

공업통계학 데이터 EDA 정리

main subject:

- (각 팀의 승, 패)와 column의 연관성 + ANOVA(? → 어떤 column이 승패에 가장 큰 영향을 미치는가)

상관관계 확인 후

⇒ logistic regression (인과 관계를 봐야하므로), chi-square(승패와 first~)

- (각 팀의 동일 column) 간 비교 (→ 어떤 팀일수록 column을 높이기 유리하나)

⇒ column간 t 검정? z 검정?

Insight:

팀 배정 간 column 차이의 유의미를 설명하고

승 패와 column의 연관성을 보여주어

팀 배정에 의한 column의 차이로 승패에 영향을 미치는가를 보여줌.

의미있는 차이가 나오지 않을 경우 → 선수들은 극복해낸다?

project step:

1. red와 blue를 나눔.
2. 표본선택
3. blue 팀과 red 팀 각각에 대해서 regression 처리(승리, column)
4. 각 팀의 동일 column간 t 검정
5. 각 팀의 승, 패와 column 간 regression

DataFrame 구성:

- 수치형 column:

gameduration, dragonkills, baronkills, towerkills, inhibitorkills, WardPlaced, Wardkills, kills, death, assist, championDamageDealt, TotalGold, TotalMinionKills, TotalLevel, AvgLevel, JungleMinionKills, KillingSpree, TotalHeal, ObjectDamageDealt

- 범주형 column:

wins, firstblood, firstTower, firstBaron, firstDragon, FirstInhibitor