R을 이용한 PUBG 회귀분석

2016010735 최연석

**1. 주제**

Kaggle – Competition에 있는 ‘PUBG Finish Placement Prediction’ 데이터를 이용한 배틀그라운드 게임에서 최종 승리하기 위한 조건들을 솔로, 듀오, 스쿼드 3가지 모드를 나눠서 분석

**2. 데이터 및 설명**

이 데이터는 2018년 10월에 Kaggle에서 PUBG Developer API를 통해 수집된 데이터이며, 100명의 플레이어가 빈손으로 섬에 떨어져 각종의 아이템을 파밍 후 서로 싸워서 한 명(한 팀)이 남을 때까지 생존하는 배틀 로얄 게임이다. 게임에 승리하게 될 조건은 많은 변수가 될 수 있지만 최종적으로 총으로 쏴서 적을 잡는 경우가 많은 게임이고 경기지역이 좁아지는 블루존(블루존에 있을 경우 유저의 피가 조금씩 깎이며 시간이 가면 갈수록 깎이는 속도가 더 심해진다. 그리고 블루존은 랜덤으로 위치가 지정된다.)이 있어 결국에는 모든 유저가 싸우게 되는 총 게임(FPS)이다.

train, test 데이터 이렇게 두가지 있지만 train 데이터만 다룰 생각이다.

※ 데이터 수









※ 데이터 필드

Id – 플레이어 ID

groupId – 그룹을 식별하기 위한 ID

matchId – 매치를 식별하기 위한 ID

**assists** – 팀원의 킬에 도움을 준 횟수

**boosts** – 사용된 부스트 아이템의 수

**damageDealt** – 총 피해량

**DBNOs** – 기절시킨 적 플레이어 수

**headshotKills** – 헤드샷으로 죽인 적 플레이어 수

**heals** – 사용한 치유 아이템의 수

killPlace – 나를 죽인 적 플레이어의 순위

killPoints – 플레이어의 킬 기반 외부 순위

**kills** – 죽인 적 플레이어의 수

**killStreaks** – 짧은 시간에 죽인 적 플레이어의 최대 수 (연속 킬)

**longestKill** – 사망시 사망한 플레이어와 플레이어 사이의 최장 거리

matchDuration – 게임에서 생존한 시간(초)

matchType – 게임 모드(크게 솔로, 듀오, 스쿼드 이렇게 나뉨)

numGroups – 경기에서 데이터가 있는 그룹 수

rankPoints – 플레이어의 외부 점수 순위

revives – 팀원을 부활시킨 횟수

**rideDistance** – 미터 단위로 측정된 차량으로 이동한 총 거리

roadKills – 차량으로 킬 한 수

swimDistance – 미터 단위로 측정한 수영으로 이동한 총 거리

teamKills – 플레이어가 팀원을 죽인 횟수

vehicleDestroys – 파괴한 차량 수

**walkDistance** – 미터 단위로 측정한 발로 이동한 총 거리

weaponAcquired – 총 획득한 무기 수(총은 최대 주무기-(AR, SMG, LMG, SG), 보조무기-(SR, DMR, 석궁) 즉, 주무기 보조무기 포함 2개까지 소지 가능하며 추가로 권총을 소지할 수 있다.)

winPoints – 플레이어의 승리 기반 외부 순위

**winPlacePerc** – 백분위로 측정된 등수(숫자가 높을수록 1등에 가까움, 1등은 1이라고 적혀있다.)

▶ 게임에 직접적으로 영향을 위 빨간색 변수들을 선정하여 분석하고자 한다.

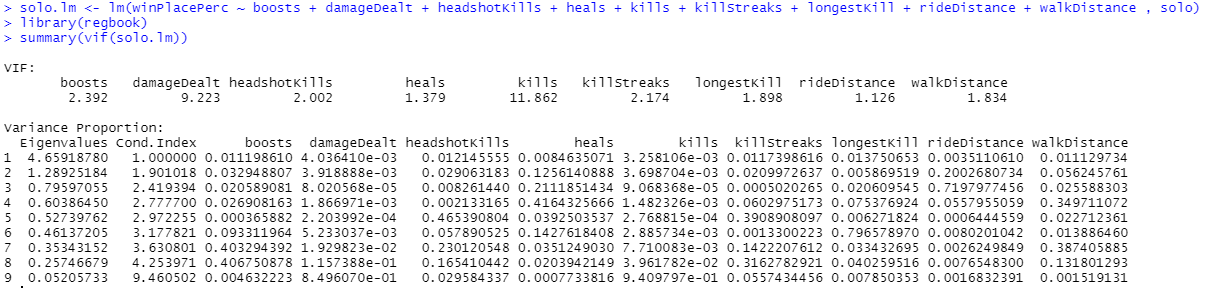
**3. 분석**

배틀그라운드 모드는 크게 총 3가지 모드(솔로, 듀오, 스쿼드)가 있지만 그 외에 ‘사용자 지정’ 모드에서 이뤄지는 모드가 있지만 그런 모드는 제외하고 1인칭, 3인칭은 모두 통합하여 솔로, 듀오, 스쿼드 3가지 모드를 중점으로 승리하기 위한 조건에 큰 영향을 주는 것들이 무엇인지 분석하고자 한다.

**3-1. 솔로 모드**

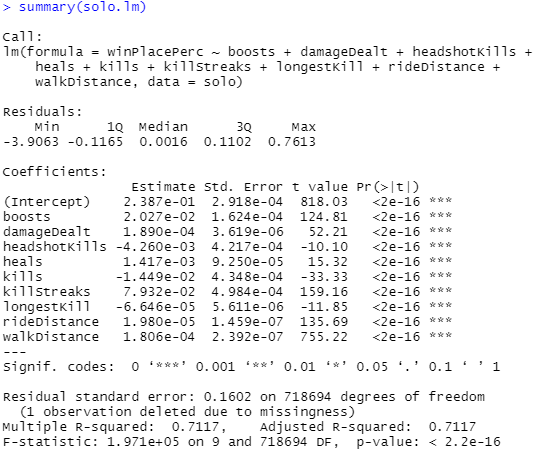
**3-1-1. 솔로 전체 범위**

**1) 다중공선성**



분석을 하기 앞서 다중공선성이 존재하는지 확인해보기 위해 regbook() 패키지에 있는 vif() 함수를 사용하였다. 첫번째 줄에는 VIF 값, 첫번째 열은 고유값, 두번째 열은 조건지수, 세번째 열부터는 분산비가 나옴을 알 수 있고 조건지수를 보았을 때, 15보다 큰 값이 없으므로 다중공선성의 가능성은 없어 보인다. 그리고 VIF값을 보았을 때, ‘kills’변수가 10보다 커서 다중공선성이 있을것 같지만 심하게 높게 측정되지 않고 조건지수에서도 다중공선성이 측정되지 않았으므로 다중공선성이 없다고 하고 분석을 해보고자 한다.

**2) t 검정과 결정계수**



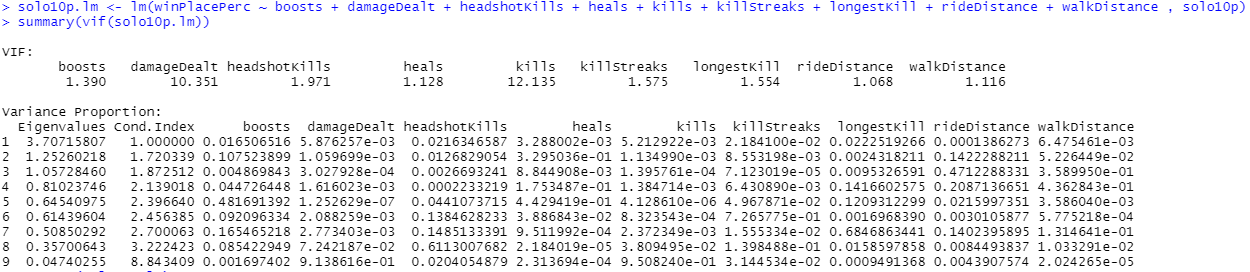
결정계수가 0.7117로 이 분석은 71.17%의 높은 설명력을 보여주고 각 설명변수들의 계수들을 t 검정한 결과 전부 귀무가설을 기각한 것을 알 수 있다. (즉, 각 계수는 0이 아님) 현재 회귀계수의 추정값을 보아

**killstreaks > boosts > heals > damageDealt > walkDistance > rideDistance > longestkill > headshotkills > kills**

크기순으로 나타낼 수 있다. 즉, 크기가 큰 값들이 상위권으로 갈수록 영향력이 큼을 알 수 있다. 이로 해석한 결과 어느 정도의 상위권을 가기 위해서는 연속킬을 잘 할 만큼의 실력이 되어야 하고 부스트나 회복 아이템이 많이 필요하다는 것을 알 수 있다. 반면 킬을 무조건 많이하는 것, 헤드샷, 장거리 킬은 크게 중요하지 않은 것을 알 수 있다.

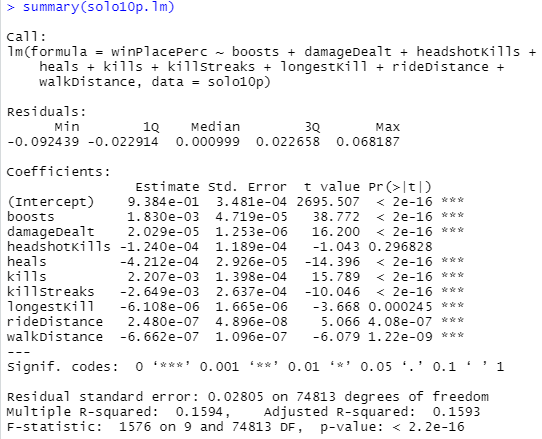
**3-1-2. 솔로 상위 10%범위**

**1) 다중공선성**

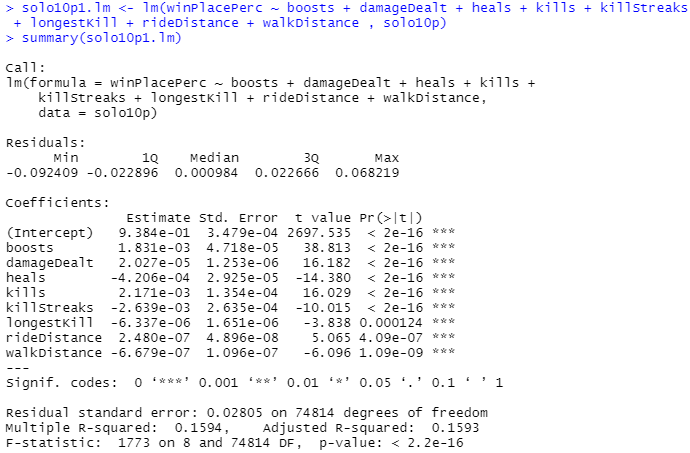
****

이것 또한 VIF값이 10보다 넘는 변수가 2가지 있지만 조건지수가 15보다 넘는 변수가 없으므로 다중공선성이 없다고 하고 분석을 해보고자 한다.

**2) t 검정과 결정계수**



결정계수가 0.1594로 이 분석은 15.94%의 낮은 설명력을 보여주고 각 설명변수들의 계수들을 t 검정한 결과 귀무가설을 하나 채택한 것을 알 수 있다. headshotkills가 채택이 되었고 headshotkills를 제외하고 다시 분석한다.



전과 후 결정계수는 0.1594로 동일하고 15.94% 낮은 설명력을 보여주지만 이 또한 t 검정 결과 전부 귀무가설을 기각한 것을 알 수 있다. (즉, 각 계수는 0이 아님) 현재 회귀계수의 추정값을 보아

**kills > boosts > damageDealt > rideDistance > walkDistance > longestkill > heals > killStreaks**

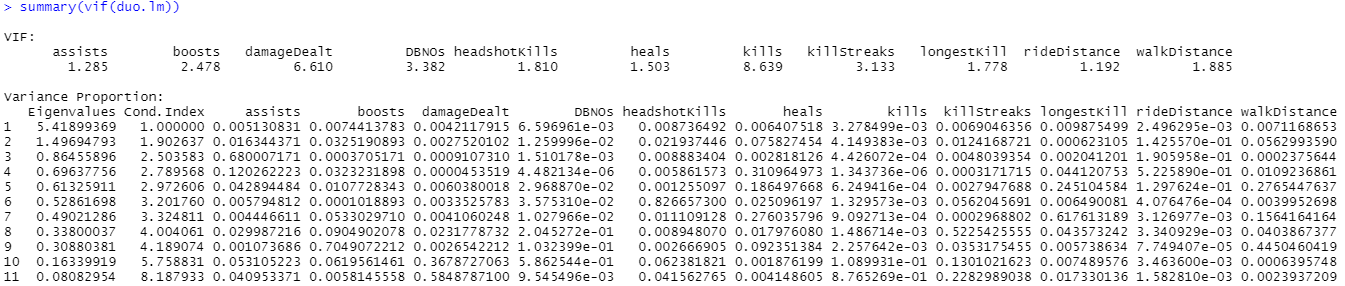
크기순으로 나타낼 수 있다. 이로 해석한 결과 상위 10%에서는 평소에 킬을 많이 하고 부스트 아이템을 많이 사용, 대미지를 많이 넣은 유저가 우승할 가능성이 높아 보이는 반면 단순회복이나 연속 킬 보다는 다른 유저들 몰래 오랫동안 생존하면서 킬 하는 것이 중요한 것을 알 수 있다.

**3-2. 듀오 모드**

**3-2-1. 듀오 전체 범위**

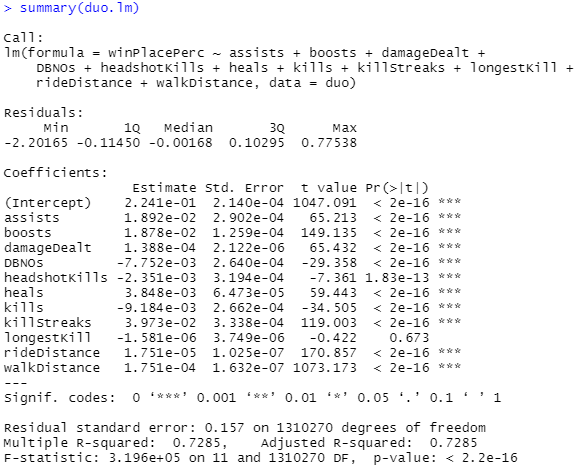
솔로와 다르게 듀오는 설명변수에 서로 도와가면서 킬을 할 때 카운팅 되는 ‘assists’, 듀오에서 둘 중 한 명이 살아 있고 한 명이 체력이 0이 되었을 때 발생하는 기절(기절되었을 때, 팀원이 10초동안 기대어 부활시킬 수 있다. 기절상태에 있을 때는 기절상태의 체력(이 체력은 damageDealt에 측정되지 않음)이 존재하며 체력이 조금씩 감소한다. 여러 번 기절하면 체력이 더 빨리 감소하며 팀원이 기대어 부활시키는 동안에는 감소하지 않는다.)이 되며 상대방을 기절시키면 카운팅 되는 ‘DBNOs’가 추가된다. 이것 또한 나중에 스쿼드에서도 동일하게 적용된다.

**1) 다중공선성**

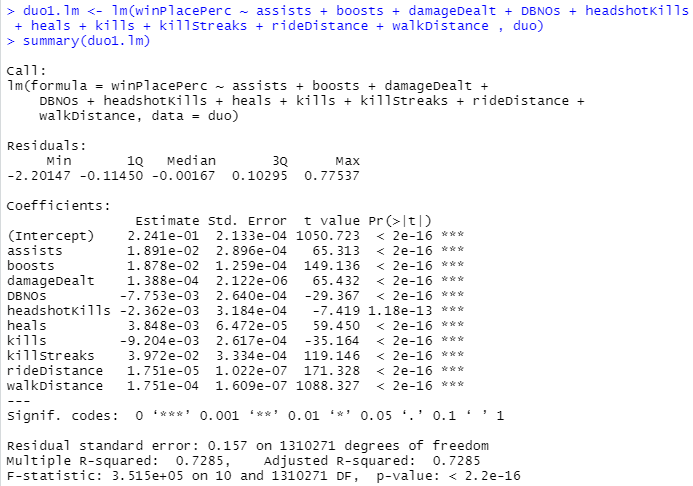


VIF값을 봤을 때, 모든 변수들은 10을 넘지 않고 조건지수를 보았을 때도 15를 넘지 않아 다중공선성이 없는 것을 알 수 있다.

**2) t 검정과 결정계수**



결정계수가 0.7285로 이 분석은 72.85%의 높은 설명력을 보여주고 각 설명변수들의 계수들을 t 검정한 결과 귀무가설을 하나 ‘longestkill’이 채택되었고 이 변수를 제외하고 다시 분석한다.



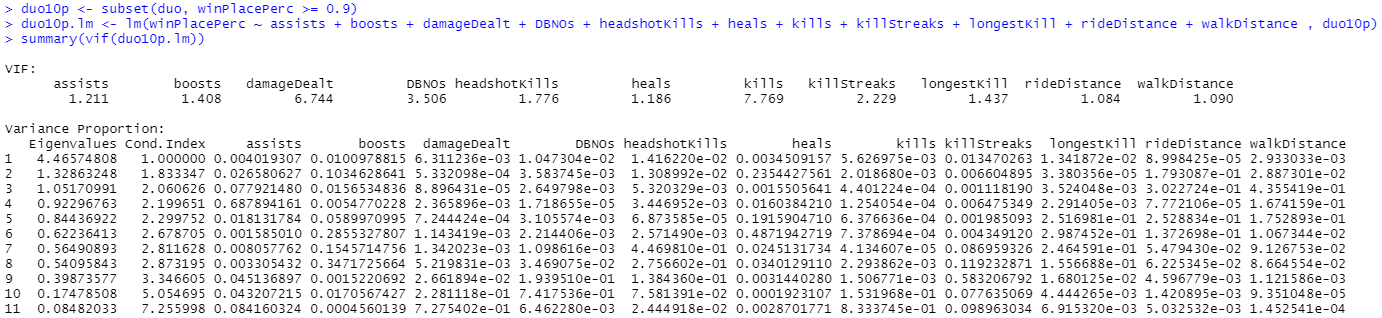
앞과 동일하게 결정계수 0.7285로 72.85%의 높은 설명력을 보여주고 t 검정 결과 전부 귀무가설을 기각한 것을 알 수 있다. 현재 회귀계수의 추정값을 보아

**killStreaks > assists > boosts > heals > walkDistance > damageDealt > rideDistance > headshotkills > DBNOs > kills**

크기순으로 나타낼 수 있다. 이로 해석한 결과 어느 정도의 상위권에 가기 위해서는 평소에 단순하게 킬 하거나 기절시키는 것보다 서로 도와가면서 연속적으로 킬을 하는 것이 중요하고 부스트나 회복 아이템을 많이 사용하는 것도 알 수 있다.

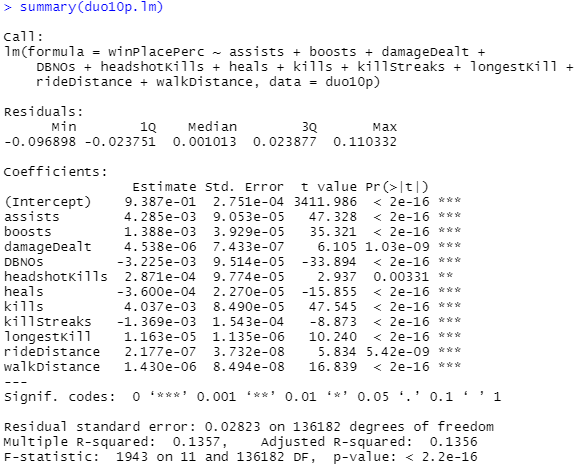
**3-2-2. 듀오 상위 10% 범위**

**1) 다중공선성**



VIF값을 봤을 때, 모든 변수들은 10을 넘지 않고 조건지수를 보았을 때도 15를 넘지 않아 다중공선성이 없는 것을 알 수 있다.

**2) t 검정과 결정계수**



결정계수는 0.1357로 13.57%로 낮은 설명력을 보여주지만 계속 분석한다. t 검정 결과 전부 귀무가설을 기각한 것을 알 수 있다. 현재 회귀계수의 추정값을 보아

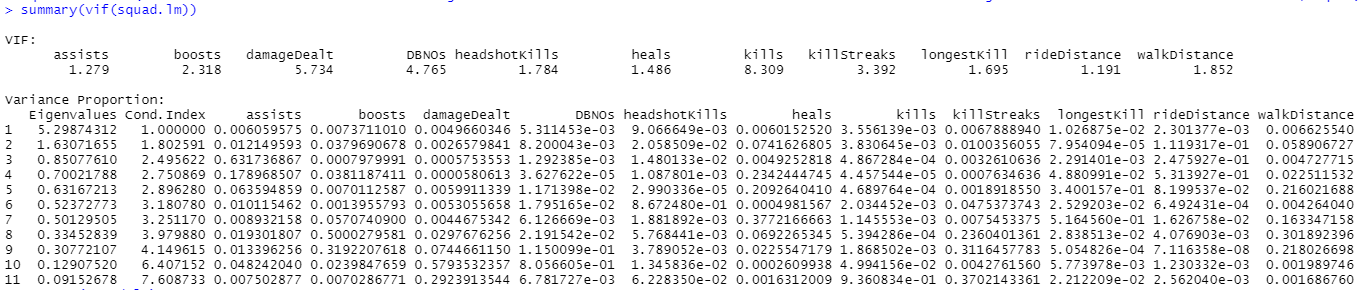
**assists > kills > boosts > headshotskills > longestkill > damageDealt > walkDistance > rideDistance > killStreaks > heals > DBNOs**

크기순으로 나타낼 수 있다. 이로 해석한 결과 상위 10%에서는 평소에 연속킬, 단순회복, 기절로만은 우승할 가능성이 낮아보이는 반면 서로 도와주면서 킬을 하고 부스트 아이템 사용이 중요한 것을 알 수 있다.

**3-3. 스쿼드 모드**

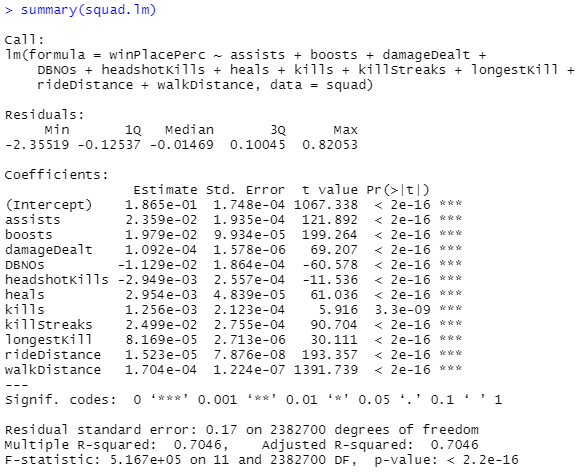
**3-3-1. 스쿼드 전체 범위**

**1) 다중공선성**

****

VIF값을 봤을 때, 모든 변수들은 10을 넘지 않고 조건지수를 보았을 때도 15를 넘지 않아 다중공선성이 없는 것을 알 수 있다.

**2) t 검정과 결정계수**



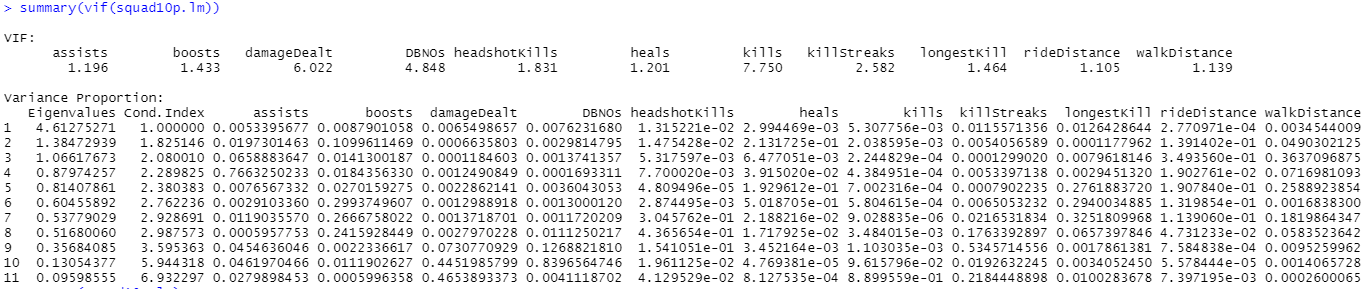
솔로, 듀오 보다는 결정계수가 0.7046으로 떨어지는 편이지만 70.46%의 높은 설명력을 보여주고 있다. t 검정 결과 전부 귀무가설을 기각한 것을 알 수 있다. 현재 회귀계수의 추정값을 보아

**killStreaks > assists > boosts > heals > kills > walkDistance > damageDealt > longestkill > rideDistance > headshotkills > DBNOs**

크기순으로 나타낼 수 있다. 이로 해석한 결과 어느 정도의 상위권에 가기 위해서는 굳이 킬을 하기위해 헤드샷을 맞추지 않고 몸샷으로 잡아도 상관없음을 알 수 있고 기절만 시키는 것은 중요하지 않음을 알 수 있다. 오직 서로 도와가며 연속적으로 킬을 많이 하고 부스트 아이템을 많이 사용해야 한다는 점을 알 수 있다.

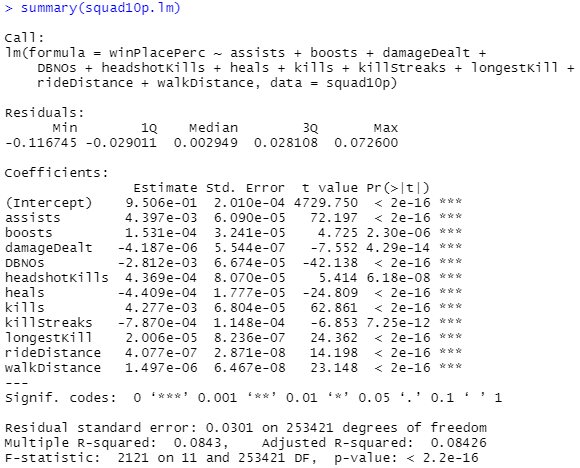
**3-3-2. 스쿼드상위 10% 범위**

**1) 다중공선성**



VIF값을 봤을 때, 모든 변수들은 10을 넘지 않고 조건지수를 보았을 때도 15를 넘지 않아 다중공선성이 없는 것을 알 수 있다.

**2) t 검정과 결정계수**



결정계수는 0.0843로 8.43%로 매우 낮은 설명력을 보여주지만 계속 분석한다. t 검정 결과 전부 귀무가설을 기각한 것을 알 수 있다. 현재 회귀계수의 추정값을 보아

**assists > kills > headshotskills > boosts > longestkill > walkDistance > rideDistance > damageDealt > heals > killStreaks > DBNOs**

크기순으로 나타낼 수 있다. 이로 해석한 결과 상위 10%에서는 평소에 연속킬, 단순회복, 기절로만은 우승할 가능성이 낮아보이는 반면 서로 도와주면서 킬을 하고 진짜 상위권에서는 헤드샷이 중요하고 부스트 아이템 사용이 중요한 것을 알 수 있다.

**4. 결과**

전반적으로 우승에 가깝게 도달하기 위해서는 킬하는 것이 중요하고 듀오나 스쿼드 같은 경우는 서로 함께 도와가며 킬을 하는 것이 중요한 것을 알 수 있었다. 대체적으로 부스트 아이템을 많이 사용하는 것이 중요한 것도 알 수 있고 전반적으로 차를 타고 이동하는 것보다는 걸어서 이동하는 것이 조금 더 중요하다는 것도 알 수 있었다. 반면에 기절만 많이 시키고 킬을 못할 경우에는 대부분 우승에는 힘들어 보이고 무리하게 연속킬을 너무 많이 한다고 우승에는 가기 어려우나 상위권은 유지할 수 있다는 점을 알 수 있었다. 무엇보다도 우승하기위해서는 이러한 분석도 중요하지만 그만큼 본인의 실력이 중요하다.

다소 부족한 부분은 많았고 그래프 그리는 것 등등 데이터가 너무 많아 오류가 나기도 했으나 시도를 많이 해보았습니다. 데이터를 줄여서 분석하는 것도 한가지 방법이긴 했으나 그래도 모든 데이터를 이용하여 분석하고 싶어서 모두 사용하였고 그래프 그리는 plot() 함수는 제외하고 분석을 하였습니다.

**5. R 전체적인 코드**

|  |
| --- |
| data = read.csv('C:/Users/GaeJu/Desktop/21년/1학기/회귀분석/기말 과제/train.csv')  head(data)  data$matchType <- gsub("solo-fpp", 'solo', data$matchType)  data$matchType <- gsub("duo-fpp", 'duo', data$matchType)  data$matchType <- gsub("squad-fpp", 'squad', data$matchType)  # solo mode  solo <- subset(data, matchType == 'solo')  # head(solo)  solo.lm <- lm(winPlacePerc ~ boosts + damageDealt + headshotKills + heals + kills + killStreaks + longestKill + rideDistance + walkDistance , solo)  library(regbook)  summary(vif(solo.lm))  summary(solo.lm)  solo10p <- subset(solo, winPlacePerc >= 0.9)  solo10p.lm <- lm(winPlacePerc ~ boosts + damageDealt + headshotKills + heals + kills + killStreaks + longestKill + rideDistance + walkDistance , solo10p)  summary(vif(solo10p.lm))  summary(solo10p.lm)  solo10p1.lm <- lm(winPlacePerc ~ boosts + damageDealt + heals + kills + killStreaks + longestKill + rideDistance + walkDistance , solo10p)  # summary(vif(solo10p1.lm))  summary(solo10p1.lm)  # duo mode  duo <- subset(data, matchType == 'duo')  # head(duo)  duo.lm <- lm(winPlacePerc ~ assists + boosts + damageDealt + DBNOs + headshotKills + heals + kills + killStreaks + longestKill + rideDistance + walkDistance , duo)  summary(vif(duo.lm))  summary(duo.lm)  duo1.lm <- lm(winPlacePerc ~ assists + boosts + damageDealt + DBNOs + headshotKills + heals + kills + killStreaks + rideDistance + walkDistance , duo)  summary(duo1.lm)  duo10p <- subset(duo, winPlacePerc >= 0.9)  duo10p.lm <- lm(winPlacePerc ~ assists + boosts + damageDealt + DBNOs + headshotKills + heals + kills + killStreaks + longestKill + rideDistance + walkDistance , duo10p)  summary(vif(duo10p.lm))  summary(duo10p.lm)  # squad mode  squad <- subset(data, matchType == 'squad')  # head(squad)  squad.lm <- lm(winPlacePerc ~ assists + boosts + damageDealt + DBNOs + headshotKills + heals + kills + killStreaks + longestKill + rideDistance + walkDistance , squad)  summary(vif(squad.lm))  summary(squad.lm)  squad10p <- subset(squad, winPlacePerc >= 0.9)  squad10p.lm <- lm(winPlacePerc ~ assists + boosts + damageDealt + DBNOs + headshotKills + heals + kills + killStreaks + longestKill + rideDistance + walkDistance , squad10p)  summary(vif(squad10p.lm))  summary(squad10p.lm) |

참고 데이터: <https://www.kaggle.com/c/pubg-finish-placement-prediction>