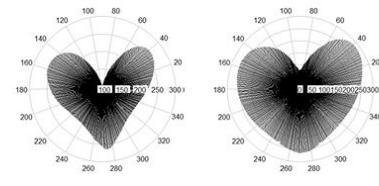


하트 프로젝트

가정 및 목표

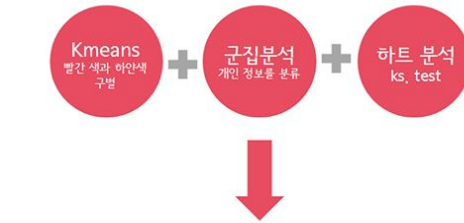


사랑에 대한 개인요소가 하트 모양에 영향을 줄 것이다

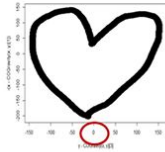
Data 설명



분석 방법 및 순서



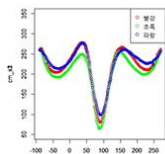
Step1. 맨 처음에 그림을 불러왔을 때 회색과 빨간색만 있다고 생각했지만 실제 값은 두가지 색이 아닌 여러가지 색이 있었습니다. 붉은색도 강 붉은색 검붉은색 등등이 있었죠. 분석시에는 문제가 생겨 불편할 것 같아서 2개로 Kmeans를 해 바탕과 하트 그림을 구별해냈습니다.



Step2. 그 다음에는 하트 그림의 무게중심을 구해 그 무게중심을 (0, 0)으로 놓고 각 각도마다 여러개 점이 있어서 극좌표계로 만들어 각 각도마다의 길이를 평균을 내서 플롯팅을 했다.

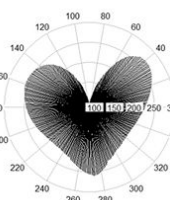
성별	연령	연애	취업	학력	소득	종교	가족	친구	취미	스포츠	여행	음악	영화	게임	독서	운동	수면	음식	주거	교통	환경	기타
1	20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Step3. 개인정보 중에서 연속형 변수 값을 클러스터링 하여 연속형 변수를 적절한 이산형 변수로 만들었다.



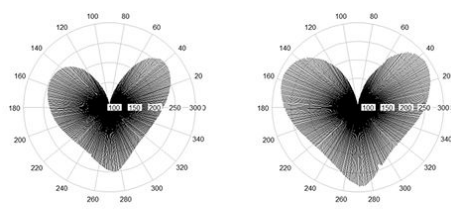
Step4. 하트의 좌표를 플롯팅 한 후에 그룹간 하트 모양의 차이가 있는 지를 Kolmogorov-Smirnov 검정 방법으로 유의수준 0.01을 기준으로 차이 여부를 검정하였다.

분석 결과



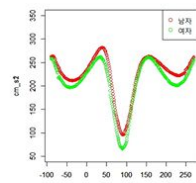
총 N=116명의 하트의 전체 평균입니다. 오른쪽이 더 위쪽에 위치한 것을 알 수 있는데 이는 오른손잡이가 많아서 그럴 것 같다는 판단을 했습니다.

성별의 차이에 따른 하트 모양



남성의 하트 평균치

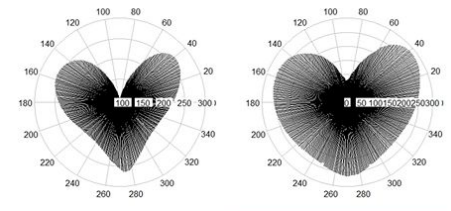
여성의 하트 평균치



```
> # Two-sample Kolmogorov-Smirnov test
> #
> # data: cm.s1 and cm.s2
> # D = 0.2687, p-value = 9.557e-12
> # alternative hypothesis: two-sided
> # 유의수준 0.01기준 기각할 수 있다.
> # 두 분포는 다른 분포라 생각된다.
> # 하트가 다르다고 생각된다.
```

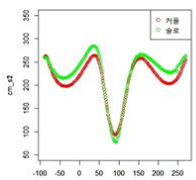
즉, 성별에 따라 그리는 하트의 모양이 다르다는 결론

현재 연애 여부에 따른 하트 모양



연애 중

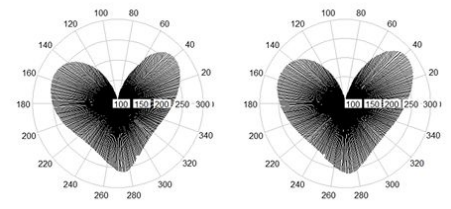
솔로



```
> # Two-sample Kolmogorov-Smirnov test
> #
> # data: cm.s1 and cm.s2
> # D = 0.2936, p-value = 6.073e-14
> # alternative hypothesis: two-sided
> # 유의수준 0.01기준 기각할 수 있다.
> # 두 분포는 다른 분포라 생각된다.
> # 하트가 다르다고 생각된다.
```

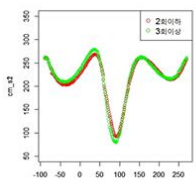
하트의 모양과 현재의 연애 여부는 관련이 있다고 보여짐!

연애 횟수에 따른 하트 모양



2회 이하

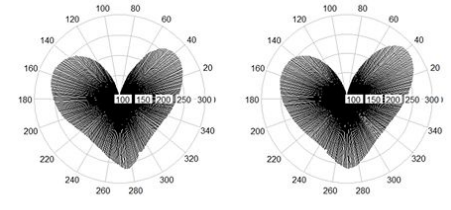
3회 이상



```
> ## 연애 횟수
> #ks.test
> #ks.test(cm.s1, cm.s2)
> # Two-sample Kolmogorov-Smirnov test
> #
> # data: cm.s1 and cm.s2
> # D = 0.1025, p-value = 0.04509
> # alternative hypothesis: two-sided
> # 유의수준 0.01기준 기각할 수 없다.
> # 두 분포는 같은 분포라 생각된다.
> # 하트가 같다고 생각된다.
```

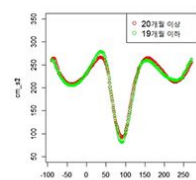
하트의 모양과 연애 횟수는 관련이 없음

최장 연애 길이에 따른 하트 모양



20개월 이상

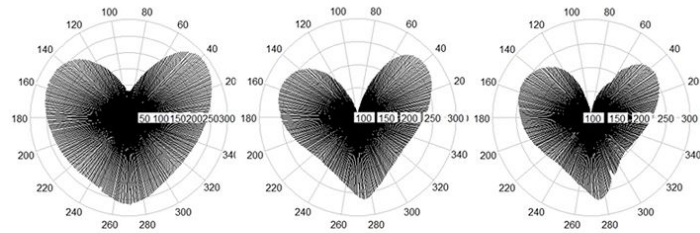
19개월 이하



```
> ## 최장 연애길이
> #ks.test
> #ks.test(cm.s1, cm.s2)
> # Two-sample Kolmogorov-Smirnov test
> #
> # data: cm.s1 and cm.s2
> # D = 0.0893, p-value = 0.3521
> # alternative hypothesis: two-sided
> # 하트가 같다고 생각된다.
```

하트의 모양과 최장 연애기간은 관련이 없음

평균 연애기간에 따른 하트 모양



짧다

중간

길다

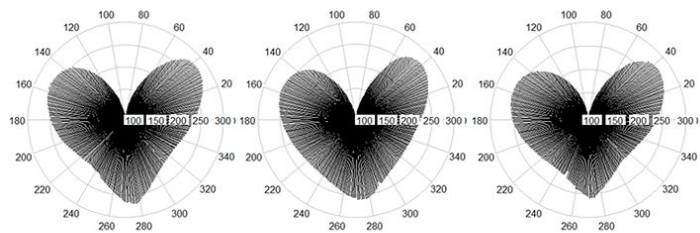
```
> ## 평균 연애 길이
> #ks.test
> #짧은, 긴
> #ks.test(cm_s1, cm_s2)
> # Two-sample Kolmogorov-Smirnov test
> #
> # data: cm_s1 and cm_s2
> # D = 0.2078, p-value = 3.42e-07
> # alternative hypothesis: two-sided
> # 유의수준 0.01기준 기각할 수 있다.
> # 두 분포는 다른 분포라 생각된다.
> # 하트가 다르다고 생각된다.
```

```
> #긴, 중간
> #ks.test(cm_s2, cm_s3)
> # Two-sample Kolmogorov-Smirnov test
> # data: cm_s2 and cm_s3
> # D = 0.169, p-value = 6.676e-05
> # alternative hypothesis: two-sided
> # 유의수준 0.01기준 기각할 수 있다.
> # 두 분포는 다른 분포라 생각된다.
> # 하트가 다르다고 생각된다.
```

```
> #짧은, 중간
> # Two-sample Kolmogorov-Smirnov test
> # data: cm_s1 and cm_s3
> # D = 0.1634, p-value = 0.0001298
> # alternative hypothesis: two-sided
> # 유의수준 0.01기준 기각할 수 있다.
> # 하트가 다르다고 생각된다.
```

즉, 평균 연애 기간과 하트의 모양은 관련이 있다.

자신의 외모 점수에 따른 하트 모양



고득점

중간 득점

저득점

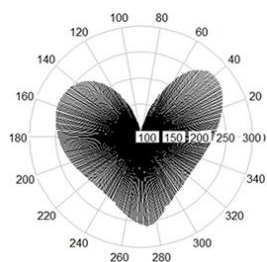
```
> ## 외모점수
> #고, 중
> # Two-sample Kolmogorov-Smirnov test
> # data: cm_s1 and cm_s2
> # D = 0.1219, p-value = 0.009375
> # alternative hypothesis: two-sided
> # 유의수준 0.01기준 기각할 수 있다.
> # 하트가 다르다고 생각된다.
> # 큰 차이는 없다.
```

```
> #중, 저
> # Two-sample Kolmogorov-Smirnov test
> # data: cm_s2 and cm_s3
> # D = 0.0997, p-value = 0.05519
> # alternative hypothesis: two-sided
> # 귀무가설 기각못함. 같은 분포라 생각된다.
```

```
> #고, 저
> # Two-sample Kolmogorov-Smirnov test
> # data: cm_s1 and cm_s3
> # D = 0.0942, p-value = 0.08134
> # alternative hypothesis: two-sided
> # 귀무가설 기각못함. 같은 분포라 생각된다.
```

자신의 외모에 대한 생각은 하트의 모양에 크게 영향을 미치지 않는 모습이다.

보완점



1. 심리적인 요인 추가
2. 데이터 수의 부족
3. 하트 분석에 대한 이론적 뒷바침 연구

방성수 고려대학교 통계대학원

서덕성 서울시립대 수학과

김상진 인하대학교 통계학과

김현주 숙명여자대 경제학과