2022-2 Dry Assignment: 통계학

22100748 최윤영

1. 양성종앙과 악성종앙 모집단를 가장 잘 구분하기 위해 사용할 수 있는 파라미터 3개를 찾고 이를 얻기 위해 하였던 과정을 설명하여라.

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

* Shapiro-Wilk검정을 시행했을 때, 데이터가 정규성을 만족하지 않는다고 판단하여 Wilcoxon 검정을 시행하였다.
* Wilcoxon 검정을 시행했을 때, fractal\_dimension\_mean 변수만 p-value가 0.05보다 크게 나와서 평균 차이가 유의미하지 않다고 판단했다.
* 나머지 변수들에 대해서는 평균 차이의 비율을 비교하여 어떤 것이 가장 연관성이 큰지 비교하였다.

텍스트이(가) 표시된 사진

자동 생성된 설명

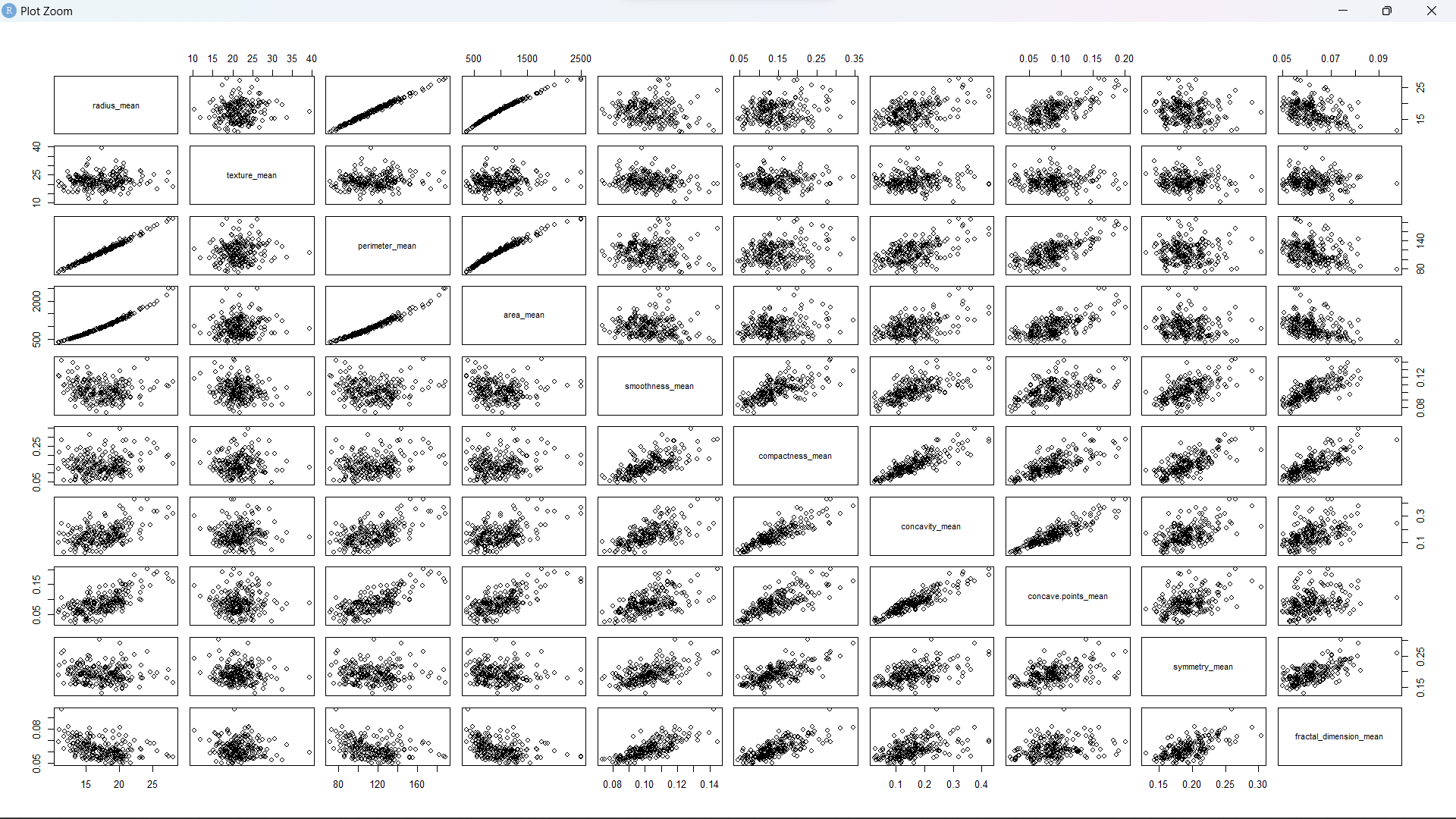
텍스트이(가) 표시된 사진

자동 생성된 설명

* 3번째 열의 ratio을 비교했을 때, 클수록 평균 차이의 비율이 크다고 할 수 있다.
* 큰 순서대로 3개 나열하면 concavity, concave.points, area이다.

따라서 양성종양과 악성종양을 구별하는 3개의 파라미터는 concavity, concave.points, area이다.

1. 악성종앙의 경우 어떤 파라미터가 평균 면적과 가장 큰 연관성을 가지고 있는지를 찾고 이를 얻기 위해 하였던 과정을 설명하여라.



텍스트이(가) 표시된 사진

자동 생성된 설명

* 악성종양만 따로 데이터를 만들어, 상관관계를 분석하였다.
* Scatter plot에서 볼 수 있듯이, area 변수와 가장 큰 연관성을 갖는 변수는 radius와 perimeter이다. (상관분석 표에서도 0.9가 넘는 값을 가지는 것을 알 수 있다)
* 따라서 평균면적과 가장 큰 연관성을 가지고 있는 변수들은 radius와 perimeter이다.