

<매치업 심화과정 : 통계를 활용한 탐색적 데이터 분석>

강의계획서 (2023년 2차)

1. 수업개요

- 데이터분석의 첫단계이자 데이터분석 결론에 대한 해석의 기준이 되는 탐색적 데이터분석의 전체적인 과정을 실무에서 적용할 수 있는 방법론을 학습합니다.
- 학습 진행 과정은 과제 제출과 실습 위주 및 피드백 제시로 학습자가 실제 통계를 활용하여 탐색적 데이터분석을 수행하는 모습과 가장 비슷한 형태로 진행됩니다.

2. 수업목표

- 데이터의 주요 특성을 요약하고 시각화할 수 있습니다.
- 탐색적 데이터 분석의 방법론을 적용할 수 있습니다.
- 탐색적 데이터 분석을 수행하는 프로그램이나 툴을 선택하여 활용할 수 있습니다.
- 실제 데이터를 수집하여 데이터 탐색을 수행하고 분석 문제를 도출할 수 있습니다.

3. 사전 지식(학습자)

- 기초 통계 역량
- 사이킷런 패키지를 이용한 머신러닝 알고리즘 적용 경험

4. 실습환경

- 통계 프로그램 실습은 R project와 R Studio를 사용합니다.

5. 주차계획* 실습은 학생들이 스스로 실습해보는 것으로 평가에는 포함하지 않음

주차	주차명 및 내용			강의유형 강의자료
1	오리엔테이션	1차시	- 강의소개, 교수자, 조교 소개 - 수업진행 방식, 평가방식 안내 - 질문답변	실시간강의 (온/오프 병행)
		2차시	- 실습환경 설명 및 구성 ① R설치 / R studio 설치	강의록 실습파일
		학습 활동	- 퀴즈 : 없음 - 과제 : 없음	
2	EDA이해와 개념	1차시	EDA이해	온라인 동영상
		2차시	EDA기본개념	
		3차시	EDA방법	
		학습 활동	- 퀴즈 : 객관식 5문제 (30분) - 과제 : EDA 필요한 사례/방법 (60분)	

3	기술통계와 확률분포	1차시	기술통계	온라인 동영상
		2차시	이상치탐색	
		3차시	확률분포	
		학습 활동	- 퀴즈 : 객관식 5문제 (30분) - 과제 : ① 과제1 : 실제 데이터 기술통계 작성(30분) ② 과제2 : 이상치 탐색(30분) ③ 과제3 : 확률분포 실생활용(60분)	강의록 실습파일
4	모수추정과 가설검정	1차시	모수추정	온라인 동영상
		2차시	가설검정	
		학습 활동	- 퀴즈 : 객관식 5문제 (30분) - 과제 : ① 과제1 : 모평균, 모비율에 대한 구간추정 문제 (40분) ② 과제2 : 단일모집단 평균을 아는 경우, 모르는 경우 각각 가설검정과 판정 문제 (60분)	강의록 실습파일
5	변수간 관계	1차시	범주형 변수간 관계	온라인 동영상
		2차시	연속형 변수간 관계	
		3차시	종속변수/독립변수간 관계	
		학습 활동	- 퀴즈 : 객관식 5문제 (30분) - 과제 : ① 과제1 : 독립성검정 실전 적용 문제(20분) ② 과제2 : 연속형 변수간 상관관계 실전상황 문제 (20분) ③ 과제3 : 잔차검토 실전상황 문제(20분)	강의록 실습파일
6	온라인 세미나	1차시	- 데이터에 대한 통찰을 위한 체크리스트 - 과제질문에 대한 해설(2주~5주차)	실시간 세미나 (Zoom)
		학습 활동	- 퀴즈 : 없음 - 과제 : 없음	강의록
7	EDA실무(I) - 예측모델을 위한 EDA	1차시	데이터셋 이해	온라인 동영상
		2차시	EDA순서	
		3차시	R을 이용한 EDA	
		학습 활동	- 과제 : 실제상황 데이터셋(서울시 따릉이 대여데이터)활용 따릉이 대여대수 예측모델을 위한 EDA보고서 작성(80분)	강의록 실습파일
8	EDA실무(2) - 가설설정을 위한 EDA	1차시	데이터셋 이해	온라인 동영상
		2차시	EDA순서	
		3차시	R을 이용한 EDA	
		학습 활동	- 과제 : 실제상황 데이터셋(은행마케팅 데이터)활용 가설설정을 위한 EDA보고서 작성(80분)	강의록 실습파일

9	데이터 시각화	1차시	R시각화 패키지	온라인 동영상
		2차시	정형데이터 시각화	
		3차시	비정형데이터 시각화	
		학습 활동	- 퀴즈 : 객관식 4문제 (30분) - 과제 : ① ggplot2를 이용한 베이커리 판매량의 시간대별 추이 그래프 작성(20분) ② ggplot2를 이용한 베이커리 아이템별의 판매량 그래프 작성(20분) ③ 다음 제시한 그래프를 R의 ggplot2패키지를 이용하여 작성(20분)	강의록 실습파일
10	기말 프로젝트 (실시간)	1차시	프로젝트 주제 및 접근 방법 설명	실시간 세미나 (Zoom)
		학습 활동	- 퀴즈 : 없음 - 과제 : 기말 프로젝트	강의록

* 퀴즈, 과제, 프로젝트는 평가에 포함됨

4. 학습기간

- 2023. 11. 18(토) ~ 2023. 12. 23(토)

5. 이수기준

- 총점 70점 이상 시 수료됩니다.**

- ① 퀴즈 30%
- ② 과제 40%
- ③ 기말프로젝트 30%

6. 강사소개

강애띠 교수님

주요경력



- 현 (주)한컴인스페이스 연구소장
- 현 고려사이버대학교 융합정보대학원 외래교수
- 전 한국국토정보공사 공간정보연구원 선임연구원
- 전 주소공유체계 및 주소기반 지식그래프 연구 PM
- 전 주소부여대상 개체 및 안내시설 현장조사용 단말기 활용방안 연구 PM
- 전 블록체인기술기반 LX주소혁신플랫폼 구축방안연구 PM
- 전 주소의 국제표준화 대응방안 연구 PM
- 전 주소체계고도화 및 4차산업창출 시범사업 PM
- 전 스마트 공간정보 기술개발 기획연구 PL
- 전 공간정보SW활용을 위한 오픈소스 가공기술개발 참여연구원

7. 조교소개

조은영 조교님



주요경력

- 현 성균관대학교 일반대학원 통계학과 석사과정 대학원생
- 전 고려사이버대학교 논문세미나 TA
- 전 성균관대학교 통계학원론 TA
- 전 식약처 적응적 임상시험 가이드라인 안 참여연구원
- 전 부경대학교 자연과학대학 통계학과 학부과정 대학생