

다변량통계분석방법론 기말 레포트
-스피드 데이팅 분석-

2018150408 이충은

2018150453 곽동호

1. 서론

스피드 데이팅 데이터¹는 kaggle에 공개되어 있는 데이터로 컬럼비아 비즈니스 스쿨의 교수인 Ray Fisman과 Sheena Iyengar가 그들의 논문인 " Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment"를 위해 작성되었습니다. 2002년부터 2004년 사이에 진행된 실험적 스피드 데이트 이벤트로부터 수집되었습니다. 이벤트 동안 참석자들은 다른 성별의 참가자들과 4분간의 짧은 데이팅을 하는 것이 실험의 주요 요소였습니다. 데이터의 변수들은 각 참가자들의 개인적인 정보와 실험 전후와 실험 중간 설문조사의 결과를 포함하고 있습니다. 저희는 이 데이터를 활용하여 다음과 같은 두 가지 분석을 실시하고자 합니다.

첫 번째로 실시하는 분석은 다변량 회귀 분석을 통해 상대방 성별의 데이트내에서의 결정과 양측의 호감 표시로 인한 매칭 성사에 영향을 미치는 변수들을 찾는 것입니다. 설문조사를 통해 수집된 다양한 정보들을 통해 어떠한 변수들이 얼마나 데이팅의 결과에 영향을 미치는 지를 알아보하고자 하였습니다.

두 번째 분석은 다변량 분산 분석(MANOVA)을 사용해 남성과 여성 간의 각 항목별 자기 객관화 정도의 차이를 분석하고 위의 첫번째 분석과 같이 이 차이를 나타내는 변수들이 상대방 성별의 데이트내에서의 결정과 매칭 성사에 어떤 영향을 미치는 지를 알아보하고자 합니다. 이 분석을 통해 성별에 따른 자기 인식의 차이가 데이팅에서의 선택에 어떻게 작용하는 지를 탐구하고자 합니다.

이 보고서는 스피드 데이팅의 상황에서 발생하는 상호간의 선택과 이러한 선택의 일치에 영향을 미치는 변수들이 무엇인지 알아보고, 이를 통해 우리 사회에서 성별 역할, 매력 인식, 그리고 대인 관계의 형성에 대한 더 깊은 이해에 도움이 되고자 합니다. 이러한 연구는 단순히 데이팅 상황뿐만 아니라, 이를 확장하여 일반적인 대인 관계 및 사회적 상호작용에 대한 이해를 넓히는데 기여할 수 있을 것이라 생각합니다.

2. 데이터 시각화

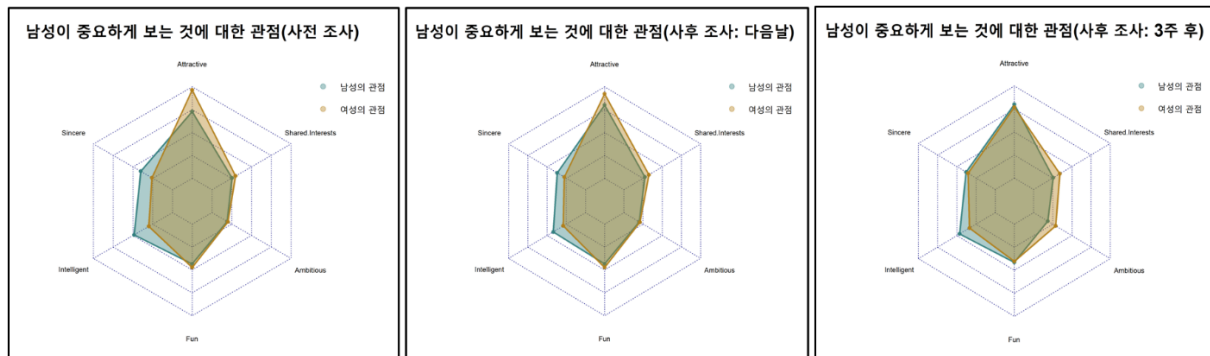
본격적으로 서론에서 설명한 분석을 하기에 앞서 데이터의 주요 변수에 대하여 시각화를 진행하여 분석의 목적과 방향을 확인하였습니다.

이성 선택간 주요 요소 및 이성의 예측

¹ <https://www.kaggle.com/datasets/annavictoria/speed-dating-experiment>

첫번째로 데이팅 전과 데이팅 다음날, 데이팅 3주 후 이렇게 3번의 설문조사에서 각 성별이 이성을 볼 때 중요하게 생각하는 것과 이성의 생각에 대한 예측 설문 결과를 시각화하여 실제 무엇을 중요하게 생각하는지와 그것에 대한 이성의 생각을 알아보았고 후에 분석 결과와 비교해볼 수 있었습니다.

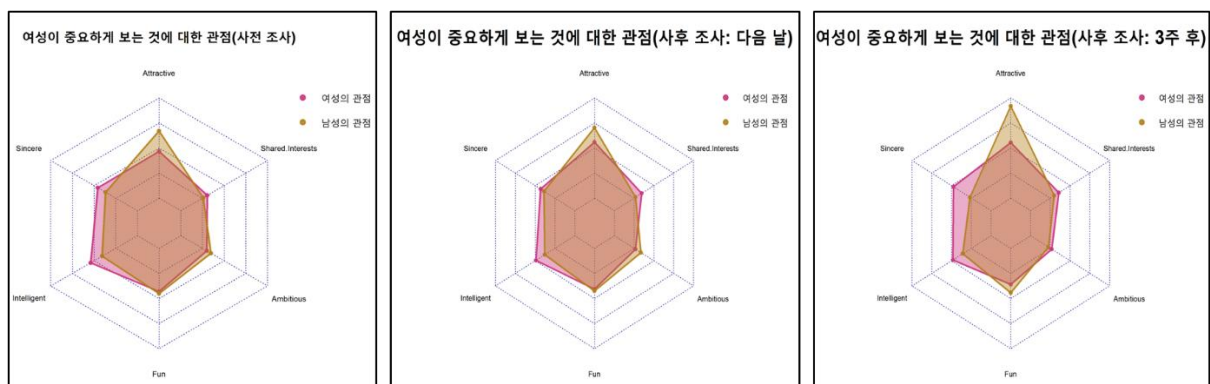
우선 남성의 이성 선택간 주요 고려 요소와 이를 여성이 예측한 것을 시각화 하여 살펴보았습니다.



< Figure 1:설문조사의 타임라인 별 남성의 이성 선택 주요 고려요소 >

시각화 결과 남성이 중요하다 생각하는 것에는 데이팅 진행 전과 후에 뚜렷한 변화가 보이지 않음을 알 수 있었습니다. 특히 외모에 대해 가장 중요시하고 그 다음 유머와 공통 관심사가 뒤따랐습니다. 여성이 바라본 관점의 경우 데이팅 전에는 남성들의 주요 고려 요소가 외모에 실제보다 더 치우쳐져 있을 거라 생각하나 데이팅 후 시간이 지날 수록 점차 남성들의 실제 생각과 비슷하게 가는 모습을 보였습니다.

다음으로는 여성의 경우에 대해 아래와 같이 시각화를 해보았습니다.

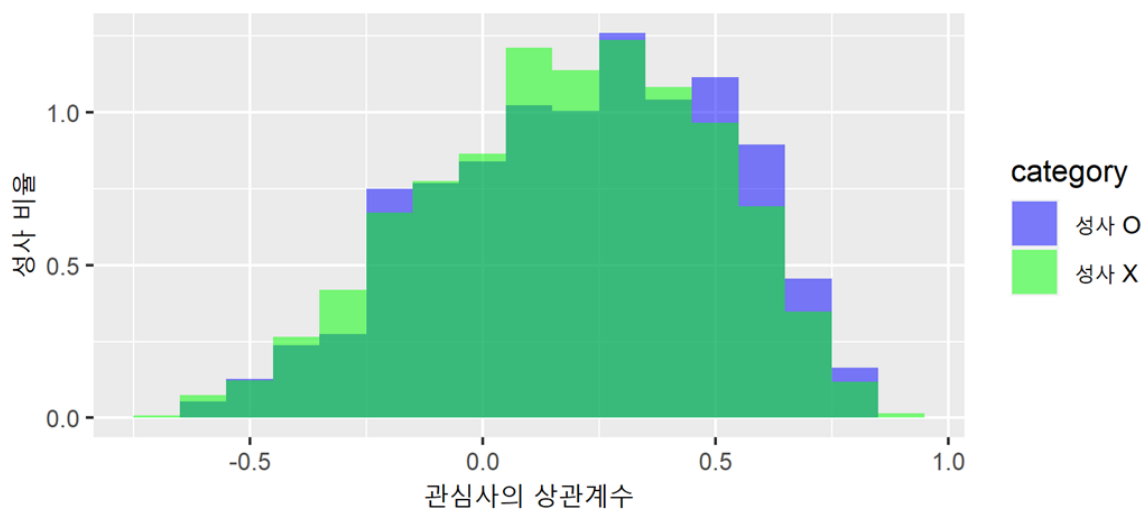


< Figure 2:설문조사의 타임라인 별 여성의 이성 선택 주요 고려요소 >

시각화하여 살펴본 결과 여성의 경우에도 데이팅 진행 전후로 이성 선택에 있어 중요하게 생각하는 것이 변하지 않는다는 것을 확인할 수 있었습니다. 그러나 남성의 경우 여성과 다르게 데이팅 진행 후 시간이 지나면서 여성의 주요 고려에 가까워지는 것이 아닌 더 멀어지는 모습을 보였고 위의 시각화 분석과 같이 뒤의 데이터 분석을 통해 이들과 실제 데이트에서의 영향을 미치는 요소를 비교할 수 있었습니다.

관심사의 유사 정도

다음으로는 흔히 사람들이 생각하는 관심사 및 취미가 겹치는 사람에게 호감을 느낀다는 것이 실제로 그러한 지 살펴보았습니다. 매칭이 성사된 사람들과 아닌 사람들을 나누어 둘의 관심사의 상관 계수와 그에 따른 성사 비율을 시각화 하였습니다.



< Figure 3:관심사의 유사 정도와 성사 여부 >

결과적으로 평소 당연하게 생각하던 것과 달리 성사된 경우 성사되지 않은 경우의 관심사의 상관 계수 차이는 눈으로 보기에 큰 차이가 없는 것으로 나왔습니다. 이렇게 시각화한 결과를 토대로 실제로도 분석에서도 그러한 지 직접 데이터를 분석하여 알아보았습니다.

3. 데이터 전처리

변수 제거

앞서 언급하였듯이 데이터는 데이팅 참가자들에게 여러 번에 걸쳐 설문조사한 항목을 그대로 옮겨 놓은 것이기 때문에 모든 변수들을 분석에 사용하지 않았습니다. 따라서 여러 변수들을 제거하고 데이터 분석을 실시하였습니다. 아래 표는 제거한 변수명과 해당 변수를 제거한 이유에 대하여 정리한 표입니다.

제거 변수명	변수 제거 이유	제거 변수명	변수 제거 이유
id	iid 변수에 정보 포함	4_1 설문 결과들	앞선 설문의 답변과 중복되는 변수
idg		5_1 설문 결과들	
round		1_s 설문 결과들	
pid		3_s 설문 결과들	
partner		7_2 설문 결과들	
condtn	변수의 정보가 불명확	race_o	개인별 정보에 포함
position		pf_o_sin	
poistin1		pf_o_int	
field	encoding된 변수 존재 (field_cd)	pf_o_fun	
undergrad	분석에 필요하지 않은 변수	pf_o_amb	
mn_sat		pf_o_sha	
tuition		dec_o	
from		attr_o	
zipcode		sinc_o	
income		intel_o	
career		fun_o	
career_c		amb_o	
exhappy		shar_o	
met		like_o	
numdat_3		prob_o	
num_in_3		met_o	

< Table 1:제거 변수명과 제거 이유 >

범주형 변수 처리

그 다음으로는 범주형 변수들을 R 프로그램 내에서 'character'가 아닌 'factor'변수로 바꾸어 분석에 용이하도록 바꾸어 주었습니다. 바꾼 변수들을 정리한 표는 다음과 같습니다:

변수명	Gender	Samerace	Race	dec	date
변수 설명	성별	동일 인종 여부	인종	결정	데이트 여부

< Table 2:처리된 범주형 변수들 >

데이터 표준화

그 다음 데이터 표준화 작업을 진행하였습니다. 기존 데이터가 장기간에 걸친 여러 번의 데이팅의 데이터임에 따라 설문조사의 점수 산정 방법이 데이팅 회차에 따라 달라지는 모습을 보이는 항목들이 존재하였습니다. 따라서 데이터 분석을 위해 이렇게 같은 내용의 질문에 대한 답이지만 설문의 측정 단위가 회차에 따라 달라지는 변수들을 표준화해주는 작업을 진행하였습니다.

새로운 변수 생성

마지막으로 차후 진행될 분석을 위해 새로운 변수들을 만들었습니다. 기존의 데이터는 본인과 파트너와의 조합에 따라 각각 항목이 존재하여 한 개인에 대한 점수를 살펴보기 어려웠고 중복된 데이터도 많았습니다. 이에 따라 개인에 대한 분석을 진행하기 위해 파트너에 따라 별개로 수집된 데이터들을 개인을 기준으로 통합시켜 개인별 점수를 보여주는 변수들을 새롭게 만들어 분석에 사용하였습니다.

이러한 과정들을 거쳐 전처리된 데이터들을 하나의 데이터셋으로 만들어주었고 이를 바탕으로 알고자 하는 내용을 위한 데이터 분석을 실시하였습니다.

4. 방법론

저희는 크게 두 가지의 방법을 사용하여 데이터를 분석했습니다.

- 1) 다변량회귀분석을 통한 매칭 여부와 상대 성별의 결정에 영향을 미치는 변수 찾기
- 2) MANOVA를 사용해 각 성별 간 자기 객관화 정도의 차이 분석

4.1. 다변량회귀분석을 통한 영향변수 식별

개요

저희의 주요 관심사는 참가자들이 파트너에게 선택받은 비율(per_par_dec)과 매칭된 비율(per_match)입니다. 이 두 변수를 다변량 회귀분석의 종속 변수로 설정하였습니다. 첫 번째 설문과 두번째, 세 번째 설문에 응답한 사람의 수에 큰 차이가 존재하기 때문에, 가장 많은 데이터가 수집된 첫 번째 설문조사와 데이팅 중간에 수집된 파트너의 평가 등의 변수들을 독립 변수로 선정하여 포함시켰습니다. 이 독립 변수들은 참가자의 개인적 특성, 선호도, 사회적 활동 등 다양한 측면을 포함합니다. 저희는 이 독립 변수들에 대해 다변량 회귀 모델을 적합시켜 변수를 선택하고자 했습니다.

다변량 회귀분석 및 모델 구축

다변량 회귀 모델을 적합하기 위해 R의 lm() 함수를 사용했고, 종속변수를 cbind()함수를 사용했습니다. 이 모델은 per_par_dec와 per_match를 종속 변수로 하며, 위의 독립변수들 간의 관계를 추정합니다. 모델 구축의 첫 단계로 모든 관련 변수를 포함하는 포괄적인 모델을 제작했습니다.

MANOVA와 변수 선택

포괄적인 모델을 적합시킨 후, 모델의 유효성을 평가하고 중요한 변수만을 선택하기 위해 다변량 분산분석(MANOVA)을 실시합니다. 이를 위해 R의 `Anova()` 함수를 사용하여 모델에 대한 MANOVA를 실행합니다. 여기서 $tr[H(H + E)^{-1}]$ 로 정의되는 Pillai Statistic을 사용해 유의성을 검정합니다. 이 분석을 통해 각 독립 변수가 종속 변수에 미치는 영향의 유의성을 평가하고, 유의하지 않은 변수들을 제거합니다.

모델 단순화와 비교

이후 유의한 변수들만을 포함한 새로운 간소화된 모델을 만들고, 이를 원래의 포괄적인 모델과 비교합니다. 이를 위해 R의 `update()` 함수를 사용하여 변수가 제거된 새로운 모델을 생성하고, `anova()` 함수를 사용하여 두 모델 간의 유의한 차이를 검정합니다. 유의한 차이가 없다면, 더 간단한 모델을 최종 모델로 채택합니다.

4.2. MANOVA 활용 자기 객관화의 차이 분석

개요

이 분석에서는 스피드 데이팅 이벤트 참가자들의 자기 객관화 정도가 이성의 결정 및 매칭에 어떻게 영향을 미치는지를 다룹니다. 여기서 자기 객관화는 참가자가 스스로 평가한 점수와 실제 데이팅에서 파트너들이 평가한 점수의 평균의 차이로 정의합니다. 성별에 따른 평균적인 자기 객관화 정도의 차이를 비교하기 위해 다변량 분산분석(MANOVA)을 활용합니다.

데이터 전처리 및 변수화 변수 선택

참가자의 자기 평가 점수와 파트너들에 의한 평가 점수의 평균을 선택합니다. 두 점수 간의 차이를 차이(gap)으로 정의하고, 데이터를 성별에 따라 분할하여 남성과 여성 데이터셋을 구분합니다.

정규성 검정 및 데이터 변환

Shapiro-Wilk 검정을 사용하여 `rating_difference`의 정규성을 검정합니다. 여기서 정규성을 만족하지 않는 경우, Yeo-Johnson 변환을 적용하여 데이터를 정규화합니다. Yeo-johnson 변환은 다음과 같습니다:

$$y_i^{(\lambda)} = \begin{cases} ((y_i + 1)^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, y \geq 0 \\ \log(y_i + 1) & \text{if } \lambda = 0, y \geq 0 \\ -\frac{(-y_i + 1)^{2-\lambda} - 1}{2 - \lambda} & \text{if } \lambda \neq 2, y < 0 \\ -\log(y_i + 1) & \text{if } \lambda = 2, y < 0 \end{cases}$$

<Figure 4: Yeo-johnson Transformation>

다변량 분산분석(MANOVA)

lm() 함수를 사용하여 성별(gender)을 독립 변수로, 차이(gap)를 종속 변수로 하는 모델을 설정합니다. 여기에 manova() 함수를 사용하여 성별에 따른 자기 객관화 정도의 차이에 대한 다변량 분산분석을 수행합니다.

유의성 검정 및 변수 추출

MANOVA 결과에 대해 Bonferroni Interval을 계산하여 유의한 변수를 식별하고, 결과를 해석합니다

5. 분석 결과 및 해석

각 항목에 대한 실제 분석을 진행한 결과는 다음과 같습니다.

5.1. 다변량 회귀분석을 통한 영향변수 식별

저희가 회귀분석에 설명변수로 적합 시킨 변수들은 다음과 같습니다:

변수 명	변수 설명	변수 명	변수 설명
attr1_1	각 요소 별 본인이 중요하게 생각하는 정도	dining	항목 별 관심의 정도
sinc1_1		museums	
intel1_1		art	
fun1_1		hiking	
amb1_1		gaming	
attr2_1	각 요소 별 이성이 중요하게 생각할 것 같은 정도	clubbing	
sinc2_1		reading	
intel2_1		tv	
fun2_1		theater	
field_cd	Encoding된 전공	movies	
attr3_1	각 요소 별 본인이 평가한 스스로의 점수	concerts	
sinc3_1		music	
intel3_1		shopping	
fun3_1		yoga	
amb3_1		match_es	예상한 매칭 횟수
imprace	인종/윤리관의 중요도	mean_int_corr	둘의 관심사의 상관계수
date	평소 데이트의 빈도	mean_attr_o	상대가 평가한 외모점수 평균
go_out	평소 외출 빈도	mean_sinc_o	상대가 평가한 진실성 평균
gender	성별	mean_intel_o	상대가 평가한 지성 평균
age	나이	mean_fun_o	상대가 평가한 유머 평균

goal	데이팅의 목표	mean_amb_o	상대가 평가한 야망 평균
sports	항목 별 관심의 정도	mean_shar_o	상대가 평가한 공통 관심사 평균
tvsports		mean_like	상대가 좋다고 한 정도의 평균
exercise		mean_prob	상대의 선택 받을 확률 예상 평균

< Table 3:설명변수로 사용된 변수들 >

해당 변수들에 대한 적합 결과는 너무 많은 변수를 포함하고 있기 때문에, MANOVA를 이용해 두 종속 변수 모두에 유의한 영향을 미치는 변수를 식별했습니다. MANOVA의 결과는 다음과 같습니다:

Type II MANOVA Tests: Pillai test statistic

	Df	test stat	approx F	num Df	den Df	Pr(>F)
attr1_1	1	0.01558	1.187	2	150	0.30795
sinc1_1	1	0.00661	0.499	2	150	0.60818
intel1_1	1	0.02107	1.614	2	150	0.20247
fun1_1	1	0.02188	1.678	2	150	0.19031
amb1_1	1	0.00883	0.668	2	150	0.51424
attr2_1	1	0.03916	3.057	2	150	0.04997 *
sinc2_1	1	0.01029	0.780	2	150	0.46026
intel2_1	1	0.03617	2.815	2	150	0.06308 .
fun2_1	1	0.00027	0.020	2	150	0.98017
amb2_1	1	0.00804	0.608	2	150	0.54593
field_cd	1	0.02563	1.973	2	150	0.14264
attr3_1	1	0.00142	0.107	2	150	0.89875
sinc3_1	1	0.01507	1.147	2	150	0.32029
intel3_1	1	0.03561	2.769	2	150	0.06593 .
fun3_1	1	0.03956	3.089	2	150	0.04845 *
amb3_1	1	0.04016	3.138	2	150	0.04622 *
imprace	1	0.00572	0.431	2	150	0.65038
date	1	0.00152	0.114	2	150	0.89247
go_out	1	0.00122	0.092	2	150	0.91258
age	1	0.02761	2.129	2	150	0.12248
goal	1	0.00225	0.169	2	150	0.84461
sports	1	0.01960	1.499	2	150	0.22657
tvsports	1	0.02230	1.711	2	150	0.18424
exercise	1	0.03403	2.643	2	150	0.07449 .

dining	1	0.01978	1.513	2	150	0.22356
museums	1	0.02020	1.546	2	150	0.21643
art	1	0.02027	1.552	2	150	0.21530
hiking	1	0.00550	0.415	2	150	0.66103
gaming	1	0.00380	0.286	2	150	0.75172
clubbing	1	0.01753	1.338	2	150	0.26537
reading	1	0.01832	1.400	2	150	0.24985
tv	1	0.01381	1.050	2	150	0.35235
theater	1	0.02831	2.185	2	150	0.11599
movies	1	0.01657	1.264	2	150	0.28558
concerts	1	0.00012	0.009	2	150	0.99073
music	1	0.00729	0.551	2	150	0.57755
shopping	1	0.01506	1.147	2	150	0.32039
yoga	1	0.01396	1.062	2	150	0.34843
match_es	1	0.14802	13.030	2	150	6.055e-06 ***
mean_int_corr	1	0.00022	0.017	2	150	0.98344
mean_attr_o	1	0.32175	35.579	2	150	2.260e-13 ***
mean_sinc_o	1	0.02109	1.616	2	150	0.20218
mean_intel_o	1	0.02282	1.752	2	150	0.17700
mean_fun_o	1	0.03379	2.623	2	150	0.07594 .
mean_amb_o	1	0.01689	1.288	2	150	0.27876
mean_shar_o	1	0.01067	0.809	2	150	0.44731
mean_like	1	0.04124	3.226	2	150	0.04250 *
mean_prob	1	0.00908	0.687	2	150	0.50462

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

<Table 4:여성 데이터셋에 대한 MANOVA 검정 결과>

Type II MANOVA Tests: Pillai test statistic

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
attr1_2	1	0.010491	0.4347	2	82	0.648959				
sinc1_2	1	0.010453	0.4331	2	82	0.649970				
intel1_2	1	0.012059	0.5005	2	82	0.608082				
fun1_2	1	0.003394	0.1396	2	82	0.869898				
amb1_2	1	0.010227	0.4236	2	82	0.656095				
attr2_2	1	0.013481	0.5603	2	82	0.573214				
sinc2_2	1	0.008872	0.3670	2	82	0.693950				

intel2_2	1	0.042748	1.8310	2	82	0.166751
fun2_2	1	0.030647	1.2963	2	82	0.279098
amb2_2	1	0.019092	0.7980	2	82	0.453683
field_cd	1	0.003712	0.1528	2	82	0.858569
attr3_2	1	0.050348	2.1737	2	82	0.120267
sinc3_2	1	0.011330	0.4699	2	82	0.626756
intel3_2	1	0.012328	0.5118	2	82	0.601336
fun3_2	1	0.002096	0.0861	2	82	0.917572
amb3_2	1	0.045288	1.9449	2	82	0.149542
imprace	1	0.007884	0.3258	2	82	0.722875
date	1	0.000840	0.0345	2	82	0.966135
go_out	1	0.007672	0.3170	2	82	0.729228
age	1	0.009380	0.3882	2	82	0.679507
goal	1	0.008561	0.3540	2	82	0.702920
sports	1	0.006405	0.2643	2	82	0.768410
tvsports	1	0.009943	0.4118	2	82	0.663834
exercise	1	0.001676	0.0689	2	82	0.933520
dining	1	0.003383	0.1392	2	82	0.870295
museums	1	0.001741	0.0715	2	82	0.931032
art	1	0.020100	0.8410	2	82	0.434961
hiking	1	0.038419	1.6381	2	82	0.200637
gaming	1	0.090901	4.0996	2	82	0.020093 *
clubbing	1	0.026709	1.1251	2	82	0.329579
reading	1	0.020009	0.8371	2	82	0.436618
tv	1	0.027963	1.1795	2	82	0.312598
theater	1	0.039928	1.7051	2	82	0.188133
movies	1	0.084990	3.8082	2	82	0.026210 *
concerts	1	0.003728	0.1534	2	82	0.858015
music	1	0.004430	0.1824	2	82	0.833582
shopping	1	0.040992	1.7525	2	82	0.179765
yoga	1	0.020335	0.8510	2	82	0.430712
match_es	1	0.154165	7.4728	2	82	0.001044 **
mean_int_corr	1	0.042418	1.8162	2	82	0.169129
mean_attr_o	1	0.224898	11.8963	2	82	2.909e-05 ***
mean_sinc_o	1	0.032077	1.3588	2	82	0.262705
mean_intel_o	1	0.038868	1.6580	2	82	0.196835
mean_fun_o	1	0.005645	0.2328	2	82	0.792853
mean_amb_o	1	0.031189	1.3199	2	82	0.272773

mean_shar_o	1	0.064826	2.8421	2	82	0.064060 .
mean_like	1	0.054006	2.3407	2	82	0.102663
mean_prob	1	0.019538	0.8170	2	82	0.445302

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

<Table 5:남성 데이터셋에 대한 MANOVA 검정 결과>

여성의 경우 attr2_1, intel2_1, intel3_1, fun3_1, amb3_1, exercise, match_es, mean_attr_o, mean_fun_o, mean_prob의 변수가 유의하게 나타났고, 남성의 경우 gaming, movies, match_es, mean_attr_o, mean_shar_o 변수가 유의하게 나타났습니다. 이후 유의한 변수들을 설명변수로 사용해 다변량 회귀모형을 각각 새로 적합하고, anova()함수를 이용해 전체 모델과 새로 적합한 reduced model을 서로 비교했습니다. 그 결과는 다음과 같습니다:

Analysis of Variance Table

Model 1: cbind(per_par_dec, per_match) ~ attr1_1 + sinc1_1 + intel1_1 + fun1_1 + amb1_1 + attr2_1 + sinc2_1 + intel2_1 + fun2_1 + amb2_1 + field_cd + attr3_1 + sinc3_1 + intel3_1 + fun3_1 + amb3_1 + imprace + date + go_out + age + goal + sports + tvsports + exercise + dining + museums + art + hiking + gaming + clubbing + reading + tv + theater + movies + concerts + music + shopping + yoga + match_es + mean_int_corr + mean_attr_o + mean_sinc_o + mean_intel_o + mean_fun_o + mean_amb_o + mean_shar_o + mean_like + mean_prob

Model 2: cbind(per_par_dec, per_match) ~ intel2_1 + attr2_1 + intel3_1 + fun3_1 + amb3_1 + exercise + match_es + mean_attr_o + mean_fun_o + mean_prob

Res.Df	Df	Gen.var.	Pillai approx	F	num Df	den Df	Pr(>F)
--------	----	----------	---------------	---	--------	--------	--------

1	151	154.45					
---	-----	--------	--	--	--	--	--

2	189	38	153.15	0.38688	0.95301	76	302	0.5899
---	-----	----	--------	---------	---------	----	-----	--------

<Table 6: 남성에 대한 Reduced model과 Full model의 Anova table>

Analysis of Variance Table

Model 1: cbind(per_par_dec, per_match) ~ attr1_1 + sinc1_1 + intel1_1 + fun1_1 + amb1_1 + attr2_1 + sinc2_1 + intel2_1 + fun2_1 +

```

amb2_1 + field_cd + attr3_1 + sinc3_1 + intel3_1 + fun3_1 +
amb3_1 + imprace + date + go_out + age + goal + sports +
tvsports + exercise + dining + museums + art + hiking + gaming +
clubbing + reading + tv + theater + movies + concerts + music +
shopping + yoga + match_es + mean_int_corr + mean_attr_o +
mean_sinc_o + mean_intel_o + mean_fun_o + mean_amb_o + mean_shar_o +
mean_like + mean_prob
Model 2: cbind(per_par_dec, per_match) ~ gaming + movies + match_es +
mean_attr_o + mean_shar_o

```

	Res.Df	Df	Gen.var.	Pillai approx	F	num Df	den Df	Pr(>F)
1			152		127.59			
2	195	43	132.99	0.50207	1.1848	86	304	0.152

<Table 7: 여성에 대한 Reduced model과 Full model의 Anova table>

Anova 검정 결과 남성의 경우 p-value 0.5899, 여성의 경우 p-value 0.152로 유의수준 5% 아래 귀무가설을 기각하지 못하고, Full model과 reduced 간에 유의한 차이가 없음을 결론내리고 reduced model을 활용해 변수를 선택하고 간단한 회귀모형을 적합할 수 있습니다. 간소화된 모델 만으로도 매칭률과 파트너 선택률에 대한 설명력이 전체 모델과 비교해 차이가 없다고 해석할 수 있습니다. 따라서 일련의 과정을 통해 최종 적합된 회귀 모형은 다음과 같습니다:

Response per_par_dec :

Call:

```

lm(formula = per_par_dec ~ intel2_1 + attr2_1 + intel3_1 + fun3_1 +
amb3_1 + exercise + match_es + mean_attr_o + mean_fun_o +
mean_prob, data = reg_data_w)

```

Residuals:

Min	1Q	Median	3Q	Max
-39.826	-8.852	-0.882	8.142	48.233

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-58.92386	12.75772	-4.619	7.13e-06 ***
intel2_1	-0.11757	0.24821	-0.474	0.63629
attr2_1	-0.03223	0.07814	-0.412	0.68047
intel3_1	-0.89277	1.00566	-0.888	0.37581

fun3_1	0.06791	0.77727	0.087	0.93047
amb3_1	-0.46198	0.63515	-0.727	0.46791
exercise	-0.26285	0.41467	-0.634	0.52693
match_es	-0.05755	0.47014	-0.122	0.90270
mean_attr_o	14.49313	1.35030	10.733	< 2e-16 ***
mean_fun_o	4.81226	1.61218	2.985	0.00321 **
mean_prob	-0.96752	0.79690	-1.214	0.22622

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.16 on 189 degrees of freedom
 Multiple R-squared: 0.6434, Adjusted R-squared: 0.6245
 F-statistic: 34.1 on 10 and 189 DF, p-value: < 2.2e-16

<Table 8: 여성에 대한 reduced per_par_dec 회귀 모형>

Response per_match :

Call:

lm(formula = per_match ~ intel2_1 + attr2_1 + intel3_1 + fun3_1 +
 amb3_1 + exercise + match_es + mean_attr_o + mean_fun_o +
 mean_prob, data = reg_data_w)

Residuals:

Min	1Q	Median	3Q	Max
-23.292	-7.056	-0.545	5.990	51.789

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-39.73690	10.10817	-3.931	0.000119 ***
intel2_1	0.39941	0.19666	2.031	0.043664 *
attr2_1	0.12739	0.06191	2.058	0.040994 *
intel3_1	1.20751	0.79681	1.515	0.131331
fun3_1	-0.86450	0.61585	-1.404	0.162029
amb3_1	-0.76726	0.50324	-1.525	0.129020
exercise	-0.86959	0.32855	-2.647	0.008813 **
match_es	2.19859	0.37250	5.902	1.63e-08 ***
mean_attr_o	3.61855	1.06987	3.382	0.000874 ***

```

mean_fun_o    3.35919    1.27736    2.630 0.009248 **
mean_prob     0.55591     0.63140    0.880 0.379739

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.22 on 189 degrees of freedom
Multiple R-squared: 0.3972, Adjusted R-squared: 0.3653
F-statistic: 12.45 on 10 and 189 DF, p-value: < 2.2e-16

<Table 9: 여성에 대한 reduced per_match 회귀 모형>

Response per_par_dec :

Call:

```
lm(formula = per_par_dec ~ gaming + movies + match_es + mean_attr_o +
    mean_shar_o, data = reg_data_m)
```

Residuals:

Min	1Q	Median	3Q	Max
-36.182	-8.006	-0.262	6.872	42.139

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-64.08519	6.94017	-9.234	< 2e-16 ***
gaming	0.17623	0.36487	0.483	0.630
movies	0.22212	0.53430	0.416	0.678
match_es	-0.06563	0.43806	-0.150	0.881
mean_attr_o	10.17424	1.04849	9.704	< 2e-16 ***
mean_shar_o	7.47469	1.20963	6.179	3.68e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.28 on 195 degrees of freedom
Multiple R-squared: 0.669, Adjusted R-squared: 0.6605
F-statistic: 78.84 on 5 and 195 DF, p-value: < 2.2e-16

<Table 10: 남성에 대한 reduced per_par_dec 회귀 모형>

Response per_match :

Call:

```
lm(formula = per_match ~ gaming + movies + match_es + mean_attr_o +  
    mean_shar_o, data = reg_data_m)
```

Residuals:

Min	1Q	Median	3Q	Max
-24.505	-6.916	-1.628	6.340	46.728

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-21.0918	6.0120	-3.508	0.000560	***
gaming	0.2977	0.3161	0.942	0.347469	
movies	-0.4004	0.4628	-0.865	0.388054	
match_es	2.0930	0.3795	5.516	1.1e-07	***
mean_attr_o	3.2140	0.9083	3.539	0.000503	***
mean_shar_o	2.4643	1.0479	2.352	0.019685	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.5 on 195 degrees of freedom

Multiple R-squared: 0.3226, Adjusted R-squared: 0.3052

F-statistic: 18.57 on 5 and 195 DF, p-value: 4.51e-15

<Table 11: 남성에 대한 reduced per_match 회귀 모형>

여성에 대한 파트너의 선택 비율을 종속 변수로 한 회귀 모형에서 유의한 변수는 파트너가 평가한 평균 외모 점수(mean_attr_o)와 파트너가 평가한 평균 재미 점수(mean_fun_o)였습니다. 두 변수 모두 유의한 양의 회귀계수를 가지고 있고, 특히 외모가 높은 회귀 계수를 지닌 것으로 미루어 보았을 때 외향적으로 매력적인 여성이 남성 파트너에게 선택받는 경향이 있고, 그 다음으로 중요한 요소는 재미임을 알 수 있습니다. 회귀 모델의 p-value가 매우 유의하고, R-squared가 0.6245로 해당 모델이 파트너의 선택 비율을 어느 정도 잘 설명하고 있다고 볼 수 있습니다.

여성에 대한 매칭 비율을 종속 변수로 한 회귀 모형에서 유의한 변수는 자신이 평가한 지능의 중요도(intel2_1), 외모의 중요도(attr2_1), 자신의 운동에 대한 관심(exercise), 예상하는 매칭 횟수

(match_es), 파트너가 평가한 평균 외모 점수(mean_attr_o), 파트너가 평가한 평균 재미 점수(mean_fun_o)입니다. exercise를 제외한 유의한 모든 변수는 양의 회귀계수를 가지고 있습니다. 따라서 특히 외모와 재미가 뛰어나며, 본인 스스로 많은 파트너와 매칭 될 것으로 예상한 여성이 매칭 비율이 높음을 알 수 있습니다. 해당 모델은 유의하지만, R-squared 가 0.5 이하로 매칭 비율을 잘 설명하고 있다고 할 수는 없습니다.

남성에 대한 파트너의 선택 비율을 종속 변수로 한 회귀 모형에서 유의한 변수는 파트너가 평가한 평균 외모 점수(mean_attr_o)와 파트너가 평가한 평균 취미 공유 정도(mean_shar_o)였습니다. 두 변수 모두 유의한 양의 회귀계수를 가지고 있음을 고려하면 외향적으로 매력적이고, 여성들과 취미가 맞는 남성이 여성 파트너에게 선택받는 경향이 있음을 알 수 있습니다. 회귀 모델의 p-value가 매우 유의하고, R-squared가 0.6605로 해당 모델이 파트너의 선택 비율을 어느 정도 잘 설명하고 있다고 볼 수 있습니다.

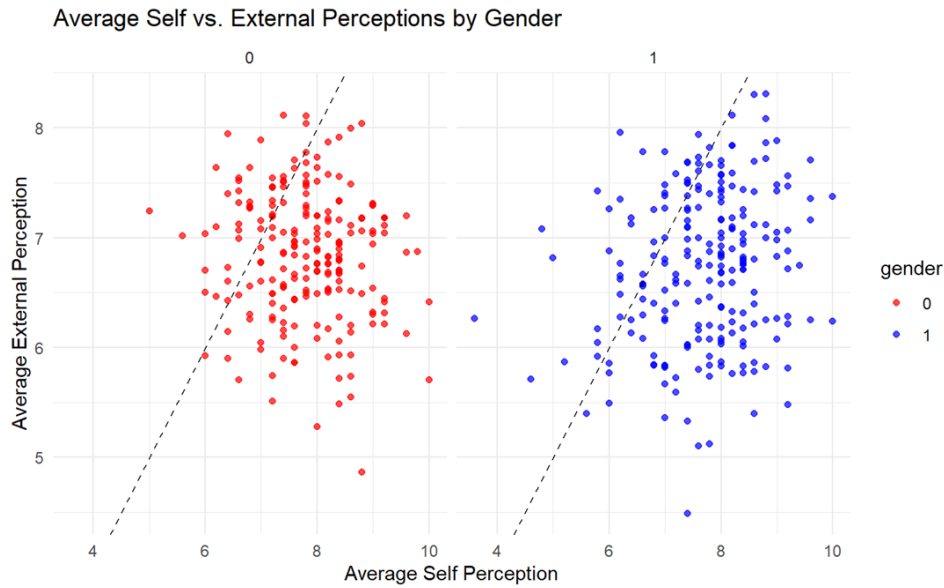
남성에 대한 매칭 비율을 종속 변수로 한 회귀 모형에서 유의한 변수는 예상하는 매칭 횟수(match_es), 파트너가 평가한 평균 외모 점수(mean_attr_o), 파트너가 평가한 평균 취미 공유 정도(mean_shar_o) 입니다. 유의한 모든 변수는 양의 회귀계수를 가지고 있고 특히 외모가 뛰어나며 여성의 취미에 잘 공감하고, 본인 스스로 많은 파트너와 매칭 될 것으로 예상한 남성의 매칭 비율이 높음을 알 수 있습니다. 해당 모델은 유의하지만, R-squared 가 0.5 이하로 매칭 비율을 잘 설명하고 있다고 할 수는 없습니다.

전반적으로 남성과 여성 모두 외모가 파트너 선택 및 매칭과 큰 연관성이 있다고 할 수 있으나, 매칭 비율의 경우 적합한 회귀 변수 외에도 다른 요인들이 영향을 미칠 것으로 예상됩니다.

5.2. MANOVA 활용 성별 간 자기 객관화의 차이 분석

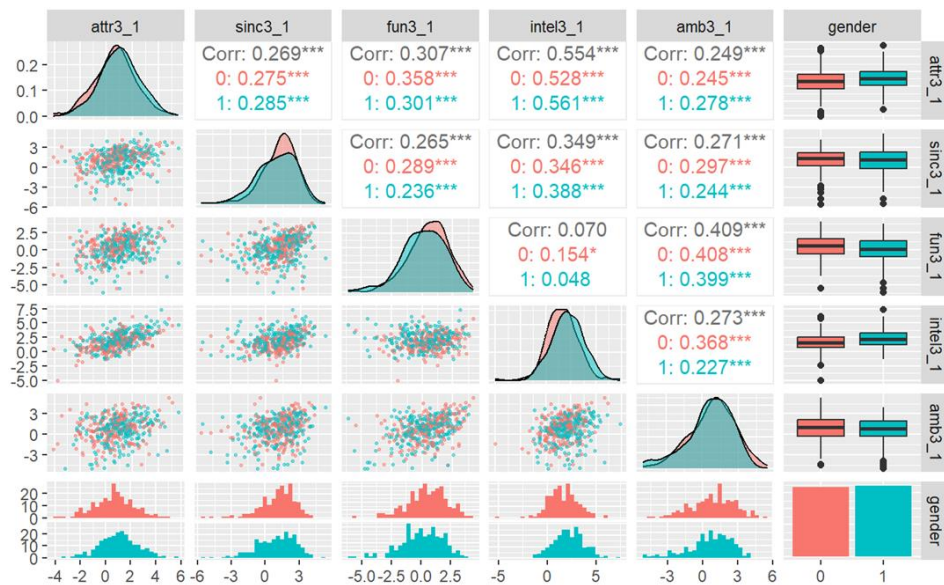
회귀모형을 이용한 분석 결과 파트너의 선택에 스스로의 자존감 여부와 관련이 있는 match_es 변수가 유의한 변수로 양의 영향을 미치는 것으로 나타났습니다. 따라서 '자기 객관화'의 척도를 제시하고, 이 척도가 성별 간 어떤 차이가 있는지 그룹 다중비교를 통해 알아보고자 했습니다.

먼저 저희가 정의한 자기 객관화의 척도 'gap'은 본인이 평가한 5가지 항목(외모, 재미, 지능, 야망, 진실됨)과 스피드 데이팅 파트너가 평가한 각 항목의 평균 점수의 차이로 정의됩니다. 이를 통해 각 항목별로 본인에 대한 과대평가 혹은 과소평가 여부를 파악할 수 있을 것으로 예상했습니다. 저희는 먼저 각 항목별 gap 변수를 계산하고, 그것의 평균과 실제 평가 점수의 평균을 scatterplot으로 나타내어 실제로 과대평가가 일어나고 있는지 확인했습니다. 그 결과 $y=x$ 선 우측 아래에 대다수의 관측치가 분포하고 있어, 실제로 과대평가가 일어나는 경향이 많다는 것을 알 수 있었습니다.



<Figure 4: 남성과 여성별 실제 점수와 본인이 평가한 점수>

각 성별마다 항목별 gap의 분포를 시각화한 결과, 남성과 여성의 분포가 amb를 제외한 모든 변수에서 차이를 보이고 있음을 알 수 있었으며, 대다수의 항목의 gap의 분포가 정규분포에 유사한 형태임을 알 수 있었습니다. 항목간 가장 높은 상관관계가 0.5로, multicollinearity 문제는 없을 것으로 예상했습니다. 따라서 MANOVA 검정을 통해 통계적으로 두 그룹간 gap의 유의한 차이가 있는지 검정해보고자 했습니다.



<Figure 5: 각 항목별 두 성별의 gap 분포와 correlation>

먼저 MANOVA의 3가지 가정인 정규성, 등분산성, 독립성을 검정을 통해 확인하고자 했습니다. 독립성의 경우 각 관측치는 한 명의 사람을 나타내고, 한 사람의 설문은 다른 사람의 설문에는 영향을 주지 않는 것이 보장되기 때문에 만족한다고 할 수 있기에 정규성과 등분산성을 만족하는지

검정을 실시했습니다.

먼저 각 gap들에 대한 MANOVA를 실시하기 위해 그룹별로 다변량 정규성(multivariate normality)를 만족하는지 검정하기 위해 MVN()함수를 사용해 Henze-zirkler 검정을 실시했습니다. 그 결과 남성의 경우 Attr과 Fun gap을 제외한 모든 변수에서 정규성을 만족하지 않았고, 다변량 정규성 역시 만족하지 않았습니다. 여성의 경우 Intel을 제외한 모든 변수에서 정규성을 만족했지만, 다변량 정규성은 만족하지 않았습니다.

Variable	P-value	Normality
Multivariate	2.338554e-06	No
Attr_gap	0.9883	Yes
Sinc_gap	0.0044	No
Intel_gap	0.0387	No
Fun_gap	0.8752	Yes
Amb_gap	<0.001	No

<Table 12: 남성에 대한 정규성 검정 결과>

Variable	P-value	Normality
Multivariate	0	No
Attr_gap	0.4964	Yes
Sinc_gap	<0.001	Yes
Intel_gap	0.0815	No
Fun_gap	0.5028	Yes
Amb_gap	0.1233	Yes

<Table 13: 여성에 대한 정규성 검정 결과>

Gap의 경우 값이 음수부터 양수까지 분포되어 있기에, 모든 범위의 숫자에 대한 정규성 확보에 도움이 되는 Yeo-Johnson Transformation을 사용해 정규성을 확보하고자 했습니다. 여 존슨 변환 결과 남성의 경우 모든 변수에 대한 정규성을 확보했으며, 다변량 정규성 역시 p-value 0.1로 귀 무가설을 기각하지 못해 확보할 수 있었습니다. 여성의 경우 여전히 다변량 정규성 검정을 통과 하지는 못했지만, 모든 변수에 대한 정규성을 확보했으며, 200개 이상의 sample size로 어느 정도 다변량 정규성 가정을 완벽하게 적용하지 못하더라도 근사하게 생각하고 분석을 진행할 수 있었습니다.

Variable	P-value	Normality
Multivariate	0.1015034	Yes
Attr_gap	0.9883	Yes
Sinc_gap	0.2443	Yes
Intel_gap	0.3882	Yes
Fun_gap	0.8020	Yes

Amb_gap	0.7984	Yes
---------	--------	-----

<Table 14: 변환 이후 남성에 대한 정규성 검정 결과>

Variable	P-value	Normality
Multivariate	2.518763e-05	No
Attr_gap	0.6458	Yes
Sinc_gap	0.8629	Yes
Intel_gap	0.9488	Yes
Fun_gap	0.2954	Yes
Amb_gap	0.9024	Yes

<Table 15: 변환 이후 여성에 대한 정규성 검정 결과>

이후 두번째 가정인 공분산행렬의 동질성을 검정하기 위해 box-m test를 이용했습니다. 그 결과 p-value 0.4498로 귀무가설을 기각하지 못하고, 각 그룹 간 공분산행렬의 동질성을 만족한다는 결론을 내릴 수 있었습니다.

Chisq	df	p-value
15.023	15	0.4498

<Table 16: Box-m test 결과>

정규성을 근사적으로 만족하고, 독립성과 공분산행렬의 동질성을 모두 만족하기에 모든 gap 변수를 종합했을 때 MANOVA 검정을 통해 각 그룹간 유의미한 차이가 있는지 알아보았습니다. Pillai Statistics를 활용해 검정을 실시한 결과 p-value 1.353e-06으로 귀무가설을 기각하고, 두 집단 간 유의미한 차이가 존재한다는 결론을 내렸습니다.

```

Df    Pillai approx F num Df den Df    Pr(>F)
gender  1 0.077543    7.3133      5   435 1.353e-06 ***
Residuals 439

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

<Table: MANOVA 결과>

더 나아가 어떤 항목의 gap 변수가 그룹 간 유의한 차이가 있는지 개별적으로 알아보기 위해 Bonferroni interval을 계산했습니다. 각 항목별로 Simultaneous Bonferroni interval을 계산한 결과, Attractive, Funny, Intelligent의 신뢰구간이 0을 포함하지 않기 때문에 성별 간 유의한 차이가 존재하며, 남자의 경우 본인의 외모, 유머 감각을 상대적으로 유의미하게 과대평가하며, 여자의 경우 남자에 비해 상대적으로 본인의 지능을 유의미하게 과대평가한다는 결론을 내렸습니다.

변수	Lower_ci	Upper_ci
Attractive gap	-0.771089375	-0.005020311
Sincere gap	-0.1953552	0.6059984

Funny gap	-0.9548967	-0.2082906
Intelligent gap	0.01895008	0.85784481
Ambitious gap	-0.1708693	0.7523608

<Table 17: 각 Gap 별 Bonferroni Confidence Interval>

더 나아가 해당 gap들만을 종속변수로 하여 per_match와 per_par_match 변수(여기서는 dec)에 대한 다변량 회귀 모델을 적합시키고, 5.1. 다변량 회귀에서와 같은 방식으로 MANOVA를 이용한 변수선택을 한 결과 최종 회귀모델은 다음과 같이 적합되었습니다:

Response dec :

Call:

```
lm(formula = dec ~ (attr_gap + sinc_gap + intel_gap + fun_gap +
  amb_gap + dec2) - dec2 - amb_gap - sinc_gap, data = d)
```

Residuals:

Min	1Q	Median	3Q	Max
-48.837	-12.604	-1.095	10.325	62.758

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	53.9591	2.1933	24.602	< 2e-16 ***
attr_gap	-4.2587	1.0601	-4.017	8.10e-05 ***
intel_gap	3.0067	0.7545	3.985	9.19e-05 ***
fun_gap	-5.5002	0.9908	-5.551	8.16e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.72 on 219 degrees of freedom

Multiple R-squared: 0.3418, Adjusted R-squared: 0.3328

F-statistic: 37.91 on 3 and 219 DF, p-value: < 2.2e-16

Response match :

Call:

```
lm(formula = match ~ (attr_gap + sinc_gap + intel_gap + fun_gap +
  amb_gap + dec2) - dec2 - amb_gap - sinc_gap, data = d)
```

```

Residuals:
    Min       1Q   Median       3Q      Max
-0.93966  0.08805  0.15613  0.20525  0.42937

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.874983   0.044613   19.613  <2e-16 ***
attr_gap    -0.042458   0.021564   -1.969   0.0502 .
intel_gap     0.022801   0.015348    1.486   0.1388
fun_gap     -0.005587   0.020154   -0.277   0.7819
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 0.3807 on 219 degrees of freedom
Multiple R-squared: 0.03299, Adjusted R-squared: 0.01974
F-statistic: 2.49 on 3 and 219 DF, p-value: 0.06118
<Table 18: 남성에 대한 다변량 회귀 적합 결과>

Response dec :

```

Call:
lm(formula = dec ~ (attr_gap + sinc_gap + intel_gap + fun_gap +
  amb_gap + dec2) - dec2 - intel_gap - sinc_gap, data = d2)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-41.929 -12.548  -0.339  12.885  49.282

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  58.5391     1.9576   29.904  < 2e-16 ***
attr_gap     -3.4383     0.9670   -3.556 0.000464 ***
fun_gap      -5.4691     1.0833   -5.049 9.5e-07 ***
amb_gap      -1.3185     0.7619   -1.731 0.084981 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 19.2 on 214 degrees of freedom

Multiple R-squared: 0.3174, Adjusted R-squared: 0.3078
F-statistic: 33.17 on 3 and 214 DF, p-value: < 2.2e-16

Response match :

Call:

```
lm(formula = match ~ (attr_gap + sinc_gap + intel_gap + fun_gap +  
  amb_gap + dec2) - dec2 - intel_gap - sinc_gap, data = d2)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.94004	0.06832	0.15865	0.23647	0.38080

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.8512892	0.0400091	21.277	<2e-16 ***
attr_gap	-0.0263064	0.0197644	-1.331	0.1846
fun_gap	-0.0003705	0.0221403	-0.017	0.9867
amb_gap	-0.0347520	0.0155721	-2.232	0.0267 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3924 on 214 degrees of freedom

Multiple R-squared: 0.04538, Adjusted R-squared: 0.032

F-statistic: 3.391 on 3 and 214 DF, p-value: 0.01889

<Table 19: 여성에 대한 다변량 회귀 적합 결과>

두 그룹 모두에서 매우 낮은 R-squared 값과 매우 낮은 회귀계수들을 통해 매칭 여부는 gap 변수들로 잘 설명되지 못하는 것을 알 수 있습니다. 한편 남성의 경우 유의한 attr, intel, fun gap을 통해 본인의 외모와 재미에 대한 과대평가가 심한 남성이 여성에게 선택받지 못하는 경향성이 있음을 알 수 있으며, 지능에 대해 과대평가가 심할수록 여성에게 선택받은 비율이 증가함을 알 수 있습니다. 여성의 경우 attr과 fun gap만이 유의해, 외모와 재미에 대해 과대평가가 심한 여성은 남성에게 선택받는 비율이 감소하는 경향성을 보여줍니다. 이는 곧 외모와 재미에 대한 자기 객관화가 이성과의 만남에 있어 상당히 중요한 요소임을 방증합니다.

6. 결론, 한계점 및 제언

결론

우선 첫번째로 파트너의 선택과 매칭에 영향을 준 변수를 찾고자 한 다변량 회귀 분석의 결과를 요약하면 총 5 가지의 매력 평가 항목인 외모, 재미, 진실성, 지성, 야망 중 성별에 따라 유의한 변수들과 그 비중이 다르고 또한 파트너의 선택과 매칭 간에도 유의한 변수들에서 차이가 존재함을 확인할 수 있었습니다. 이를 통해 데이팅에 있어 성별 별로 이성의 선택을 받고 최종적으로 매칭되기 위해서 중요하게 생각해야 할 요소를 알 수 있었고 이는 데이팅에 나가고자 하는 사람들에게 도움이 될 것이라 생각합니다.

그 다음 MANOVA를 활용하여 자기 객관화 분석의 결과를 요약하면 성별에 따라 각 매력 항목 별 자기 객관화의 정도가 파트너의 선택에 유의한 영향을 미치는지와 그 정도가 다른 것을 볼 수 있었습니다. 그러나 앞선 분석과는 달리 해당 지표들이 매칭 여부에는 두 성별 모두에서 유의한 영향을 미치지 못하는 것을 볼 수 있었습니다.

이러한 분석 결과들을 통해 데이팅에 있어 어떠한 요소들이 파트너의 선택과 최종 매칭에 영향을 주는 지 알 수 있었습니다. 이러한 결과들은 단순히 데이팅이라는 특수한 상황뿐만이 아닌 다양한 사람들과의 상호작용에서 4분 데이팅처럼 첫인상에서 상대방에게 긍정적인 인상을 심어줄 수 있는 방법으로도 확장할 수 있는 여지가 존재한다고 생각합니다. 따라서 이러한 분석들을 대인관계에서의 긍정적인 인상을 심어주기 위한 연구 중 데이팅이라는 특수한 경우에 맞춰 실험한 것이라 보고 다양한 방향으로 연구를 확장할 수 있을 것이라 생각합니다.

한계점 및 제언

본 연구의 가장 큰 한계점은 소개팅과 같이 일반적으로 생각하는 남녀 간의 일상적인 만남이 아닌 특수한 상황(4분 데이팅)에서의 데이터를 사용하여 데이터 분석을 실시하였기 때문에 소개팅과 같은 일반적 상황에서도 그대로 적용하기에는 한계가 있을 수 있다는 것입니다. 첫인상이 중요하다고는 하지만 단지 4분만으로는 한 사람의 장단점이 많이 들어 나지 않을 경우도 많기 때문에 명백한 한계가 존재한다고 할 수 있습니다. 또한 모든 데이터가 스스로 점수를 매긴 설문이기 때문에 개인들의 주관이 들어가 객관적으로 평가한 데이터가 되기에는 힘들다는 것이 주어진 스피드 데이팅 데이터를 이용한 위 분석들의 한계라 볼 수 있습니다.

제언할 것은 크게 두가지입니다. 우선 첫번째로, 분석 결과에서 볼 수 있듯이 성별 간 중요하게 생각하는 것이 다를 수 있음을 인지하는 태도를 가져야 한다는 것이며 파트너의 선택을 받기 위해서는 본인을 가꾸는 노력이 필요하다는 것입니다. 데이팅을 통해 연인을 찾고자 하는 사람들에게 큰 도움이 될 수 있으리라 생각합니다. 두번째로는 이러한 4분 데이팅이 아닌 좀 더 일반적인 긴 시간의 데이팅에 대한 데이터와 분석할 기회가 있으면 더 보편적인 분석과 적용이 가능할 것이라는 점입니다. 이는 실생활에서의 사람들에게 대한 데이팅에서의 상호작용을 더 잘 분석할 수 있으리라 생각됩니다.

7. References

- [1] Raymond Fisman, Sheena S. Iyengar, Emir Kamenica, Itamar Simonson, Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment, *The Quarterly Journal of Economics*, Volume 121, Issue 2, May 2006, Pages 673–697
- [2] Yeo, I.-K., & Johnson, R. A. (2000). A New Family of Power Transformations to Improve Normality or Symmetry. *Biometrika*, 87(4), 954–959.
- [3] Raymond Fisman, Sheena S. Iyengar, Emir Kamenica, Itamar Simonson, Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment, *The Quarterly Journal of Economics*, Volume 121, Issue 2, May 2006, Pages 673–697,