

# 033.020: 통계학실험

Course Project, 2021년 봄학기

## 프로젝트 : 와인품질 데이터

수업시간에 배운 여러가지 다양한 분석방법을 통해 실제 데이터를 분석해 보자. 데이콘 (<https://dacon.io/>)에서 진행하는 와인품질 예측 경진대회 자료를 사용한다. 여러분은 이 프로젝트를 진행하는 과정에서 탐색적 자료분석, 통계적 추론, 모델링, 그리고 예측 등 현재까지 배운 기법들과 앞으로 배운 기법들을 활용하게 될 것이다. 분석에 필요한 데이터는 데이콘 홈페이지에서 얻을 수 있다.

## 데이터 개요

데이콘 경진대회 홈페이지(<https://dacon.io/competitions/open/235610/data/>)에 제시되어 있는 2개의 데이터 파일을 사용할 수 있다.

## 보고서

- **제안서: 4월 2일 금요일** 까지 다음 내용을 R markdown을 사용하여 두장으로 요약하여 제출할 것. 모든 보고서는 R markdown을 사용한다. 보고서의 모든 내용은 재현가능하도록 작성하고, rmd 파일, html 파일 뿐 아니라 기타 컴파일을 위해 필요한 모든 파일을 함께 제출한다. 제안서를 위한 R markdown template은 eTL에 게시되어있다.
  1. 데이터에 대한 설명(Codebook 작성 - 각 데이터 셋에 포함되어 있는 변수들과 그에 대한 설명)
  2. 연구가설 (와인품질예측과 이와 관련된 흥미있는 연구가설을 하나 추가할 것)
  3. 각 연구가설와 관련된 탐색적 자료 분석
  4. 분석에 대한 날짜별 계획(timetable)
  5. 각 조원의 예상 역할
- **경과보고서 5월 7일 금요일**까지 3장이내로 경과보고서를 다음과 같은 형식으로 제출할 것
  1. 연구가설에 대한 자세한 소개 (예: 와인 품질에 영향을 미치는 중요한 성분은 무엇인가, 예측의 우수성을 평가하는 측도가 적절한지 여부 등)
  2. 중간분석 결과
  3. 향후분석 계획

- **최종보고서: 6월 4일 금요일**까지 10장이내로 최종보고서를 다음과 같은 형식으로 제출할 것

1. 서론
2. 분석의 주요목적 : 이 자료분석을 통하여 얻고자 하는 바와 분석의 어려움등
3. 방법 : 분석에 사용된 방법
4. 분석결과
5. 결론
6. 각 조원이 보고서 및 발표에 기여한 바