

운동 수행 능력 기준의 성별 집단 세분화

4조 : 송호영 이승준 선주영 김민준

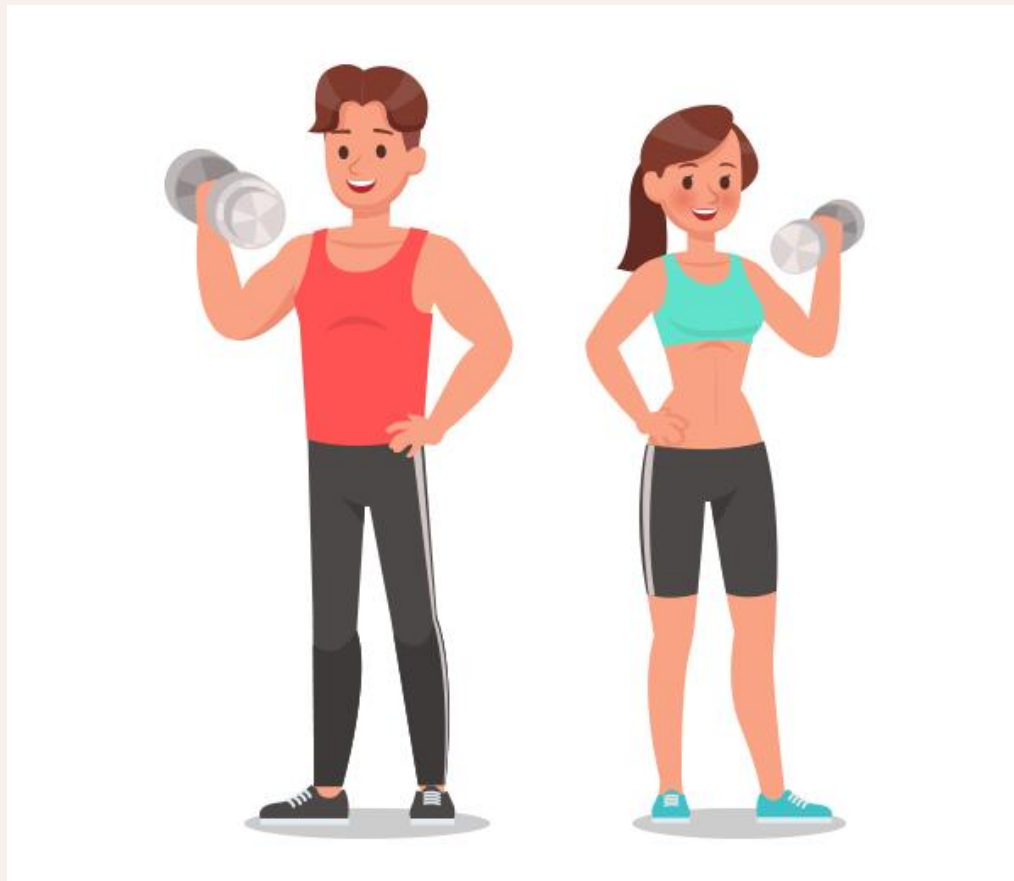
CONTENTS



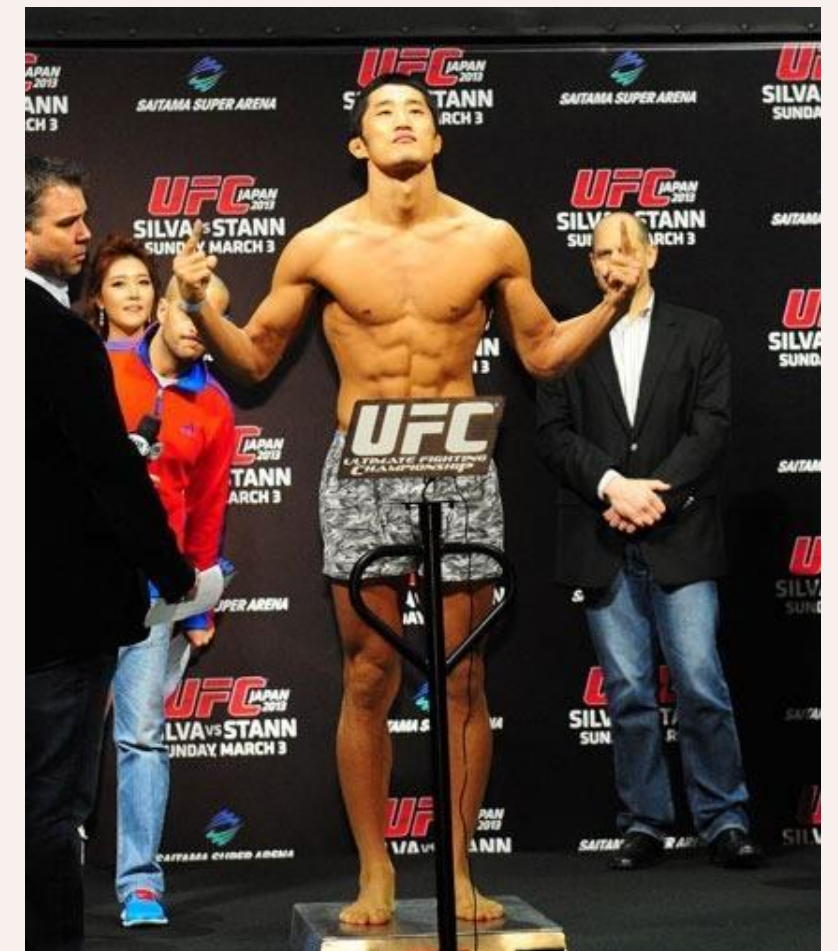
- 01 분석 배경 및 목적
- 02 자료 소개
- 03 이상치 처리
- 04 주성분분석
- 05 다변량분산분석
- 06 결론

01 분석 배경 및 목적

일상 생활 속 남녀의 체력적 차이

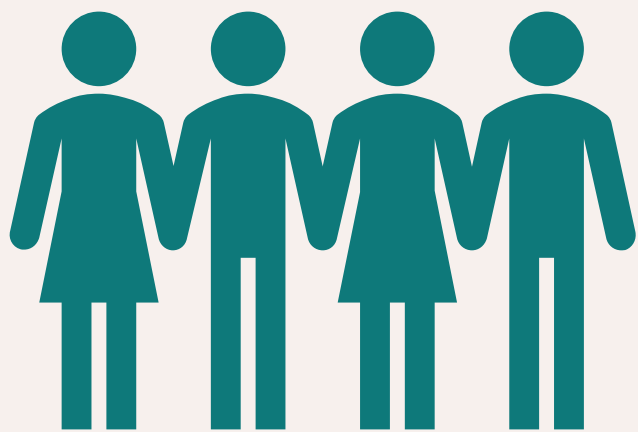


+ 성별 외의 추가적인 차이



예시) 체급에 의한 구분

02 자료 소개



대한민국 성인 남녀 13,393명
[전국 체력 인증 센터]

kaggle

<Body performance Data>

변수 종류	변수 이름	변수 순번	변수 설명
신체 변수	age	1	나이
	gender	2	성별
	height_cm	3	키(cm)
	weight_kg	4	몸무게(kg)
	body.fat_.	5	체지방률(%)
	diastolic	6	이완기 혈압
	systolic	7	수축기 혈압
운동 수행 변수	gripForce	8	악력
	sit.and.bend.forward_cm	9	앉아 윗몸 앞으로 굽히기(cm)
	sit.ups.counts	10	윗몸 일으키기
	broad.jump_cm	11	제자리 멀리뛰기(cm)
등급 변수	class	12	운동 수행 능력 등급

02

자료 소개

변수 소개 - 요약 통계량

변수 이름	표본평균	표본 표준편차	최솟값	중위수	최댓값
age	36.775	13.626	21	32	64
height_cm	168.560	8.427	125	169.2	193.8
weight_kg	67.447	11.950	26.3	67.4	138.1
body.fat_.	23.240	7.257	3	22.8	78.4 ✓
diastolic	78.797	10.742	0 ✓	79	156.2
systolic	130.235	14.714	0 ✓	130	201
gripForce	36.964	10.625	0 ✓	37.9	70.5
sit.and.bend.forward_cm	15.209	8.457	-25	16.2	213 ✓
sit.ups.counts	39.771	14.277	0 ✓	41	80
broad.jump_cm	190.130	39.868	0	193	303

02

자료 소개

변수 소개 - 상위, 하위 10개 관측치

body.fat_.	78.4	54.9	53.5	50.6	50.3	50.2	49.8	49.3	49.2	48.9
sit.and.bend.forward	213.0	185.0	42.0	40.0	40.0	37.0	35.2	35.2	35.2	35.2
diastolic	0	6	8	30	37	40	41	41	42	42
systolic	0.0	14.0	43.9	77.0	82.0	84.0	86.0	86.0	86.0	88.0
gripForce	0.0	0.0	0.0	1.6	2.1	3.5	4.4	5.3	6.7	7.9
broad.jump	0	0	0	0	0	0	0	0	0	0

03 이상치 처리

체지방률

b_fat	gender	h_cm	w_kg	diastolic	class
78.4 ✓	M	177.6	74.5	69	A
54.9	M	172.8	95.0	90	D
53.5	F	162.5	113.3	82	D
50.6	F	160.0	109.2	94	D
50.3	F	160.0	76.9	82	D

<체지방률 50% 이상 5명>



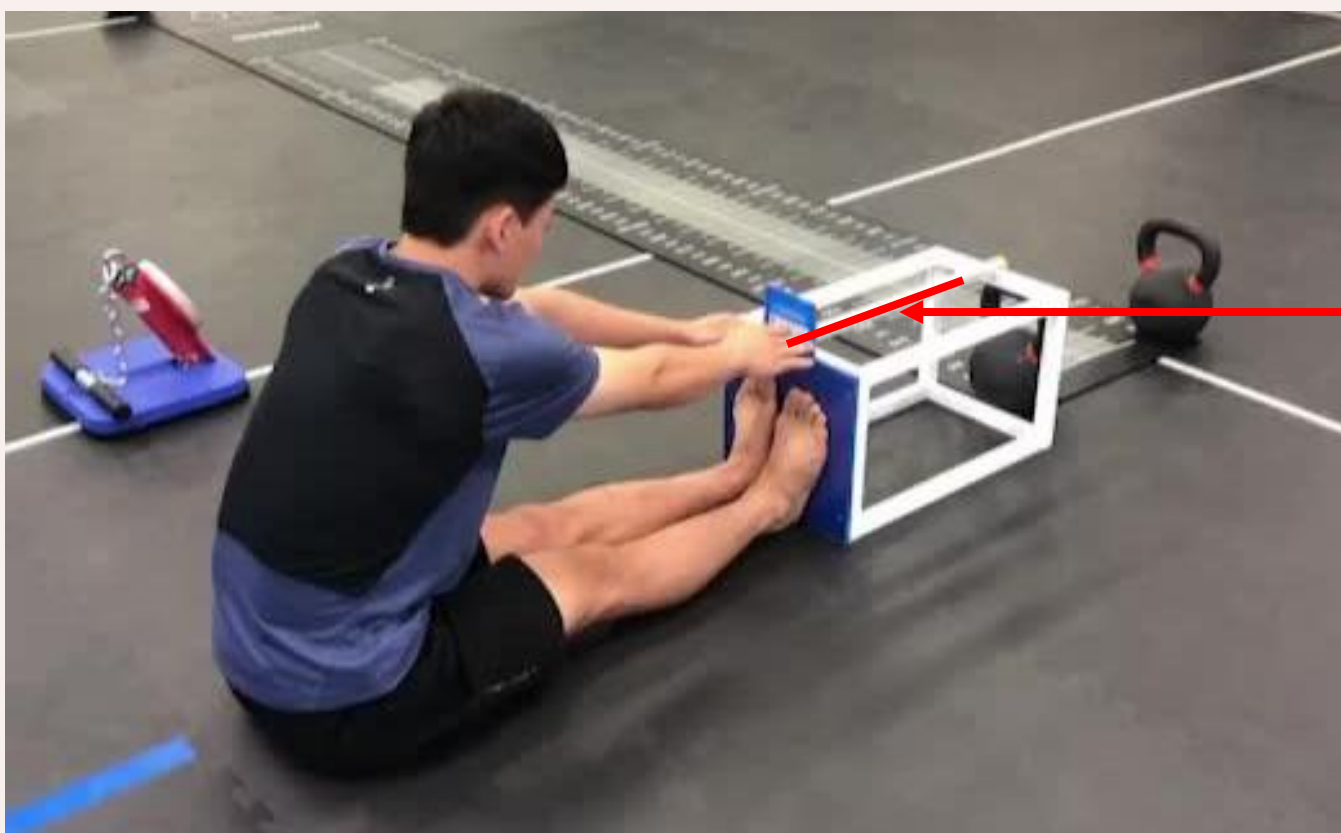
A등급, 177.6cm, 74.5kg

다른 사람들에 비해 유독 체지방률과
어울리지 않는 신체 기록을 보유

03

이상치 처리

앉아 윗몸 앞으로 굽히기



(팔 길이 + 상체 길이) - 하체 길이

✓	✓								
213.0	185.0	42.0	40.0	40.0	37.0	35.2	35.2	35.2	35.2

<앉아 윗몸 앞으로 굽히기 상위 10명 (cm) >

213cm, 185cm

이론적으로 나올 수 없는 관측치

○
○
○

13,393개 자료 -> (이상치 19개 제거) -> 13,374개 자료

04 주성분분석

분석 방향

변수 종류	변수 이름	변수 순번	변수 설명
신체 변수	age	1	나이
	gender	2	성별
	height_cm	3	키(cm)
	weight_kg	4	몸무게(kg)
	body.fat_.	5	체지방률(%)
	diastolic	6	이완기 혈압
	systolic	7	수축기 혈압
운동 수행 변수	gripForce	8	악력
	sit.and.bend.forward_cm	9	앉아 윗몸 앞으로 굽히기(cm)
	sit.ups.counts	10	윗몸 일으키기
	broad.jump_cm	11	제자리 멀리뛰기(cm)

주성분분석

성별을 제외한 6개의 신체 변수 이용
새로운 척도 생성

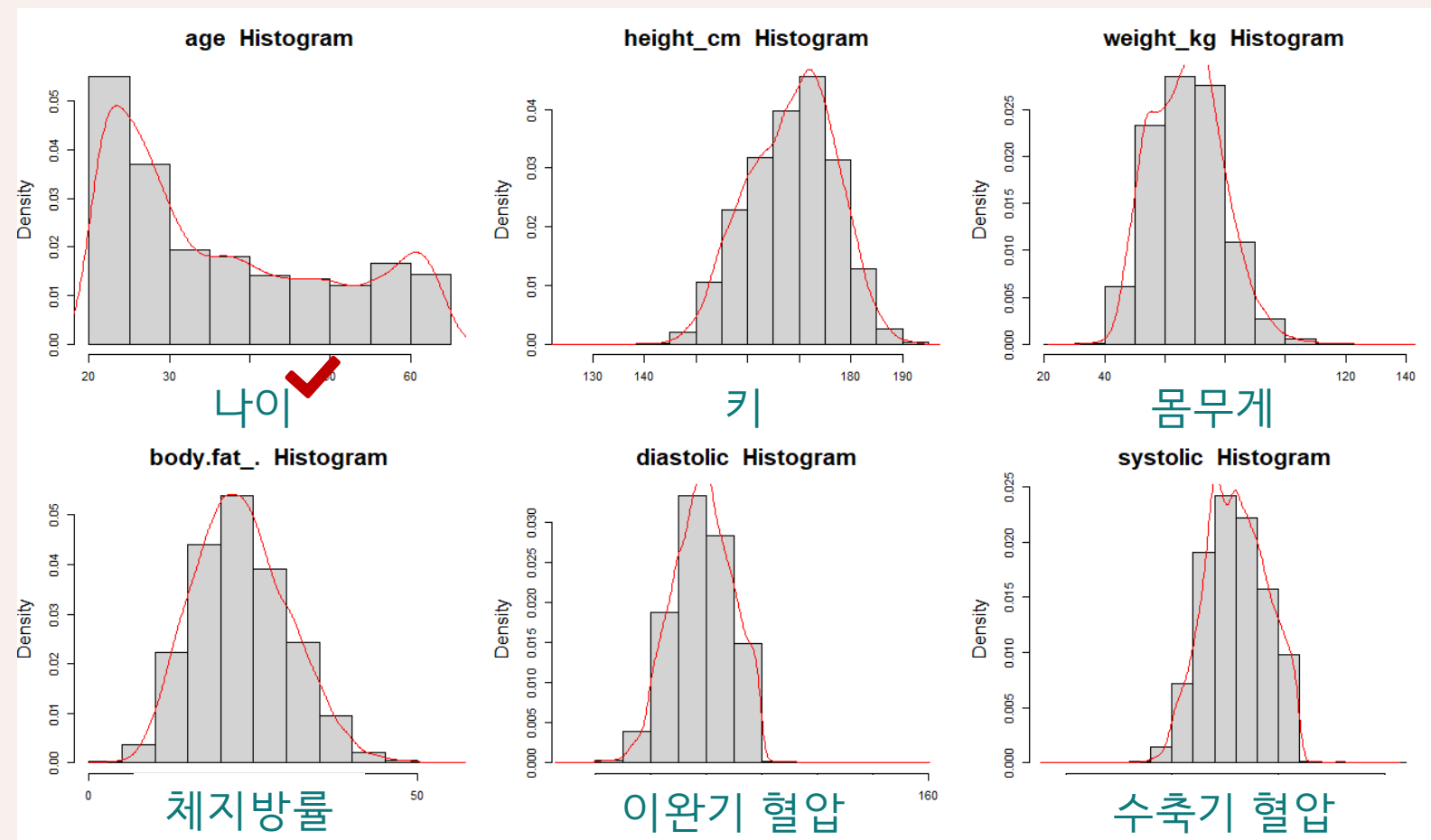
다변량분산분석

생성한 척도가 운동 변수에 대하여
잘 구분하는지 평가

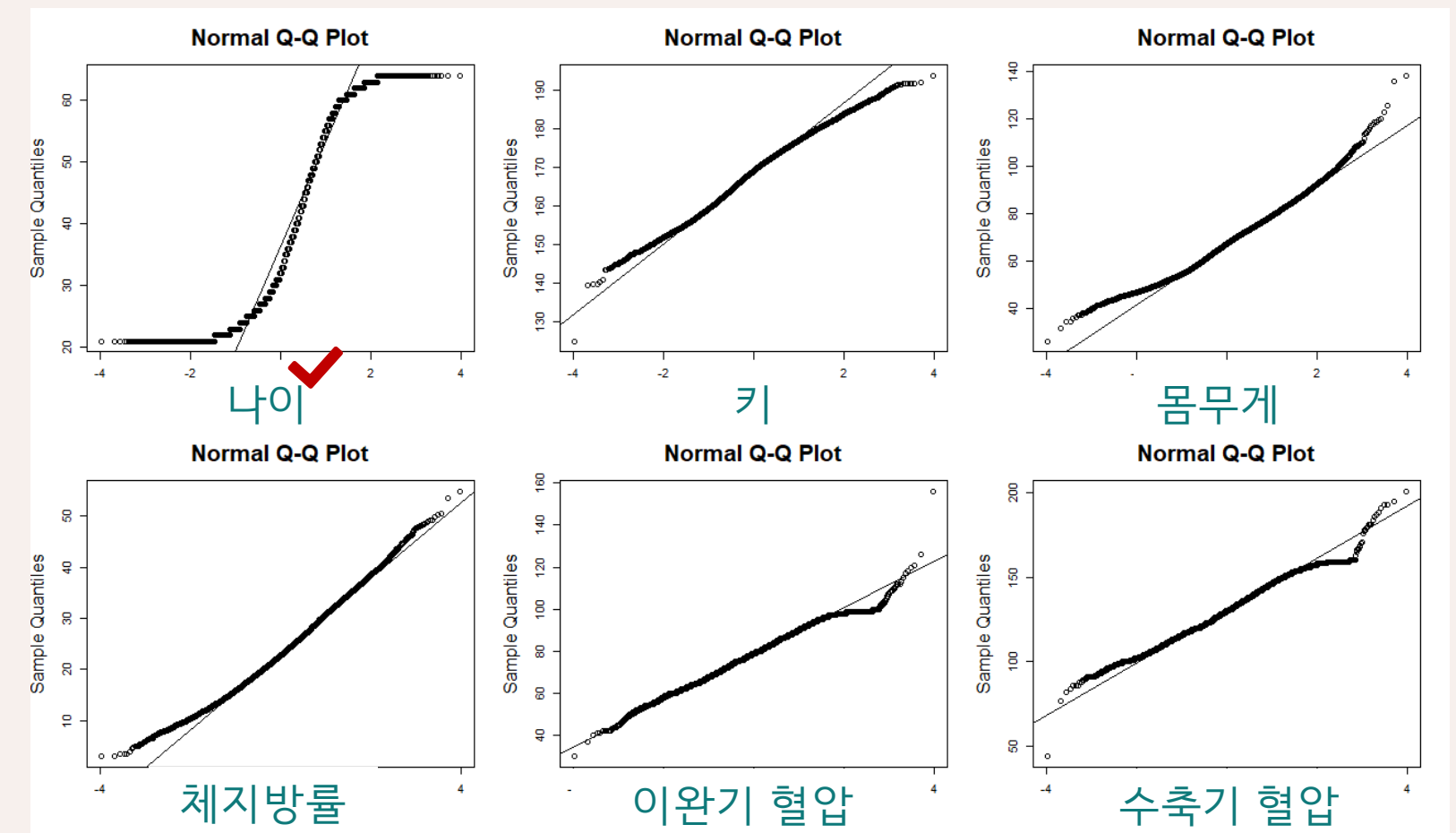
04 주성분분석

정규성 확인

성별을 제외한 6개의 신체 변수



히스토그램



Normal Q-Q Plot

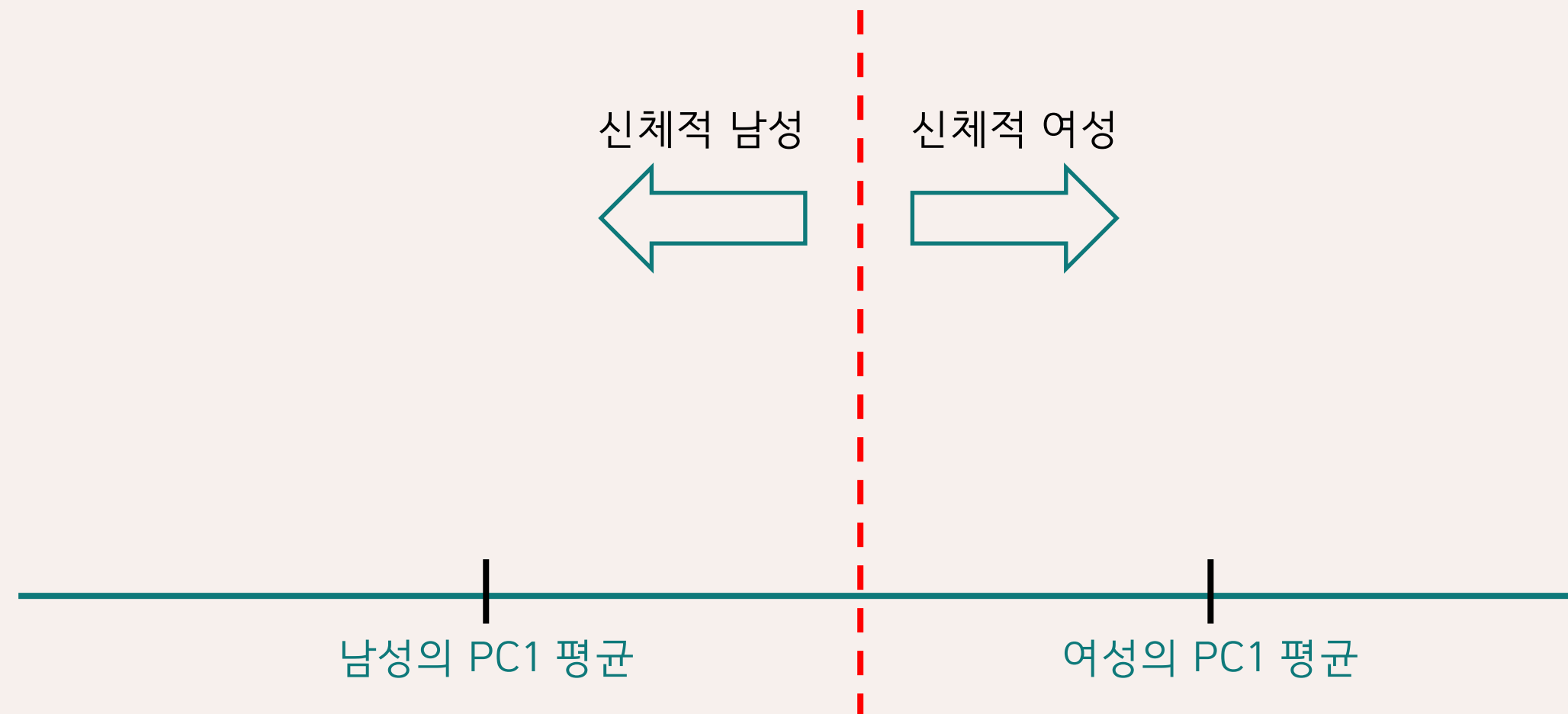
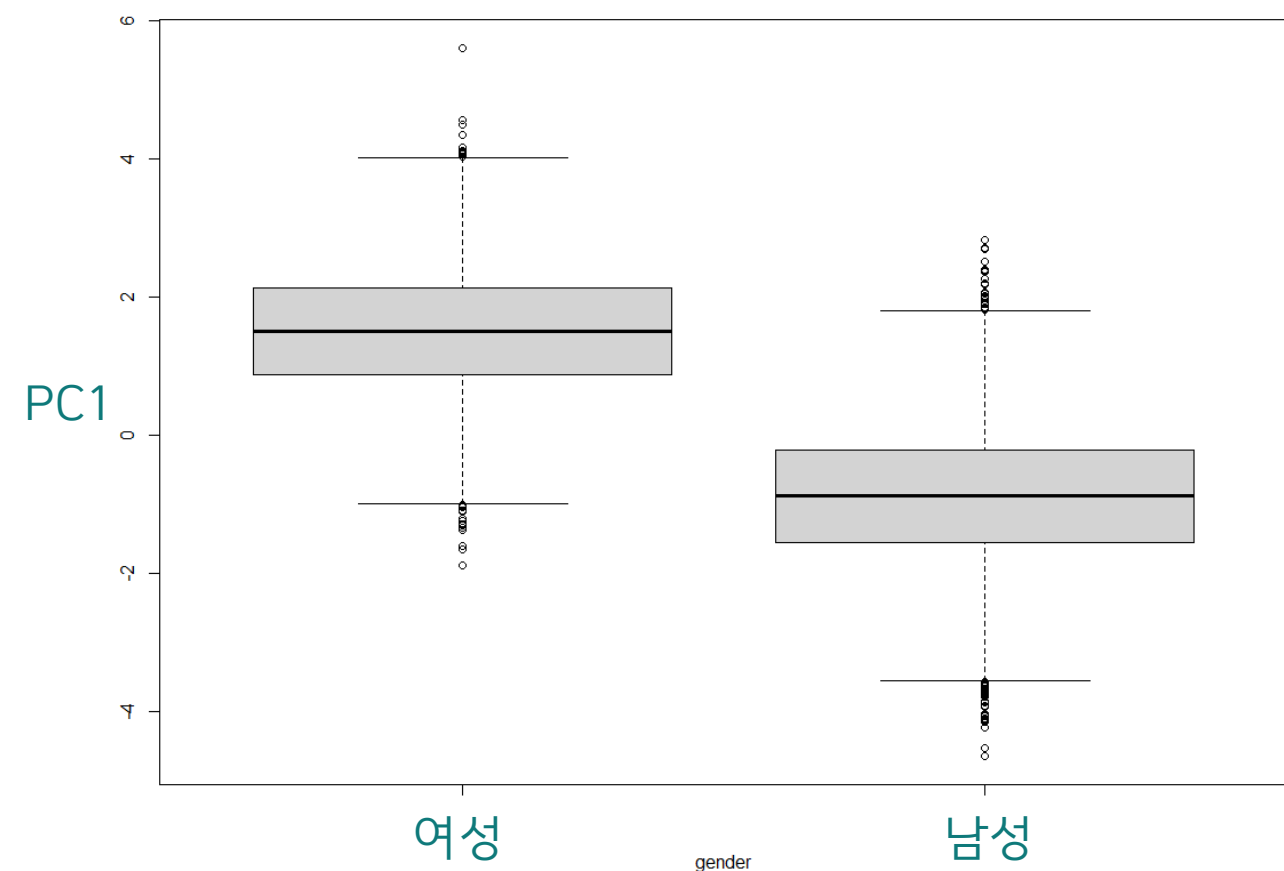
04 주성분분석 (상관 계수 행렬)

주성분 계수 해석

신체 변수		age	height_cm	weight_kg	body.fat_.	diastolic	systolic		
변수 설명		나이	키	몸무게	체지방률(%)	이완기 혈압	수축기 혈압	새로운 척도로 사용!	
80.598%	PC1 (37.986%)	0.106	-0.556	-0.535	0.276	-0.374	-0.421	→	<u>신체적 여성 성향</u>
	PC2 (28.890%)	0.515	-0.323	-0.052	0.419	0.486	0.464	→	노화 정도, 신체 나이
	PC3 (13.721%)	0.121	-0.105	-0.561	-0.759	0.185	0.221	→	반(反) 비만도
		

04 주성분분석 (상관 계수 행렬)

새로운 변수 생성

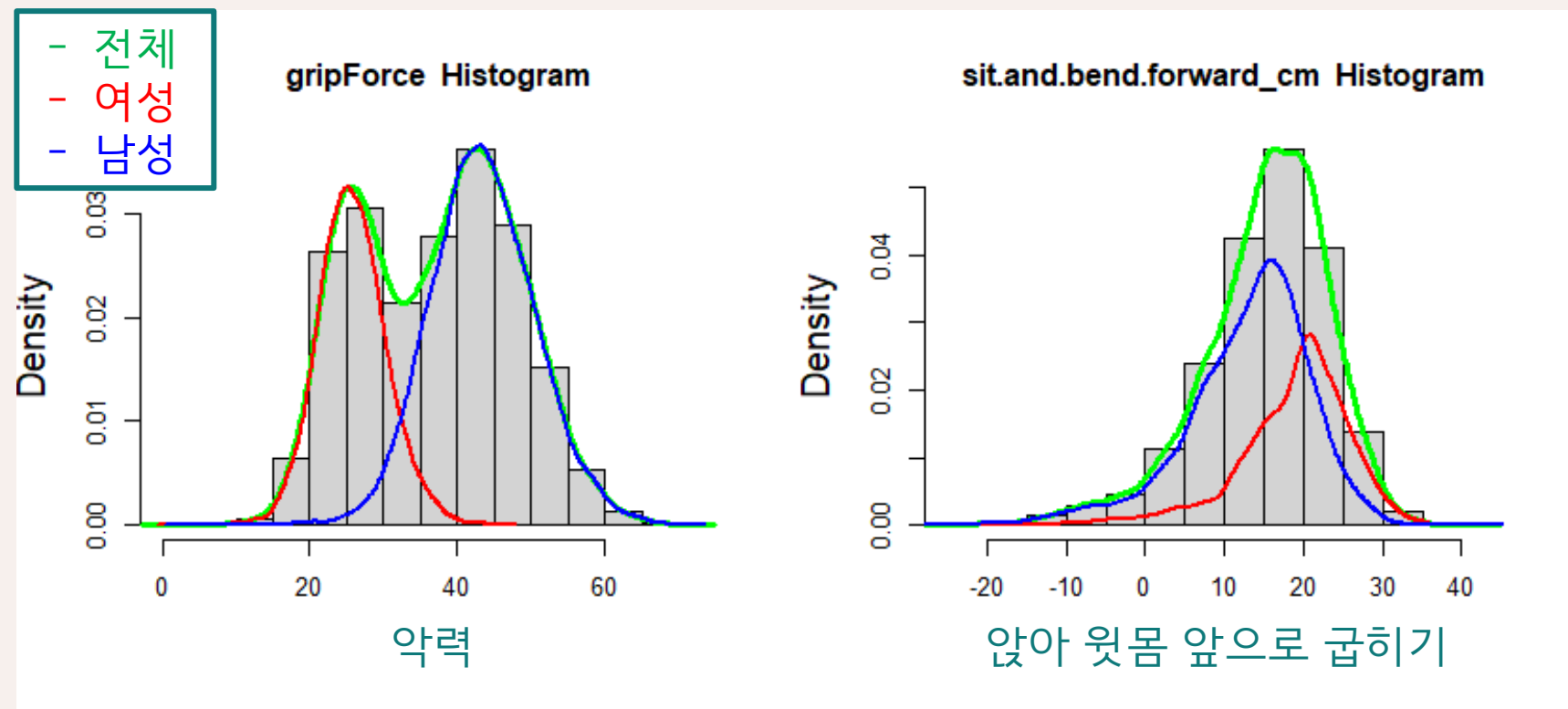


본인의 실제 성별과의 일치 여부에 따라 Yes(True)와 No(False)로 구분되는

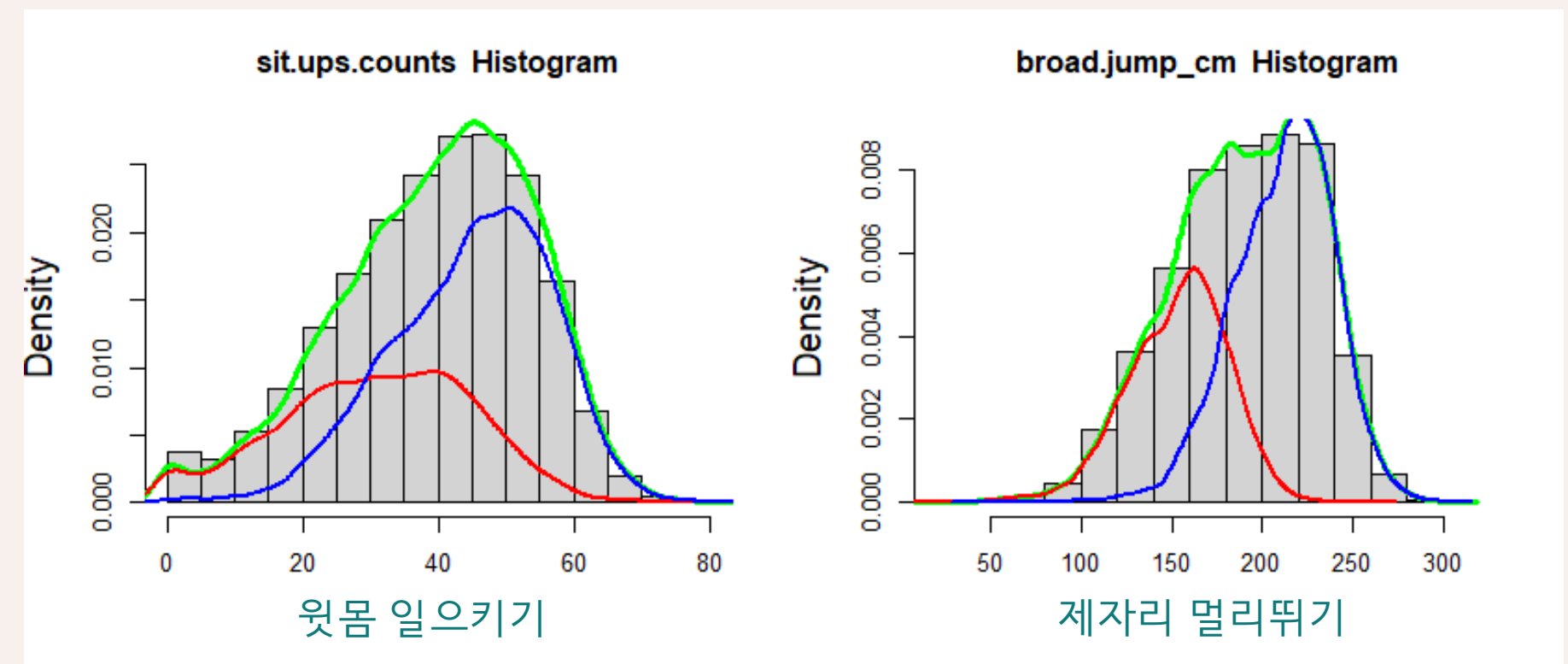
`gender.test(성향 일치)` 변수 생성

05 다변량분산분석

운동 변수 시각화



남성과 여성의 분포 차이가 뚜렷함



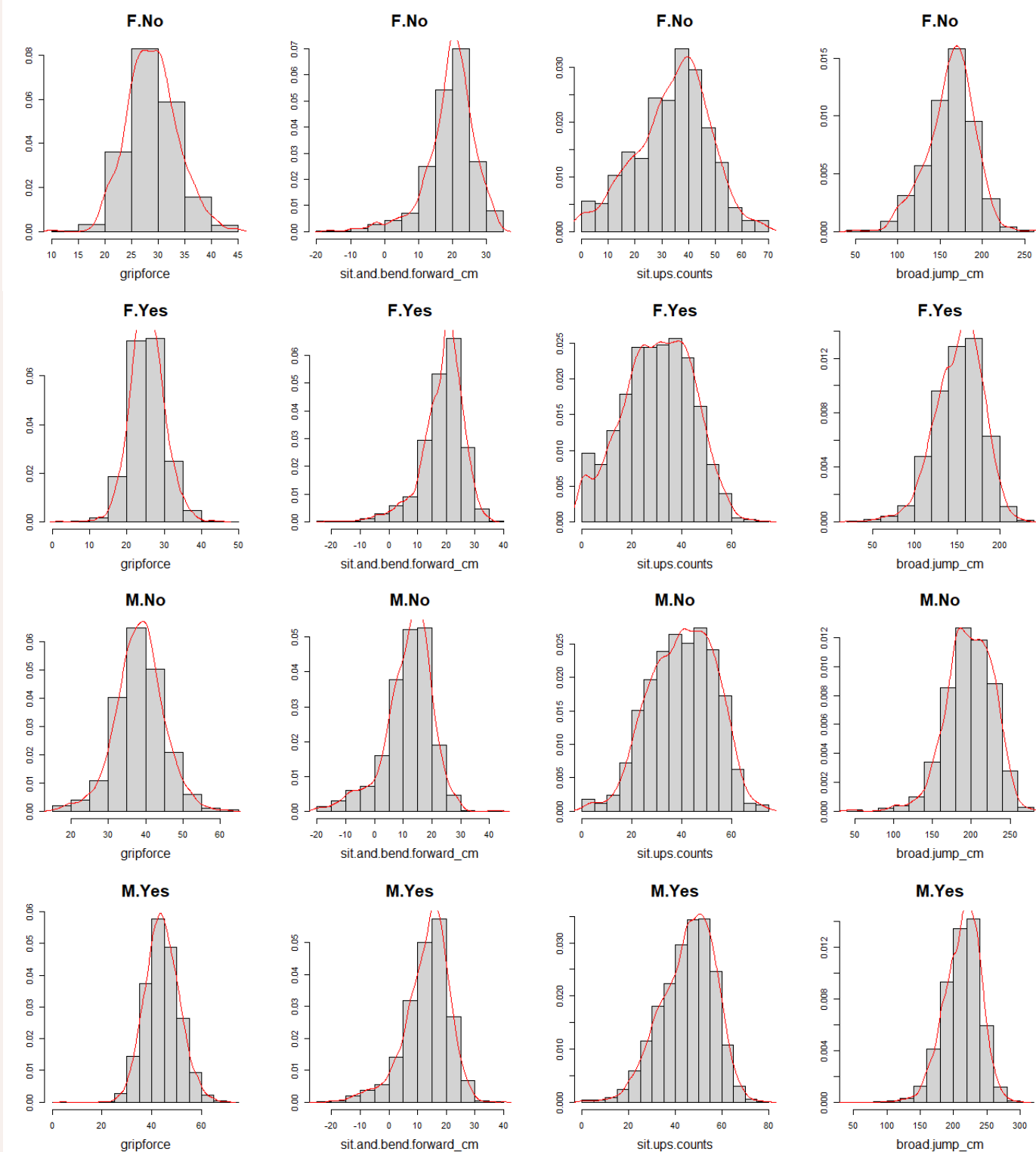
여성의 경우 왼쪽 벽 존재
- 0보다 작은 값이 나올 수 없기 때문

전체적으로 우측으로 치우친 분포
- 반동을 이용하는 등 정확하지 않은 자세로 측정되었을 가능성

4가지 운동 변수가 성별에 따라 모두 다른 경향성을 보여주기 때문에, 성별에 더한 추가적인 척도를 평가하기 위해서는 다변량분산분석(MANOVA)이 필요함

05 다변량분산분석

정규성 확인



악력

앉아 윗몸
앞으로 굽히기

윗몸 일으키기

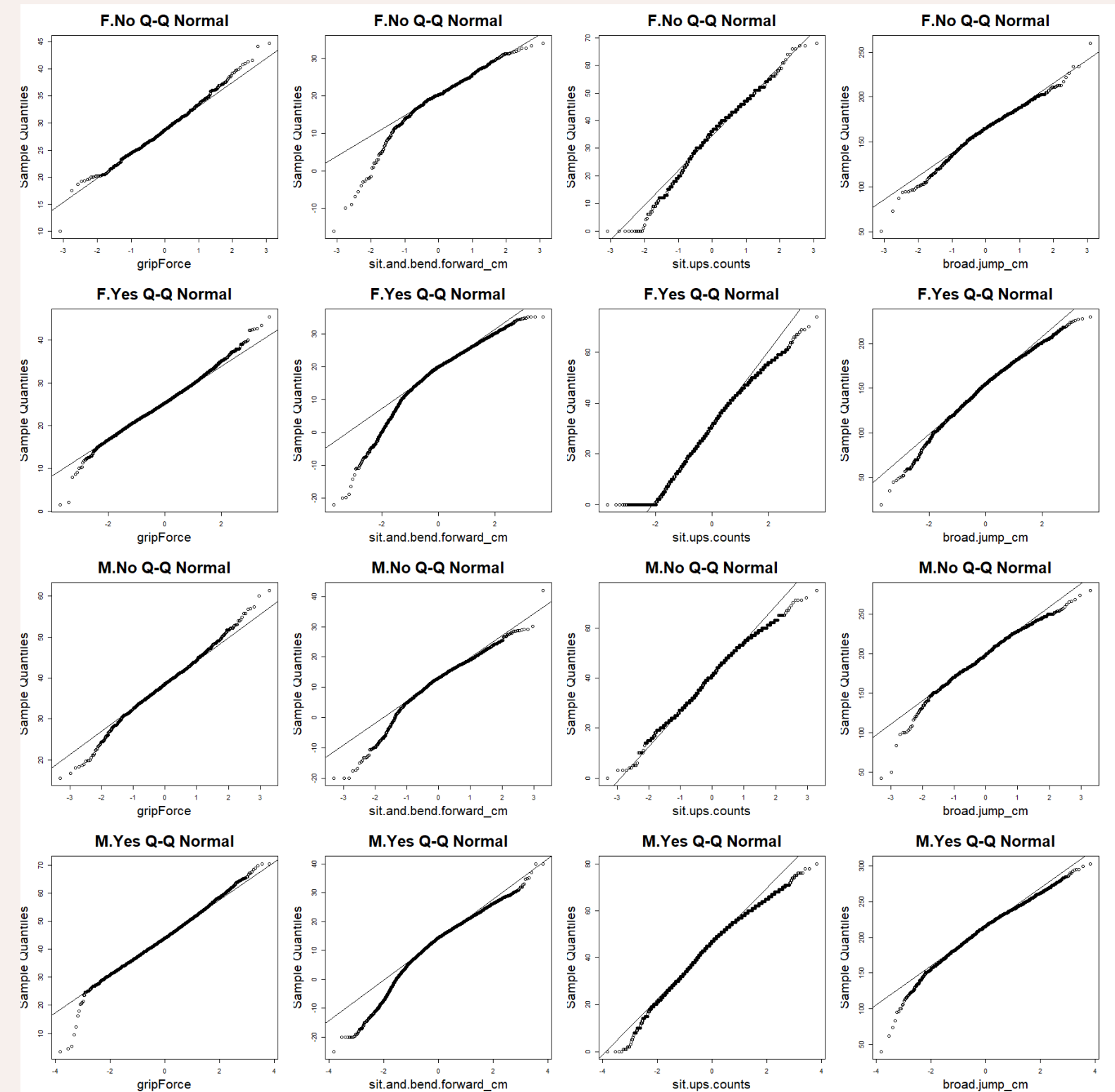
제자리 멀리뛰기

신체적 남성
여성

신체적 남성
여성

신체적 남성
여성

신체적 남성
여성



악력

앉아 윗몸
앞으로 굽히기

윗몸 일으키기

제자리 멀리뛰기

05 다변량분산분석

등분산성 확인

	gripForce	sit.and.bend.forward	sit.ups.counts	broad.jump
gripForce	22.025181	7.478549	24.01158	55.58584
sit.and.bend.forward	7.478549	50.612232	47.40818	88.45411
sit.ups.counts	24.011578	47.408182	184.94446	254.24128
broad.jump	55.585836	88.454109	254.24128	762.19697

<신체적 남성인 여성>

	gripForce	sit.and.bend.forward	sit.ups.counts	broad.jump
gripForce	41.39383	12.51143	34.65981	97.88991
sit.and.bend.forward	12.51143	68.83297	46.35013	98.40913
sit.ups.counts	34.65981	46.35013	166.83374	271.02304
broad.jump	97.88991	98.40913	271.02304	898.89471

<신체적 여성인 남성>

	gripForce	sit.and.bend.forward	sit.ups.counts	broad.jump
gripForce	20.617354	9.915044	25.84738	55.74165
sit.and.bend.forward	9.915044	51.047939	40.00303	76.42456
sit.ups.counts	25.847382	40.003033	191.78479	274.94692
broad.jump	55.741647	76.424562	274.94692	764.37110

<신체적 여성인 여성>

	gripForce	sit.and.bend.forward	sit.ups.counts	broad.jump
gripForce	48.30149	15.17152	27.72683	76.25209
sit.and.bend.forward	15.17152	62.07235	37.63402	86.36512
sit.ups.counts	27.72683	37.63402	130.41797	190.99446
broad.jump	76.25209	86.36512	190.99446	738.46132

<신체적 남성인 남성>

4개의 집단의 공분산행렬
큰 차이가 없이 비슷하다.

05 다변량분산분석

자료 나열

		성별	성향 일치	반응 변수				
y	id	gi	gk	gj	gikj	gij	gkj	gik
34.1	1	F	No	1	F.No.1	F.1	No.1	F.No
19.0	1	F	No	2	F.No.2	F.2	No.2	F.No
30.0	1	F	No	3	F.No.3	F.3	No.3	F.No
155.0	1	F	No	4	F.No.4	F.4	No.4	F.No
32.1	2	F	No	1	F.No.1	F.1	No.1	F.No
...

같은 ID 내에서 반응 변수간 독립성을 가정할 수 없음 ---> GLS 모델 사용

05 다변량분산분석

가설 검정

H_{0a} : 성별 간의 운동 수행 능력에 차이가 없다.

Full Model : $Y_{ijd} = g_{ij} + \varepsilon_{ijd}$

Reduced Model : $Y_{jd} = g_j + \varepsilon_{jd}$

	Model	df	AIC	BIC	logLik	Test	Ratio	p-value
RM	1	11	454387.7	454485.5	-227182.9			
FM	2	15	428028.0	428161.3	-213999.0	1 vs 2	26367.71	<u><.0001</u>

Reject H_{0a}

유의 수준 : 0.05

H_{0b} : 성별에 추가적으로 성향 일치 변수를 추가하는 것은 도움이 되지 않는다.

Full Model : $Y_{ikjd} = g_{ikj} + \varepsilon_{ikjd}$

Reduced Model : $Y_{ijd} = g_{ij} + \varepsilon_{ijd}$

	Model	df	AIC	BIC	logLik	Test	Ratio	p-value
RM	1	15	428028.0	428161.3	-213999.0			
FM	2	23	427038.6	427243.0	-213496.3	1 vs 2	1005.472	<u><.0001</u>

Reject H_{0b}

유의 수준 : 0.05

H_{0c} : 성향 일치에 따라, 남성과 여성의 반응 변수 차이가 동일하다.

Full Model : $Y_{ikjd} = g_{ikj} + \varepsilon_{ikjd}$

Reduced Model : $Y_{ikjd} = (g_i + g_k) * g_j + \varepsilon_{ikjd}$

	Model	df	AIC	BIC	logLik	Test	Ratio	p-value
RM	1	19	427781.3	427950.2	-213871.7			
FM	2	23	427038.6	427243.0	-213496.3	1 vs 2	750.7741	<u><.0001</u>

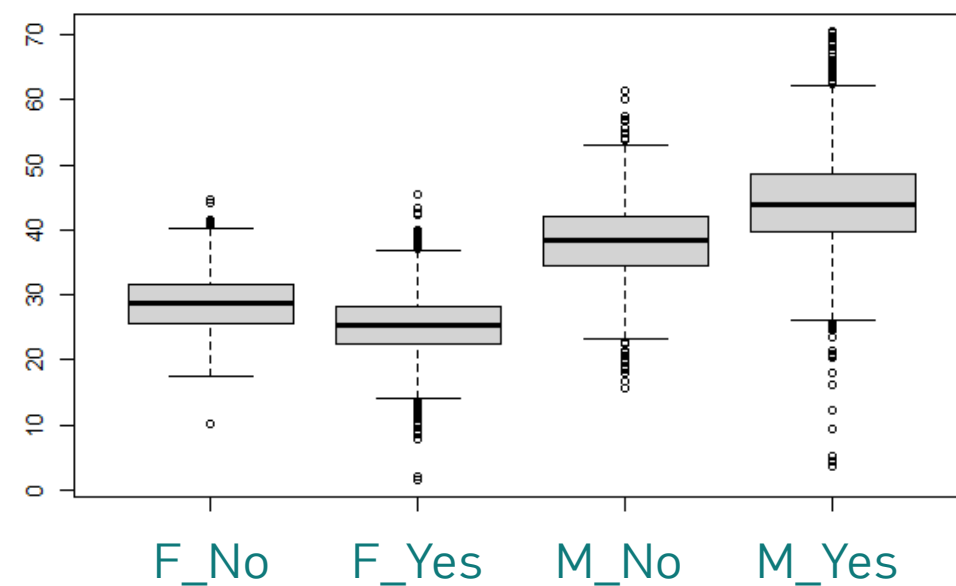
Reject H_{0c}

유의 수준 : 0.05

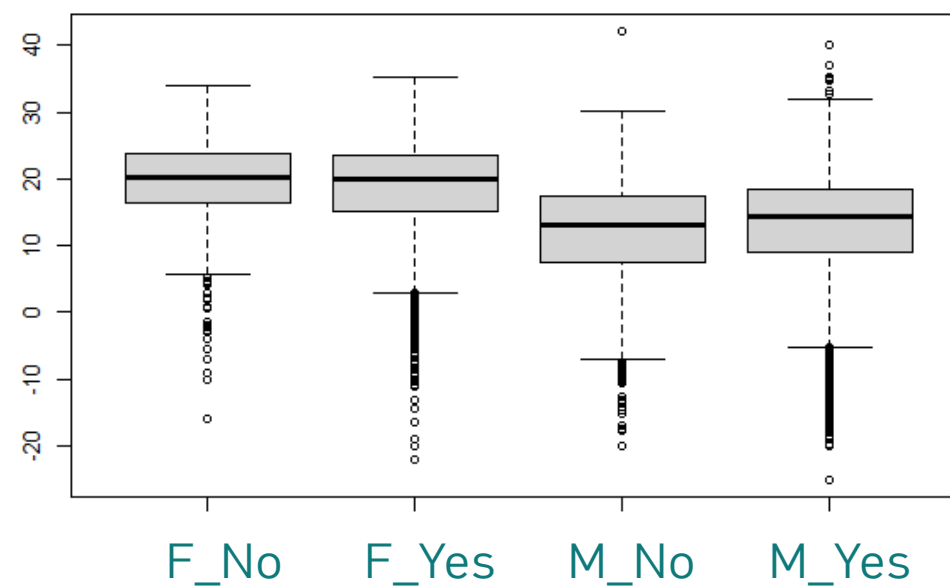
06 결론

검정 결과 시각화

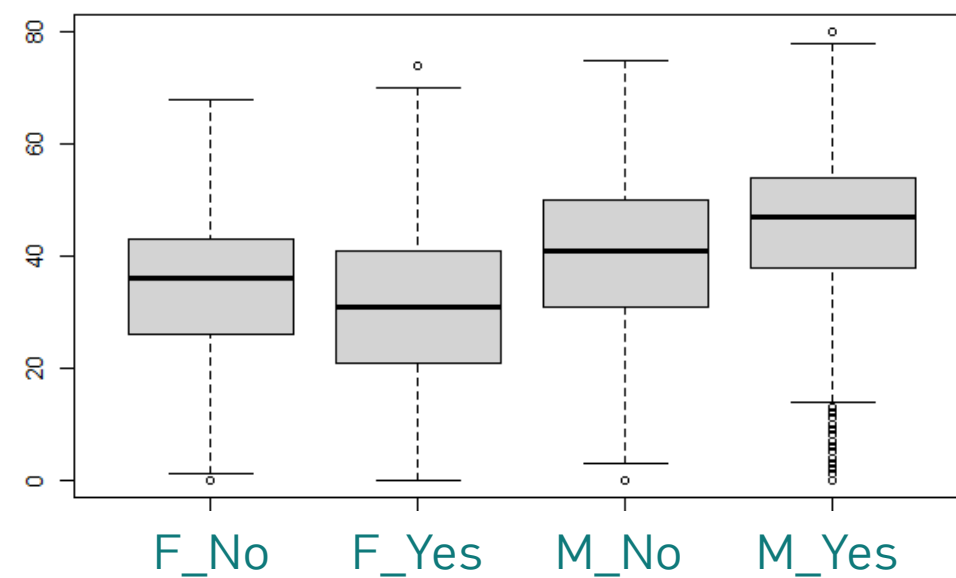
악력 BoxPlot



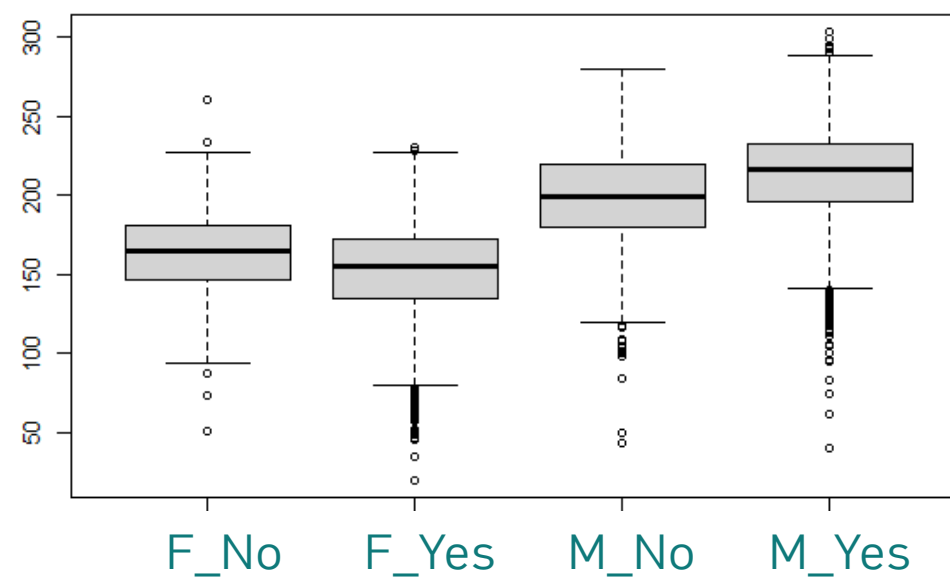
앉아 윗몸 앞으로 굽히기 BoxPlot



윗몸 일으키기 BoxPlot



제자리 멀리뛰기 BoxPlot



앉아 윗몸 앞으로 굽히기를 제외하면
유의미한 차이 존재

특히 악력은 눈에 보이는 큰 차이

06

결론 한계점

변수 종류 부족 : 신체 변수 - 허리 둘레, 발 길이 등..
운동 수행 변수 - 오래 달리기 등..

자료 측정의 신뢰성 부족 : 터무니없는 이상치 존재, 변수 측정 조건 명시 X

추가적으로 수집할 수 있는 다양한 정보 활용
외적 요인 고정 후, 자료 수집 및 분석

감사합니다