**[과제] 통계분석**

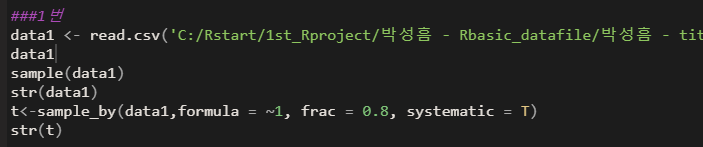
* 아래의 문제들을 해결하는 코드와 콘솔 실행 결과를 제출하세요.

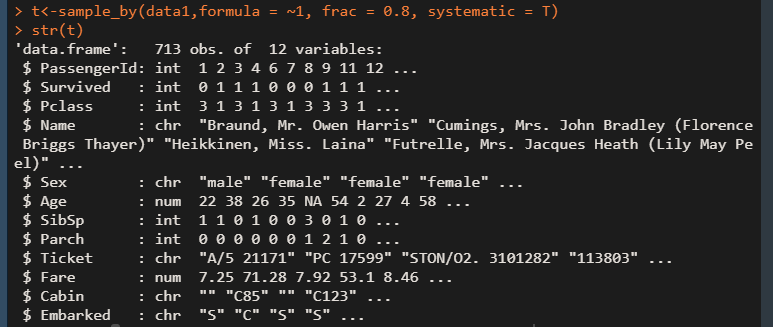
이름 : 박성흠

이메일 : bad\_wing@naver.com

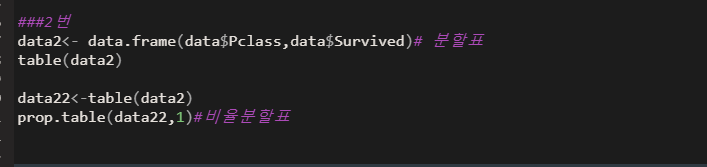
**주어진 ‘titanic\_data.csv’ 파일을 다운받아 아래의 과제를 수행하시오.**

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



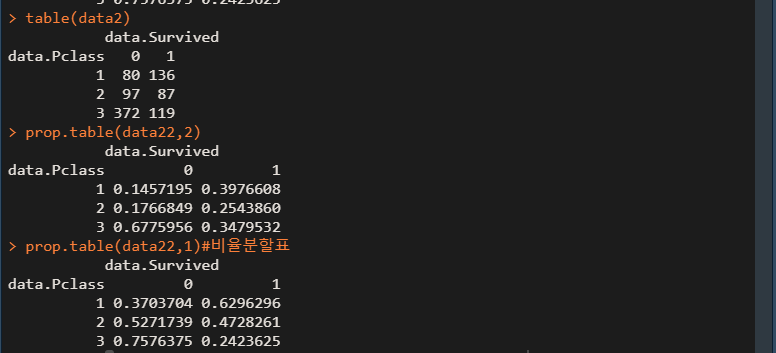


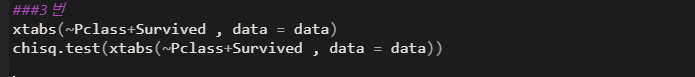
1. Pclass (선실 등급)과 Survived(생존 여부) 변수로 분할표와 비율 분할표를 작성하여 결과를 해석하세요.

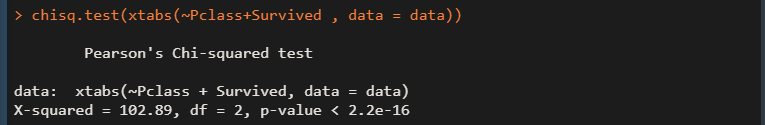


비생존자는 확실히 3class가 가장 높았고

생존자는 1class가 가장 높았다-> 비싼 값을 하는것 같다



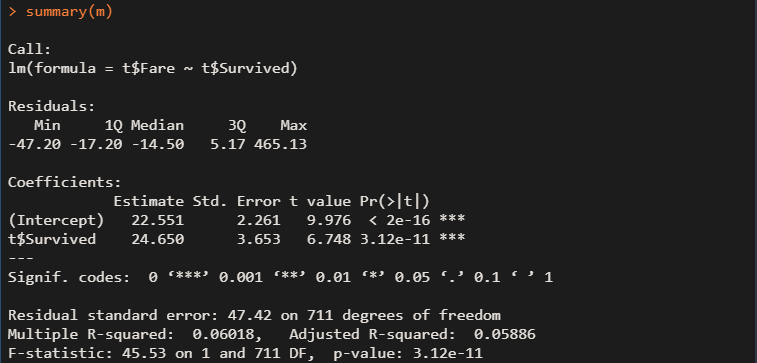
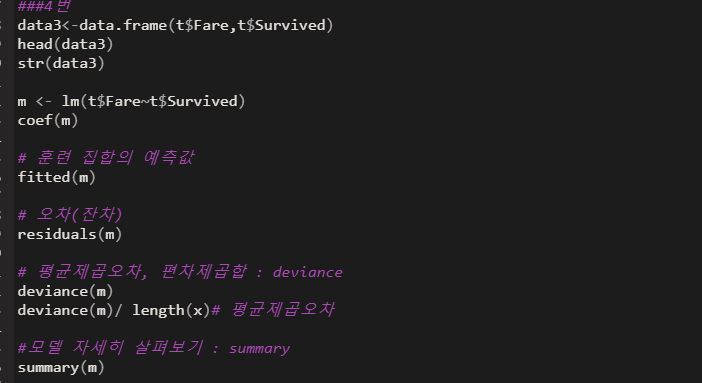
1. 위 분할표를 카이 제곱 검정하여 두 변수간 독립성 여부를 확인하세요.
2. 



Class와 survived는 독립이다 라는 귀무가설이

p-value=0.05기준보다 작으므로 기각된다

따라서 class와 survivied는 관계가 있다. 서로 영향을 준다

1. Fare(운임요금)이 Survived(생존 여부)에 영향을 주는지를 알아보기 위한 선형 모델을 구축하고, 결과를 해석하세요.
2. 
3. 
4. fare와 survived는 독립이다 라는 귀무가설이
5. p-value=0.05기준보다 작으므로 기각된다
6. 따라서 fare와 survivied는 관계가 있다. 서로 영향을 준다