

연구입문



AI로 만드는 더 나은 스포츠
Center for Sports
Analytics·AI

박재현
2025.01.16.

지식 소비자
지식 생산자

·**지식**·이란 무엇인가?



지식 知識 Knowledge

사실지 know-what
존재지 know-why
방법지 know-how

기계가 쉽게 할 수 있는 일

input



복잡한 함수

단언컨대, 과학적 지식은 차이를 확인하는 것입니다.

output

기계가 어려워 하는 일

인간은 쉽게 할 수 있는 일



학위논문

초등학생의 태권도 수련이 인성발달 및 신체적 자신감에 미치는 영향 The Effects of Taekwondo Training for Human nature Development and Physical Confidence

첫째. 인구통계특성 및 수련정도에 따라 태권도 수련을 통한 인성발달에 차이가 있는가.

- 1-3. 수련여부에 따라 인성발달에 차이가 있을 것이다.
- 1-4. 수련빈도에 따라 인성발달에 차이가 있을 것이다.
- 1-5. 수련기간에 따라 인성발달에 차이가 있을 것이다.
- 1-6. 수련강도에 따라 인성발달에 차이가 있을 것이다.

지식생산의 방법을 익힌다는 것은...

차이를 식별할 수 있는 눈(도구)을 갖는 다는 것을 의미합니다.

단언컨대, 새로운 과학지식은 모두 차이를 확인하는 것입니다.



이 켤에서...

제가 여기저기 강의하다보면, 이켤에서 이런 질문 꼭 나오던데...

모든 과학연구가 차이만 확인하는 것이 아니던데요?

"A가 B에 미치는 영향"

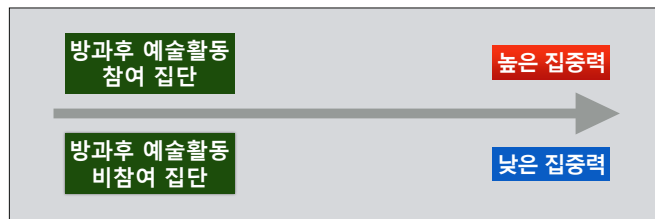
예) 초등학생의 신체활동이 비만에 미치는 영향

"A와 B의 관계"

예) 초등학생의 신체활동과 비만의 관계

연구가설의 기술에서 차이, 영향, 관계의 검증 논리는 동일

- 변수A에 따라 변수B는 차이가 있을 것이다.
- 변수A는 변수B에 영향을 미칠 것이다.
- 변수A는 변수B와 관계가 있을 것이다.



변수와 척도에 대한 개념을 정확히 알면 연구주제 설정에 엄청난 도움

집단구성원이 어떤 특성에 있어서 서로 다른 가치를 가지고 있을 때 그 특성을 변수

속성에 따른 변수구분(척도)

연속변수/불연속변수

(명명척도)

(서열척도)

(동간척도)

(비율척도)

기능에 따른 변수구분

독립변수

종속변수

(매개변수)

(조절변수)

(통제변수)

변수의 속성에 따른 분류

변수의 속성(명명, 서열, 동간, 비율)

성(Sex) = {남자, 여자} : 유한

불연속 변수

신장 = {cm | cm > 0} : 무한

연속 변수

변수의 기능에 따른 분류

- 독립변수 (Independent Variable : IV)
연구자의 통제 하에 있는 변수
- 종속변수 (Dependent Variable : DV)
독립변수 조작에 따른 효과변인

[문제]
앞의 연구에 대하여 다음의 질문에 답하세요.



독립변수는? 답:

종속변수는? 답:

연구가설은? 답:

연구설계



연구(research)란?

어떤 현상(사건, 문제 등)을 체계적으로 관찰(측정)하고 그 결과를 기술, 설명, 예언하기 위한 과정

연구
research



과학적 방법을
적용한 문제해결

과학적 방법
scientific method



형식화된 사고 또는
이론에 근거를 둔 문제해결

실험노트

말을 알아듣는 훈련된 바퀴벌레

바퀴다리가 청력에 미치는 영향

실험방법

바퀴다리를 하나씩 떼어내며
'출발'이라고 명령한다.



X 왜 과학적인 문제해결이 아닌가?

첫째, 형식화된 사고과정에 부합하지 않음.

(논리에 맞지 않음)

둘째, 이론에 기초를 두고 있지 않음

(어떤 메커니즘에 의해서 청력을 잃게 되는지에 대한 설명 불가능)

그래서 논리에 맞는, 말이 되는 연구결과 해석이 중요합니다.

[측정 관련 용어]

- 신뢰도(reliability)
- 타당도(validity)
- 조작화(operationalization)

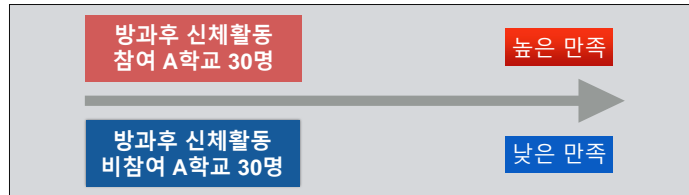
[연구설계 관련 용어]

- 내적타당도(internal validity)
- 외적타당도(external validity)
- 일반화(generalization)

방과 후 신체활동 참여가 학교생활 만족에 미치는 영향

방과 후 신체활동

학교생활 만족



[결론]

-방과 후 신체활동 프로그램 참가는 우리나라 중학생들의 학교생활 만족도를 높인다.

[논쟁 사항]

-A학교 결과를 우리나라 중학생의 결과로 해석할 수 있는지 여부 (일반화 오류: 외적타당도 문제)
-학교생활에 만족하는 학생들이 방과 후 신체활동 프로그램에 참가하였을 가능성 여부 (인과관계 잘못 설정: 내적타당도 문제)

방과 후 신체활동 참여가 학교생활 만족에 미치는 영향

진실험설계(true experimental design)

[통제집단 사전-사후 측정 설계]

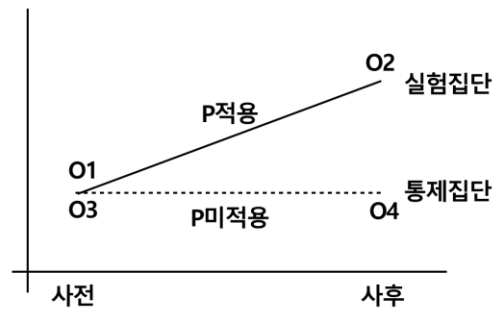
		사전 측정	처치	사후 측정	효과
무선 배정	실험집단	O1	P	O2	O2-O1
	통제집단	O3	-	O4	O4-O3

P: 방과 후 신체활동 프로그램 적용
독립변수 효과: (O2-O1) - (O4-O3)

방과 후 신체활동 참여가 학교생활 만족에 미치는 영향

진실험설계(true experimental design)

[통제집단 사전-사후 설계]



P: 방과 후 신체활동 프로그램

방과 후 신체활동 참여가 학교생활 만족에 미치는 영향

진실험설계(true experimental design)

[통제집단 사후 측정 설계]

		처치	사후 측정
무선 배정	실험집단	P	O1
	통제집단	-	O2

P: 방과 후 신체활동 프로그램 적용
독립변수 효과: O1-O2

방과 후 신체활동 참여가 학교생활 만족에 미치는 영향

진실험설계(true experimental design)

[솔로몬 4집단 측정 설계]

무선 배정		사전 측정	처치	사후 측정	효과
	실험집단	O1	P	O2	O2-O1
	통제집단	O3	-	O4	O4-O3
	실험집단	-	P	O5	$O5-(O1+O3)/2$
	통제집단	-	-	O6	$O6-(O1+O3)/2$

P: 방과 후 신체활동 프로그램 적용

방과 후 신체활동 참여가 학교생활 만족에 미치는 영향

진실험설계(true experimental design)

[플라시보 통제집단 사후 측정 설계]

무선 배정		처치	사후 측정
	실험집단	P	O1
	통제집단	-	O2
	플라시보집단	xP	

P: 방과 후 신체활동 프로그램 적용

xP: 플라시보 프로그램(예: 미술 프로그램)

방과 후 신체활동 참여가 학교생활 만족에 미치는 영향

유사 실험설계(quasi experimental design)

[비 동등 통제집단 사전-사후 측정 설계]

임의 배정		사전 측정	처치	사후 측정	효과
	실험집단	O1	P	O2	O2-O1
	통제 집단	O3	-	O4	O4-O3

P: 방과 후 신체활동 프로그램 적용
독립변수 효과: (O2-O1)-(O4-O3)

방과 후 신체활동 참여가 학교생활 만족에 미치는 영향

유사 실험설계(quasi experimental design)

[비동등 통제집단 사후 측정 설계]

임의 배정		처치	사후 측정
	실험집단	P	O1
	통제 집단	-	O2

P: 방과 후 신체활동 프로그램 적용
독립변수 효과: O1-O2

* 시계열 설계, 반복측정 설계 등, 체육 및 스포츠과학 분야에서 자주 사용하는 연구설계에 익숙해야

통계설계



$$12 + 10 = ?$$

각 숫자 및 기호의 특성과 기능을 이해한다면,
답을 구하는 일은 식은 죽먹기~

변수 (Variable)

집단구성원이 어떤 특성에 있어서 서로 다른 가치를 가지고 있을 때 그 특성을 변수

우리 반 학생들의 성(gender)은 어떻게 측정?

성(Sex) = {남자, 여자}

변수

측정치(변수수준)

우리 반 학생들의 신장(height)은 어떻게 측정?

신장 = {cm | cm > 0}

변수

측정치(변수수준)

변수의 속성에 따른 분류

변수의 속성(명명, 서열, 동간, 비율)

성(Sex) = {남자, 여자} : 유한

불연속 변수

신장 = {cm | cm > 0} : 무한

연속 변수

변수의 기능에 따른 분류

- 독립변수 (Independent Variable : IV)
연구자의 통제 하에 있는 변수
- 종속변수 (Dependent Variable : DV)
독립변수 조작에 따른 효과변인

Q

노인의 신체활동 참여에 따른 삶의 질

1. 변수는?
2. 독립변수? 개수는? 하위수준은?
3. 종속변수? 개수는? 하위수준은?

Q

프로야구 경기결과별 홈구장 관중 수

1. 변수는?
2. 독립변수? 개수는? 하위수준은?
3. 종속변수? 개수는? 하위수준은?

경기결과의 측정치

(승, 패, 무승부)

1. 독립변수 개수는?
2. 독립변수 하위수준의 개수는?

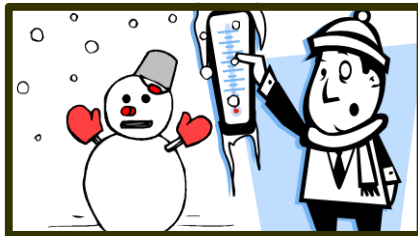
홈구장 관중 수

0명부터 무한대

1. 종속변수 개수는?
2. 종속변수 하위수준의 개수는?

변수통제

홈구장 관중 수를
날씨가 상이한 날 측정한 경우



패한날 날씨

변수통제

홈구장 관중 수를
날씨가 상이한 날 측정할 경우

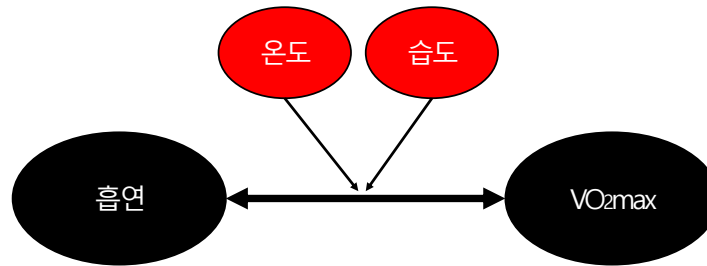


승리한날 날씨

변수통제 방법

- 실험적 통제
- 통계적 통제

외적변수의 통제방법 : 실험처치



온도와 습도를 동일한 조건으로 검사 실시
(변수의 상수화)

통계처리방법 표 참고



AI융합 스포츠분석센터

AI로 만드는 더 나은 스포츠
Center for Sports
Analytics·AI

통계처리방법 선택

독립변수			종속변수			통제변수 및 반복측정	검정가설	통계처리 방법
변수속성	개수	하위수준	변수속성	개수	하위수준			
불연속	1개	2수준	연속	1개	연속수준	해당 없음	집단간 차이	T-test
연속	1개	2회 반복	반복 측정시 해당 없음			반복측정	시행간 차이	종속 T-test
불연속	1개	2수준이상	연속	1개	연속수준	해당 없음	집단간 차이	1원 ANOVA
불연속	2개	2수준이상	연속	1개	연속수준	해당 없음	집단간 차이	다원 ANOVA
연속	1개	2회 이상 반복	반복 측정시 해당 없음			반복측정	시행간 차이	반복측정 ANOVA
불연속	1개 이상	2수준이상	연속	1개	연속수준	통제필요	집단간 차이	ANCOVA
불연속	1개 이상	2수준이상	연속	2개 이상	연속수준	해당 없음	집단간 차이	MANOVA
불연속	1개 이상	2수준이상	연속	2개 이상	연속수준	통제필요	집단간 차이	MANCOVA
연속	1개	연속수준	연속	1개	연속수준	해당 없음	변인간 관계	상관(회귀)분석
연속	2개 이상	연속수준	연속	1개	연속수준	해당 없음	변인간 관계	다중회귀분석
연속	1개 이상	연속수준	연속	1개	연속수준	통제필요	변인간 관계	부분상관분석
연속	1개 이상	연속수준	불연속	1개	2수준이상	해당 없음	변인의 변별	판별분석
불연속	1개 이상	2수준이상	불연속	1개 이상	2수준이상	해당 없음	명목의 차이	χ^2 test
불연속	1개	명목 및 서열	불연속	1개	명목 및 서열	해당 없음	변인간 관계	서열 종류의 상관

통계분석 설계

• 흡연이 심폐지구력에 미치는 영향

- 독립변인은 흡연
(흡연무, 20개피 이하, 20개피 이상)
- 종속변인은
심폐지구력(VO_{2max})

연습

One-way ANOVA

상관 or 회귀 분석

- 흡연이 심폐지구력에 미치는 영향

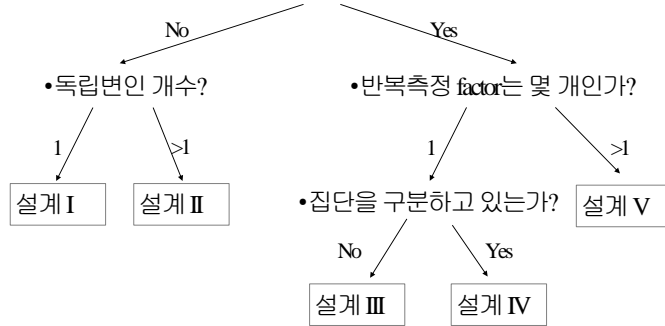
- 독립변인은 흡연
(니코틴 농도: 0mm~)
- 종속변인은
심폐지구력(VO_{2max})

📁 변량분석의 구분

- 종속변인 개수에 의한 구분
 - ❖ 1개(univariate) – 변량분석(ANOVA)
 - ❖ 2개 이상(multivariate) – 다변량분석(MANOVA)
- 독립변인 개수에 의한 구분
 - ❖ 1개(1-way) – 일원 변량분석(1-way ANOVA)
 - ❖ n개(n-way) – n원 변량분석(n-way ANOVA)
- 반복측정 포함
 - ❖ Repeated Measures ANOVA
- 공변인 조정(Covariance)
 - ❖ 공변인 조정이 필요 – 공변량 분석
(Analysis of Covariance)

변량분석설계 선택하기

• 동일 대상자를 반복하여 측정하였는가?



변량분석의 종류

Independent Groups ANOVA

• One-way ANOVA

설계 I

• Factorial ANOVA

설계 II

Repeated Measures ANOVA

• One factor repeated measures ANOVA

설계 III

• Two factor mixed measures ANOVA

설계 IV

• Two factor repeated measures ANOVA

설계 V

with covariate ANOVA

XXX factor (repeated measures) ANCOVA

설계 VI

초등학생의 통학형태별 신체활동량

통학형태

도보
버스/학원차
승용차

신체활동량

도보수

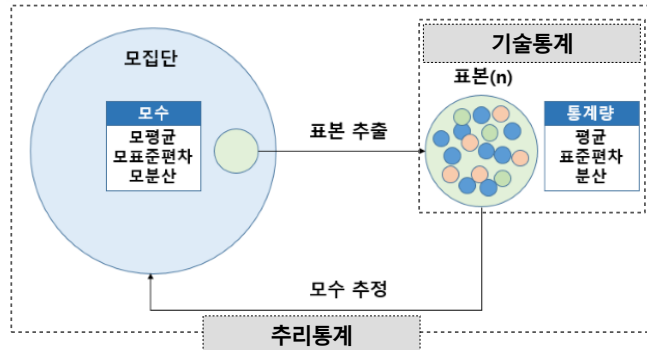
-
- 일원변량분석(one-way ANOVA)

가설검증



AI로 만드는 더 나은 스포츠
Center for Sports
Analytics·AI

✕ 모집단과 표본집단의 관계



- 연구결과, 표본집단에서 나타나는 차이는 연구자의 궁극적 관심 대상이 아님.
- 연구자는 독립변수에 따른 종속변수의 차이(영향, 관계)가 있는지 여부에 대한 관심은 표본집단이 아닌 모집단에 두어야 함.

[대상선택 관련 용어]

- 무선표집(random sampling)
- 무선배정(random assignment)

✕ 뚝배기 된장국과 어머니 장맛



된장국 맛 추정의 오류

오류의 종류

		독배기 속 된장국(모집단 특성)	
		맛 있음	맛 없음
숫가락 속 된장국 (수집된 자료의 결과)	맛 있음	해피 식사	망함
	맛 없음	망함	해피 식사



이 연구결과에 따르면 치매환자는 후각기능이 떨어진다고 하네요.



독립변수는?

답:

종속변수는?

답:

연구가설은?

답:

이 주장이 사실인지를 확인하기 위하여
200명 대상(치매환자 100명, 정상인 100명)의 연구수행

10cm 거리에서 냄새 식별가능 여부

	가능	불가능
치매환자 (100명)	49 명	51 명
정상인 (100명)	51 명	49 명

연구결론

치매환자와 정상인 간의 후각기능은 차이가 있다.

가설검증의 오류

가설을 검증하는 연구자의 결정에는 오류가 있을 수 있음

오류의 종류

		연구문제의 실제상황	
		$A=B(H_0 \text{ 참})$	$A \neq B(H_0 \text{ 거짓})$
연구자 결정	$A=B(H_0 \text{ 참})$	정확한 결정	제2종 오류 β
	$A \neq B(H_0 \text{ 거짓})$	제1종 오류 α	정확한 결정 $1-\beta$

α 는 영가설을 부정할 경우 연구자의 결정이 잘못된 결정일 확률
-알파를 유의도 수준이라고도 합니다.

연구가설의 채택과 기각(연습 1)

- 유의도 수준(α) = .05

차이가 있다는 결정을 내렸을 경우 그 결정이
잘못된 결정일 가능성(확률)을 5%까지 허용

- 자료분석 결과 얻어진 확률치

$p\text{-value}(\text{sig}) = .086$

차이가 있다는 결정을 내렸을 경우 그 결정이
잘못된 결정일 확률이 8.6%임으로 허용범위를 벗어남

연구입문



AI로 만드는 더 나은 스포츠
Center for Sports
Analytics·AI

박재현
2025.01.16.