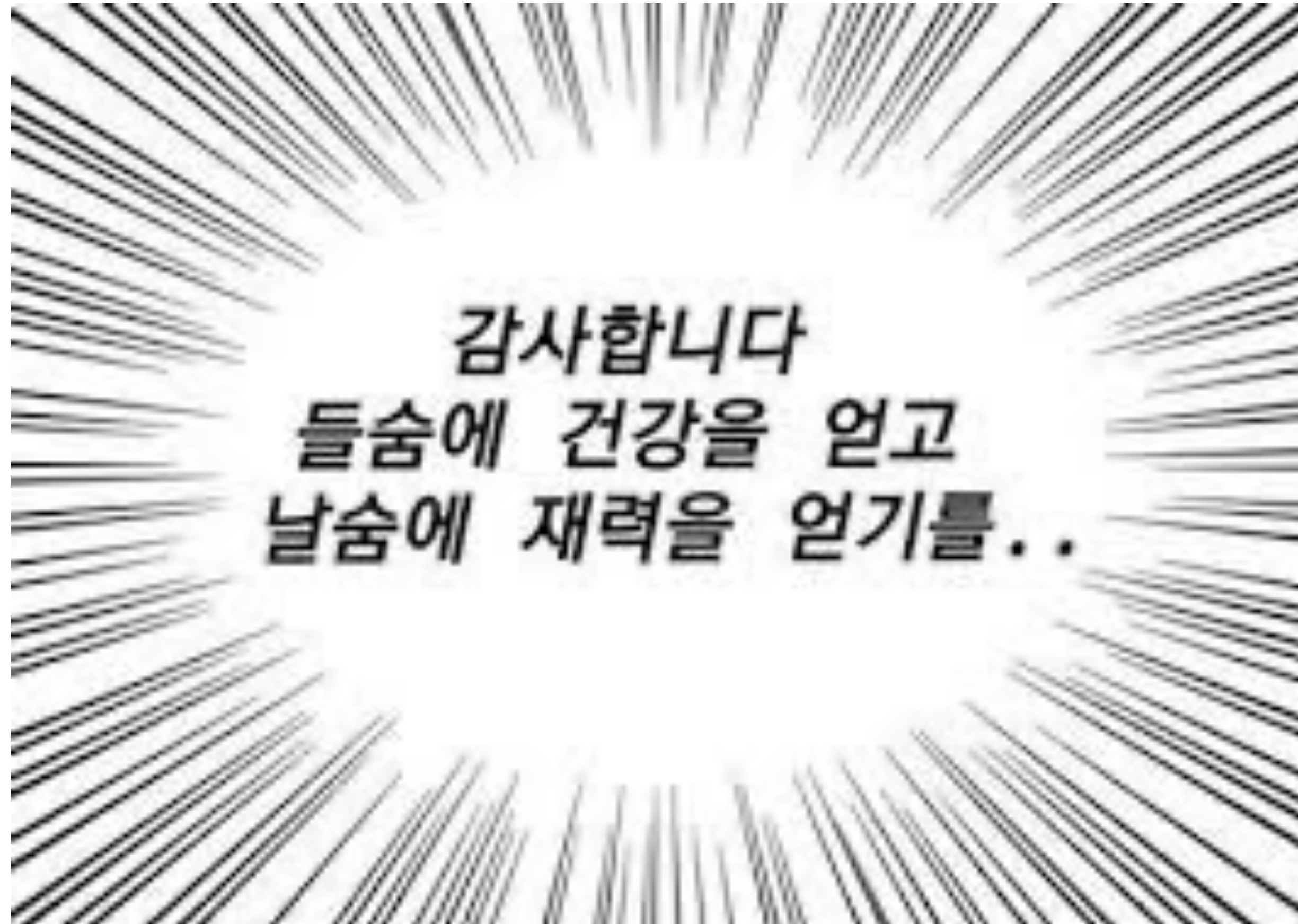


실험심리언어학도를 위한 기초 확률과 통계

0. 시작하기 앞서

2022. 07. 13.
박기효



참여해주셔서 감사합니다!



박기호 [pak.gi.ɕo]

코넬대학교 언어학 박사과정

코넬대학교 계산심리언어학 연구그룹 및 언어의미 연구실

건국대학교 뇌인지연구센터 방문연구원

수업 내용 - 7월 13일(수)

- 시작하기 앞서...
- 통계와 확률이 줄 수 있는 것과 못 주는 것
- R과 통계학
- 기술통계와 추론통계
 - ➔ (여러분께) 바라는 것: 통계학과 R이랑 친해지기

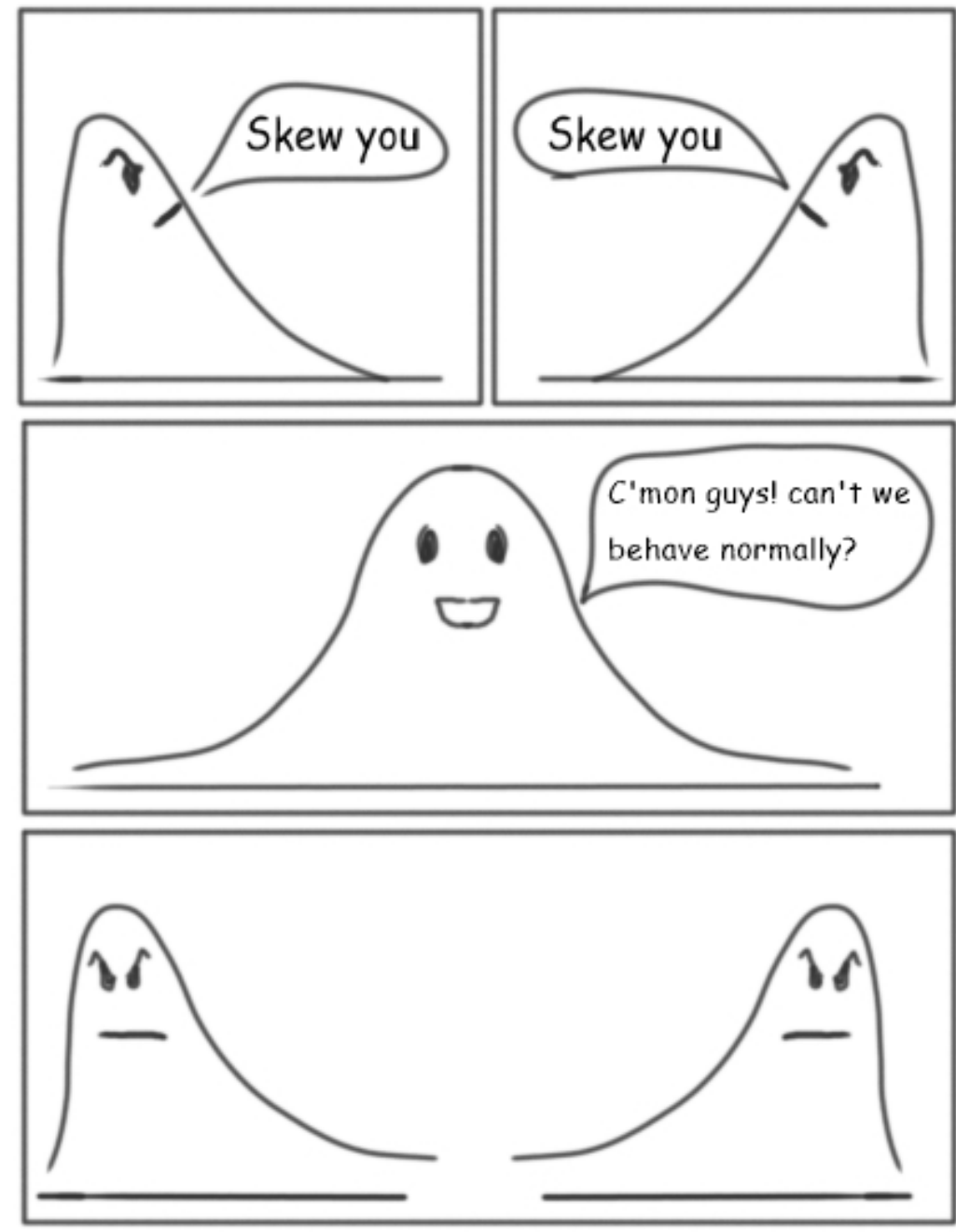
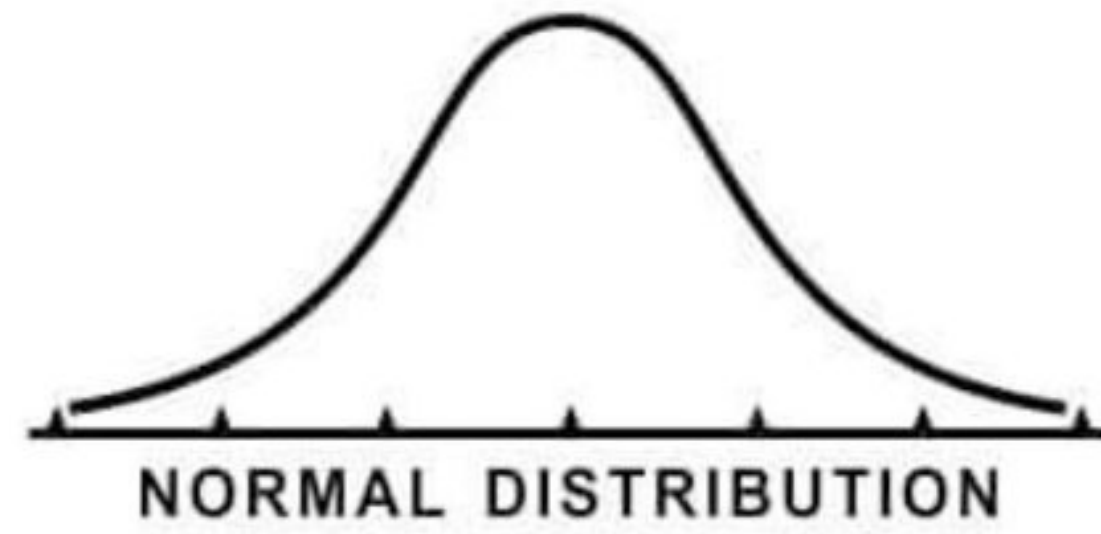
수업 내용 - 7월 14일(목)

- 확률
- 확률분포
- 표본분포와 중심극한정리

➔ (여러분께) 바라는 것: 통계학의 근간이 되는 분야 중 하나인 확률론과 친해지기

수업 내용 - 7월 15일(금)

- 추정
- 가설검정(I, II)
 - ➔ (여러분께) 바라는 것: 추정과 가설검정이랑 친해지기



실험심리언어학도를 위한 기초 확률과 통계

1-1. 통계와 확률이 줄 수 있는 것과 못 주는 것

2022. 07. 14.
박기효

과학과 이론, 그리고 통계

- 한 걸음 잠깐 물러나서 이야기를 살짝 넓게 바라봅시다.

과학과 이론, 그리고 통계

- 한 걸음 잠깐 물러나서 이야기를 살짝 넓게 바라봅시다.
- 우리는 종종 관심 있어하는 현상을 발견하면 그것이 어떤 패턴을 보이는지 궁금해합니다.

과학과 이론, 그리고 통계

- 한 걸음 잠깐 물러나서 이야기를 살짝 넓게 바라봅시다.
- 우리는 종종 관심 있어하는 현상을 발견하면 그것이 어떤 패턴을 보이는지 궁금해합니다.
- 예를 들어...

과학과 이론, 그리고 통계

- 한 걸음 잠깐 물러나서 이야기를 살짝 넓게 바라봅시다.
- 우리는 종종 관심 있어하는 현상을 발견하면 그것이 어떤 패턴을 보이는지 궁금해합니다.
- 예를 들어...
 - ✓ 은지는 경찰에게 ().
 - ✓ 은지는 신문지를 ().

과학과 이론, 그리고 통계

- 한 걸음 잠깐 물러나서 이야기를 살짝 넓게 바라봅시다.
 - 우리는 종종 관심 있어하는 현상을 발견하면 그것이 어떤 패턴을 보이는지 궁금해합니다.
 - 예를 들어...
 - ✓ 은지는 경찰에게 ().
 - ✓ 은지는 신문지를 ().
- ➡ 일반화 해보기 (칠판)

cf. Yun and Hong (2014) - Journal of Cognitive Science, 15(3).

과학과 이론, 그리고 통계

- 방금 우리가 한 것
 - ✓ 일반화를 통해 질서가 없어보이는 것들에 질서를 부여
 - ✓ 즉, 특정 현상 뒤에 숨겨진 질서를 찾으려고 생각했고, 행동함
- 이 일반화라는 것은 좀 더 멋드러지게 표현하면 ‘이론’이라 함
- 그리고 이 이론이라는 것들을 지닌 집합을 우리는 ‘과학’이라 부름
 - ➡ 결국 이론과 과학이란 것은 ‘질서를 발견하려는 노력’

과학과 이론, 그리고 통계

- 질서를 발견해서 무엇을 할까?
- 예측
 - ✓ 특정 결과값 예견
 - ✓ 특정 기제(mechanism)의 상태와 조건 예상
- 이해
 - ✓ 특정 기제를 이루는 단위들이 서로 어떻게 상호작용하는가?
 - ✓ 그리고 그 상호작용은 어떤 과정을 통해서 일어나는가?

과학과 이론, 그리고 통계

- 도구이자 방법으로서의 통계학
 - ✓ 질서 발견에 도움을 주는 방법!
 - ✓ 주어진 데이터를 분석하게끔 해주는 도구이자 곧 방법!
- 하지만 통계의 맹점 또한 존재
 - ✓ 전혀 관련이 없어 보이는 데이터도 관련이 있는 것처럼 보이는 경우가 종종 있음.
 - ➡ 허위 관련성(spurious correlation)

과학과 이론, 그리고 통계

- 통계학이 과학적 도구이자 방법으로 제 기능을 하려면 허무맹랑한 주장이 아니어야함
 - ✓ 즉, 기존 제시된 다양한 이론, 관찰, 그리고 연구자의 (훈련된) 직관과 통찰을 기반으로 한 논리적 주장이 있어야지 통계학은 제 기능을 함

과학과 이론, 그리고 통계

- 이는 구체적인 선언문(statement)으로서 가설이 있어야지만이 통계학을 올바르게 사용할 수 있다는 뜻!
- 가설의 예
 - ✓ “한국어의 기본어순은 주어-목적어-술어다.”
 - ✓ “문법적인 문장보다 비문법적 문장에서 읽기 시간이 더 늘어난다.”

과학과 이론, 그리고 통계

- 다시 아래 예를 살펴보자.
- 은지는 경찰에게 ().
- 은지는 신문지를 ().

➡ 가설?

과학과 이론, 그리고 통계

- 결국 과학, 이론, 통계는 서로 맞물리는 관계
- 이를 좀 더 도식적으로 표현하면 아래와 같음
 - ✓ 관찰 ↔ 경험 ↔ 실험
- 무엇보다 위 과정은 어느 누구나 할 수 있는 것!
- 그리고 지금 이 슬라이드를 보고 있는 여러분들도 충분히 하고도 남는 것!

확률론과 통계학

- 확률

- ✓ 일반 → 특정

- ✓ 모집단 → 표본

- ✓ 모형 → 자료

- 통계학

- ✓ 일반 ← 특정

- ✓ 모집단 ← 표본

- ✓ 모형 ← 자료

확률론과 통계학

- 확률론은 통계학의 근간을 이루는 학문으로서 반드시 살펴보아야하는 분야
- 확률론 없이 통계학을 배운다는 건 물 없이 밥 먹다가 결국 체하는 거랑 똑같음
- 따라서 확률론에 대한 기본적인 지식은 반드시 필요! (목요일에 배울 예정)

통계학에 대한 오해들 1



자료 싹 다 가져왔고
사례비도 드리겠습니다!
분석해주세요!

?!



통계학에 대한 오해들 1



구체적인 목적이 없으시면
도와드릴 수가 없습니다.



통계학에 대한 오해들 1



통계학은 자동으로 의미 있는
패턴을 찾아주는 것이 아님!



통계학에 대한 오해들 1



통계 분석 진행 전 연구 질문 설정 필요!

→ 가설의 중요성!

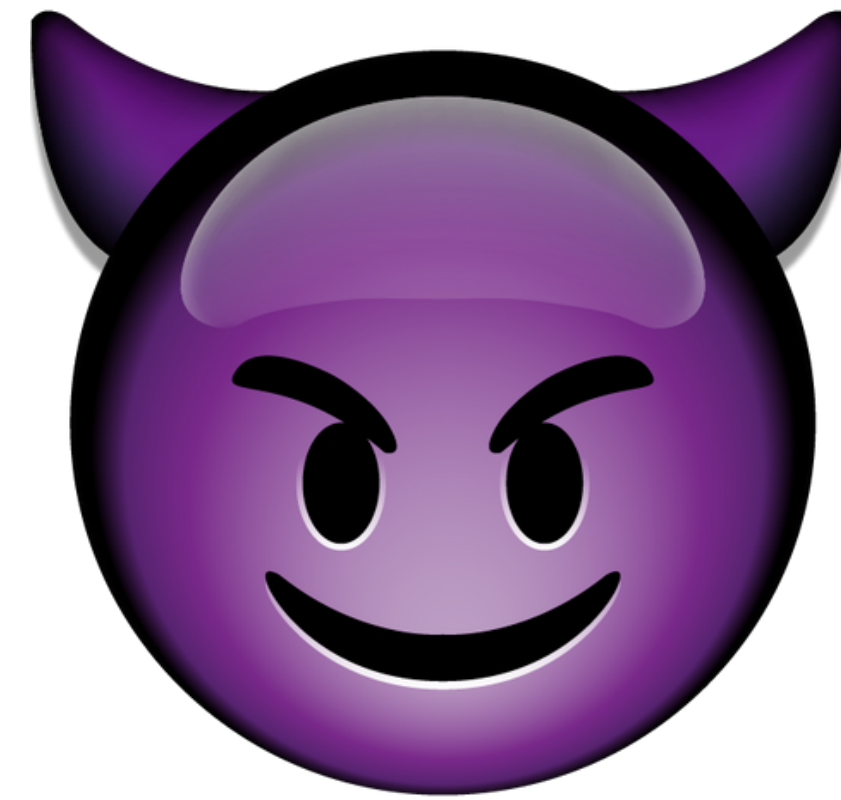


통계학에 대한 오해들 2

통계는 데이터를 넣으면 정답을 알려주는 요술봉!



통계는 그럴 듯하게 사기를 치는 방법!



통계학에 대한 오해들 2

통계는 데이터를 보면 정답을
알려주는 마법봉!



통계는 그럴듯한 사기를 치는



통계학에 대한 오해들 2

통계는 데이터를 보면 정답을 알려주는 마법봉!

통계는 그럴듯한 사기를 치는

통계학은 양적 비판을 하기 위한 과학적 도구!

→ 과연 이 분석이 충분히 타당한가?

→ 가장 합리적인 분석인가?



통계학에 대한 오해들 3

- 통계학은 언제나 단 하나의 답을 주는 것은 아님
- 많은 경우 대안적 분석 혹은 대안적 모형 존재
- 그러나 명백한 오류에 대한 비판이나, 최선에 가까운 합리적 추론은 통계학을 통해서 가능

(도대체) 왜 통계를 배워야 하나요?

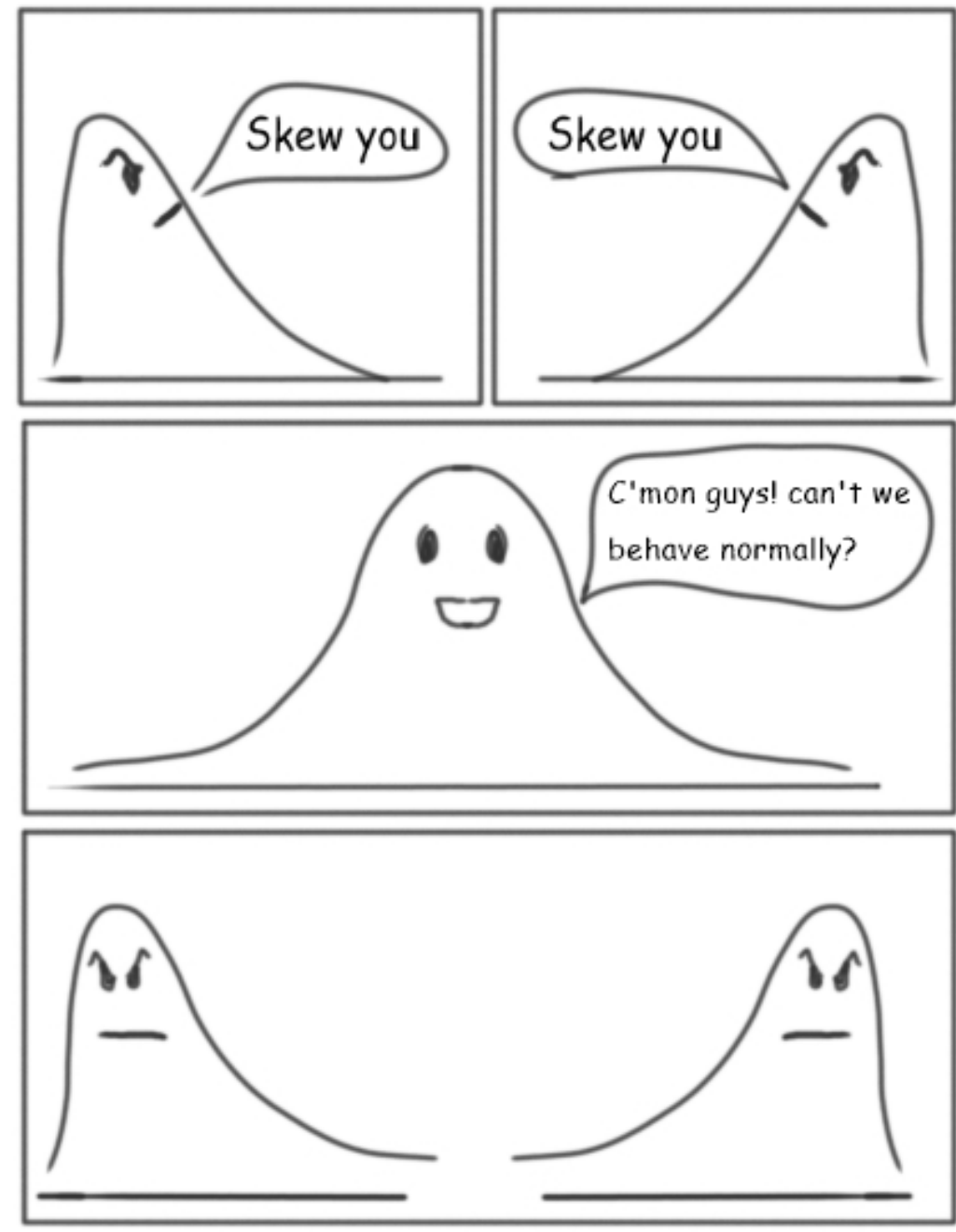
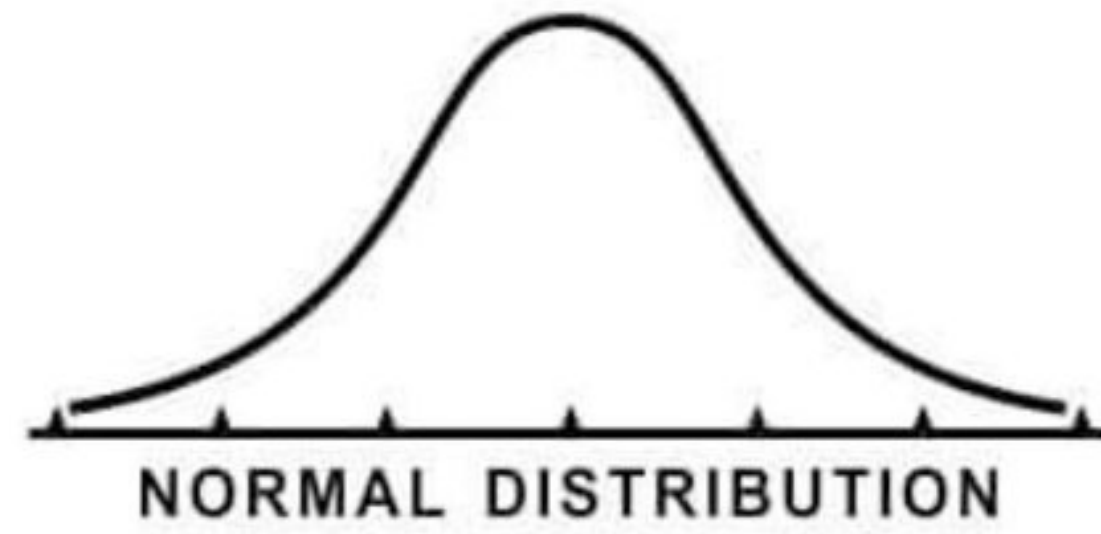
- 양적 연구에서 기본적인 설득력을 갖추기 위해 (=실험연구자는 숫자로 말한다)
- 다른 사람의 데이터에 근거한 논증을 이해하기 위해
- 모르고 사기를 안 치기 위해 (잘못된 분석의 사용)
- 남들의 사기나 헛점을 발견하기 위해

(도대체) 왜 통계학을 배워야하나요?

- 스스로와 타인의 연구를 비판적으로 바라보기 위해!

(도대체) 왜 통계를 배워야 하나요?

- 결국 통계학이란 것은 합리적 판단과 의사결정을 위해 데이터를 활용하는 과학적 도구!
- 특정 언어 현상에 대한 질서를 찾기 위해 가설을 세우고, 실험을 통해 관련 데이터를 얻고 분석하기 위해서 통계학은 반드시 필요한 도구!
- 따라서 이 도구를 제대로, 적절하게 사용함으로써 가설을 과학적으로 검증함으로써 인간의 언어능력 연구에 기여!



실험심리언어학도를 위한 기초 확률과 통계

2. R과 통계학

2022. 07. 14.
박기효

(하필) 왜 R이예요?

- 기본적으로 컴퓨터(computer)의 기본적인 뜻은 ‘계산기’
- 하지만 계산기도 보통 크게 두 가지로 나눔
 - ✓ 일반 계산기
 - ✓ 공학용 계산기
 - ➡ R은 통계학적 계산을 위한 공학용 계산기

SPSS 쓰면 안돼요?

- 써도 상관 없습니다.
- 저도 처음엔 SPSS로 시작했습니다.
- 하지만 SPSS는 기본적으로 학교계정을 쓰지 못하면 유료입니다.
- 저는 유료가 아니라 무료인 R이 좋습니다. (그냥 공짜면 전부 좋습니다.)
- 무엇보다 R은 이미 수많은 통계학자들이 사용하고 있습니다.
- 즉, 통계학자들이 직접 제작한 통계 알고리즘 대다수가 R을 활용해 만들어졌고, 지금도 만들어지고 있습니다.

코딩 어려워요 ㅠㅠ

- 저도 어렵습니다 ㅠㅠ
- 하지만 코딩, 혹은 프로그래밍이란 것은 컴퓨터라는 기계와 대화를 하는 것과 동일합니다.
- 인간: “지뢰찾기 찾아서 틀어줘”
- 컴퓨터: 시작-프로그램-보조 프로그램-게임-지뢰찾기
- R 또한 이의 연장선입니다.
- 대신 R은 데이터 분석과 통계적 계산이란 대화방식에 특화된 것일뿐이죠.

R과 RStudio 설치하기

- 윈도우 계정명 영어로 바꾸기
- <https://comeinsidebox.com/change-windows-10-account-name/>
- R과 RStudio 설치하기
- <https://iamoverthemoon.tistory.com/64>

RStudio로 하는 간단한 산수

- 이제부터 우리는 RStudio만 쓸 겁니다.
- 왜? 그냥 R보다 훨씬 편하니까요.

RStudio로 하는 간단한 산수 (스크립트)

- 더하기
- 빼기
- 빼기? (3 -)
- 나누기
- 곱하기
- 제공
- 괄호

RStudio로 하는 간단한 산수 (스크립트)

- 제공근
- 제공근?
- 절댓값

여러분 스스로가 만드는 첫 R 스크립트

- 주석처리 (#)
- 변수 설정하기
- 스크립트 저장하기

여러분 스스로가 만드는 첫 R 스크립트

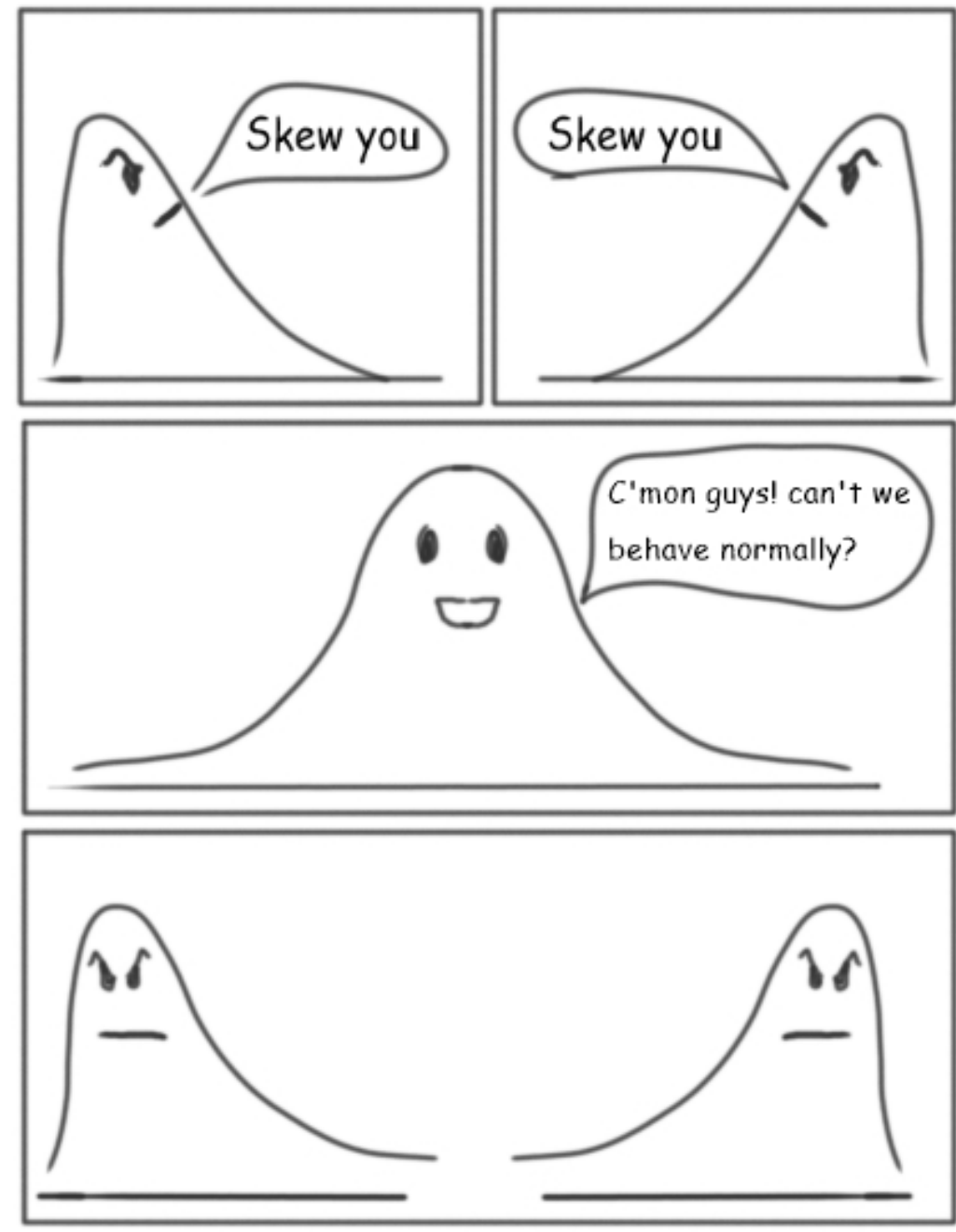
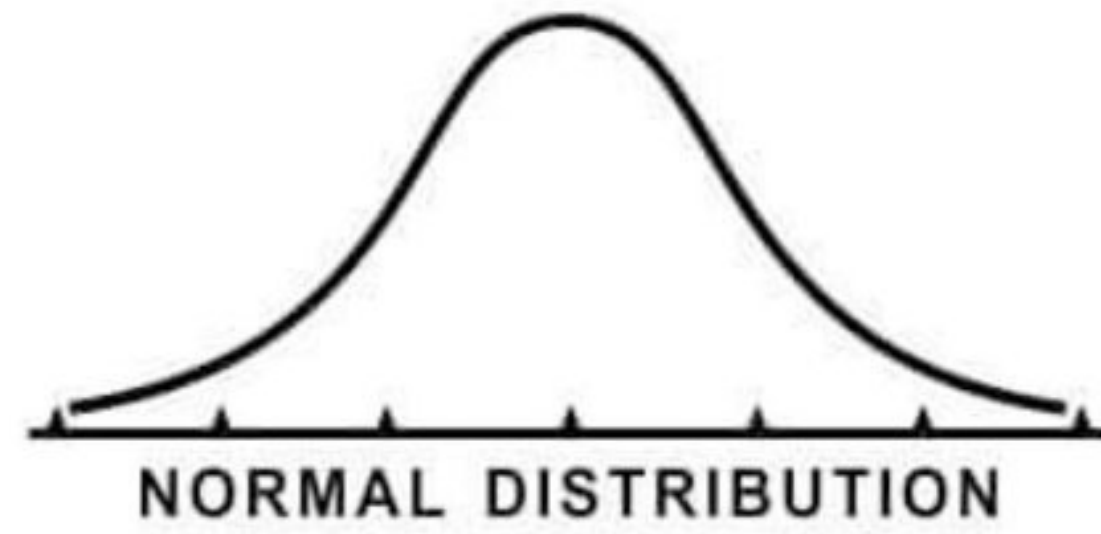
- 아니 이렇게 간단한 걸 갖고서 언제 통계분석을 해요?
- 물론 못합니다!
- 하지만 이 간단한 걸 모르면 앞으로 RStudio를 활용해서 여러분의 졸업 논문 데이터를 분석하는게 불가능해지고, 윤희옥 교수님께서서 (아마) 나중에 수업 때 공유해주시는 R 스크립트에 담긴 코드들이 무슨 말인지 하나도 이해하지 못하실 겁니다.
- 가장 간단한 것일수록 가장 기본적인 것!
- 그리고 가장 기본적인 것일수록 모르면 안되는 것!
- 계속 이 기본적인 것들을 살펴볼까요?

R 세상 속 공용어 (스크립트)

- 벡터(Vectors)
- R에서는 그냥 숫자, 문자 등을 나열한 걸 벡터라고 부릅니다. R 세상 속 공용어에 담긴 약속 그 이상 그 이하도 아님.
- 수치 벡터(Numeric vectors)
- 색인(Indexing)
- 논리 벡터(Logical vectors)
- 문자 벡터(Character vectors)
- 요인 벡터(Factor vectors)

R 세상 속 공용어 (스크립트)

- 데이터 프레임(Data Frames)
- 패키지 설치하고 부르기
- 파일 부르기



실험심리언어학도를 위한 기초 확률과 통계

1-3. 기술 통계와 추론 통계

2022. 07. 14.
박기효

기술통계와 추론통계

- 기술통계

- ✓ 관측된 데이터를 일목요연하게 요약하는 방법

- ➡ 표, 그림, 그래프 등을 활용해 수치를 정리하고 요약

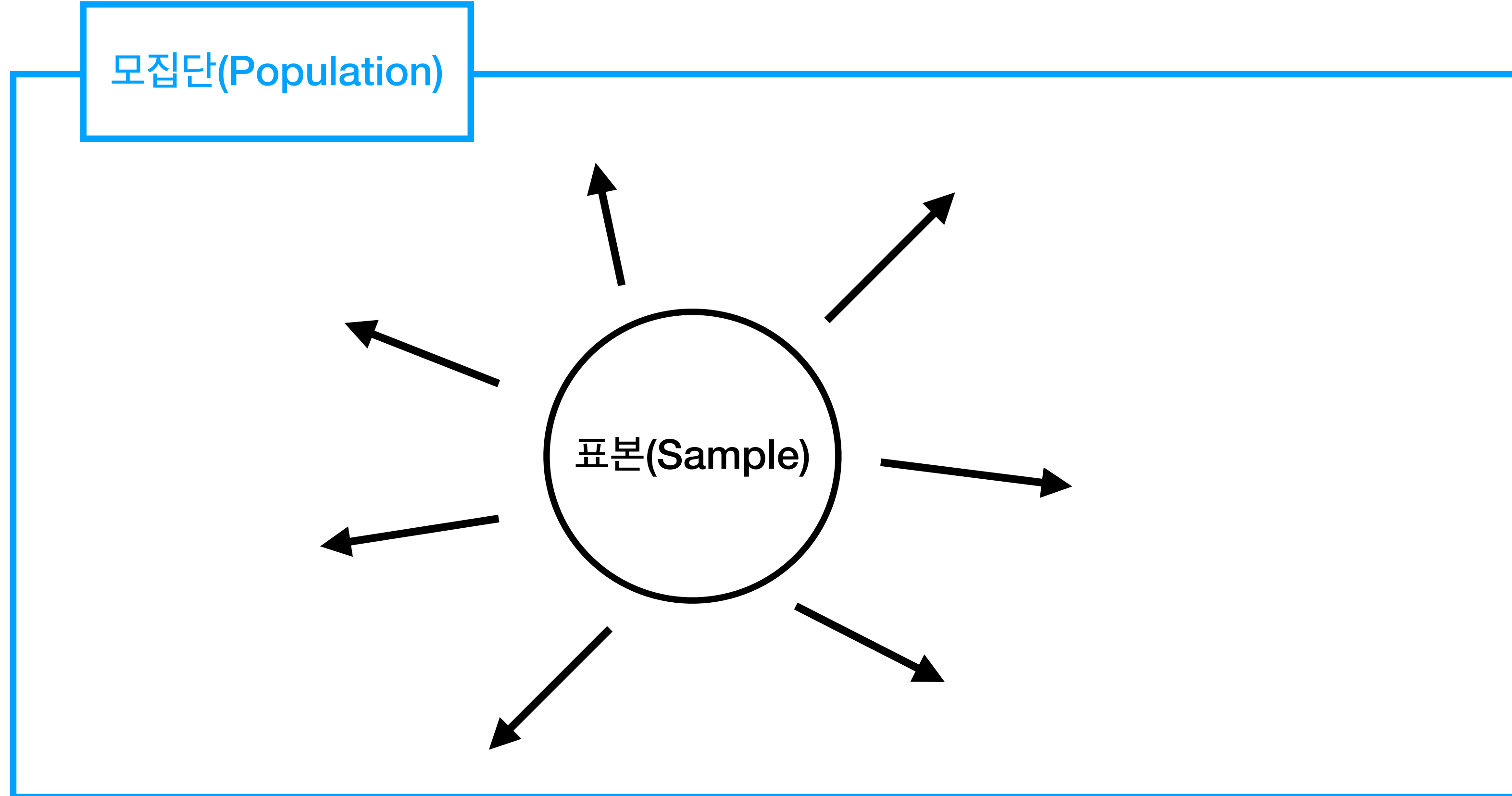
- 추론 통계

- ✓ 관측된 데이터를 바탕으로 관측되지 않은 영역(예: 모집단)에 대한 정보를 추론하는 방법

- ➡ 보통 통계학 공부는 추론 통계

기술통계와 추론통계

- 추론 통계



데이터 요약과 시각화 - 기술통계

- 수치형 기술통계
- 수치를 이용해서 데이터를 정리 요약하는 것
- 평균, 표준편차, 분산 등등..

데이터 요약과 시각화 - 위치 관련 수치

- “비문법적인 문장에 대한 1번 실험참여자의 읽기 시간은 전체 데이터에서 어느 정도에 위치할까?”
- 어떤 관측값 x_1, x_2, \dots, x_i 가 N 개 있을 때...
 - ✓ 평균(mean) (혹은 산술평균(arithmetic mean))

$$\rightarrow \frac{1}{N} \sum_{i=1}^N x_i$$

데이터 요약과 시각화 - 위치 관련 수치

- 평균의 대한 계산은 기본적으로 모집단에 대한 것이나, 표본에 대한 것이나 동일하다.
- 그러나 모평균과 표본평균은 각각 다음과 같이 표기한다.

$$\checkmark \text{ 모평균: } \mu = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\checkmark \text{ 표본평균: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

데이터 요약과 시각화 - 위치 관련 수치

- 한편, 평균은 특정 관측값(들)에 의해 크게 좌우되는 경향이 있다.
- 세 명의 실험참여자의 읽기 시간이 250ms, 256ms, 260ms이라 할 때, 500ms 읽기 시간을 가진 실험참여자의 관측값이 추가 되면 평균은 상승한다.
- 실제 데이터 분석에서 우리는 보통 이렇게 평균 등에 영향을 끼치는 관측값을 특이치 (outlier)라고 부른다.
- 그렇담 특이치들의 영향을 받지 않고 데이터들의 위치를 분석할 수 있는 수치가 있을까?
✓ 있다!

데이터 요약과 시각화 - 위치 관련 수치

- 중위수(median)

✓ 데이터를 크기 순으로 나열했을 때 가운데에 해당하는 관측값

➡ N 이 홀수일 때의 중위수: $\frac{N+1}{2}$ 번째 수(값)

➡ N 이 짝수일 때의 중위수: $\frac{N}{2}$ 번째 수(값) & $\frac{N}{2} + 1$ 번째 수(값)

- median()

데이터 요약과 시각화 - 위치 관련 수치

- 한편, 주어진 데이터에서 어떤 값이 가장 자주 나오는지 궁금해할 수도 있음.
- “한국어 단어에서 가장 많이 사용되는 단어는 무엇일까?”
- 최빈값(mode)
 - ✓ 가장 많은 빈도를 가지고 있는 수치
- table()

데이터 요약과 시각화 - 변동성 관련 수치

- 관측값이 서로 얼마나 다를까?
- 주어진 영어 문장이 비문법적이다라는 걸 맞춘 정답률 데이터가 있다고 해보자.
- 그리고 평균 정답률이 0.8(모두 다 맞추었으면 1.0이므로)이라고 해보자.
- 앗! 근데 알고보니 이 정답률 데이터에서 수집된 실험참여자들이 사실은 영어영문학과 전공자와 비전공자로 이루어져 있다는 걸 알게 됐다.
- 그렇다면 전공자의 정답률과 비전공자의 정답률이 서로 다를 수도 있다고 추론해볼 수 있으며, 이를 뒷받침해줄 수 있는 수치가 있지 않을까?

✓ 변동성 수치

데이터 요약과 시각화 - 변동성 관련 수치

- 범위(range)
 - ✓ 최소값과 최대값의 차이 = 최대값 - 최소값
- sort()
- range()

데이터 요약과 시각화 - 변동성 관련 수치

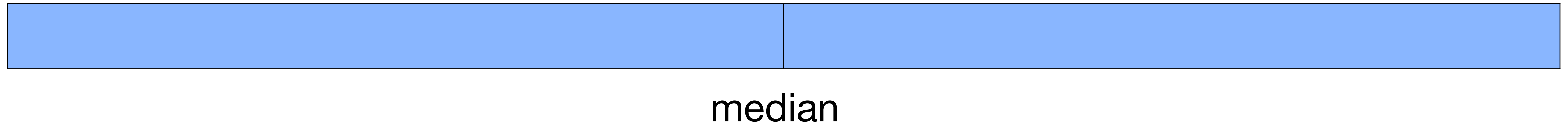
- 분위수(quantile)
 - 데이터를 오름차순 또는 내림차순으로 정렬한 뒤, 전체 데이터의 수를 특정의 개수로 나눌 때 기준이 되는 수
 - ✓ 2분위수: 전체 데이터의 개수를 2개로 나눔
 - ✓ 4분위수: 전체 데이터의 개수를 4개로 나눔
- ➡ 이와 같이 나누는 방식을 q 분위수(q -quantiles)라고 함.

데이터 요약과 시각화 - 변동성 관련 수치

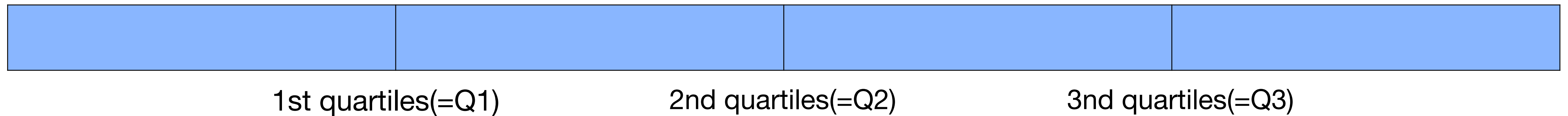
- 2분위수(2-quantiles) = median
- 4분위수(4-quantile) = quartiles
- 10분위수(10-quantiles) = deciles
- 100분위수(100-quantiles) = percentiles

데이터 요약과 시각화 - 변동성 관련 수치

- 2분위수(2-quantiles) = median



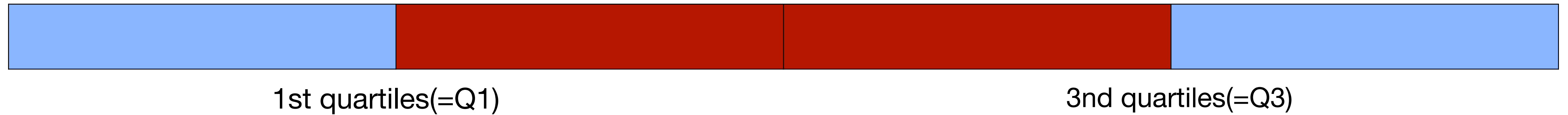
- 4분위수(4-quantiles) = quartiles



- `sort()`
- `quantile(data, type = 3)`

데이터 요약과 시각화 - 변동성 관련 수치

- 4분위간 범위(interquartile range) = $Q3 - Q1$



- `IQR(data, type = 3)`
- `boxplot()`

데이터 요약과 시각화 - 변동성 관련 수치

- 분산(Variance)

- ✓ 관측값들이 평균으로부터 얼마나 벗어나 있는가?

- ✓ 데이터의 분포가 어떻게 이루어져 있는가? (위에보다 더 중요한 정보!)

➔ 모분산: $\sigma^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N}$

➔ 표본분산: $s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}$

데이터 요약과 시각화 - 변동성 관련 수치

- 어? 근데 왜 표본분산에선 N 이 아니라 $n - 1$ 인가요?
- 표본분산은 모집단을 추정하기 위한 통계량(statistic)임.
(cf. 표본평균도 똑같이 통계량에 속함.)
- 결국 표본이란 것은 결국 모집단의 성질을 (어느정도) 갖고 있으나 그 크기가 훨씬 작은 것.
- 표본분산과 모분산 공식을 다시 한 번 비교해보자.

데이터 요약과 시각화 - 변동성 관련 수치

- $\sum_{i=1}^N \frac{(x_i - \mu)^2}{N}$ vs. $\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$

- 위 각각의 식에서 분자는 편차제곱합이라고 한다.
- 이 때, 모집단과 표본의 편차제곱합이 서로 동일하다고 가정해보자.
- 표본은 천성적으로 모집단보다 그 크기가 작을 수밖에 없다.

데이터 요약과 시각화 - 변동성 관련 수치

- 하지만 우리는 표본을 이용해서 모집단을 ‘추정’하고자 한다.
- 그렇다면 우리는 표본을 ‘뺑튀기’를 시켜서 모집단과 크기가 비슷하게 보여야한다.
- 만약 이렇게 하지 않으면 우리는 과소추정의 오류를 범하게 된다.
- 어차피 숫자만 뺑튀기시키는 것이지 관측대상을 조작하는 것이 아니기 때문에 괜찮다.
 - ✓ 그래서 하는 게 바로 분모의 값을 줄이는 것!

데이터 요약과 시각화 - 변동성 관련 수치

- 그렇담 왜 하고 많은 것 중에 $n - 1$ 인가? (예: 그냥 분모를 2로 고정하면 되지 않나요?)
- 이는 우리가 표본에 대한 계산을 할 때 항상 ‘추정치’가 들어가기 때문!
- 표본분산에서 추정치는 평균이다.
- 그리고 이 추정치는 항상 특정한 값을 가지게 되므로, 그 자체로서 하나의 ‘제약’이 된다.
- 따라서 우리는 관측된 대상의 개수 n 에서 추정치의 개수만큼 빼줘야하는데, 표본분산의 경우 추정치가 평균 1개이므로 $n - 1$ 을 하는 것이다.
- 그리고 우리는 이를 자유도(degree of freedom)라 부른다.

데이터 요약과 시각화 - 변동성 관련 수치

- 표준편차(Standard deviation)

- ✓ 원론적으로 분산의 기능과 동일

- ✓ 대신 분산의 값이 제곱으로 뱅튀기 되어있어서 루트를 씌워 보기 쉽게하는 것.

→ 모표준편차: $\sigma = \sqrt{\sum_{i=1}^N \frac{(x_i - \mu)^2}{N}}$

→ 표본표준편차: $s = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}}$

(추론)통계로 우리가 결국하게 되는 것들

1. 모형 만들기 (추정)
2. 여러 모형 중 좋은 거 고르기 (모형 선택)
3. 모형이 제대로 된 건지 확인 (모형진단)
4. 모형이 틀렸는지 보기 (가설 검정)
5. 모형으로 모르는 것 예측하기 (예측)
6. 예측한 내용으로 의사결정하기 (의사결정)

통계적 모형

- 표본 데이터의 생성에 대한 통계적 가정을 구체화하는 (수학적) 모형
- 즉, “데이터가 어떠한 (확률) 분포로 만들어진다.”라는 가정
- 실제로 ‘가설’은 수학적으로 위와 같은 방식으로 표현
- ‘수학적으로’란 말에 겁 먹지 말자. 사실 별 거 없다.

통계적 모형

- 다음 중 통계적 모형은 뭘까?

통계적 모형

- 다음 중 통계적 모형은 뭘까?
- 임의의 한국 남성을 뽑으면 체중이 72.7kg이다.

통계적 모형

- 다음 중 통계적 모형은 뭘까?
- 임의의 한국 남성을 뽑으면 체중이 72.7kg이다.
- 한국 남성의 체중은 평균 72.7kg이다.

통계적 모형

- 다음 중 통계적 모형은 뭘까?
- 임의의 한국 남성을 뽑으면 체중이 72.7kg이다.
- 한국 남성의 체중은 평균 72.7kg이다.
- 한국 남성의 체중은 평균 72.7kg, 표준 편차 9.53이다.

통계적 모형

- 다음 중 통계적 모형은 뭘까?
- 임의의 한국 남성을 뽑으면 체중이 72.7kg이다.
- 한국 남성의 체중은 평균 72.7kg이다.
- 한국 남성의 체중은 평균 72.7kg, 표준 편차 9.53이다.
- 한국 남성의 체중은 평균 72.7kg, 표준 편차 9.53의 정규 분포를 따르며 각각 서로 독립이다.

통계적 모형

- 다음 중 통계적 모형은 뭘까?
- 임의의 한국 남성을 뽑으면 체중이 72.7kg이다.
- 한국 남성의 체중은 평균 72.7kg이다.
- 한국 남성의 체중은 평균 72.7kg, 표준 편차 9.53이다.
- $w_i \stackrel{iid}{\sim} N(72.7, 9.53), i = 1, \dots, n$

과학적 지식 획득을 위한 모형 vs. 의사결정 모형

- 과학적 지식 획득을 위한 모형
 - ✓ 적은 수의 해석 가능한 변수들

- 실용적 의사결정을 위한 모형
 - ✓ 예측정확도에 기여하기만 하면 해석 가능성이 낮아도 충분히 실용적 (기계학습, 경영, 공학 등)

과학적 지식 획득을 위한 모형 vs. 의사결정 모형

- 과학적 지식 획득을 위한 모형
 - ✓ 적은 수의 해석 가능한 변수들
 - ➡ 우리가 하는 통계학은 이것!
- 실용적 의사결정을 위한 모형
 - ✓ 예측정확도에 기여하기만 하면 해석 가능성이 낮아도 충분히 실용적 (기계학습, 경영, 공학 등)

모수 통계와 비모수 통계

- 모수 통계
 - ✓ 한정된 수의 모수(parameter)를 가정해 특정 확률 분포를 활용하는 모델링
 - ✓ 올바른 가정에서 출발하는 것이 중요

모수 통계와 비모수 통계

- 비모수 통계

- ✓ 한정된 수의 모수를 많이 가정하지 않고 모델링

- ✓ 가정이 약하고 적으므로 웬만해선 다 작동. 그러나...

- ➡ 낮은 검정력 (참인 주장도 확실히 입증하지 못할 확률이 높음)

- ➡ 얇은 해석 가능성 (모수 통계에 비해 의미부여가 어려운 경우가 상대적으로 많음)

모수 통계와 비모수 통계

- 우리가 이번에 하는 건 모수 통계!
- 모수적 모형을 만든다는 것은 곧 모형의 종류를 고르고, 모수를 추정하는 것임.

끝