[빅데이터 기반 다변량 자료 분석]

제목 : 결혼정보회사 설문지 분석을 통한 커플 성사방법 제시

응용통계학과
32142221- 서근태

1. 프로젝트 배경

일반적으로 결혼정보회사는 고객이 사전에 작성한 프로필과 설문지를 기반으로 수많은 소개팅을 주선합니다. 하지만 소개팅 횟수 대비 커플 성사율이 높지 않아 결혼정보회사와 고객 모두 경제·시간적 손실이 증가하고 있는 추세입니다. 실제로 결혼정보회사들도 커플 성사율을 높이기 위해 많은 방법을 찾고 있는데 저는 머신러닝 모델을 이용해서 분석해보고자 합니다.

2. 프로젝트 목표

빅데이터 기반의 다양한 머신러닝 모델을 이용한 커플 성사율 증가

3. 본문

다변량 자료 분석

3-1. 요인분석 / K-mean 군집분석 / 다차원 척도법

1. KMO와 Bartlett의 검정
2. 회전된 성분 행렬
3. 평균연결법을 사용한 덴드로그램(집단 내)
4. 스크리 도표
5. 설명된 총분산
6. 성별, 학년별 행렬 산점도(성별과 학년에 따른 차이)
7. 요인분석, K-mean 군집분석, 다차원 척도법의 정보량이 결합된 그래프
   (※ 맨 뒤에 후첨)

KMO와 Bartlett의 검정

표본 적절성의 Kaiser-Meyer-Olkin 측도.		.591
Bartlett의 구형성 검정	근사 카이제곱	39.983
	자유도	36
	유의확률	.298

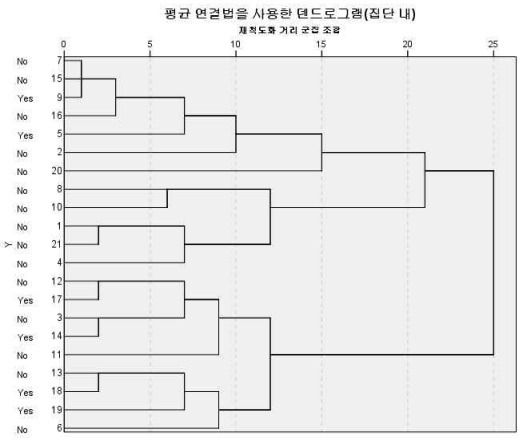
1. KMO와 Bartlett의 검정

회전된 성분행렬[a]

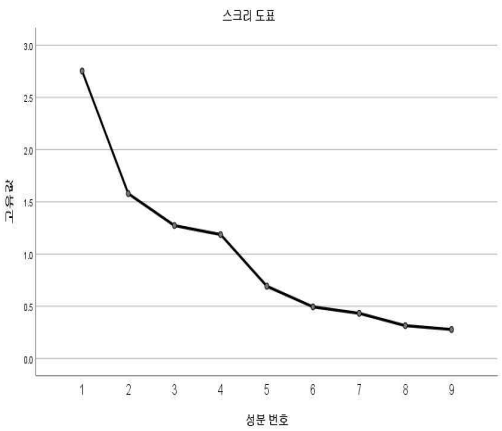
	성분			
	1	2	3	4
학력	.812	.232	-.123	-.078
외모	.647	.218	.219	-.498
재력	.783	-.051	.271	.353
직업	.311	.654	.198	.225
성격	-.120	.823	.002	-.008
장래성	.295	.825	-.006	-.066
종교	-.051	.046	.862	-.165
가정환경	.182	.057	.851	.057
자녀계획	.038	.103	-.069	.920

추출 방법: 주성분 분석.
회전 방법: 카이저 정규화가 있는 베리맥스.

2. 회전된 성분 행렬



3. 평균연결법을 사용한 덴드로그램



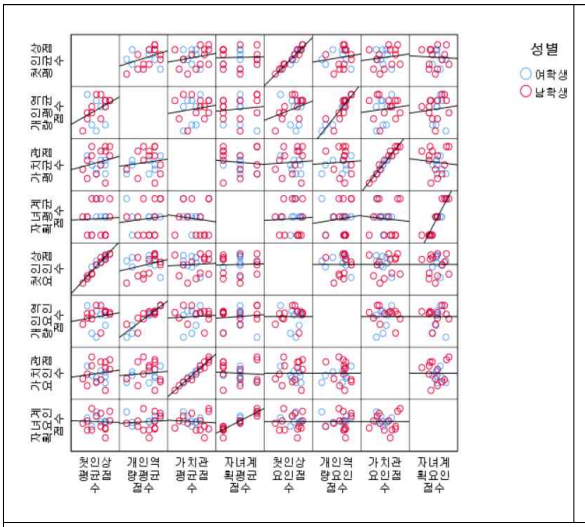
4. 스크리 도표

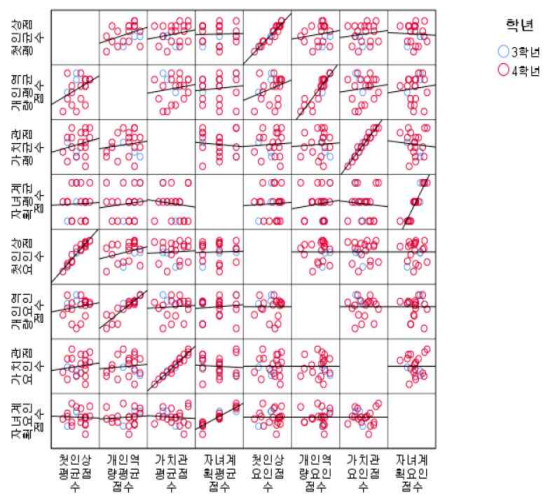
설명된 총분산

성분	초기 고유값			추출 제곱합 적재량			회전 제곱합 적재량		
	전체	% 분산	누적 %	전체	% 분산	누적 %	전체	% 분산	누적 %
1	2.754	30.595	30.595	2.754	30.595	30.595	1.926	21.395	21.395
2	1.578	17.539	48.134	1.578	17.539	48.134	1.906	21.179	42.574
3	1.273	14.140	62.274	1.273	14.140	62.274	1.649	18.318	60.892
4	1.186	13.181	75.456	1.186	13.181	75.456	1.311	14.564	75.456
5	.692	7.688	83.144						
6	.494	5.485	88.629						
7	.432	4.803	93.432						
8	.314	3.489	96.920						
9	.277	3.080	100.000						

추출 방법: 주성분 분석.

5. 설명된 총분산





6. 성별, 학년별 행렬 산점도(**성별**과 **학년**에 따른 차이)

3-2. 판별분석

1. 표준화 정준 판별함수 계수(**성별**과 **학년**에 따른 차이)
2. 함수의 집단 중심값
3. 2번을 이용한 실제 분류 결과 : Dis_1이 예측한 분류값
4. 분류 결과
5. 정준 판별함수

표준화 정준 판별함수 계수

	함수 1
학력	.586
외모	-.655
재력	.165
직업	.499
성격	-.345
장래성	.402
종교	.415
가정환경	.132
자녀계획	-.362

1. 표준화 정준 판별함수 계수(성별과 학년에 따른 차이)

함수의 집단 중심값

결혼여부_2점	함수 1
NO	-.421
YES	1.052

표준화하지 않은 정준 판별함수가 집단 평균에 대해 계산되었습니다.

2. 함수의 집단 중심값

결혼여부2점	Dis_1	Dis1_1
NO	YES	.70853
NO	NO	.24746
NO	NO	-.17246
NO	NO	-2.62730
YES	YES	2.09871
NO	NO	-.35979
NO	YES	.34499
NO	NO	-2.06592
YES	NO	-.48559
NO	NO	-.88765
NO	YES	1.05873
NO	NO	-.35357
NO	NO	-.66114
YES	YES	1.72539
NO	YES	.48334
NO	NO	-.87767
YES	YES	1.24411
YES	YES	.91059
YES	YES	.81951
NO	NO	-1.43650
NO	NO	.28623

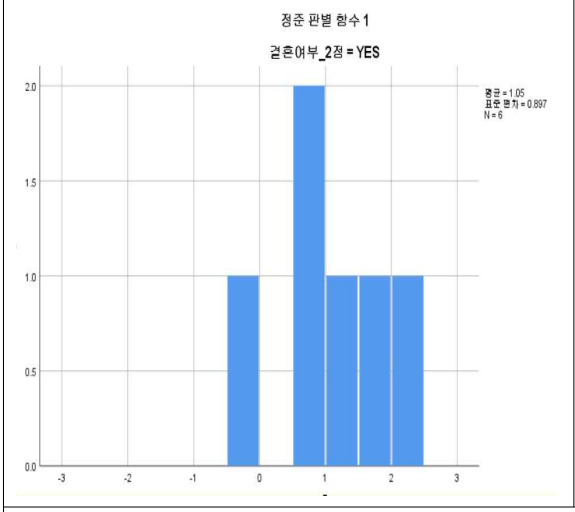
3. 2번을 이용한 실제 분류 결과 : Dis_1이 예측한 분류값

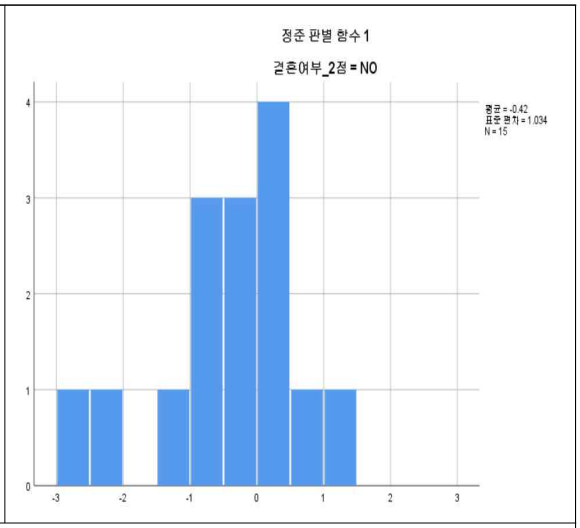
분류 결과[a]

			예측 소속집단		
		결혼여부_2점	NO	YES	전체
원래값	빈도	NO	11	4	15
		YES	1	5	6
	%	NO	73.3	26.7	100.0
		YES	16.7	83.3	100.0

a. 원래의 집단 케이스 중 76.2%이(가) 올바로 분류되었습니다.

4. 분류 결과





5. 정준 판별함수

3-3. 로지스틱 회귀분석

1. 종속변수 인코딩, 분류표, 범주형 변수 코딩
2. 범주형 변수(**성별**, **학년**)를 제외한 분류표, 방정식의 변수
3. 모든 변수를 포함한 분류표, 방정식의 변수(사용) (**성별**과 **학년**에 따른 차이)
4. 방법을 후진선택법으로 사용한 경우

종속변수 인코딩

원래 값	내부 값
No	0
Yes	1

블록 0: 시작 블록

분류표[a,b]

관측됨			예측 결혼2점 No	예측 결혼2점 Yes	분류정확 %
0 단계	결혼2점	No	15	0	100.0
		Yes	6	0	.0
	전체 퍼센트				71.4

a. 모형에 상수항이 있습니다.
b. 절단값은 .500입니다.

범주형 변수 코딩

		빈도	모수 코딩 (1)
학년	3학년	3	1.000
	4학년	18	.000
성별	여학생	8	1.000
	남학생	13	.000

1. 종속변수 인코딩, 분류표, 범주형 변수 코딩

8 단계[a]	직업	1.868	1.201	2.420	1	.120	6.476
	성격	-6.216	3.770	2.718	1	.099	.002
	장래성	3.006	1.735	3.002	1	.083	20.200
	학년(1)	-22.436	20122.633	.000	1	.999	.000
	상수항	9.182	9.977	.847	1	.357	9717.499

a. 변수가 1: 학력, 외모, 재력, 직업, 성격, 장래성, 종교, 가정환경, 자녀계획, 성별, 학년 단계에 입력되었습니다.

2. 범주형 변수(성별, 학년)를 제외한 분류표, 방정식의 변수

분류표[a]

관측됨		예측 결혼2점 No	Yes	분류정확 %
1 단계	결혼2점 No	15	0	100.0
	Yes	1	5	83.3
전체 퍼센트				95.2

a. 절단값은 .500입니다.

방정식의 변수

		B	S.E.	Wald	자유도	유의확률	Exp(B)
1 단계[a]	학력	1.092	1.200	.827	1	.363	2.979
	외모	-1.265	1.641	.594	1	.441	.282
	재력	.456	1.122	.166	1	.684	1.578
	직업	1.852	1.325	1.954	1	.162	6.375
	성격	-4.235	4.139	1.047	1	.306	.014
	장래성	1.940	1.764	1.209	1	.272	6.956
	종교	.398	.527	.569	1	.451	1.488
	가정환경	.261	.844	.095	1	.758	1.298
	자녀계획	-1.275	1.425	.800	1	.371	.279
	상수항	5.575	11.568	.232	1	.630	263.831

a. 변수가 1: 학력, 외모, 재력, 직업, 성격, 장래성, 종교, 가정환경, 자녀계획 단계에 입력되었습니다.

분류표[a]

관측됨		예측 결혼2점 No	Yes	분류정확 %
1 단계	결혼2점 No	14	1	93.3
	Yes	1	5	83.3
전체 퍼센트				90.5

a. 절단값은 .500입니다.

방정식의 변수

		B	S.E.	Wald	자유도	유의확률	Exp(B)
1 단계[a]	학력	.443	1.552	.082	1	.775	1.558
	외모	-2.538	3.578	.503	1	.478	.079
	재력	1.294	1.954	.439	1	.508	3.648
	직업	1.084	1.389	.609	1	.435	2.956
	성격	-3.614	4.871	.550	1	.458	.027
	장래성	2.603	2.083	1.562	1	.211	13.506
	종교	-.446	1.082	.170	1	.680	.640
	가정환경	.839	1.152	.530	1	.467	2.313
	자녀계획	-1.277	2.065	.382	1	.536	.279
	성별(1)	-.633	2.577	.060	1	.806	.531
	학년(1)	-23.194	18642.051	.000	1	.999	.000
	상수항	8.168	13.506	.366	1	.545	3526.912

a. 변수가 1: 학력, 외모, 재력, 직업, 성격, 장래성, 종교, 가정환경, 자녀계획, 성별, 학년 단계에 입력되었습니다.

3. 모든 변수를 포함한 분류표, 방정식의 변수(사용) (성별과 학년에 따른 차이)

4. 방법을 후진선택법으로 사용한 경우

3. 본문

다변량 자료 전략
3-4. 요인분석 / K-mean 군집분석 / 다차원 척도법

<요인분석 / K-mean 군집분석 / 다차원 척도법-1>
요인분석을 가능 여부를 알기 위해 여러 검정 중 KMO와 Bartlett의 검정을 실행해본 결과, 측도가 0.591, 유의확률이 0.298로 요인분석을 실행하기에 부적절하다. 하지만 표본의 개수가 작아 나타난 것으로 가정하고 분석을 실행하겠다.

<요인분석 / K-mean 군집분석 / 다차원 척도법-2>
학력, 외모, 제력을 하나의 변수로 묶고 **첫인상 요인점수**라 이름 붙였다. 처음 남녀가 만날 때 이름과 나이를 제외하고 통상적으로 처음 물어보는 질문이 학력, 제력, 외모이기 때문에 첫인상 요인점수라는 이름을 달았다.

다음으로 직업, 성격, 장래성을 하나의 변수로 묶고 **개인역량 요인점수**로 종교, 가정환경을 **가치관 요인점수**로 자녀계획은 **자녀계획 요인점수**로 도출되지만 사실 기존의 자녀계획과 같은 값을 가지게 된다. 그래서 정확히 말하면 데이터의 요인은 4개이지만 새롭게 만들어진 요인은 3개이다.

<요인분석 / K-mean 군집분석 / 다차원 척도법-3,4,5>
모두 요인을 4개로 선택하게끔 만들어 주었다. 누적 75.456%로 충분한 분산값을 가진다.

<요인분석 / K-mean 군집분석 / 다차원 척도법-6>
각각의 요인점수와 평균점수가 비례하고 있음을 보여준다. 즉, 정보량이 겹치는 모습을 볼 수 있다.

단순하게 각 4개의 요인에 해당하는 점수의 평균을 내서 어떤 요인에 더 점수를 많이 주었는지를 비교해보았다. 그 결과 **첫인상 평균점수**를 높게 준 사람은 1명, **개인역량 평균점수** 10명, **가치관 평균점수** 4명, **자녀계획 평균점수** 6명으로 개인역량 평균점수가 높은 회원이 좋은 결혼상대임을 알 수 있었다.

다음으로 요인점수를 보았다. **첫인상 요인점수**를 높게 준 사람은 5명, **개인역량 요인점수** 3명, **가치관 요인점수** 7명, **자녀계획 요인점수** 6명으로 가치관 요인점수가 높은 회원이 좋은 결혼상대임을 알 수 있었다.

<요인분석 / K-mean 군집분석 / 다차원 척도법-7>

먼저 K-mean 군집개수는 3개로 지정했고 그래프를 보면 4번째, 10번째 응답자가 직업과 장래성에 낮은 점수를 부여했다. 성격과 합쳐 개인역량 요인점수로 묶은 파트이다. 종교에 있어서 19번째, 6번째 응답자가 높은 점수를 줬고 2번째 응답자가 가장 낮은 점수를 줬다. 이는 가정환경과 함께 가치관 요인점수로 묶은 파트다. 11번째, 3번째 응답자는 외모와 재력에 높은 점수를 줬다. 학력과 함께 첫인상 요인점수로 묶은 파트다. 5번째 응답자는 자녀계획에 5점을 줬다. 자녀계획 요인점수로 사용할 수 있다.

다변량 자료 전략

3-5. 판별분석

<판별분석-1>

표본의 개수가 너무 적어 단계선택법으로 시행할 시에 유의한 변수가 나오지 않아 모든 변수에 유의하다고 가정했다.

설문에 참여한 학생들은 학력에 제일 높은 점수를 줬고 다음으로 직업, 종교, 장래성에 대해 높은 점수를 준다면 결혼여부에 대해 yes라 분류되기 쉽다. 반대로 외모와 자녀계획 성격에 대해 점수가 높을수록 결혼여부에 대해 no라고 분류되기 쉬웠다. 다른 분석과 다르게 성격이 높은 점수일수록 yes라 하기 좋다고 생각하지 않은 것은 표본이 적어서 그렇게 분류되었다고 생각한다.

성별에 의한 분류

선택변수를 **여학생**으로 한 경우, 여학생은 직업에 가장 높은 점수를 줬고 동률로 성격과 장래성에 대해 점수가 높을수록 결혼여부에 yes로 분류되기 쉬워졌다. 반대로 자녀계획의 점수가 가장 낮았고 재력 학력 순으로 결혼여부에 대해 no로 분류되기 쉽다. 아마 자녀계획에 대해 부정적인 현대사회에 딱 맞는 결과가 아닌가 생각했다. 의외로 남성들이 걱정하는 외모는 결혼여부에 대한 생각에 큰 영향을 미치지 못했다.

선택변수를 **남학생**으로 한 경우, 남학생은 성격에 가장 높은 점수를 줬고 다음은 외모에 점수를 주었다 여성과 반대로 자녀계획은 결혼여부 yes에 분류되기 쉬운 요소였다. 장래성과 종교, 학력에 가장 낮은 점수를 주어 결혼여부 no라 할 요소가 되었는데 남성들의 경우 자신보다 똑똑한 이성에 매력을 느끼기보단 오히려 부담스러워 하는 것을 볼 수 있었다.

학년에 의한 분류

선택변수를 **3학년**으로 한 경우, 표본의 너무 적어 나오지 않았고 선택변수를 **4학년**으로 한 경우, 장래성과 학력에 높은 점수를 주면 결혼여부에 yes로 분류되기 쉬웠다. 반대로 외모를 보면 가장 낮은 점수로 결혼여부에 no라 분류되기 쉬웠다. 3학년에 대한 결과를 알지 못해 애매하지만 감히 말하자면 학년이 높아질수록 장래성과 학력에 높은 점수를 주어 고학력자인 이성에 끌려하는 것을 알 수 있었다.

<판별분석-2>

평균은 (Yes-No)/2 = (1.502-0.421)/2 = 0.3155이므로 0.3155 이상이면 결혼여부에 대해 yes로 분류하고 0.3155 이하면 no라고 분류했다.

<판별분석-3>

앞에서 구한 평균을 기준으로 실제 분류 결과는 다음과 같이 나왔다.

<판별분석-4>

예측한 것과 기존 분류를 비교해보면 76.2%로 잘 분류 되었다.

<판별분석-5>

정준 함수를 보면 결혼여부 2점 No는 -0.421을 중심으로 모여 있고 Yes는 1.052를 중심으로 잘 모여 있다는 것을 볼 수 있다.

다변량 자료 전략

3-6. 로지스틱 회귀분석

<로지스틱 회귀분석-1>

오즈비를 해석할 때 분자가 내부값이 0인 결혼여부 no이고 분모는 내부값이 1인 결혼여부 yes가 된다. 첫 시작 블록 분류표의 분류정확도는 71.4%에서 시작한다. 마찬가지로 학년과 성별에 대해서 3학년, 여학생이 분자로 가고 4학년, 남학생이 분모로 가서 오즈비를 구하게 된다.

<로지스틱 회귀분석-2>

범주형 변수를 제외하고 95.2%의 분류 정확도를 보여주었다.

<로지스틱 회귀분석-3>

설명력이 기존에 71.4%에서 범주형 변수(성별, 학년)를 제외한 모든 변수를 넣고 로지스틱 회귀분석을 시행하면 95.2%로 분류정확도가 올라간다. 설명력이 기존에 71.4%에서 모든 변수를 넣고 로지스틱 회귀분석을 시행하면 90.5%까지 분류정확도가 올라간다. 즉, 범주형 변수(성별, 학년)를 넣으면 95.2%에서 90.5%로 내려가지만 유의하게 비교할 수 있다.

성별에 관한 오즈비 해석

먼저 범주형 변수인 성별에 대해 오즈비를 분석해보면 남학생에 비해 여학생이 결혼의사에 대해 no라 할 확률은 0.531배가 된다. 즉, 여학생인 경우 결혼여부에 대해 no라 할 가능성은 낮다.

학년에 관한 오즈비 해석

두 번째 범주형 변수인 학년에 대해 오즈비 분석을 해보면 오즈비가 0이므로 분석하기 어렵다. 3학년 모두가 결혼여부에 대해 no로 대답했기에 오즈 하나가 0이 되어버렸다.

실험 : 4학년의 결혼여부 no 대 yes의 비율이 2대 1이므로 3학년도 이와 같이 2대 1 비율로 맞춰 3명의 설문조사 중 한 명을 랜덤 추출해 결혼 여부를 yes로 바꾸고 분석하였다. Exp(B) 값은 0.575가 나와 4학년에 비해 3학년이 결혼의사에 대해 no라 할 확률은 0.575배가 된다. 즉, 3학년의 경우 결혼여부에 대해 no라 할 가능성이 낮다.

종속형 변수에 관한 오즈비 해석

다음으로 오즈비가 가장 작은 성격에 대해 해석하자면, 성격에 대한 평가가 한 단계 높아질 때(ex) 긍정->매우 긍정) 결혼의사에 대해 No라 할 오즈(가능성)가 0.027배 낮다. 즉, 성격에 대해 한 단계 높은 평가일 때 결혼의사에 대해 No라 대답할 경향이 더 낮다는 것을 보여준다. 다시 말해 성격에 대해 높은 평가를 준 사람일수록 결혼의사에 대해 Yes로 생각하고 있다는 것을 알 수 있었다. (결혼 의사에 대해 Yes라 대답한 사람 중에는 성격에 높은 평가를 준 사람이 많다고 볼 수 있다. -> 오즈가 다르게 된다.) 따라서 성격에 대해 높은 평가를 준 사람들끼리 이어주는 방법을 생각해 볼 수 있다.

반대로 오즈비가 가장 큰 장래성에 대해 해석하자면, 장래성에 대한 평가가 한 단계 높아질 때(ex) 긍정->매우 긍정) 결혼의사에 대해 No라 할 오즈(가능성)가 약 13배 가량 높다. 다시 말해 장래성에 대해 높은 평가를 준 사람일수록 결혼의사에 대해 Yes로 생각하고 있다는 것을 알 수 있었다.

성격 다음으로 오즈비가 낮은 외모, 자녀계획, 성별, 종교 순으로 높은 점수를 준 사람일수록 결혼의사에 대해 긍정적으로 생각하고 있기 때문에 위와 같은 변수에 높은 점수를 준 사람끼리 연결해준다면 결혼할 확률이 높아지겠다. 즉, 가치관이 같은 사람끼리 연결해주는 방식이다.

위에서 언급한 변수를 제외한 나머지 변수인 학력, 직업, 재력, 가정환경에 높은 점수를 준 사람들은 한 단계 낮은 점수를 준 사람들에 비해 결혼의사에 대해 No라고 대답하는 경향이 크다.

3. 본문

최종 전략

판별 분석

No : 외모, 자녀계획, 성격

Yes : 학력, 직업, 종교, 장래성, 재력, 가정환경

ex) 성격이라는 지표에 높은 점수를 매길수록 결혼여부에 대해 no로 분류되기 쉽다.

로지스틱 분석 - 오즈 비

No : 장래성, 재력, 직업, 가정환경, 학력

Yes : 성격, 외모, 자녀계획, 종교

ex) 성격에 대한 점수가 높아지면 그만큼 결혼 여부에 대해 yes라 할 가능성이 커짐

위의 경우와 같이 데이터가 너무 작아 분석을 시도했지만 서로 반대의 결과가 나왔다. 그래서 가장 상식적으로 생각했을 때 나올 수 있는 분석(요인, 로지스틱 분석)을 기준으로 해석 및 해결방안을 제시하려 한다.

먼저 모두가 관심 있어 하고 보통 결혼의 중요요인 1순위로 뽑히는 **성격**에 관해 얘기보자.

판별분석 결과, 성격에 높은 점수가 매겨질수록 결혼여부에 관해서는 no로 분류되기 쉽다는 결론을 얻었다. 하지만 설문조사결과 다른 변수들과 달리 크게 점수에서 차이가 나지 않았다. 대부분의 사람들이 5점 만점을 줬기 때문에(평균 4.8점) <판별분석-1>에서 -0.386*X7(성격)은 거의 고정된 값이라고 봐도 무방했다.

따라서 결혼여부에 대한 yes와 no의 분류에 크게 유의미한 차가 없다고 생각했다. 즉, 결과해석에는 크게 염두 해두지 않았다.

이와 반대로 평균점수와 로지스틱 회귀분석에서는 유의미한 차이가 있고 결혼여부에 대해 yes가 되기에 좋은 요소임을 알 수 있었다.

위의 내용과는 별개로 보통 사람마다 성격이 다르고 그 유형이 다르다. 따라서 회원들에게 최근 유행하는 MBTI 성격유형검사를 제공한다. 이는 성격에 높은 점수를 주었지만 자신과 맞는 성격을 찾는 것이 힘들다고 생각하여 결혼여부가 No로 나왔을 가능성도 있기에 MBTI 성격유형검사 통해 같은 부류의 성격을 가진 사람을 소개시켜준다.

두 번째로 **종교**에 관해 얘기해보자.

종교는 다른 8개의 요인 중에서 가장 점수의 범위가 다양하게 나온 요인이었다. 더불어

모든 분석에서 종교에 대해 높은 점수를 부여하면 결혼여부에 대해 Yes로 기대되는 유일한 요인이기도 한다.

따라서 회원가입서에 종교가 있으면 기재하게 하고 중요도와 상대방 종교의 자유도에 대해 적게 한다. 만약 중요도가 높고 상대방의 종교유무 및 타종교에 대해 호의적이면 결혼상대를 빨리 찾아줄 수 있고 타종교에 대해 부정적이면 같은 종교를 가진 사람을 소개시켜준다.

세 번째로 **판별 분석에서 결혼 여부에 대해 No**라 대답한 요소들을 다시 한 번 살펴보자.
앞서 첫 번째로 말한 성격과 비슷하게 외모와 자녀계획은 다른 변수들과 달리 응답자 모두 3점 이상을 부여한 변수들이었다. 즉. -0.794*X4(외모), -0.338*X9(자녀계획)의 값은 평점에 따라 다르겠지만 거의 변하지 않는 값으로 봐도 무방했다. 성격과 마찬가지로 결혼여부에 대한 yes와 no의 분류에 크게 유의미한 차가 없다고 생각했다.
결국 판별 분석에서의 결혼여부에 대한 Yes와 No의 분류는 사실상 Yes의 요인인 학력, 직업, 종교, 장래성, 재력, 가정환경에 의해 결정된다는 것으로 생각했다.

네 번째로 **로지스틱 분석에서 결혼 여부에 대해 No**라 대답한 요소들을 다시 한 번 살펴보자.

물론 오즈비가 2배 이상 차이가 나는 요인들이지만 가장 큰 장래성(13.506)과 재력(3.648)을 제외하고 현실적으로 많은 양의 데이터가 쌓였다고 가정하고 오즈 비에서 많은 차이가 나지 않을 만한 학력(1.558), 직업(2.956), 가정환경(2.313)을 결혼여부에 대해 중요한 요소로 보고 결혼전략을 짜고자 한다.

마지막으로 **요인, K-mean 군집, 다차원 척도, 판별, 로지스틱 분석에서 결혼여부에 대해 Yes**라 생각한 요인들을 바탕으로 전략을 짜보자.
물론 소개받는 상대방의 성격이나 가치관도 중요하지만 현대 사회에서 이와는 달리 직업이나 학력 같은 요인에 대해 현실적으로 고민할 수밖에 없다고 생각한다. 따라서 결혼하려는 사람들은 상대방의 여러 조건을 따질 수밖에 없게 되었다.

하지만 실제로 똑똑하거나 돈 많은 사람들은 우리 사회에서 대다수를 차지하고 있지 않기 때문에 결혼상대를 찾다 지친 회원들이 결혼의사에 대해 No라고 대답했을 가능성도 크다고 생각했다.
즉, 결혼의사에 대해 No라고 대답한 사람들한테는 우선적으로 본인이 중요하다는 요인에 부합하는 사람을 소개시켜주는 것이 최고의 전략이 되겠지만 그런 사람은 소수이기 때문에 이렇게 해서는 획기적으로 결혼확률을 높이긴 어렵다고 판단했다.

이 문제를 해결하기 위해 회원 중 반대의 가치관을 가진 사람(이성이 부담스럽다면 먼저 동성)을 연결해 줘서 새로운 사고를 가지게 하는 프로그램을 생각해볼 수 있다. 보통 사람은 스스로의 생각만으론 가치관을 바꾸기 어렵다. 하지만 본인과 다른 가치관을 가진 사람을 직접 만나 어울리다 보면 생각의 전환이 수월해진다. 즉, 새로운 사람을 만나게 하여 본인의 가치관을 변화시켜 결혼할 확률을 높이는 방법을 택하겠다.

이를 다시 결혼 여부에 관한 5점 척도로 보면 보통이라고 응답한 사람들의 마음을 긍정 혹은 매우 긍정으로 만들어 결혼시키기 위해서 첫 번째로 본인들이 중요하다고 생각한 변수를 만족하는 사람들을 소개시켜주고 안된다면 처음부터 결혼 여부에 긍정 혹은 매우 긍정으로 대답한 사람들이 중요하게 생각한 변수에 대해 높은 점수를 가진 사람을 이어줌으로써 결혼확률을 높이려 하겠다.

4. 별첨

4-1. 의사결정나무

의사결정나무는 시행했지만 제대로 된 결과가 나오기 힘들어 예외 장으로 넣었다.

결혼2점	Predicted Value	Predicted Probability_1	Predicted Probability_2
No	No	.71	.29
No	No	.71	.29
No	No	.71	.29
No	No	.71	.29
Yes	No	.71	.29
No	No	.71	.29
No	No	.71	.29
No	No	.71	.29
Yes	No	.71	.29
No	No	.71	.29
No	No	.71	.29
No	No	.71	.29
No	No	.71	.29
Yes	No	.71	.29
No	No	.71	.29
No	No	.71	.29
Yes	No	.71	.29
Yes	No	.71	.29
Yes	No	.71	.29
No	No	.71	.29
No	No	.71	.29

위험도

추정값	표준화 오류
.286	.099

성장방법: CRT
종속변수: 결혼2점

분류

관측	예측 No	예측 Yes	정확도 퍼센트
No	15	0	100.0%
Yes	6	0	0.0%
전체 퍼센트	100.0%	0.0%	71.4%

성장방법: CRT
종속변수: 결혼2점

CRT로 만들고 해석해보면 노드가 만들어지지 않았다. 그리고 모두 No로 분류되도록 예측되었다. 71%의 확률로 No라 예측했고 29%의 확률로 Yes라 예측하여 분류한 것을 볼 수 있었다. 마지막 그래프에서 분류 정확도는 71.4%로 나왔다.

4-2. 요인분석, K-mean 군집분석, 다차원 척도법의 정보가 결합된 그래프