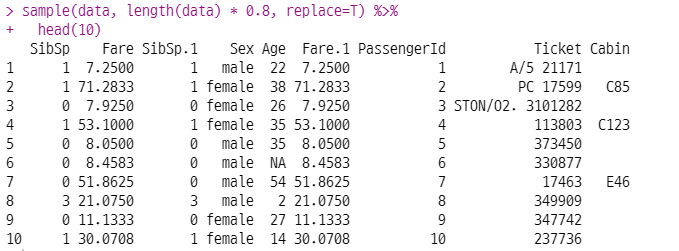
**[과제] 통계분석**

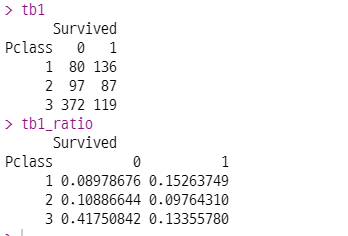
* 아래의 문제들을 해결하는 코드와 콘솔 실행 결과를 제출하세요.

**주어진 ‘titanic\_data.csv’ 파일을 다운받아 아래의 과제를 수행하시오.**

1. 전체 데이터로부터 80%의 훈련 샘플을 단순 임의 복원 추출하세요.

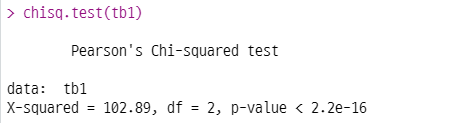


1. Pclass (선실 등급)과 Survived(생존 여부) 변수로 분할표와 비율 분할표를 작성하여 결과를 해석하세요.



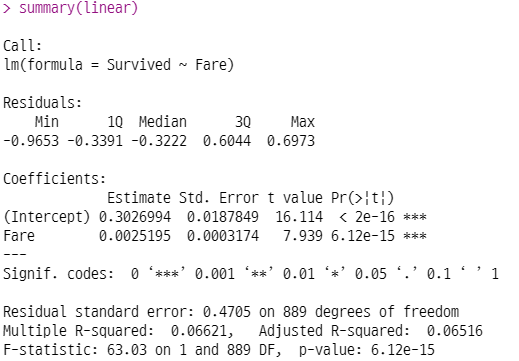
1번 클래스에 생존자가 많고 3번클래스에 사망자가 많다.

1. 위 분할표를 카이 제곱 검정하여 두 변수간 독립성 여부를 확인하세요.



P-value < 0.05 이므로 객실 등급과 생존에 관계가 없다는 귀무가설을 기각한다.

1. Fare(운임요금)이 Survived(생존 여부)에 영향을 주는지를 알아보기 위한 선형 모델을 구축하고, 결과를 해석하세요.



A0 = 0.0025195, a1 = 0.3026994

P-value < 0.05

운임요금이 생존과 관계가 없다는 귀무가설을 기각한다.