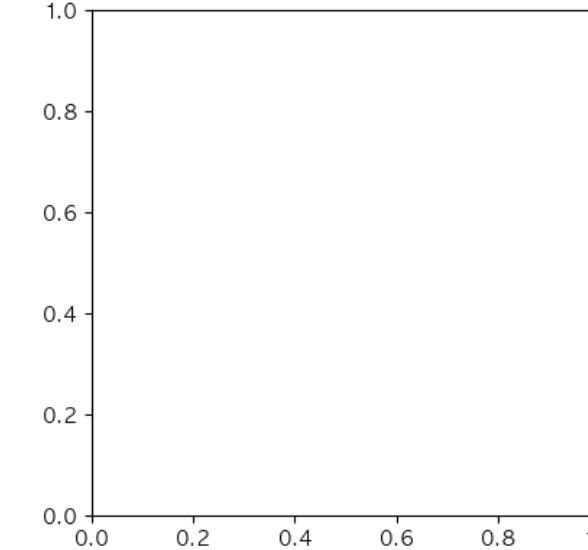
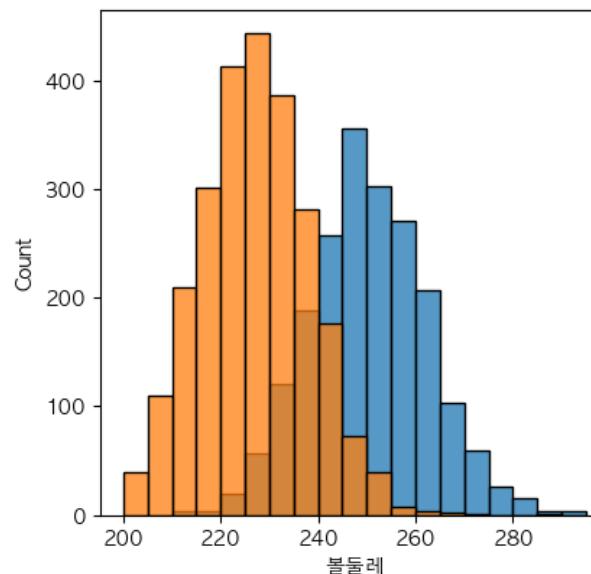
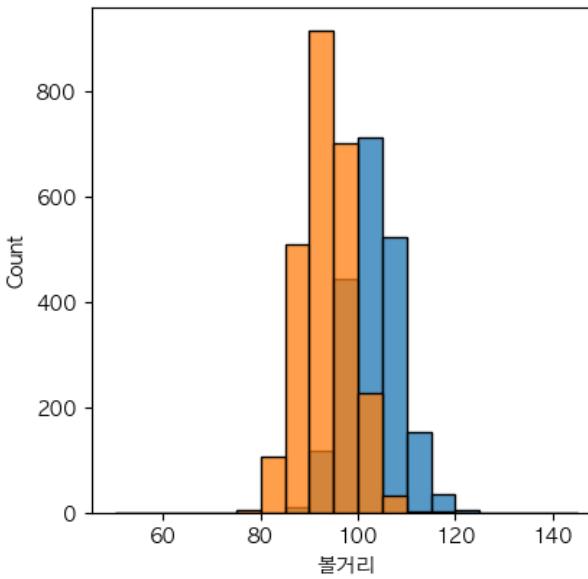
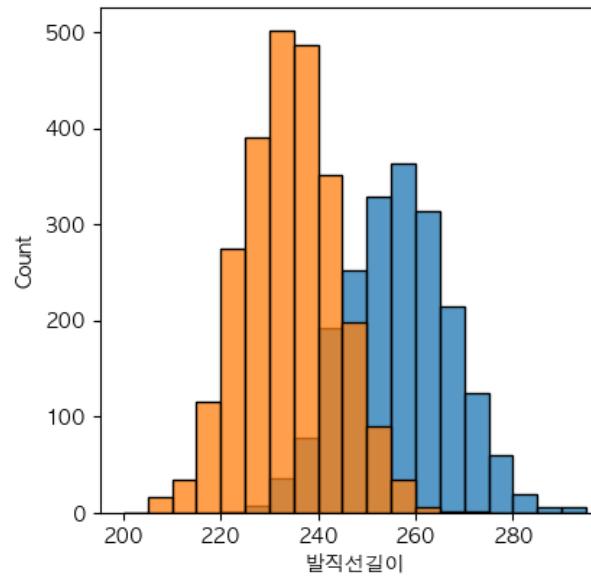


발 측정 항목 분석 및 군집 통한 발 모양 예측

권순형

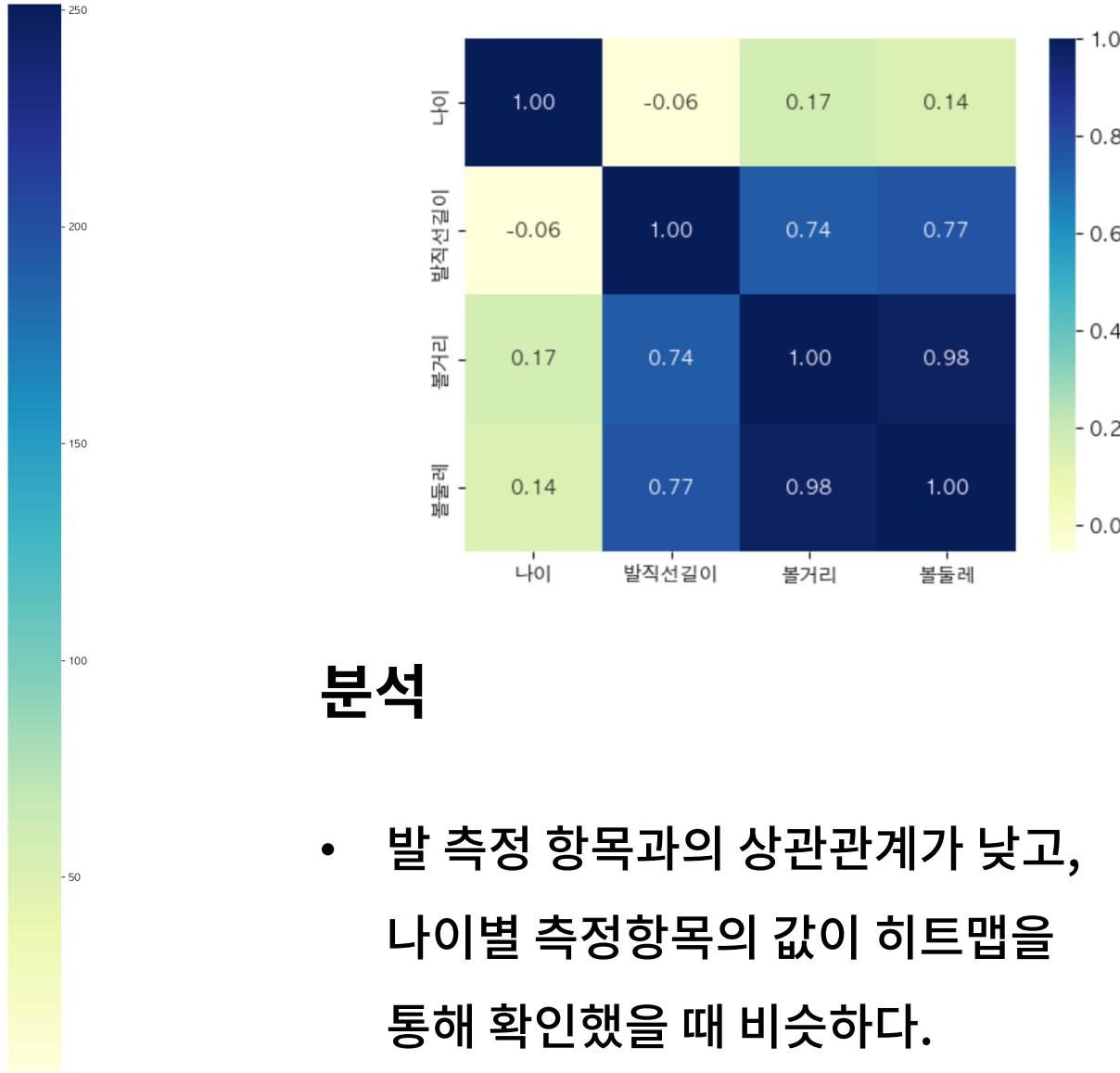
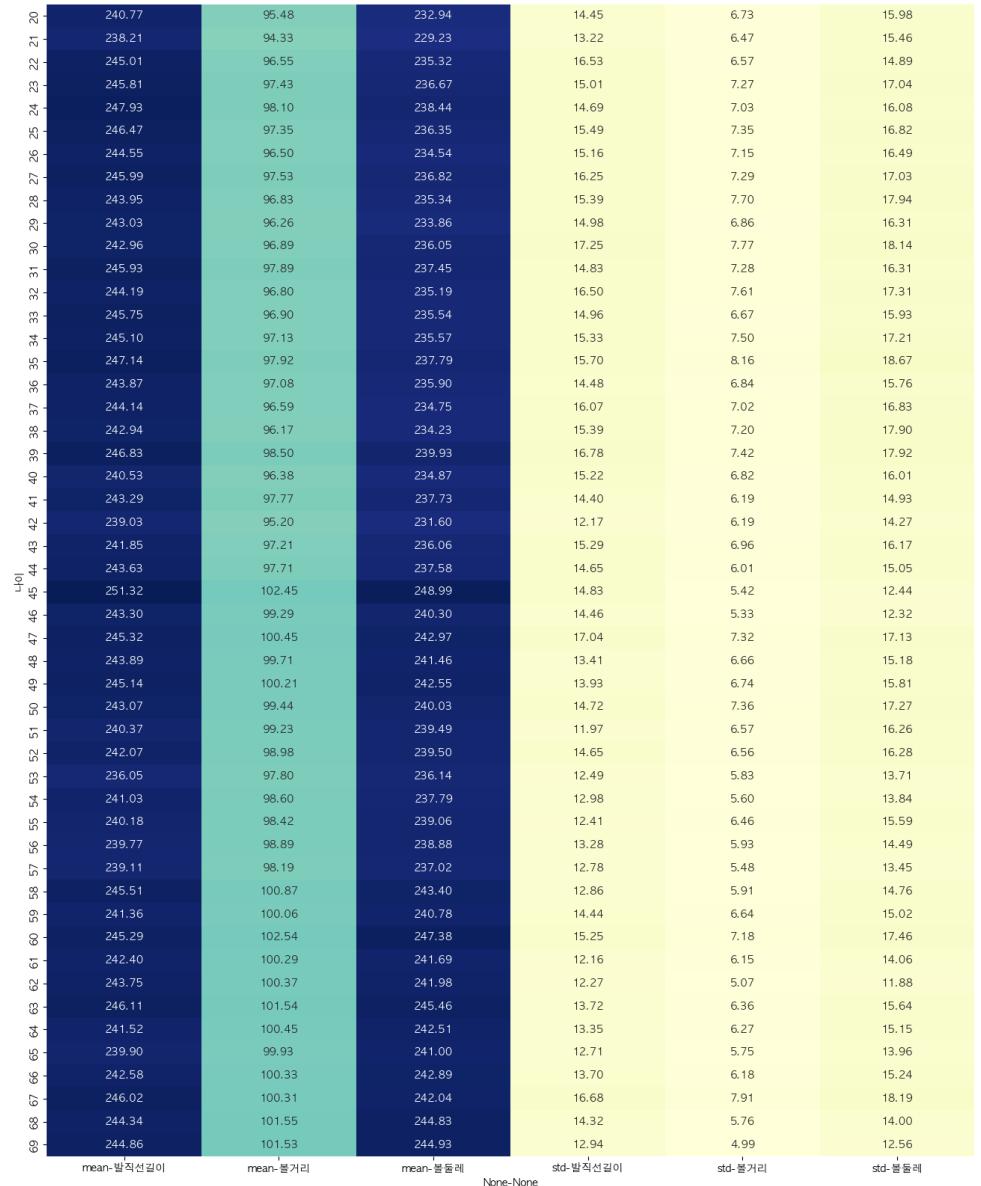
범주형 데이터에 따른 발 측정 항목 분포 - 성별



분석

- 발 측정항목의 대표적인 길이, 볼거리와 볼둘레를 비교
- 분포의 차이가 있음을 히스토그램을 통해 파악할 수 있다.

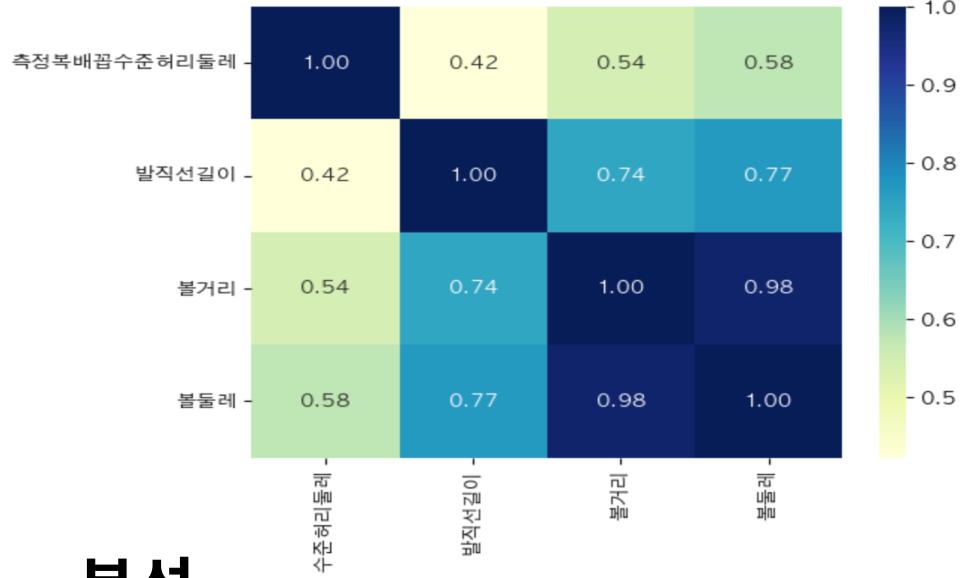
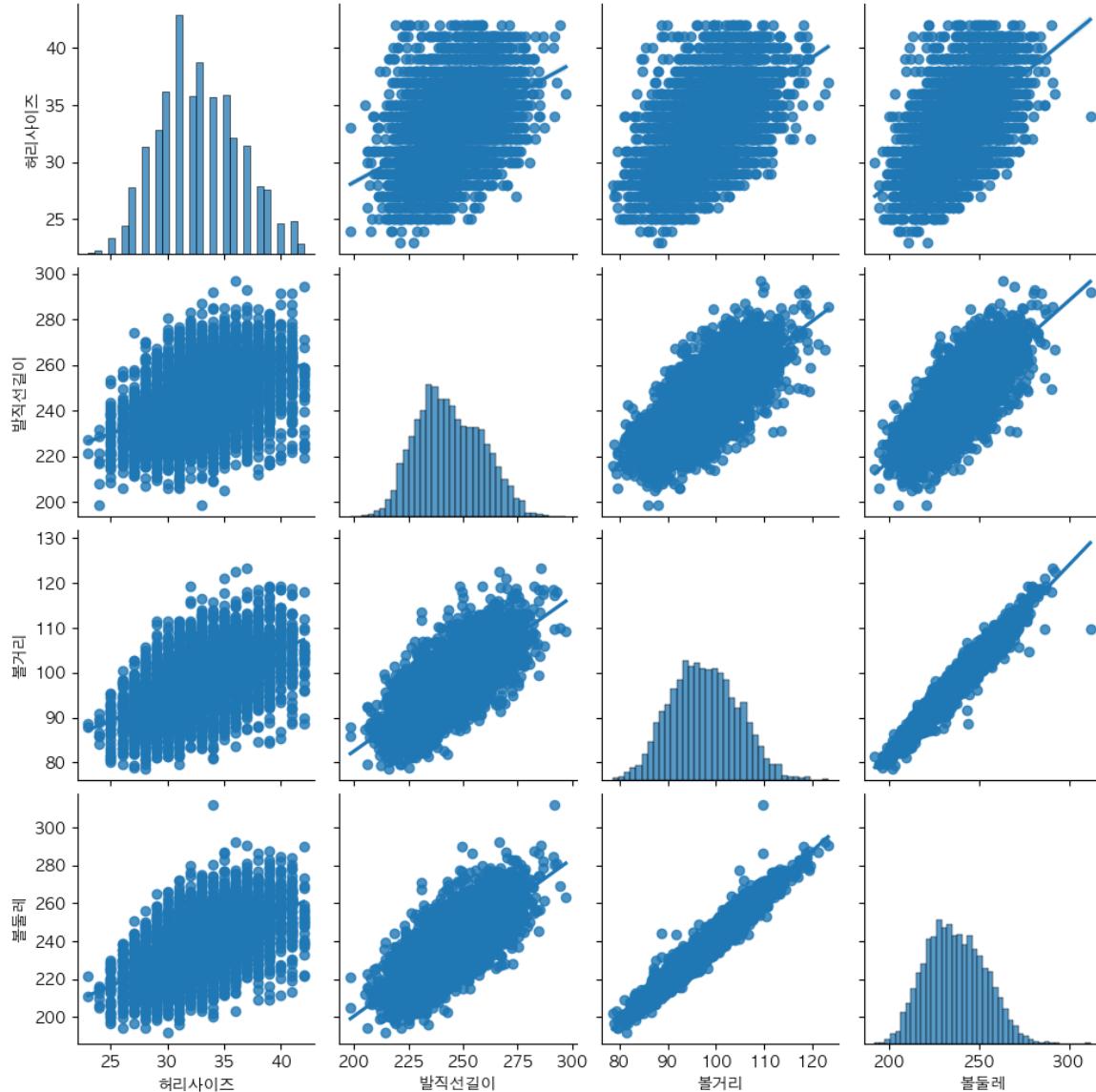
범주형 데이터에 따른 발 측정 항목 분포 - 나이



분석

- 발 측정 항목과의 상관관계가 낮고, 나이별 측정항목의 값이 히트맵을 통해 확인했을 때 비슷하다.

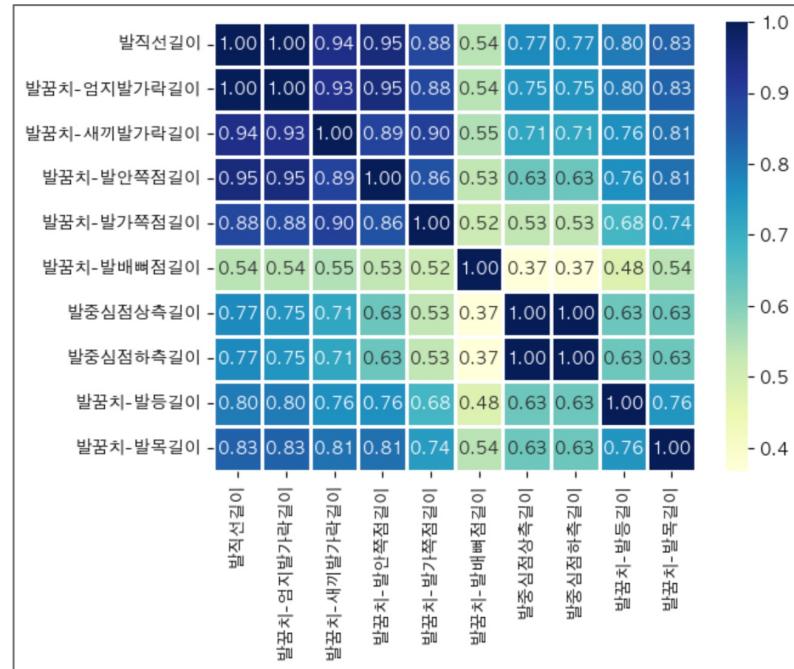
범주형 데이터에 따른 발 측정 항목 분포 - 허리둘레



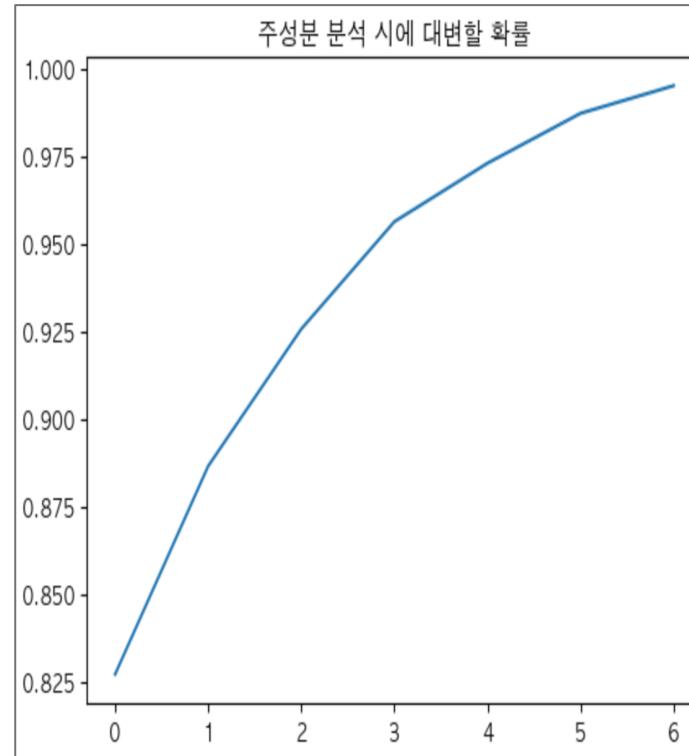
분석

- 허리 둘레를 흔히 사용하는 허리사이즈로 변환
- 선형적인 관계를 가지지만, 그 분포가 커서 데이터를 나눠지는 않겠다.

주성분 분석(PCA) - 길이 항목



길이 항목간 상관관계



	0	1
발직선길이	0.465532	-0.134146
발꿈치-엄지발가락길이	0.462061	-0.125679
발꿈치-새끼발가락길이	0.386520	-0.044874
발꿈치-발안쪽점길이	0.340432	-0.048640
발꿈치-발가쪽점길이	0.300748	0.001335
발꿈치-발배뼈점길이	0.195794	0.957115
발중심점상측길이	0.149943	-0.145889
발중심점하측길이	0.149943	-0.145889
발꿈치-발등길이	0.280335	-0.008499
발꿈치-발목길이	0.228362	0.055864

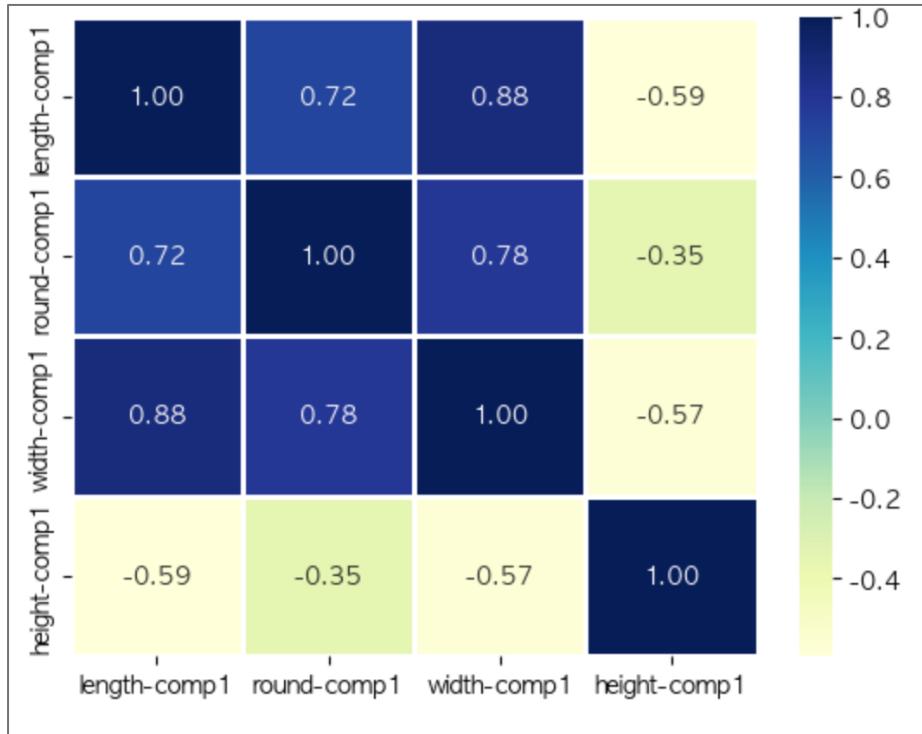
분석

- 항목간의 상관관계가 상당히 높다.
- 다중공선성의 위험이 있기에 제거 필요

분석

- 1주성분의 값들 중에 높다고 판단되는 특성들의 상관관계가 높게 판단이 되었다.
- 다중공선성의 위험이 있다고 판단하여서 VIF(분산 팽창 계수)를 통해서 10이상인지 확인 후 제거 후 다시 PCA 실행을 했다.

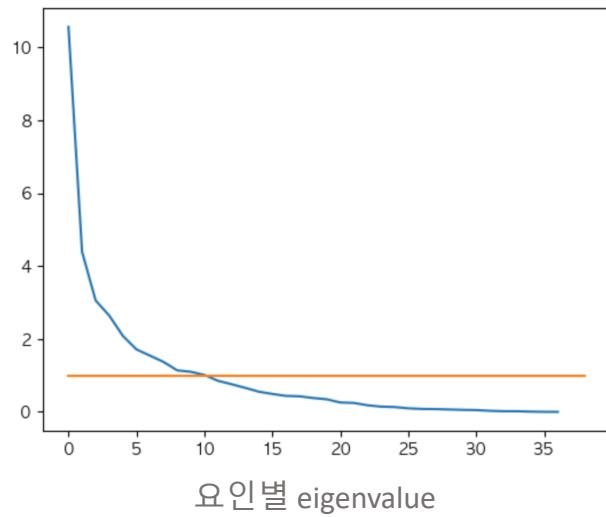
주성분 분석(PCA) - 5가지 항목 합치기



분석

- 항목별로 나누어서 다중공선성을 제거하면 변수간의 관계가 독립적이게 만들었다.
- 하지만 항목간 관계도 독립적이라고 생각을 해서 변수들을 나누었는데 이와 같지 않았던 것이라 파악이 된다.
- 그러므로, 변수 간의 관계가 있다고 가정하고 요인 분석 방법을 사용하려고 한다.

요인 분석(FA) - 남성 발 측정 항목



분석

- 고유값이 1 이상인 수를
최적의 요인 수로 찾았다.
그 수는 8로 나왔다.

	0	1	2	3	4	5	6	7
풀둘레	0.85	0.09	0.01	0.07	0.29	0.27	0.11	-0.11
발너비	0.76	0.07	0.09	0.13	0.21	0.49	0.11	0.18
발꿈치너비	0.73	-0.02	0.24	0.03	-0.04	-0.06	-0.05	0.03
발등수직둘레	0.88	0.15	0.02	0.04	0.15	0.19	0.18	0.02
발목수직둘레	0.80	0.21	0.15	0.06	0.15	0.03	0.16	-0.02
발꿈치-발목둘레	0.76	0.19	0.48	-0.00	0.04	-0.10	0.07	-0.00
기쪽복사수평둘레	0.70	-0.22	0.31	-0.07	0.02	-0.23	0.20	0.20
안쪽복사수평둘레	0.77	0.03	0.38	0.01	-0.03	-0.12	-0.03	0.07
발등높이	0.20	0.51	-0.00	-0.19	0.02	-0.19	0.42	0.20
발목높이	0.21	0.58	0.04	-0.15	0.14	-0.22	0.28	0.24
발꿈치위점높이	0.04	0.93	0.13	-0.03	0.02	0.03	0.09	-0.04
기쪽복사높이	0.08	0.89	0.10	-0.07	0.14	-0.03	0.10	-0.03
기쪽복사아래높이	-0.01	0.85	0.13	-0.01	-0.09	0.09	0.13	-0.03
발작선길이	0.29	0.11	0.87	0.07	0.31	0.03	0.04	-0.08
발꿈치-엄지발가락길이	0.29	0.09	0.87	0.09	0.27	0.03	0.04	-0.10
발꿈치-새끼발가락길이	0.27	0.10	0.83	0.03	0.26	0.01	-0.04	0.16
발꿈치-발안쪽점길이	0.27	0.09	0.86	0.01	-0.01	0.12	0.09	-0.29
발꿈치-발가쪽점길이	0.06	0.11	0.88	0.05	-0.06	-0.00	0.13	0.30
내측발너비	0.27	-0.04	0.13	0.87	0.09	0.10	0.09	-0.06
엄지발가락축각도	0.07	-0.12	-0.02	0.76	0.15	0.13	-0.04	0.18
발중심선각도	0.15	0.06	-0.07	-0.81	-0.02	0.22	-0.08	0.21
발중심점상축각이	0.23	0.06	0.30	0.16	0.86	-0.05	-0.13	-0.06
발중심점하축각이	0.23	0.06	0.30	0.16	0.86	-0.05	-0.13	-0.06
새끼발가락축각도	0.14	0.01	-0.09	-0.34	0.20	0.56	0.12	0.32
풀연령률	0.03	-0.08	0.13	0.11	-0.18	0.81	-0.10	0.01
엄지발가락높이	0.05	0.15	0.07	0.04	-0.09	0.00	0.85	-0.12
새끼발가락높이	0.12	0.11	0.02	0.10	-0.12	0.16	0.78	-0.08
첫째발 하리뼈높이	0.16	0.20	0.09	-0.01	0.02	-0.17	0.80	0.07
발뼈뼈높이(Arch높이)	0.18	0.30	-0.01	0.14	-0.25	-0.04	-0.22	0.37
발불각도	-0.02	0.00	-0.03	0.08	0.06	-0.14	0.05	-0.90



요인별 특징 구분

1요인 : 둘레

2요인 : 높이

3요인 : 발 테두리 길이

4요인 : 발 왼쪽 휘어짐 상태

5요인 : 발 안쪽 길이

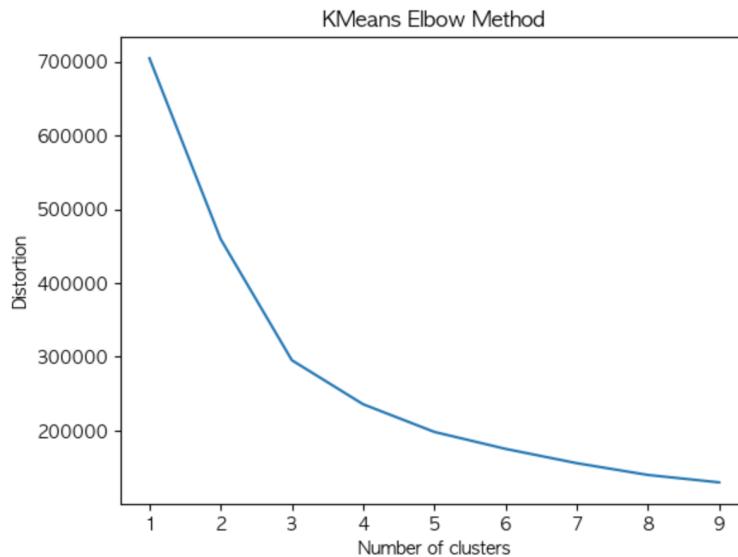
6요인 : 발 오른쪽 휘어짐 상태

7요인 : 발 앞부분 높이

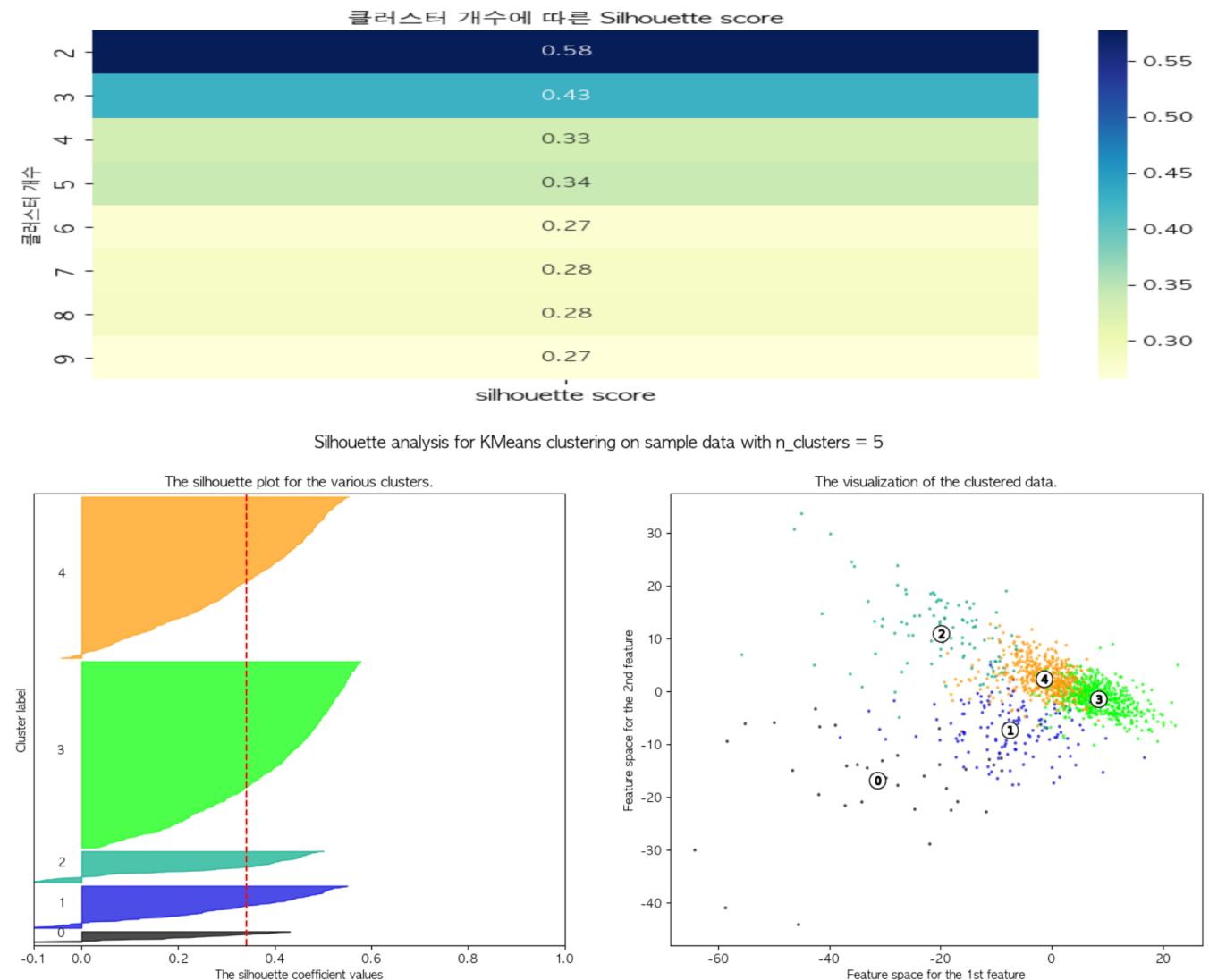
8요인 : 발 가로 각도

Clustering(Kmeans)- 최적의 K값 찾기

Elbow Method



Silhouette Score



분석

- Elbow method와 Silhouette score를 통해 클러스터 개수를 5개로 정했다.

Clustering(Kmeans)- 분산 분석 및 발 형태 분석

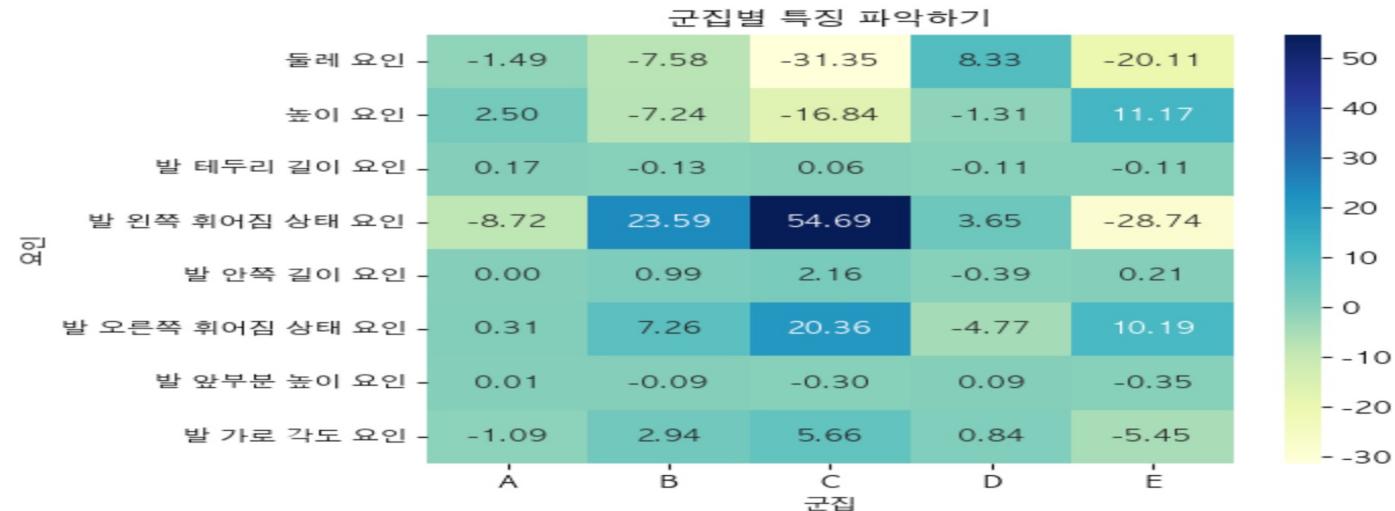
분산분석

	f-statistic	p-value
둘레 요인	826.337385	0.000000e+00
높이 요인	579.307628	1.724465e-290
발 테두리 길이 요인	6.318295	4.900107e-05
발 왼쪽 휘어짐 상태 요인	1201.811299	0.000000e+00
발 안쪽 길이 요인	112.501550	8.956155e-83
발 오른쪽 휘어짐 상태 요인	405.257512	5.011997e-229
발 앞부분 높이 요인	5.002348	5.282122e-04
발 가로 각도 요인	433.065685	9.312188e-240

분석

- p-value 값이 모두 0.05 보다 작기에 클러스터별 차이가 유의미하다고 판단했다.

군집별 특징 파악



군집별 발 모양

A : 발이 왼쪽으로 덜 휘어진 모양

D : 발 둘레가 넓은 모양

B : 발이 왼쪽으로 휘어지고,
둘레나 높이가 낮은 발 모양

E : 발 둘레는 좁고, 발 높이가 높으며
발이 오른쪽으로 휘어진 모양

C : B모양에 비해 발의 휘어짐이 세고
둘레나 높이도 더 낮은 발 모양