R을 활용한 통계 데이터 분석 보고서 작성

수강 대상

- 통계를 하나도 모르지만 숫자 자료에 대한 실무 업무를 수행하는 분
- 데이터 분석과 R을 처음 배워 보고 싶은 분

과정 특징

- 기초가 없는 수강생을 전제로 통계 이론과 데이터 분석에 대한 전체적인 틀을 잡을 수 있는 기본적 내용을 다룹니다.
- 머신러닝이나 딥러닝을 수행하기 전 수행하는 자료의 전처리와 탐색적 자료분석 과정을 학습합니다.
- 강의 내용을 기반으로 심화 과정으로 나아갈 수 있는 방향을 제시합니다.
- 데이터 분석 전담 실무 환경에서 일상적으로 수행되는 업무를 경험합니다.

목차(초안)

- 1. 1H (통계이론) 통계학 기초
 - 확률과 통계
 - 자료의 형태
 - 기술 통계
 - Appendix : 머신러닝과 통계
- 2. 2H (R) R기초와 프로그래밍
 - R 분석 환경 설정
 - (tidyverse) 라이브러리 활용
 - 데이터 형태와 연산
 - (ggplot2) 데이터 시각화
 - 조건문
 - 함수
- 3. 1H (R) 데이터베이스 다루기
 - SQL 언어의 기본 구조
 - (dplyr)을 통한 데이터 가공
 - 이상값과 결측치
- 4. 1H (통계이론+R) 확률 변수와 확률 분포
 - 확률 변수
 - 확률분포와 정규분포
- 5. 2H (통계이론+R) 가설 검정
 - 가설 검정 개요
 - 단일 집단에 대한 가설 검정
 - 다중 집단에 대한 가설 검정
- 6. 30m (실무) 분석 기획과 보고서 작성
 - R Markdonwn 사용법
 - 분석 보고서 기획 팁
- 7. 1H (실무+실습) 분석 보고서 작성 시나리오
 - 30m. Case 1
 - 30m. Case 2

약 8시간 30분 분량 기획

과정별 상세

마케팅 활동을 통해 수집된 가상의 실적 자료를 기반으로 실제로 마케팅이 효과가 있었는지 확인하는 일련의 통계 분석 보고서 작성 과정을 이해하기 위한 이론적 기초와 프로그램 활용 방법을 익힙니다.

모든 챕터에서 3~5분 정도 이번 과정의 목표와 해당 과정의 전체적인 구조, 그리고 그 중 다루게 될 일부 기초적인 내용들을 프리뷰 합니다.

• 1. 통계학 기초 (통계 이론)

데이터 분석의 전 과정에 따라 모든 챕터에 걸쳐 이번 과정에서 다루게 될 이론적 내용들을 전체적으로 소개

- 확률과 통계

수학적 확률 이론과 이를 기반으로 수집된 자료(표본)을 해석하고 나아가 예측까지 수행할 수 있는 통계 이론의 차이 및 필요성에 대해서 간략히 소개

- 자료의 형태

수치형 자료(이산형/연속형)와 범주형 자료(명목형/순서형)의 차이 및 명목/서열/구간/비율 척도에 대해서 설명

- 기술 통계

통계학의 큰 두 줄기인 기술 통계(Descriptive Statistics)와 추론 통계(Inferential Statistics)의 차이와 기술 통계의 주요 통계량인 대표값, 산포도, 왜도 및 첨도 등에 대해 소개하고 수치적인 특성으로 자료를 해석하는 방법을 이해

- (Appendix) 머신러닝과 통계

머신러닝, 딥러닝, 통계 모델링 등 무엇이 다른것인지, 서로 어떤 연관이 있는 것인지, 각 분야를 학습하기 위해 무엇을 먼저 공부해야 하는지 등 영역을 더 확장해나가기 위한 방향을 안내

2. R기초와 프로그래밍 (R 실습)

통계 분석에서 R 프로그램의 장점과 Python과의 비교 등 R 활용에 대해 전반적으로 소개

- R 분석 환경 설정

R 프로그램에 대한 소개, 설치, IDLE인 R Studio의 UI 및 Workspace 등을 활용하는 방안과 요령 소개

- 라이브러리 활용

R의 기능 확장인 라이브러리 설치 방법과 이용 방법에 대해 소개하고 tidyverse를 포함하여 이번 과정에서 활용하게 될 대표적인 라이 브러리들에 대한 소개 및 설치 진행

- 데이터 형태와 연산

스칼라, 벡터, Data Frame, tibble, List, Matrix 등 R의 데이터 형태와 관련 함수들, 연산자들을 소개하고 실습

- (ggplot2) 데이터 시각화

ggplot2 패키지에 대한 기본 활용법과 cheetsheet 활용법, 1장에서 배운 자료의 형태에 따라 적합한 그래프의 역할과 탐색적 자료분 석에서의 시각화 팁 등을 연습

- 조건문

프로그래밍에서 if~else 조건문에 대한 원리와 간단한 프로그램 구축 실습

- 함수

프로그래밍에서 함수의 역할과 function() 함수를 이용한 User Define Function 생성 방법에 대해 실습

• 3. 데이터베이스 다루기 (R 실습)

데이터 분석 전 과정 중 가장 많은 시간이 소요되는 데이터 전처리와 마트 구성 과정에 대해 소개

- (sqldf) SQL 언어의 기본 구조

데이터베이스를 다룰 때 가장 널리 사용되는 SQL 언어에 대해서 배우고 sqldf 패키지를 통해서 sql언어 활용에 대해 간략하게 실습

- (dplyr)을 통한 데이터 가공

tidyverse 생태계의 dplyr 패키지를 이용해 데이터를 원하는 형태로 선택하고 조인하여 가공하고, chain operation을 활용하는 방법 을 실습

- 이상값과 결측치

데이터 전처리에서 중요한 역할을 하는 이상값에 대한 처리와 결측에 대한 처리

4. 확률 변수와 확률 분포 (통계 이론 + R 실습)

추론통계를 구성하는 핵심인 확률 표본과 관련된 이론에 대해 전체적으로 소개하고 정규분포와 관련된 이론과 시뮬레이션 실습

- 확률 변수

모집단과 표본, 확률 변수의 개념에 대해서 학습하고, sample 함수 등을 활용하여 R 환경에서 실습

- 확률 분포와 정규 분포

분포 이론에 대해서 간략하게 설명하고, R의 시각화 기능을 이용해서 정규분포의 특성에 대해 실습

5. 가설 검정 (통계 이론 + R 실습)

사회적 현상을 수학적으로 증명할 수 있게 되는 근거인 통계적 가설 검정이 어떤 의미를 가지는지 소개하고 R을 통해 분석 수행을 실습

- 가설 검정 개요

1종 오류, 2종 오류의 개념과 귀무가설 설정을 통한 단측검정, 양측검정의 개념에 대해서 소개

- 단일 집단에 대한 가설 검정

z 검정과 one-sample t-검정을 통해 어떤 종류의 데이터가 주어질 때 어떤 가설을 검정할 수 있는지 배우고, 데이터 분석 수행을 실습

- 다중 집단에 대한 가설 검정

two-sample t-검정과 paired t-검정에 대해서 배우고, 데이터 분석 수행을 실습

• 6. 분석 기획과 보고서 작성

R Markdown을 통해 Markdown을 활용한 글쓰기 전략과 활용 방법을 배우고, 분석 기획 및 보고서 작성 실무 수행의 요령을 공유

- R Markdown 사용법

Markdown 문법을 소개하고 R Markdown 패키지를 설치하여 데이터를 분석하고 인사이트를 발굴하는 과정을 체계적으로 기록하는 요령을 소개

- 분석 보고서 기획 팁

데이터분석 실무 업무를 담당하면서 수시로 발생하는 현황파악 > 분석가설 수립 > 데이터 가공 > 탐색적 자료분석 > 가설 검정 > 분석 결과 보고의 과정에서 설득력 있는 보고서를 기획하고 작성하는 방법을 공유

• 7. 분석 보고서 작성 시나리오

지금까지 배운 내용을 모두 활용하여 실제 분석 과업을 기획하고 인사이트를 발굴하는 전 과정을 시나리오 기반으로 연습

- Case 1 / 방문 고객 수 증대를 위한 디지털 광고 마케팅 수행 성과 검증 (가안)

(실제 업무에서 수행한 간단한 프로젝트와 유사한 샘플데이터 및 시나리오 구축하여 분석 전 과정을 remind 및 보고서 작성)

- Case 2 / 넷플릭스 이용 이력 유무에 따른 TV 셋탑 서비스 활용 패턴 분석 (가안)