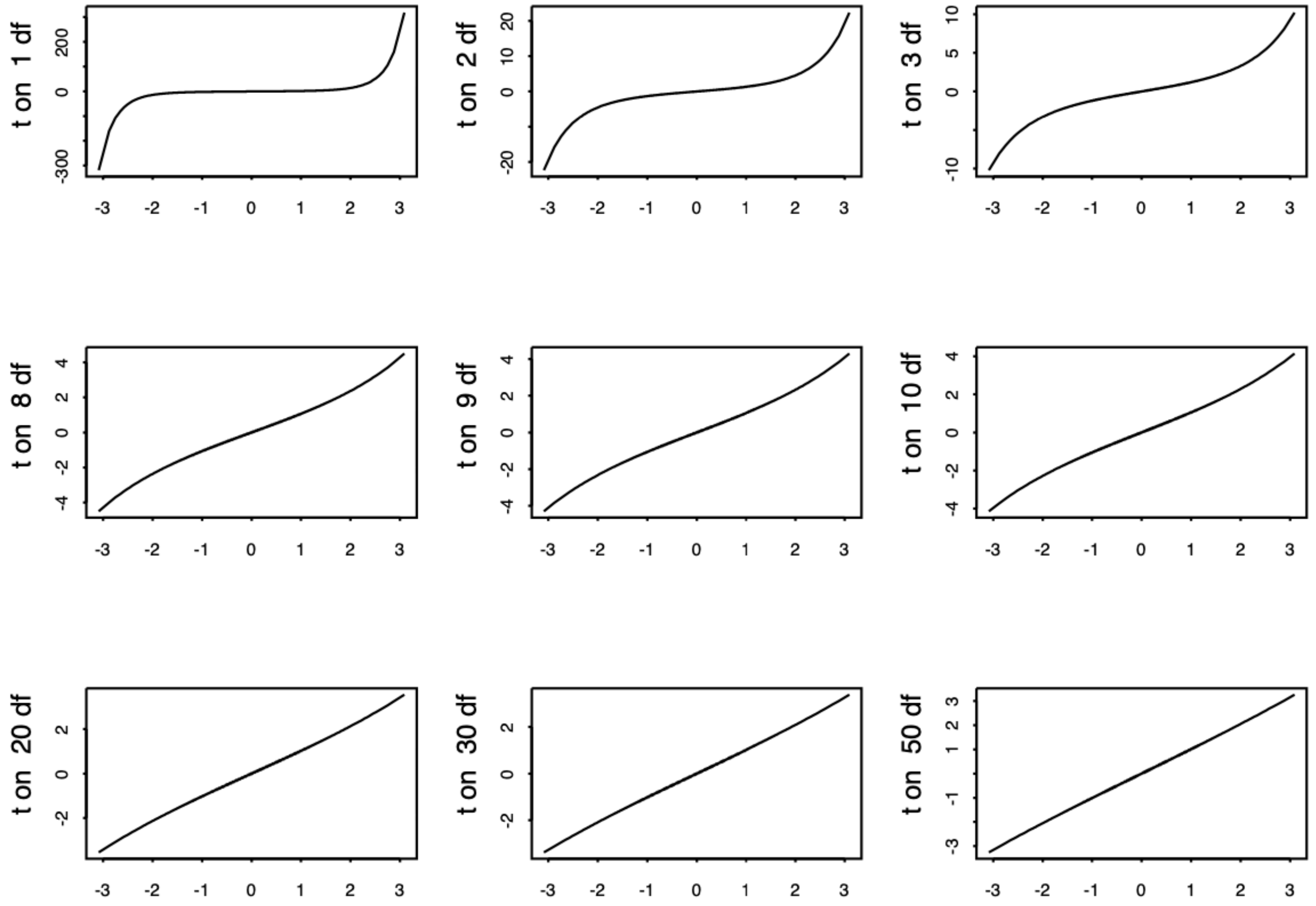
Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

p값에서 절대값 t값으로

- 이는 실제 통계학과에서도 언급되는 내용.
- 아래 내용은 미국 예일대 통계학 개론 1998년도 강의록에서 발췌한 내용임.
- http://www.stat.yale.edu/~pollard/Courses/100.fall98/pollard/lecture7.pdf



Cornell University

The plots in the last row are practically straight lines. You are pretty safe (unless you are interested in the extreme tails) in ignoring the difference between the standard normal and the t-distribution when the degrees of freedom are even moderately large.

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

p값에서 절대값 t값으로

- 참고로 실제 t 분포 표에서도 자유도(=표본수)가 높으면 높을수록 유의하다는 걸 알 수 있다.
- https://programmathically.com/t-distribution-table/ (양측/단측)

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- 결국 우리는 지금까지 t검정이란 것을 통해서 귀무가설 기각 여부를 '결정'하는 걸 들여본 것!

  ✓ 이것이 바로 가설검정(중에 하나)!

- 한편, 잊지 말자.

  ✓ 모든 빈도주의 가설 검정은 기본적으로 데이터를 수집하기 위한 실험이 '반복 가능하다'라는 가정이 만족되어야 한다!

  ✓ 이것이 만족되지 못하면 가설검정은 안 하느니만 못하다.

  ➡ 시뮬레이션을 통해 다시 한 번 살펴보자.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- 먼저 귀무가설이 기각되지 않는 경우

✓ <u>반복 표본 수집</u>해서 얻은 t값이 귀무가설에 부합하는 경우

```
n <- 100
nsim <- 10000 # 얼마나 반복?
tvals <- rep(NA, nsim)
for (i in 1:nsim) {
  y <- rnorm(n, mean = 450, sd = 100)
  SE <- sd(y) / sqrt(n)
  tvals[i] <- (mean(y) - 450) / SE
}
plot(density(tvals),
     main = "Simulated t-distribution",
     xlab = "t-values under repeated testing")
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- 먼저 귀무가설이 기각하는 경우

  ✓ <u>반복 표본 수집</u>해서 얻은 t값이 귀무가설에 부합하지 않는 경우

```
n <- 100
nsim <- 10000 # 얼마나 반복?
tvals <- rep(NA, nsim)
for (i in 1:nsim) {
  y <- rnorm(n, mean = 470, sd = 100)
  SE <- sd(y) / sqrt(n)
  tvals[i] <- (mean(y) - 450) / SE
}
plot(density(tvals),
     main = "Simulated t-distribution",
     xlab = "t-values under repeated testing")
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

- 먼저 귀무가설이 기각하는 경우
  - ✓ 반복 표본 수집해서 얻은 t값이 귀무가설에 부합하지 않는 경우
  - ✓ 실제로 값을 구해보면 t값이 2보다 훨씬 크다는 걸 알 수 있다.

```
mu <- 250
t_value <- (y_bar - mu) / SE
[1] 17.43
```

```
> # cf.
> n <- 1000
> mu <- 500
> sigma <- 100
> ## generate simulated data:
> y <- rnorm(n, mean = 500, sd = 100)
> ## compute summary statistics:
> y_bar <- mean(y)
> SE <- sd(y) / sqrt(n)
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

단일표본 t검정 (one sample t-test)

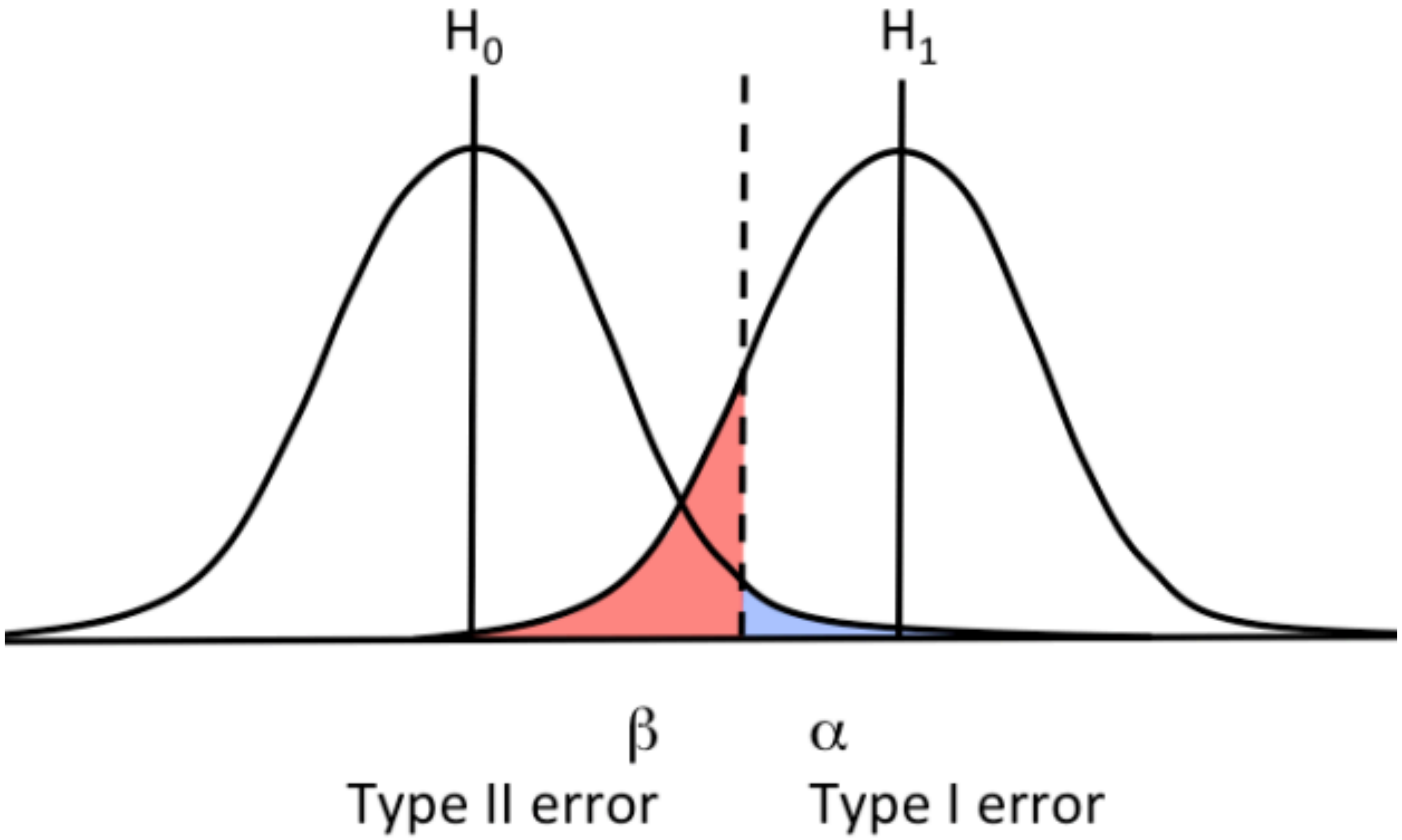
- 지금까지 우리는 이렇게 시뮬레이션을 통해서 가설 검정을 해보았다.
- 이는 비록 시뮬레이션일지라도, 실제 실험 데이터에서도 이와 동일하게 진행된다.
- 즉, 이 파트에서 우리가 지금까지 해온 것은 바로 실제 '연구문제 설정 - 귀무가설 선언 - 데이터 수집 - 가설검정'에서 일어나는 수많은 블랙박스 안을 들여다본 것이다!

Cornell University

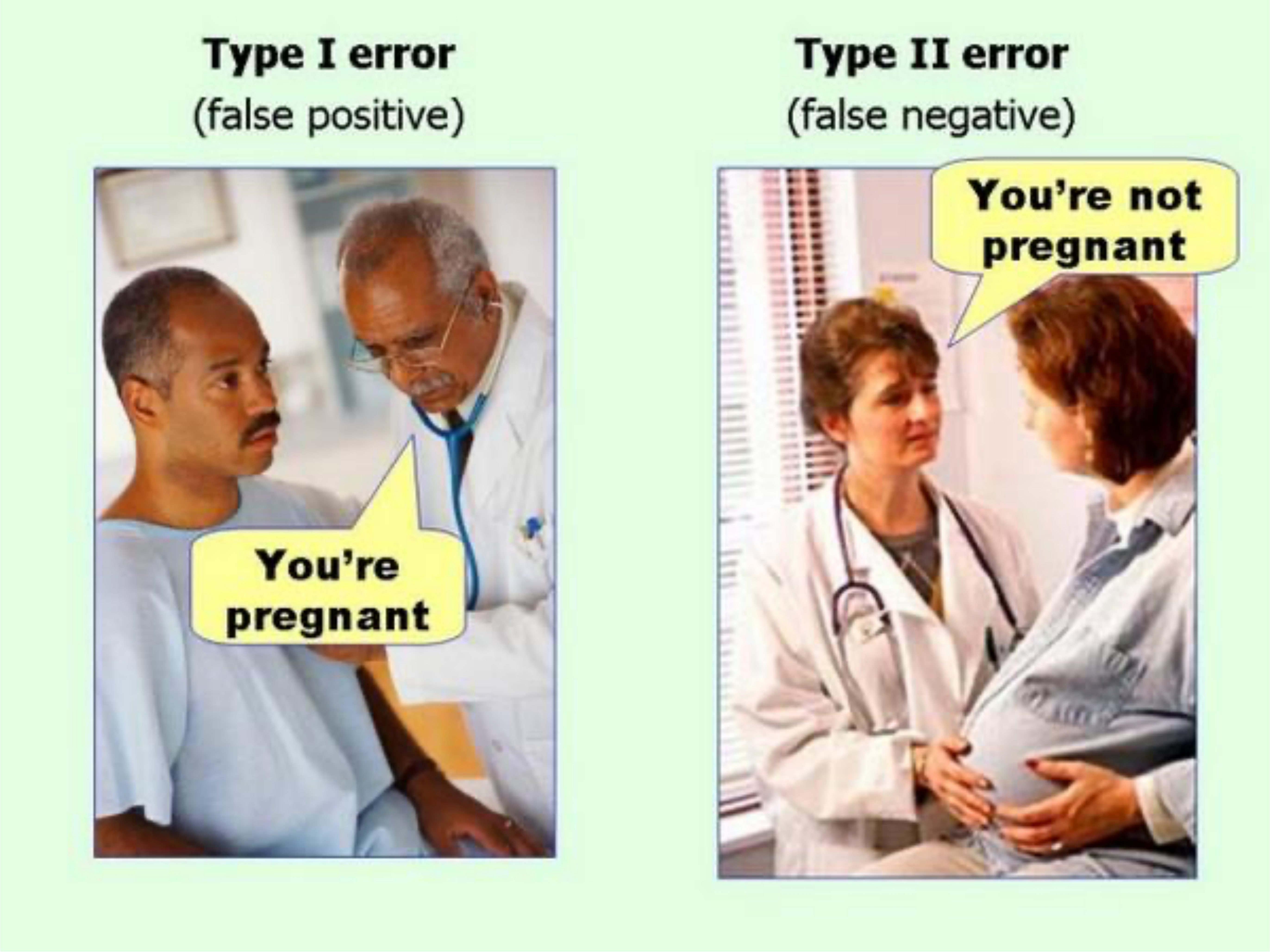
KU KONKUK UNIVERSITY

Brain & Cognition Research Center

1종 오류와 2종 오류



1종 오류와 2종 오류



Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

1종 오류와 2종 오류

- 1종 오류와 2종 오류를 유지하는 법?
- 가장 확실한 방법: 표본을 더욱 더 수집한다.

Cornell University

아래 명령어들을 복사붙여넣기 해서 본인 RStudio에 설치해주세요.

```
install.packages("devtools")
devtools::install_github("vasishth/lingpsych")
```

쉬는 시간

Cornell University

UNIVERSITÄT DES SAARLANDES

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

아래 함수를 Rstudio에 복사붙여넣기 해 실행해주세요.

```
summary_ttest <- function(res, paired = TRUE, units = "ms") {
  obs_t <- round(res$statistic, 2)
  dfs <- round(res$parameter)
  pval <- round(res$p.value, 3)
  ci <- round(res$conf.int, 2)
  est <- round(res$estimate, 2)
  if (paired == TRUE) {
    print(paste(
      paste("t(", dfs, ")=",
        obs_t,
        sep = ""
      ),
      paste("p=", pval, sep = "")))
      print(paste("est.: ", est, " [",
                  ci[1], ",", ci[2], "] ",
                  units, sep = "")
    )
  } else {
    print(paste(
      paste("t(", dfs, ")=",
        obs_t,
        sep = ""
      ),
      paste("p=", pval, sep = "")))
      print(paste(paste("est. 1: ", est[1], sep = ""),
      paste("est. 2: ", est[2], sep = ""),
      paste("CI of diff. in means: [", ci[1], ",", ci[2], "]", sep = "")))
  }
}
```

그냥 기존 내장된 t 검정 함수의 결과를 좀 더 깔끔하게 보여주는 거임.

대응 t 검정 (paired t-test)

- 앞에서 한 가설 검정 시뮬레이션을 일표본 t 검정
  - ✓ 실제론 안 쓰다시피 함.
- 실험심리언어학자가 주로 사용하는 t검정 → 대응 t 검정
  - ✓ 동일한 실험참여자에게 서로 다른 조건의 자극을 반복적으로 노출시키는 피험자 내(within-subject) 설계에 사용되는 검정!
  - ✓ 실험심리언어학에서 가장 흔한 설계
- 두 개의 서로 다른 집단 (예: 한국어 모어 화자 vs. 한국어 학습 화자) 비교를 위한 t 검정은 이표본 t 검정(two-sample t-test)을 사용.

Cornell University UNIVERSITÄT DES SAARLANDES KU KONKUK UNIVERSITY Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**
  - ✓ 영어의 주어 관계절과 목적어 관계절 처리는 서로 다르게 일어나는가?
  - ✓ 주어 관계절 조건
    - ➡ The reporter [who sent the photographer] to the editor hoped for a story.
  - ✓ 목적어 관계절 조건
    - ➡ The reporter [who the photographer sent] to the editor hoped for a story.
  - ✓ 귀무 가설: 주어 관계절, 목적어 관계절 조건 둘 다 읽기시간은 동일하다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

```
> library(lingpsych)
> data(“df_gg05e1")
> gg05e1 <- df_gg05e1
> means <- round(with(gg05e1, tapply(rawRT,
     IND = condition,
     mean
)))
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

```
> library(lingpsych)
> data(“df_gg05e1")
> gg05e1 <- df_gg05e1
> means <- round(with(gg05e1, tapply(rawRT,
    IND = condition,
    mean
)))
> means
# objgap subjgap
     471     369
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

```
> library(lingpsych)
> data("df_gg05e1")
> gg05e1 <- df_gg05e1
> means <- round(with(gg05e1, tapply(rawRT,
     IND = condition,
     mean
)))
> means
# objgap subjgap
     471     369
```

- 먼저 t 검정을 해보자.

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

```
> bysubj <- aggregate(rawRT ~ subject + condition, mean, data = gg05e1)
> byitem <- aggregate(rawRT ~ item + condition, mean, data = gg05e1)
> summary_ttest(t.test(rawRT ~ condition, paired = TRUE, bysubj))
[1] "t(41)=3.11 p=0.003"
[1] "est.: 102.29 [35.85,168.72] ms"
> summary_ttest(t.test(rawRT ~ condition, paired = TRUE, byitem))
[1] "t(15)=3.75 p=0.002"
[1] "est.: 102.29 [44.21,160.36] ms"
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

```
> bysubj <- aggregate(rawRT ~ subject + condition, mean, data = gg05e1)
> byitem <- aggregate(rawRT ~ item + condition, mean, data = gg05e1)
> summary_ttest(t.test(rawRT ~ condition, paired = TRUE, bysubj))
[1] "t(41)=3.11 p=0.003"
[1] "est.: 102.29 [35.85,168.72] ms"
> summary_ttest(t.test(rawRT ~ condition, paired = TRUE, byitem))
[1] "t(15)=3.75 p=0.002"
[1] "est.: 102.29 [44.21,160.36] ms"
```

✓ 왜 aggregate(합치기) 함수를 쓸까?

➡ t 검정의 가정 중 하나인 각각의 데이터 포인트들이 독립임을 만족시키기 위함!

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

✓ 원래 실험 데이터는 아래와 같다.

```
> t(xtabs(~ subject + condition, df_gg05e1))
         subject
condition 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
  objgap  8 8 8 8 8 8 8 8 8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
  subjgap 8 8 8 8 8 8 8 8 8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8  8
         subject
condition 30 31 32 33 34 35 36 37 38 39 40 41 42
  objgap   8  8  8  8  8  8  8  8  8  8  8  8  8
  subjgap  8  8  8  8  8  8  8  8  8  8  8  8  8
```

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

✓ 원래 실험 데이터는 아래와 같다.

```
> t(xtabs(~ item + condition, df_gg05e1))
            item
condition    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
    objgap  21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21
   subjgap  21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21
```

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

✓ 원래 실험 데이터는 아래와 같다.

```
> t(xtabs(~ item + condition, df_gg05e1))
             item
condition    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
    objgap  21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21
   subjgap  21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21
```

✓ 실험참여자와 실험재료 모두 데이터값이 독립이 아니다.

✓ 즉, 데이터 포인트가 1개 이상이다.

➡ 따라서 t 검정의 가정을 만족시키기 위해 합치기가 필요하다.

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**
- 실험참여자 및 실험재료 별로 합치기를 하면…

```
> bysubj <- aggregate(rawRT ~ subject + condition, mean, data = gg05e1)
> t(xtabs(~ subject + condition, bysubj))
         subject
condition 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
  objgap  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  subjgap 1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
         subject
condition 30 31 32 33 34 35 36 37 38 39 40 41 42
  objgap   1  1  1  1  1  1  1  1  1  1  1  1  1
  subjgap  1  1  1  1  1  1  1  1  1  1  1  1  1
```

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**
- 실험참여자 및 실험재료 별로 합치기를 하면…

```
> byitem <- aggregate(rawRT ~ item + condition, mean, data = gg05e1)
> t(xtabs(~ item + condition, byitem))
           item
condition 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
    objgap 1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
   subjgap 1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
```

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**
- 실험참여자 및 실험재료 별로 합치기를 하면…

```
> byitem <- aggregate(rawRT ~ item + condition, mean, data = gg05e1)
> t(xtabs(~ item + condition, byitem))
           item
condition 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
   objgap 1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
  subjgap 1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
```

- 데이터 포인트가 실험참여자와 실험재료 둘 다 한 개씩 됨으로써 가정을 만족하게 되었다.

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**
- 이렇게 합치기를 통해 t 검정 혹은 분산분석을 진행하는 걸 옛날 7~00년대 초중반 실험심리언어학 논문들에서 볼 수 있다.

➡ 이와 같은 분석을 흔히들 말하는 F1, F2 (혹은 quasi-F) 분석이라 한다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**
- 이렇게 합치기를 통해 t 검정 혹은 분산분석을 진행하는 걸 옛날 7~00년대 초중반 실험심리언어학 논문들에서 볼 수 있다.

  ➡ 이와 같은 분석을 흔히들 말하는 F1, F2 (혹은 quasi-F) 분석이라 한다.
- 이 때 F는 F 분포에서 따왔다.
- F1/F2 분석은 Clark (1973)에서 제안되었다.
- 사실상 실험심리언어학에서 임의 효과(random effect)를 처음 고려하는 가장 원형(prototype)적인 논문 중에 하나다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- Grodner a
- 이렇게 합치 … 중반 실험심리언
  어학 논문들

  ➡ 이와 같

- 이 때 F는 F
- F1/F2 분스
- 사실상 실험 … 원형
  (prototype

JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **12**, 335–359 (1973)

The Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research

HERBERT H. CLARK[1]

Stanford University

Current investigators of words, sentences, and other language materials almost never provide statistical evidence that their findings generalize beyond the specific sample of language materials they have chosen. Nevertheless, these same investigators do not hesitate to conclude that their findings are true for language in general. In so doing, it is argued, they are committing the language-as-fixed-effect fallacy, which can lead to serious error. The problem is illustrated for one well-known series of studies in semantic memory. With the appropriate statistics these studies are shown to provide no reliable evidence for most of the main conclusions drawn from them. A review of other experiments in semantic memory shows that many of them are likewise suspect. It is demonstrated how this fallacy can be avoided by doing the right statistics, selecting the appropriate design, and sampling by systematic procedures, or, alternatively, by proceeding according to the so-called method of single cases.

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

```
> bysubj <- aggregate(rawRT ~ subject + condition, mean, data = gg05e1)
> byitem <- aggregate(rawRT ~ item + condition, mean, data = gg05e1)
> summary_ttest(t.test(rawRT ~ condition, paired = TRUE, bysubj))
[1] "t(41)=3.11 p=0.003"
[1] "est.: 102.29 [35.85,168.72] ms”
> summary_ttest(t.test(rawRT ~ condition, paired = TRUE, byitem))
[1] "t(15)=3.75 p=0.002"
[1] "est.: 102.29 [44.21,160.36] ms"
```

✓ 하지만 합치기는 각 실험참여자가 각 조건에 대해서 ‘단 한 번의’ 측정만 갖게끔 하도록 강제하게 되는 데이터를 만들어냄.

➡ ‘반복 측정’을 하는 실험심리언어학에서는 적절하지 않은 상황!

➡ 이래서 선형 혼합 효과 모형이 사용되는 것!

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

대응 t 검정 (paired t-test)

- **Grodner and Gibson (2005) 실험1 - 자기조절읽기**

```
> bysubj <- aggregate(rawRT ~ subject + condition, mean, data = gg05e1)
> byitem <- aggregate(rawRT ~ item + condition, mean, data = gg05e1)
> summary_ttest(t.test(rawRT ~ condition, paired = TRUE, bysubj))
[1] "t(41)=3.11 p=0.003"
[1] "est.: 102.29 [35.85,168.72] ms”
> summary_ttest(t.test(rawRT ~ condition, paired = TRUE, byitem))
[1] "t(15)=3.75 p=0.002"
[1] "est.: 102.29 [44.21,160.36] ms"
```

✓ 어찌됐든, 대응 t 검정 결과 귀무가설을 기각할 수 있는 증거가 나왔다.

✓ (혹은 목적어 관계절의 효과가 읽기 시간 처리에 유의미하다는 것을 찾았다)

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

p값 제대로 해석하기

- **오해1: p 값이 (0.05보다) 작으면 작을수록 연구자가 관심 있어하는 가설을 뒷받침하는 '강력한' 증거가 생긴다.**

- p값이 유의하다는 것은 연구자가 관심 있어하는 가설을 뒷받침하는 어떤 수치적 단위가 아니며 그저 굉장히 구체적이고 좁은 영역의 귀무가설을 기각하는 것 그 이상 그 이하도 아니다.

- p값이 0.05보다 (훨씬) 작으면, 특정한 현상이 일어나거나 혹은 일어나지 않을 확률이 분포 상에서 극단치에 있다는 것을 의미한다는 것이며, 우리는 이미 0.05 보다 작아지는 그 순간부터 해당 현상이 '특이하다'라고 암묵적으로 약속하고 있을 뿐이다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

p값 제대로 해석하기

- **오해2: $p > 0.05$ 면 귀무가설은 참이다.**
- 우리는 귀무가설을 ‘기각(reject)’하거나 아니면 ‘기각하지 못하거나(fail to reject)’ 둘 중 하나다.
- 실험심리언어학뿐만 아니라 상당수 논문들이 $p > 0.05$ 일 때, ‘어떤 독립변수 X의 효과는 없는 걸로 관찰되었다’ 식의 글들을 볼 수 있는데 이는 사실 틀린 말이다.
- 우리는 그저 해당 독립변수의 효과가 있다는 증거를 찾지 못했을 뿐이다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

p값 제대로 해석하기

- **오해3: $p < 0.05$는 아니지만 '근사적으로(marginally)'하게 유의하다.**
- $p = 0.051$ 과 같은 경우가 이런 경우다.
- 하지만 근사적으로 유의하다는 말은 가설 검정에서 성립되지 않는다.
- 왜? 우리는 0.05보다 작으면 귀무가설을 기각한다고만 얘기했지, '근사적으로' 기각한다라고 얘기하지 않기 때문이다.
- 따라서 $p < 0.06$은 그냥 귀무가설을 기각하지 못한 것 그 이상 그 이하도 아니다.
- https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

정리 - 사흘 동안 다룬 것

- https://kihyo-park.github.io/stats_basic/ (수업내용)

이제 여러분이 해야하는 것 + 추천 참고문헌

- 적어도 제가 처음 통계학을 공부하기 시작했던 2014년 여름에 비하면 여러분은 그 당시 저보다 2억만배는 많이 알고 계신 상태입니다 ($p < 0.05$)
- 특히 근본적인 차원에서요! (확률론, 시뮬레이션으로 살펴보는 가설검정 과정의 블랙박스 등)
- 이제 여러분이 해야하는 것은, 선행연구들을 살펴보시면서 다양한 통계 검정법들이 어떻게 사용됐는지 직접 스스로 찾아보고(=검색해보고) 하는 것이며..
- 무엇보다 직접 본인의 데이터를 수집한 뒤에 분석을 해보는 것입니다.

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

이제 여러분이 해야하는 것 + 추천 참고문헌

- 물론 쉽지는 않을 것입니다. 그 부분은 제가 제일 잘 알고 또 공감하는 부분입니다.
- 그래서 참고문헌을 몇 개 추천해드립니다.

이제 여러분이 해야하는 것 + 추천 참고문헌

- 해외 (영어)
- http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf
- https://appliedstatisticsforlinguists.org/bwinter_stats_proofs.pdf
- 국내
- http://www.yes24.com/Product/Goods/108045614 (이번 강의에 사용한 책)
- http://www.kocw.or.kr/home/search/kemView.do?kemId=1182905&ar=relateCourse (KOCW 한국외대 언어와통계 2016년)

Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

이제 여러분이 해야하는 것 + 추천 참고문헌

- 그리고 구글!
- 아쉽지만 사실은 실험심리언어학에서 사용되는 통계분석법은 굉장히 고급통계법에 속하기 때문에, 일반적인 개론 수준의 통계학 교과서에서는 다루어지지 않는 경우가 굉장히 많습니다.
- 따라서 구글을 적극 활용하는 걸 추천드립니다.
- 주로 검색되는 키워드
- 이원분산분석(Two-way Analysis of Variance; ANOVA), 선형 회귀(Linear regression), 선형 혼합 효과 모형(Linear mixed-effect model) + “in r”, “how to do ANOVA in r” 등등..

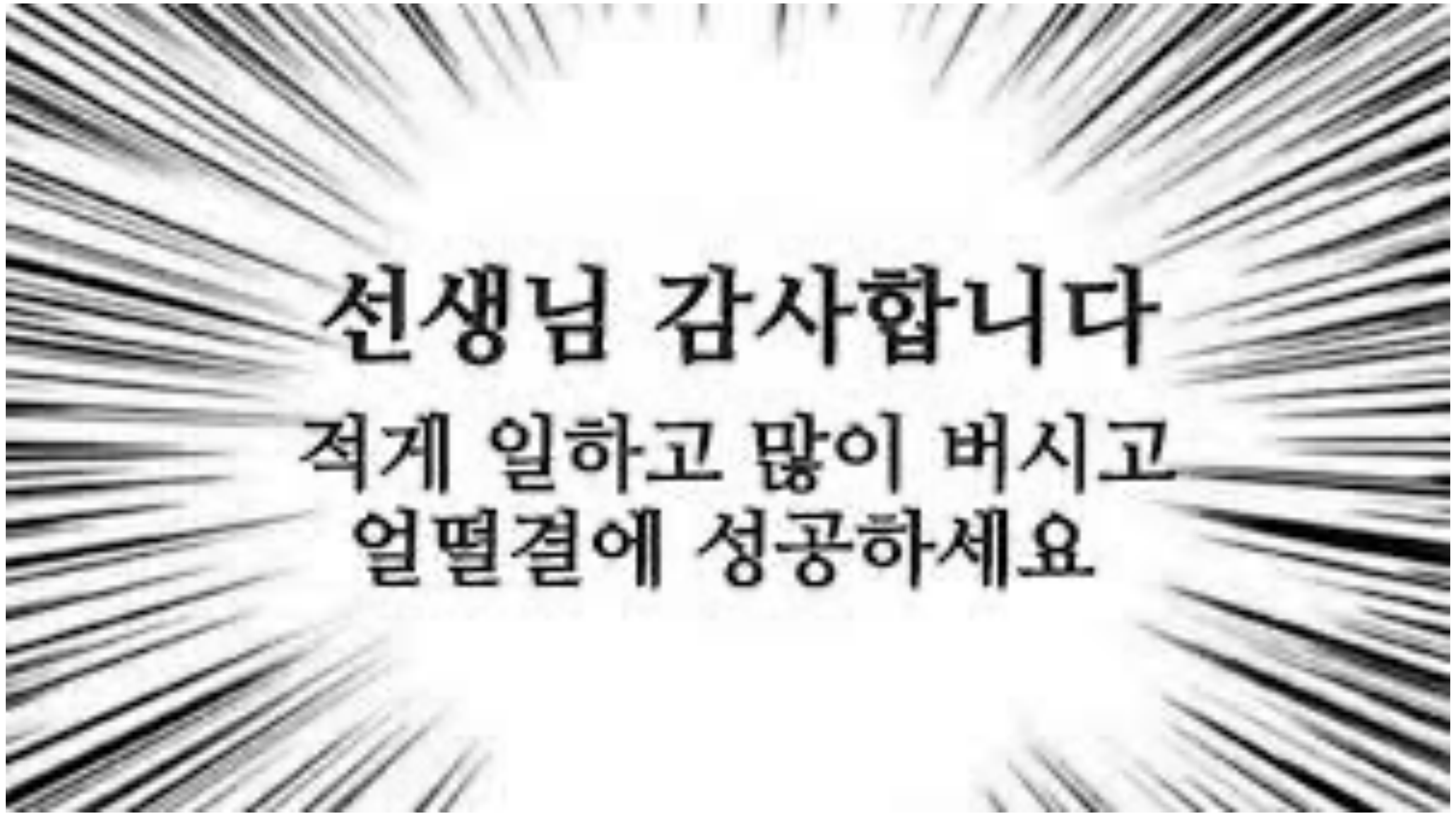
Cornell University

KU KONKUK UNIVERSITY

Brain & Cognition Research Center

이제 여러분이 해야하는 것 + 추천 참고문헌

- 무엇보다 주변 사람들과 같이 공부하고 공유하시는 걸 적극! 추천드립니다.
- 서로 가르쳐주고 배우는 것 만큼 빨리 배우는 건 없습니다.
- 그리고 윤홍옥 교수님의 통계지식을 쪽쪽(?) 빨아먹으세요(?)
- 제 것도 빨아먹으셔도(?) 됩니다. (질문 있으시면 언제든 이메일 주세요!)



귀한 시간 내주시어 끈기 있게 들어주셔서
감사드립니다!!!
수고 많으셨습니다!!