

강좌 소개

기본 정보

목표

일정

파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리

이번 시간의 목표

- 1 《Python을 활용한 언어학 통계 입문》의 목표와 일정을 설명할 수 있다.
- 2 프로그래밍 언어로서 파이썬의 특징을 설명할 수 있다.
- 3 Google Colab에서 파이썬을 계산기처럼 사용할 수 있다.



강좌 소개

기본 정보

목표

일정

파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리

1 강좌 소개

기본 정보

목표

일정

2 파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

3 마무리



강사

이름 박수민

소속 서울대학교 인문대학 연계전공 인문데이터과학 / 언어학과

전공 컴퓨터언어학

연락처 mam3b@snu.ac.kr

주교재

- B. Winter. (2019). *Statistics for Linguists: An Introduction Using R*. Routledge
 - PDF 파일: https://appliedstatisticsforlinguists.org/bwinter_stats_proofs.pdf
 - 주의사항: 교재에서는 R을 사용하나, 본 강의에서는 Python을 사용한다.

강좌 소개

기분 정보

목표

일정

파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리



강좌 소개

기초 정보

목표

일정

파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리

지금 없어도 괜찮은 것

- 파이썬에 관한 지식
- 통계학에 관한 지식

수강 요건

- 1 컴퓨터에서 웹 브라우저를 사용할 수 있다.
- 2 한국어와 영어로 된 글을 읽고 이해할 수 있다.
- 3 언어학 전공 과목을 수강했거나 관련 문헌을 읽어 본 적이 있다.
- 4 개수나 횟수를 세거나 척도의 값을 측정한 적이 있다.

웹 브라우저 Google Chrome, Safari, ...



2025 LSK-시몬느
언어학 학교

수강 요건

4 개수나 횟수를 세거나 척도의 값을 측정한 적이 있다.

⇒ 데이터로부터 숫자를 얻어 본 적이 있다.

척도의 예시

기본주파수(F0), 반응시간(RT), 전체 분절음 중 파열음의 비율, ...

이/따

수치를 얻고 나서 무엇을 어떻게 해야 하는가?

- 1 이 수치를 가지고 어떤 주장을 할 수 있는가?
- 2 내가 하고자 하는 주장이 얼마나 믿을 만한가?



연구 사례: 개수나 횟수를 세는 경우

심주희. (2023). 〈식당 내 호칭 ‘저기요’, ‘여기요’ 를 대상으로 한 사회언어학적 변이 연구〉. 《담화와인지》 30(1), 64.

남성이 여성보다 ‘저기요’를 10%가량 더 많이 쓰는 반면, 여성이 남성보다 ‘여기요’를 10%가량 더 많이 쓰고 있다. 즉, ‘저기요’는 비교적 남성에게 선택되었고 ‘여기요’는 여성에게 선택되고 있었다. 이러한 차이가 유의미한지 통계 검정하기 위하여 카이제곱 분석을 실시한 결과, 유의확률이 0.000($p < 0.05$)으로 나타났다. 남성과 여성이라는 성별이 ‘저기요’, ‘여기요’의 응답과 상관관계를 가진다는 것이 밝혀졌다.¹⁷⁾

<표 7> 성별과 응답 간 상관관계

성별	빈도(%)		전체
	저기요	여기요	
여성	1038(46.1%)	1213(53.9%)	2251(100%)
남성	799(54.3%)	673(45.7%)	1472(100.0%)
$\chi^2(p)$	23.750(0.000)*		

$p^* < 0.05$, $p^{**} < 0.01$, $p^{***} < 0.001$

(N=3723)



횟수	‘저기요’	‘여기요’
여성	1038	1213
남성	799	673

연구 사례: 척도의 값을 측정하는 경우

김수한·문성민·고연숙. (2022). 〈이름의 성별에 따라 나타나는 음성상징성: 동물의 숲 주민 캐릭터 이름을 중심으로〉. 《음성음운형태론연구》 28(1), 46.

분석 결과, **성별 간에 음절 수는 차이가 있으나, 성별 간의 분절음 수는 차이가 없는 것으로 나타났다.** 그 이유는 선행연구인 Slater and Feinman (1985)에서 사용한 영어의 음절구조와 달리, 1.2 절에서 언급한 한국어의 음절구조의 경우 하나의 모음을 기준으로 두음과 말음에 최대 1 개의 자음만 위치할 수 있다. 그래서 분절음 수가 선행연구의 결과와 다른 결과가 나온 것으로 보인다.

표 7. 이름에 들어간 음절 수와 분절음 수에 따른 성별 간의 차이의 t 검정 결과

	등분산성	t	df	p-value
음절 수	등분산	-2.570	390	< .05 *
분절음 수	등분산	-0.290	390	.772

<표 8>은 성별에 따른 음절 수와 분절음 수의 평균 값을 나타내고 있으며 이를 통해 여성 캐릭터의 음절 수와 분절음 수의 평균이 남성 캐릭터 보다 더 높기 때문에 결과에 차이가 난 것을 확인할 수 있다.

표 8. 성별에 따른 음절 수와 분절음 수의 평균

	남성	여성
음절 수	2.50	2.71
전체 분절음 수	5.53	5.58



(1)



(2)

그림 1. 동물의 숲 주민 캐릭터인 오골(1)과 철컹(2)

이름	음절	분절음
오골	2	4
철컹	2	6



평균	남성	여성
음절 수	2.50	2.71

평균	남성	여성
분절음 수	5.53	5.58

동물의 숲 주민 캐릭터 이름의 평균 분절음 수

- 남성 캐릭터(5.53) < 여성 캐릭터(5.58)

주장 2

캐릭터의 성별에 따라 이름의 평균 분절음 수가 다르다.

- 평균적으로 남성 캐릭터의 이름보다 여성 캐릭터의 이름이 더 길다.

의문 2

분절음 수의 분포가 얼마나 달라야 성별의 차이가 유의미하다고 주장할 수 있을까?



관찰 종합: 동물의 숲 주민 캐릭터 이름의 길이

1 평균 음절 수: 남성(2.50) < 여성(2.71)

2 평균 분절음 수: 남성(5.53) < 여성(5.58)

두 척도 모두 여성 캐릭터의 평균값이 더 높다.

첫 번째로 각 이름에서 포함하는 전체 자음과 모음의 개수를 나타내는 분절음과 홀로 발음할 수 있는 최소의 소리 단위를 나타내는 음절에 대한 분석 결과, **성별 간에 음절 수는 차이가 있으나, 성별 간의 분절음 수는 차이가 없는 것으로 나타났다.** 그 이유는 선행연구인 Slater and Feinman (1985)에서 사용한 영어의 음절구조와 달리, 1.2 절에서 언급한 한국어의 음절구조의 경우 하나의 모음을 기준으로 두음과 말음에 최대 1 개의 자음만 위치할 수 있다. 그래서 분절음 수가 선행연구의 결과와 다른 결과가 나온 것으로 보인다.

표 7. 이름에 들어간 음절 수와 분절음 수에 따른 성별 간의 차이의 t 검정 결과

	등분산성	t	df	p-value
음절 수	등분산	-2.570	390	< .05 *
분절음 수	등분산	-0.290	390	.772

검증 결과

평균 음절 수의 성별 차이는 유의미하지만 평균 분절음 수의 차이는 유의미하지 않다.

이 강좌의 기대 효과

수치의 차이가 어떤 경우에 통계적으로 유의미한지 판단할 수 있다.



- Python을 클라우드 환경에서 사용하기: Google Colaboratory
- Python을 계산기로 써먹기: 산술 연산

- 데이터프레임 잘 만들기: “tidy” data vs. “messy” data
- 데이터의 유형 이해하기: 수치형(이산형, 연속형), 범주형(명목형, 순서형)

- 기술통계량: 평균, 표준편차, 중앙값, 범위, 상관관계 등
- 시각화: 산점도, 히스토그램, 상자그림 등



4 통계 모델링 (1) (2025-08-12) statsmodels 패키지

- 이론적 배경: 현실의 모델링(모형화), 반응변수(종속변수)와 설명변수(독립변수)
- 선형회귀분석: 반응변수가 연속형 자료인 경우의 모델링

5 통계 모델링 (2) (2025-08-13)

- 로지스틱 회귀분석: 반응변수가 명목형 자료인 경우(이진 분류)의 모델링
- 푸아송 회귀분석: 반응변수가 이산형 자료인 경우(0 이상의 횟수)의 모델링

6 통계 모델링 (3) (2025-08-13)

- 혼합 효과 모형: 반복 측정 데이터에서 개체 간 변이를 고려하는 모델링
- 모델 선택 및 주의점



7 통계적 추론 (1) (2025-08-14)

- 이론적 배경: 가설 검정, 1종 오류, 2종 오류, 통계적 유의미

8 통계적 추론 (2) (2025-08-14)

- t -검정: 두 집단의 평균 차이가 통계적으로 유의미한지 알아보기
- 카이제곱검정: 범주형 데이터에서 두 변수의 관계가 통계적으로 유의미한지 알아보기



파이썬 공식이 말하는 파이썬

박수민

기분 정보

목표

일정

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리

<https://www.python.org>



2025 LSK-시몬느
언어학 학교

화면에 “Hello, World!”를 출력하는 프로그램

C: 총 6행; Java: 총 5행

hello.c

```
1 #include<stdio.h>
2 int main(int argc, char * argv[])
3 {
4     printf("Hello, World!\n");
5     return 0;
6 }
```

HelloWorld.java

```
1 public class HelloWorld {
2     public static void main(String[] args) {
3         System.out.println("Hello, World!");
4     }
5 }
```



화면에 “Hello, World!”를 출력하는 프로그램

```
1 print("Hello, World!")
```

파이썬의 장점

쉽고 간결하다.

파이썬의 또 다른 장점

다양한 라이브러리가 존재하므로 다른 작업으로 확장하기 편리하다.

- 텍스트 전처리
- 형태소 분석
- 기계학습, 딥러닝



파이썬 공식이 말하는 파이썬

“파이썬은 인터프리터 언어입니다. 컴파일과 링크 단계가 필요 없으므로 개발 시간을 상당히 단축해줍니다. 인터프리터는 대화형으로 사용할 수 있어서, 언어의 기능을 실험하거나, 쓰고 버릴 프로그램을 만들거나, 바닥부터 프로그램을 만들어가는 동안 함수들을 테스트하기 쉽습니다. **간편한 탁상용 계산기**이기도 합니다.”

<https://docs.python.org/ko/3/tutorial/appetite.html>



파이썬을 계산기처럼 사용하기



GO

About

Downloads

Documentation

Community

Success Stories

News

Events

1

박수민

강좌 소개

기분 정보

목표

일정

파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리



2025 LSK-시몬느
언어학 학교

파이썬 코드를 작성하는 “앱”이 따로 있나요?

문서를 다루는 응용프로그램의 종류

- 워드 프로세서
 - 한글, MS Word 등
- 텍스트 편집기
 - EmEditor, Sublime Text, Visual Studio Code 등

텍스트 편집기의 특징

- 문자(글자, 숫자, 문장부호, 특수 문자 포함)만 사용할 수 있다.
 - 워드 프로세서에서는 글자와 문단 모양을 바꾸고 도표를 삽입할 수 있다.
- 텍스트 파일은 모든 텍스트 에디터에서 읽고 쓸 수 있다.
 - 한 워드 프로세서 프로그램에서 작성한 문서는 다른 프로그램에서 열기 어렵다.

추천 Visual Studio Code(<https://code.visualstudio.com/>)



<https://colab.research.google.com>

준비물 Google 계정(무료)



파이썬을 클라우드 환경에서 사용하기

Python 시작하기

박수민

강좌 소개

기본 정보

목표

일정

파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리

Colab 시작하기

노트 열기

예 >

최근 사용 >

Google Drive >

GitHub >

업로드 >

노트북 검색

제목

Overview of Colab Features

Markdown Guide

Charts in Colab

+ 새 노트

취소



2025 LSK-시몬스
언어학 학교

지금 바로 Colab에서 할 일

- 파이썬과 인사하기
- 파이썬을 계산기처럼 사용하기

CO [2025 언어학 학교] 1강 Python 시작하기 ☆ ☁

파일 수정 보기 삽입 런타임 도구 도움말

Q 명령어 + 코드 + 텍스트 ▶ 모두 실행

✓ 0초 [1] 1 print('Hello, World!')

Hello, World!

✓ 0초 1 print(2 + 3)

5



Colab에서 더 해 볼 일

■ 자료형의 차이 알아보기

- 정수 int
- 부동소수점수 float
- 문자열 str
- 리스트 list

■ 변수 사용하기

■ 함수 사용하기

■ 오류에 익숙해지기

- SyntaxError, NameError, ZeroDivisionError, TypeError, ValueError, IndexError, ...



강좌 소개

기본 정보

목표

일정

파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리



2025 LSK-시몬느
언어학 학교

- 1 강좌 소개
기본 정보
목표
일정
- 2 파이썬 맛보기
파이썬 소개
개발 환경
텍스트 편집기
Google Colaboratory
기초 개념
- 3 마무리

■ 파이썬의 특징

- 문법이 쉽고 간결하다.
- 인터프리터 언어로서 실시간으로 한 줄씩 실행해 볼 수 있다.

■ 파이썬 기초 개념

- 산술 연산
- 자료형
- 변수
- 함수
- 오류와 예외

■ 데이터프레임과 “tidy” data

■ 데이터의 네 가지 유형

강좌 소개

기분 정보

목표

일정

파이썬 맛보기

파이썬 소개

개발 환경

텍스트 편집기

Google Colaboratory

기초 개념

마무리



2025 LSK-시몬느
언어학 학교

2025

LSK-시몬느 언어학 학교

2025 LSK-Simone Linguistic School

