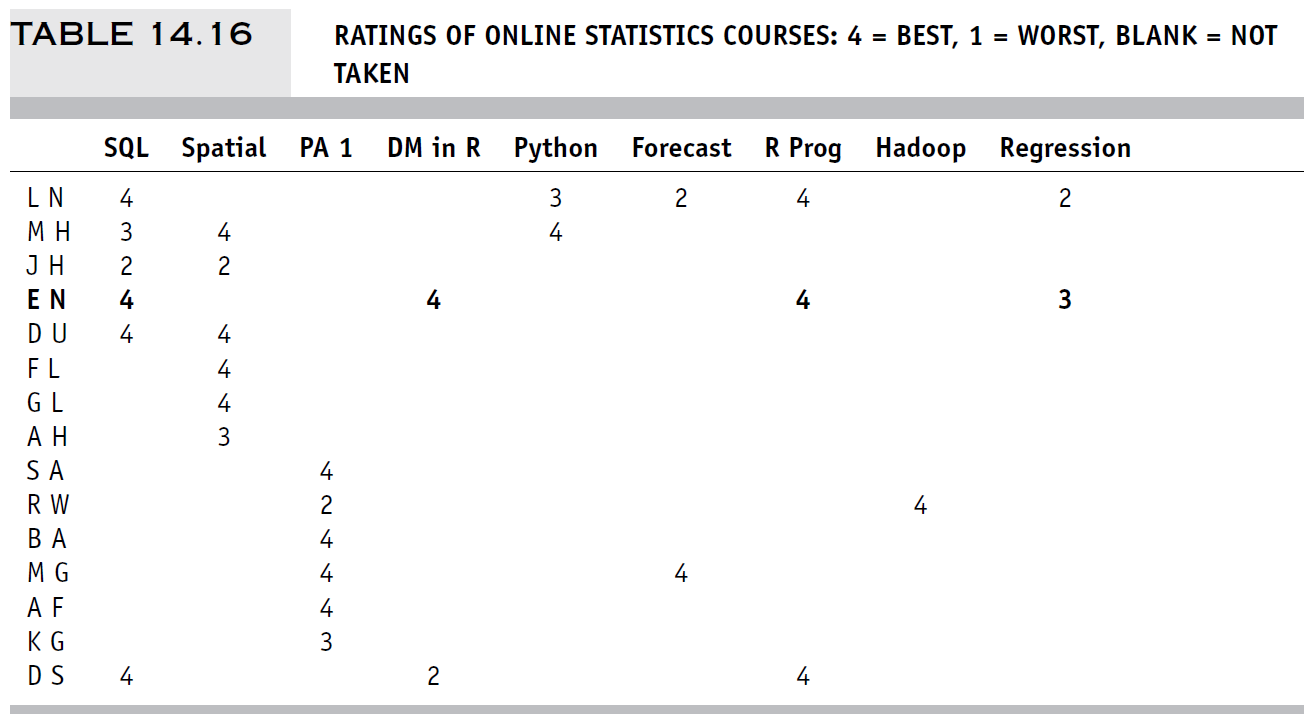
**[다섯 번째 과제]**

**IT응용시스템공학과**

**1494053**

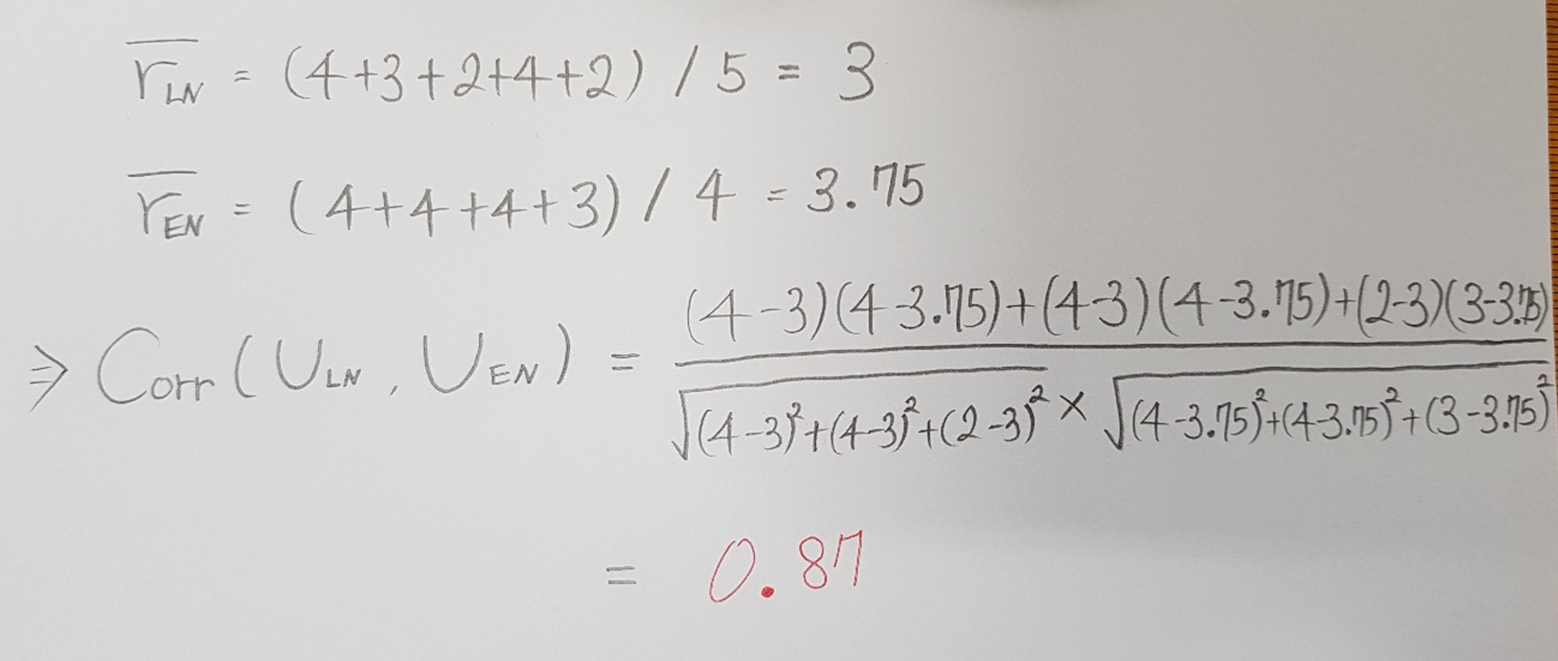
**김희택**

**14.5 강의 평가 점수 Statistics.com의 통계교육원(The Institute for Statistics Education)에서는 학생들이 한 과목을 수료하는 즉시 수료한 과목에 대해 다양한 측면에서 평가하도록 한다. 이 교육원에서는 평가가 끝난 후 학생들에게 추가적인 과목을 추천하는 시스템을 고려하고 있다. <표14.16>에 나타난 학생들의 일부 온라인 강의 평가점수와 E.N. 학생에게 어떤 과목을 추가적으로 추천할지 고려하시오.**

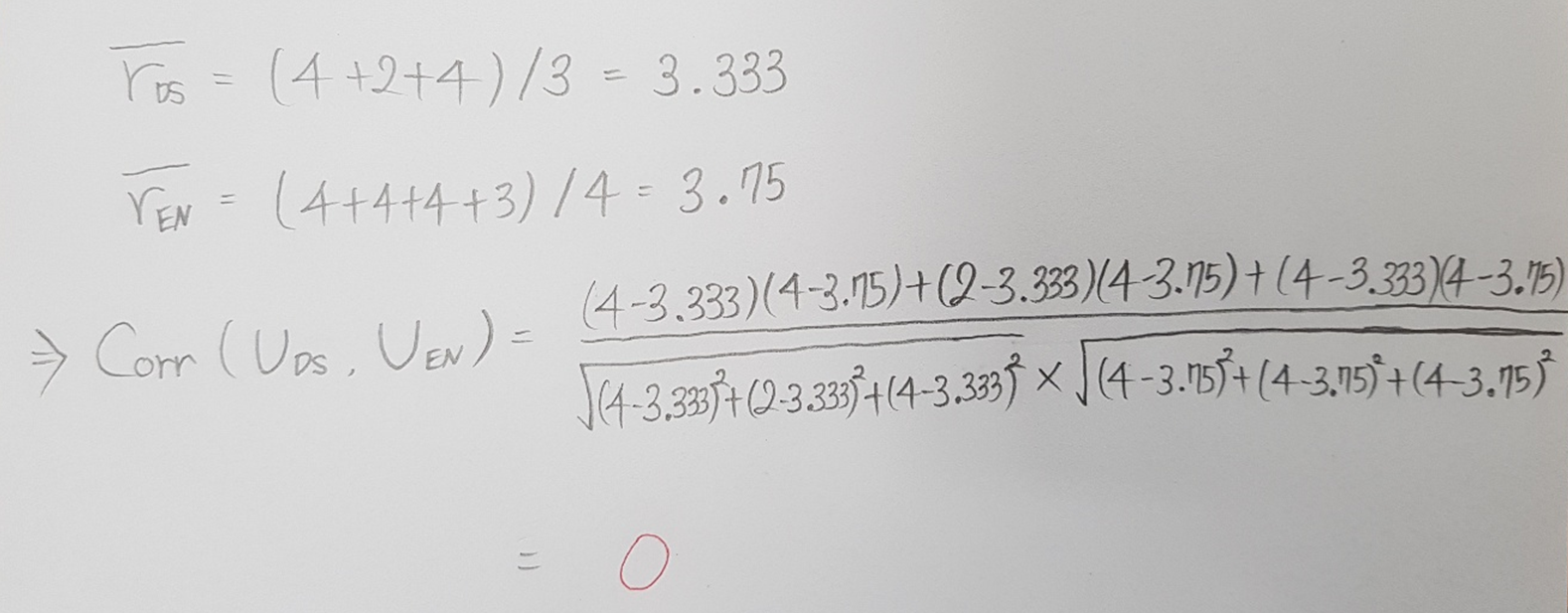


**a. 먼저 사용자 기반 협업필터를 고려해보자. 이는 모든 학생들 간의 상관계수의 계산을 요구한다. 어떤 학생들에 대해 E.N. 학생과의 상관계수를 계산할 수 있는가? 이를 계산하시오.**

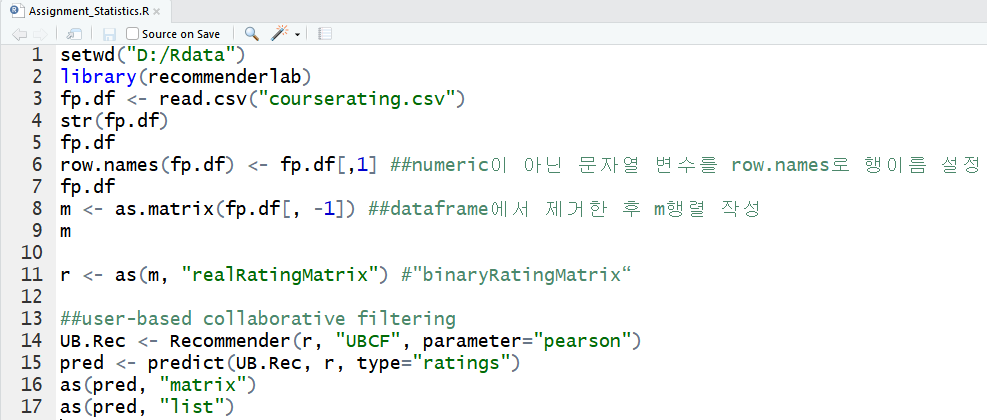
* 먼저 관심 대상의 E.N. 학생과 가장 비슷한 사용자는 L.N.과 D.S입니다.
* E.N. / L.N. co-rated ‘SQL, R Prog, Regression’
* E.N. / D.S. co-rated ‘SQL, DM in R, R Prog’
* L.N. 학생과 E.N. 학생의 상관계수는 다음과 같습니다.

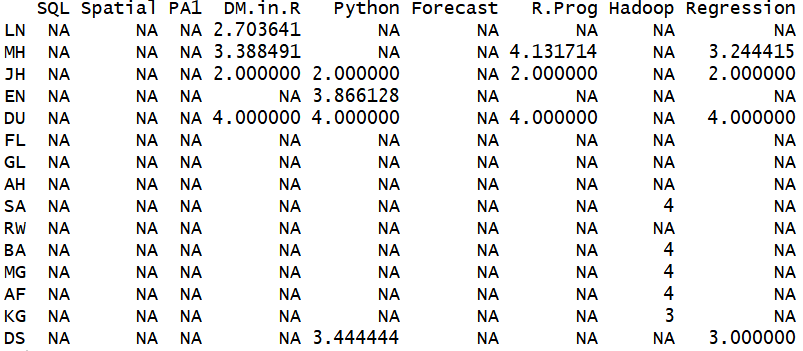


* D.S. 학생과 E.N. 학생의 상관계수는 다음과 같습니다.



**b. E.N. 학생과 가장 가까운 학생을 바탕으로 E.N. 학생에게 어떤 과목을 추천해야 하겠는가? 이유를 설명하시오. UB.Rec <- Recommender(r, “UBCF”, parameter=”pearson”)**

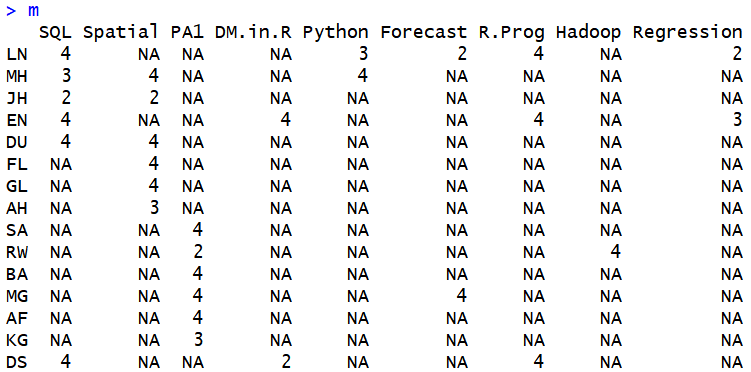
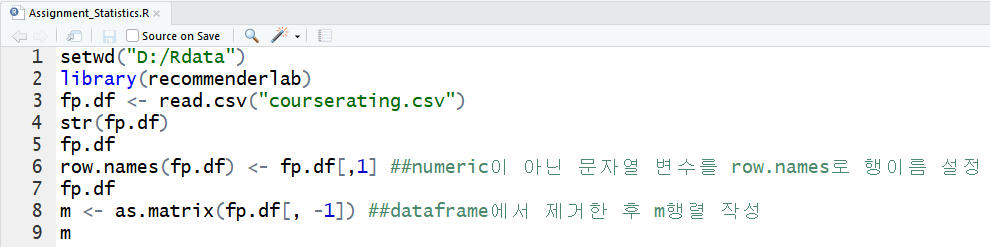




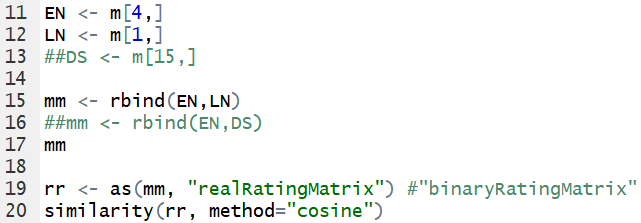
* 위 a. 에서 구한 상관계수를 통해 E.N. 학생과 가장 가까운 학생은 L.N.인 것을 알 수 있습니다.
* UB.Rec <- Recommender(r, “UBCF”, parameter=”pearson”) R을 이용하여 사용자 기반 협업 필터링 결과는 위와 같습니다. Pearson 상관 계수와 과목 추천 모델 Recommender 을 통해 E.N. 학생은 ‘Python’ 과목을 ‘3.866128’ 수치로 강한 추천을 받을 것이라 예상할 수 있습니다.
* 따라서 L.N. 학생을 바탕으로 E.N. 학생은 ‘Python’ 과목을 추천 받을 수 있습니다.
* **NA는 추천하지 말라는 것이다. 3.866128 – Python 추천 가능. (EN 한 행만 보고 판단 가능)**

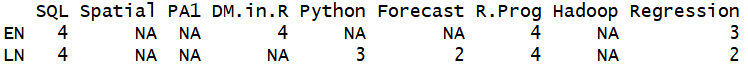
**c. R의 similarity() 함수 이용하여 EN과 고객간 코사인 유사도를 계산하시오.**

**#먼저 파일을 읽어서 numeric이 아닌 문자열 변수를 row.names로 행이름 설정하고 dataframe에서 제거한 후 m행렬 작성 (표14.8 참조)**



**#EN과 두 개 과목 이상 co-rated된 것(사용자)만을 m행렬에서 선택하여 (예를 들면 EN과 DU의 경우는 하나만 co-rated된 경우이며 이런 경우는 값에 상관없이 항상 cosine값이 1이 나온다.) mm행렬을 만든 후에 mm을 realratingMatrix인 rr로 변환 후 similarity에 사용 [similarity(rr, method=”cosine”)]**







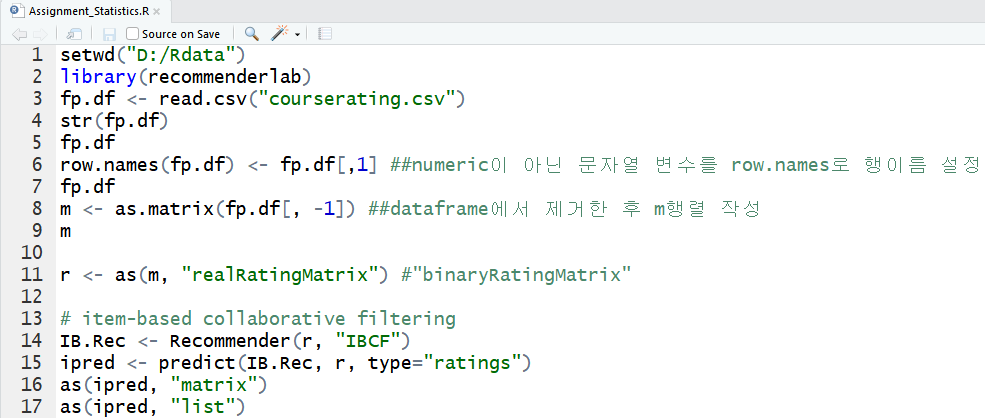
* 먼저 m행렬에서 E.N.과 두 개 과목 이상 co-rated된 것은 L.N.과 D.S. 입니다.
* m행렬에서 E.N.과 L.N. / D.S. 두 개의 행을 선택하여 위와 같이 각 각 mm행렬을 만들었습니다.
* similarity() 함수를 이용하여 E.N.과 L.N. / E.N.과 D.S. 간 코사인 유사도를 구했습니다.
* 따라서 **E.N.과 L.N. 간 코사인 유사도는 0.9891005** 이며,

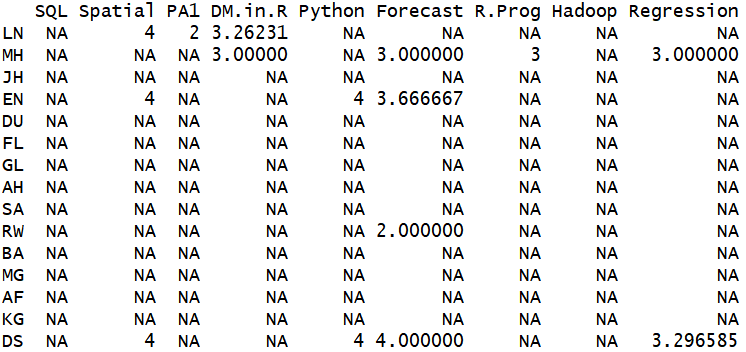
**E.N.과 D.S. 간 코사인 유사도는 0.9622504** 입니다.

**d. E.N. 학생과 코사인 유사도상으로 가장 가까운 학생을 바탕으로 E.N. 학생에게 어떤 학업과정을 추천해야 하겠는가? (similarity() 함수를 이 후 UBCF 를 이용해서 계산하면 b.의 결과와 동일하게 나온다.)**

* 위 c. 결과에 따르면 E.N 학생과 L.N. 학생 간 코사인 유사도가 ‘0.9891005’로 ‘1’에 거의 완벽히 매치되므로 L.N.이 D.S. 보다 가장 가까운 학생인 것을 알 수 있습니다.
* 따라서 L.N. 학생을 바탕으로 E.N. 학생은 ‘Python’과 ‘Forecast’ 과목을 추천 받을 수 있습니다.

**g. 항목 기반 협업필터링을 이 데이터에 작용하고 (R을 이용), 그 결과를 기반으로 E.N 학생에게 과목을 추천하시오.**





* 위 항목 기반 협업필터링을 이용하여 E.N. 학생은 ‘Spatial’, ‘Python’, ‘Forecast’ 과목을 추천 받을 수 있습니다.
* **숫자가 큰 것부터 추천해 주면 좋음 Spatial 🡪 Python 🡪 Forecast 순으로 추천 가능**