고고학 자료 통계분석

Week 3 : 기초 톰계(2)

숨실대학교 사학과 석사과정 1학기 주 찬 혁

계획

주차	제목	내용		
1	Intro	소개, R 설치		
2	<u> 기초 통계(1)</u>	모집단과 표본, 기술통계량		
3	기초 통계(2)	변수의 종류, 가설과 검정, 오류, 분석절차		
4	전처리	데이터 전처리		
5	시각화	다양한 종류의 그래프		
6	검정	t-검정, Chi-square 검정, Median Polish		
7	회귀분석	선형회귀, 다중선형회귀, 로지스틱회귀		
8	군집분석	K-means,		
9	판별분석	DA, MDA		
10	주성분분석	PCA		

복습

- 모집단과 표본이 무엇인지 안다.
- 고고학적 맥락을 살피면서 통계분석을 해야 한다.
- 기술통계량이 무엇인지 안다.
- R로 기술통계량을 산출할 수 있다.

변수의 종류

- 수량적 특성에 따른 분류
 - 질적 변수, 양적 변수
- 자료의 특성에 따른 분류
 - 범주형 변수, 연속형 변수
- 자료의 상관관계에 따른 분류
 - 독립변수. 종속변수

*변수의 종류에 따라 데이터 수집, 분석 방법 등이 설정됨

수량적 특성에 따른 분류

• 질적 변수, 양적 변수

질적 변수	서열 질적 변수 (Ordered-Qualitative Variable)	
(Qualitative Variable)	비서열 질적 변수 (Unordered-Qualitative Variable)	
양적 변수	연속형 변수 (Continuous Variable)	
(Quantitative Variable)	비연속형 변수 (UnContinuous Variable)	

질적 변수

서열 질적 변수

조사요원별 자격기준 조사단장 책임조사원 조사원 준조사원 보조원

6세기 이후 신라 묘제
횡혈식석실분
석관묘, 석곽묘

학위
박사
석사
학사

신라 골품제 성골 진골 6두품 ...

구성 요소 사이에 위계(서열)가 존재

비서열 질적 변수

성별			
남성	여성		

삼국			
고구려	백제	신라	

반려동물 종류					
개	고양이	앵무새	뱀	물고기	거북이

석기 종류				
주먹도끼	긁개	찌르개	•••	

모든 구성요소가 평등

양적 변수

연속형 변수

키
190cm ~
180~189cm
170~179cm
160~169cm
150~159cm
•••

나이
60대 ~
50대
40대
30대
20대
10대

수로 표현되고 구성요소가 연속됨 *순서는 위계가 아님

비연속형 변수

손가락 개수 10개

아파트 층수
15층

A유적의 유물 출토량					
토기 석기 청동기 철기 불상 인골					
10개	5개	1개	1개	2기	1구

정수로 표현되고 셀 수 있음

자료의 특성에 따른 분류

• 범주형 변수, 연속형 변수

범주형 변수=이산형 변수 (Categorical Variable = Discrete	명목 변수 (Nominal Variable)				
Variable = Discrete	순위 변수 (Ordinal Variable)				
연속형 변수	간격 변수 (Interval Variable)				
(Continuous Variable)	비율 변수 (Ratio Variable)				

범주형 변수

질적변수, 비연속형 변수 포함 / 모든 구성요소는 구분됨

명목형 변수

토기 돌	대 유무
있음	없음

주거지 형태								
장방형	방형	원형						

형식분류								
1	2	3	4					

*숫자는 크기의 성질이 없음

A유적 출토 유물										
토기	석기	청동기	철기	불상	인골					

묘제								
석실묘	석곽묘	석관묘	목곽묘	목관묘	•••			

구성요소가 범주에 속해지고 이름이 부여됨

순위 변수

성적
Α
В
С
D
F

학위
박사
석사
학사

6세기 이후 신라 묘제 횡혈식석실분 석관묘, 석곽묘

명목형 변수에 위계(순위)를 부여하여 순위 변수로 사용할 수 있음(반대도 O)

연속형 변수

비율 변수

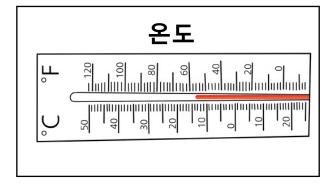
7
190cm ~
180~189cm
170~179cm
160~169cm
150~159cm
•••

나이	
60대 ~	
50대	
40대	
30대	
20대	
10대	

0과의 비율이 의미 있음

Tip: 0은 절대적

간격 변수



0과의 비율이 의미 없음

Tip: 0은 인간이 임의적으로 부여한 존재

자료의 상관관계에 따른 분류

• 독립변수, 종속변수

독립변수(Independent Variable)

종속변수(Dependent Variable)

자료의 상관관계에 따른 분류

$$y = a + bx$$

x -> 독립변수 / y -> 종속변수

x에 b만큼 변화가 가해지면 y는 그 영향으로 인해 변함

Feature Number	Excavation istitute	ВР	Error	From	То	Median	Туре
Proto-three Kingdoms~Thre e Kingdoms 4	Hanbaek Cultural Property Research Institute	1730	40	220	405	309	마한
IUI-A-26	Hanbaek Cultural Property Research Institute	1590	30	406	542	480	마한
IUI-A-28	Hanbaek Cultural Property Research Institute	1650 -	H ³⁰ 人	264	533	398	마한
IUI-A-36	Hanbaek Cultural Property Research Institute	1630	30	346	536	418	마한
I⊔t-A-63	Hanbaek Cultural Property Research Institute	1700	30	253	406	348	마한
IUF-A Pottery	Hanbaek Cultural Property Research Institute	1860	40	64	243	156	마한

독립변수: 다른 변수에 영향을 주는 변수

- 연구자가 조절하는 대상

종속변수: 다른 변수에게 영향을 받는 변수

- 연구자가 알고 싶은 대상

검토 보기 🗘 입력하세요

출처 : 박지영(2017) 편집 🕬 🕬

										<u> </u>				
E391 💂	E391 $\stackrel{\blacktriangle}{\blacktriangledown}$ \times \checkmark f_{X} Charcoal													
A	В	С	D	Е	F	G	Н	1	J	К	L	М	N	
1 X	Y	Region	No.*	material	Site	Feature Number	Excavation istitute	BP ▼	Error	From	То	Median	Type	4
2 127.098028E	37.19868N	Gyounggi-Gangwor	7	N/A	Cheonggye-ri, Hwaseong	Proto-three Kingdoms~Thr	e Hanbaek Cultural Property Research Institute	1730	40	220	405	309	마한	
3 127.098028E	37.19868N	Gyounggi-Gangwor	7	N/A	Cheonggye-ri, Hwaseong	□f-A-26	Hanbaek Cultural Property Research Institute	1590	30	406	542	480	마한	
4 127.098028E	37.19868N	Gyounggi-Gangwor	7	N/A	Cheonggye-ri, Hwaseong	마-A-28	Hanbaek Cultural Property Research Institute	1650	30	264	533	398	마한	1
5 127.098028E	37.19868N	Gyounggi-Gangwor	7	N/A	Cheonggye-ri, Hwaseong	□l-A-36	Hanbaek Cultural Property Research Institute	1630	30	346	536	418	마한	
6 127.098028E	37.19868N	Gyounggi-Gangwor	7	N/A	Cheonggye-ri, Hwaseong	□l-A-63	Hanbaek Cultural Property Research Institute	1700	30	253	406	348	마한	
7 127.098028E	37.19868N	Gyounggi-Gangwor	7	N/A	Cheonggye-ri, Hwaseong	마-A Pottery	Hanbaek Cultural Property Research Institute	1860	40	64	243	156	마한	
8 128.527565E	37.271022N	Gyounggi-Gangwor	13	Charcoal	Munsan-ri, Yeongwol	Proto-three Kingdoms	Gangwon Cultural Property Research Institute	1843	29	86	240	173	중도	
9 128.507406E	37.210748N	Gyounggi-Gangwor	14	Charcoal	Samok-ri, Yeongwol	1	Yemaek Cultural Property Research Institute	1895	20	57	208	103	중도	
10 128.507406E	37.210748N	Gyounggi-Gangwor	14	Charcoal	Samok-ri, Yeongwol	2	Yemaek Cultural Property Research Institute	1850	25	86	235	168	중도	
11 128.507406E	37.210748N	Gyounggi-Gangwor	14	Charcoal	Samok-ri, Yeongwol	3	Yemaek Cultural Property Research Institute	1880	20	71	214	115	중도	, T
12 128.507406E	37.210748N	Gyounggi-Gangwor	14	Charcoal	Samok-ri, Yeongwol	4	Yemaek Cultural Property Research Institute	1885	25	65	214	113	중도	
13 128.507406E	37.210748N	Gyounggi-Gangwor	14	Charcoal	Samok-ri, Yeongwol	5	Yemaek Cultural Property Research Institute	1840	20	125	238	176	중도	7
14 128.507406E	37.210748N	Gyounggi-Gangwor	14	Charcoal	Samok-ri, Yeongwol	6	Yemaek Cultural Property Research Institute	1840	20	125	238	176	중도	
15 128.27063E	37.268031N	Gyounggi-Gangwor	15	Charcoal	Jucheon-ri, Yeongwol	Proto-three Kingdoms 2	Yemaek Cultural Property Research Institute	1920	40	-19	214	83	중도	
6 128.27063E	37.268031N	Gyounggi-Gangwor	15	Charcoal	Jucheon-ri, Yeongwol	Proto-three Kingdoms 3	Yemaek Cultural Property Research Institute	1920	40	-19	214	83	중도	
7 128.27063E	37.268031N	Gyounggi-Gangwor	15	Charcoal	Jucheon-ri, Yeongwol	Proto-three Kingdoms 4	Yemaek Cultural Property Research Institute	1886	29	61	217	115	중도	'
8 128.265477E	37.270642N	Gyounggi-Gangwor	16	Charcoal	Jucheon-ri Sulsam, Yeongwol	1	Gangwon Archaeological Culture Research Institute	1710	50	172	428	330	중도	
128.265477E	37.270642N	Gyounggi-Gangwor	16	Charcoal	Jucheon-ri Sulsam, Yeongwol	7	Gangwon Archaeological Culture Research Institute	1770	60	93	398	262	중도	
0 128.265477E	37.270642N	Gyounggi-Gangwor	16	Charcoal	Jucheon-ri Sulsam, Yeongwol	8	Gangwon Archaeological Culture Research Institute	1690	50	231	532	351	중도	
1 128.492648E	37.14987N	Gyounggi-Gangwor	17	Charcoal	Palgoe-ri, Yeongwol	Proto-three Kingdoms pit I	nc Gangwon Cultural Property Research Institute	1720	36	240	398	323	중도	'
22 128.492648E	37.14987N	Gyounggi-Gangwor	17	Charcoal	Palgoe-ri, Yeongwol	Silla pit house	Gangwon Cultural Property Research Institute	1570	60	354	613	487	중도	
23 127.944313E	37.3848N	Gyounggi-Gangwor	18	Charcoal	Gahyeon-dong, Wonju	10	Gangwon Cultural Property Research Institute	1820	29	92	320	190	중도	
24 127.944313E	37.3848N	Gyounggi-Gangwor	18	Charcoal	Gahyeon-dong, Wonju	13	Gangwon Cultural Property Research Institute	1730	40	220	405	309	중도	
25 127.944313E	37.3848N	Gyounggi-Gangwor	18	Charcoal	Gahyeon-dong, Wonju	16	Gangwon Cultural Property Research Institute	1750	29	226	383	293	중도	
26 127.944313E	37.3848N	Gyounggi-Gangwor	18	Charcoal	Gahyeon-dong, Wonju	24	Gangwon Cultural Property Research Institute	1776	32	136	340	259	중도	
27 127.944313E	37.3848N	Gyounggi-Gangwor	18	Charcoal	Gahyeon-dong, Wonju	26	Gangwon Cultural Property Research Institute	1790	29	133	330	236	중도	
28 127.944313E	37.3848N	Gyounggi-Gangwor	18	Charcoal	Gahyeon-dong, Wonju	37	Gangwon Cultural Property Research Institute	1850	40	68	251	166	중도	
29 127.944313E	37.3848N	Gyounggi-Gangwor	18	Charcoal	Gahyeon-dong, Wonju	43	Gangwon Cultural Property Research Institute	1790	50	92	381	236	중도	
30 127.944313E	37.3848N	Gyounggi-Gangwor	18	Charcoal	Gahyeon-dong, Wonju	4	Gangwon Cultural Property Research Institute	1770	29	139	344	278	중도	
31 127.826223E	37.338016N	Gyounggi-Gangwor	19	Charcoal	Donghwa-ri, Wonju	1	Hangang Cultural Property Research Institute	1915	36	4	212	89	중도	
32 127.826223E	37.338016N	Gyounggi-Gangwor	19	Charcoal	Donghwa-ri, Wonju	3	Hangang Cultural Property Research Institute	1920	36	1	211	83	중도	· ·
33 127.826223E	37.338016N	Gyounggi-Gangwor	19	Charcoal	Donghwa-ri, Wonju	5	Hangang Cultural Property Research Institute	1970	36	-46	122	30	중도	
34 128.638004E	37.209188N	Gyounggi-Gangwor	20	Charcoal	Yemi-ri, Jeongseon	10	Gangwon Cultural Property Research Institute	1690	50	231	532	351	중도	
35 128.638004E	37.209188N	Gyounggi-Gangwor	20	Charcoal	Yemi-ri, Jeongseon	6	Gangwon Cultural Property Research Institute	1710	50	172	428	330	중도	
36 128.638004E	37.209188N	Gyounggi-Gangwor	20	Charcoal	Yemi-ri, Jeongseon	8	Gangwon Cultural Property Research Institute	1770	60	93	398	262	중도	
37 127.726466E	37.756309N	Gyounggi-Gangwor	21	Charcoal	Gunja-ri, Chuncheon	1	Yemaek Cultural Property Research Institute	1802	13	136	311	220	중도	
38 127.726466E	37.756309N	Gyounggi-Gangwor	21	Charcoal	Gunja-ri, Chuncheon	2	Yemaek Cultural Property Research Institute	1758	15	239	333	292	중도	
39 127.721289E	37.891914N	Gyounggi-Gangwor	22	Charcoal	Geunhwa-dong, Chuncheon	11	Gangwon Archaeological Culture Research Institute	2260	36	-400	-206	-289	중도	
10 127.721289E	37.891914N	Gyounggi-Gangwor	22	Charcoal	Geunhwa-dong, Chuncheon	12	Gangwon Archaeological Culture Research Institute	1640	60	253	550	412	중도	
41 127.721289E	37.891914N	Gyounggi-Gangwor	22	Charcoal	Geunhwa-dong, Chuncheon	18	Gangwon Archaeological Culture Research Institute	2190	36	-373	-168	-280	중도	1
_						To c		7 2452		262				

Gangwon Archaeological Culture Research Institute

Yemaek Cultural Property Research Institute

Gangwon Cultural Property Research Institute

Museum of Hanrim University

Museum of Hanrim University

2162

1840

1792

1984

1890

1670

1879

1810

1821

1780

41

20

32

27

40

40

39

43

32

40

-362

125

132

-43

252

55

87

131

26

Udu II-1

Udu II-3

Udu park-3

Udu park-4

Udu Ⅲ-1

Udu Ⅲ-3

37.891914N

37.891914N

37.867458N

37.922077N

37.913389N

37.913389N

37.913389N

37.913389N

37.913389N

22

Gyounggi-Gangwor

Gyounggi-Gangwor

Gyounggi-Gangwor

Gyounggi-Gangwor

Gyounggi-Gangwor Gyounggi-Gangwor

Gyounggi-Gangwor

Gyounggi-Gangwor

Charcoal

Geunhwa-dong, Chuncheon

Geunhwa-dong, Chuncheon

Sinmae-ri, Chuncheon

Udu-dong, Chuncheon

Udu-dong, Chuncheon

Udu-dong, Chuncheon

Udu-dong, Chuncheon

Udu-dong, Chuncheon

Udu-dong, Chuncheon

Samcheon-dong, Chuncheon

홈 삽입 그리기 페이지 레이아웃 수식 데이터

-224

176

233

18

116

373

128

206

190

251

중도

-92

230

530

235

333

377

127.721289E

127.721289E

127.692716E

127.711551E

127.743126E

127.743126E

127.743126E

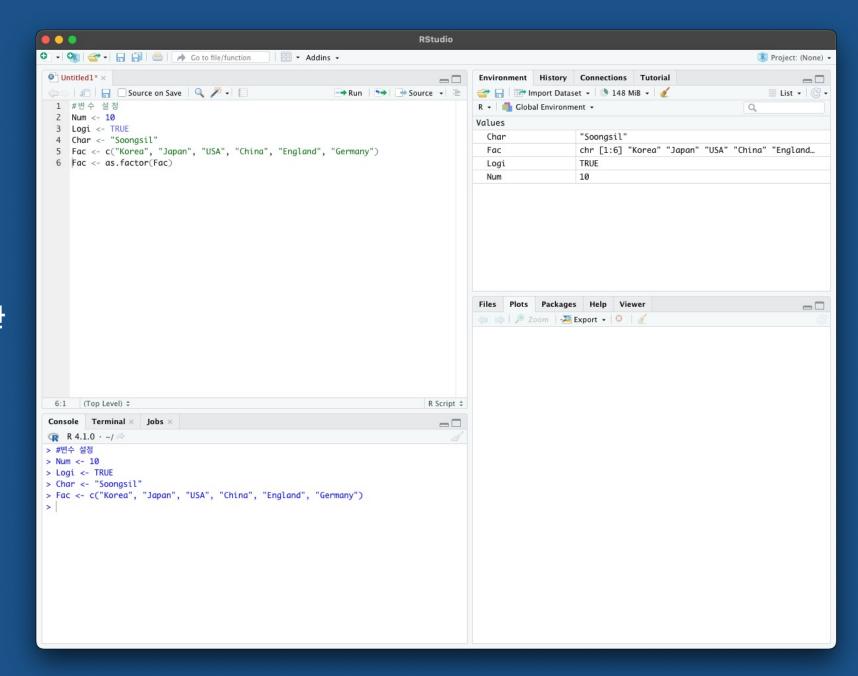
127.743126E

49 127.743126E

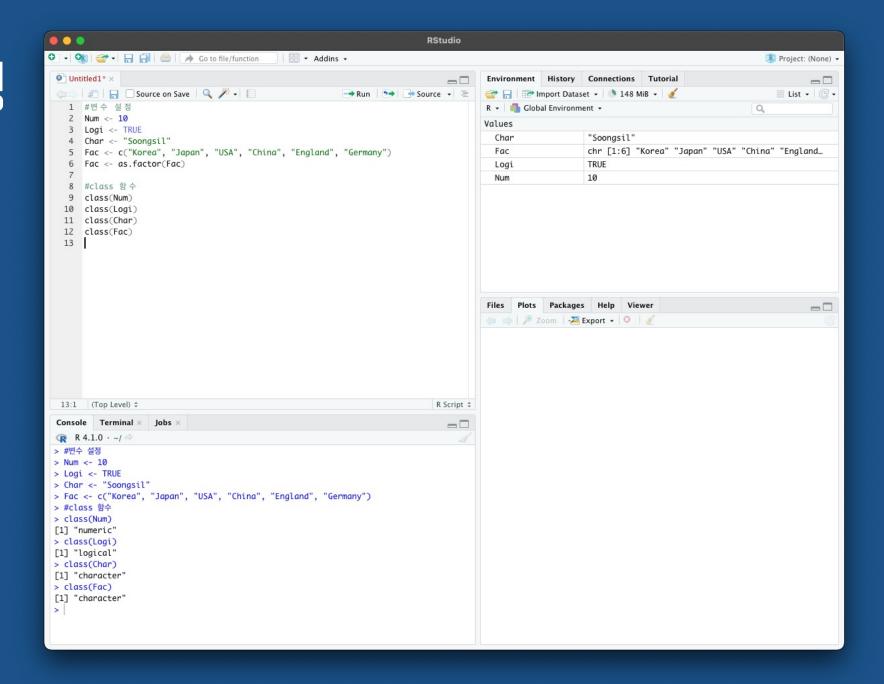
51 127.743126E

숫자형(Numeric)	논리형(Logical)
문자형(Character)	팩터형(Factor)

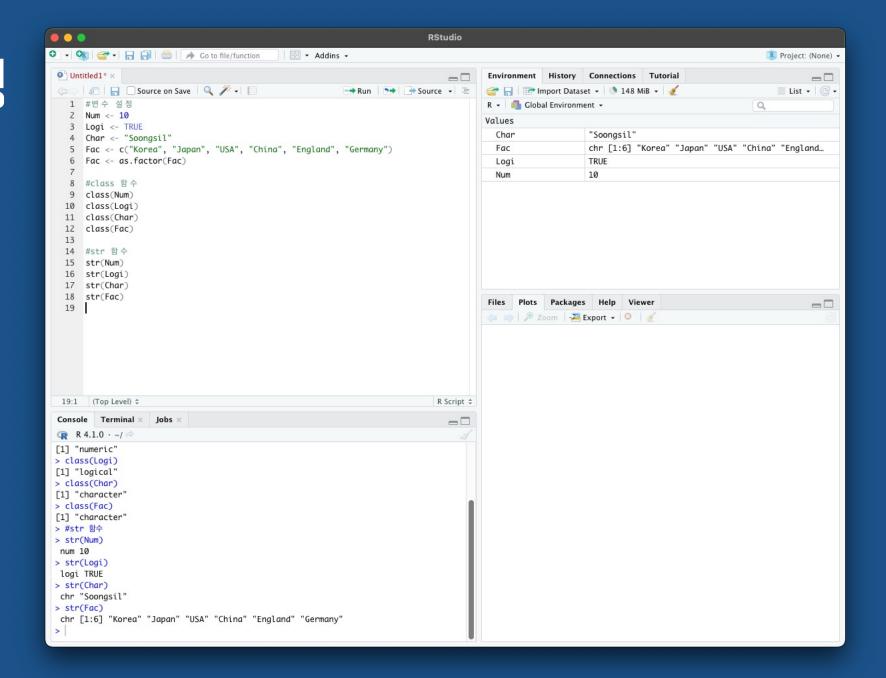
as.numeric(변수) as.logical(변수) as.character(변수명) as.facter(변수명) > as.ㅁㅁㅁ의 자료형으로 변환



*class : 자료형 출력



*str : 자료형과 구성요소 출력



가설과 검정 그리고 오류

- 가설
 - 귀무가설(영가설, H_0)
 - 대립가설(대안가설, H_1)
- 검정
 - 채택(Accept)할 것인가?, 기각(Reject)할 것인가?
- 오류
 - 1종 오류
 - 2종 오류

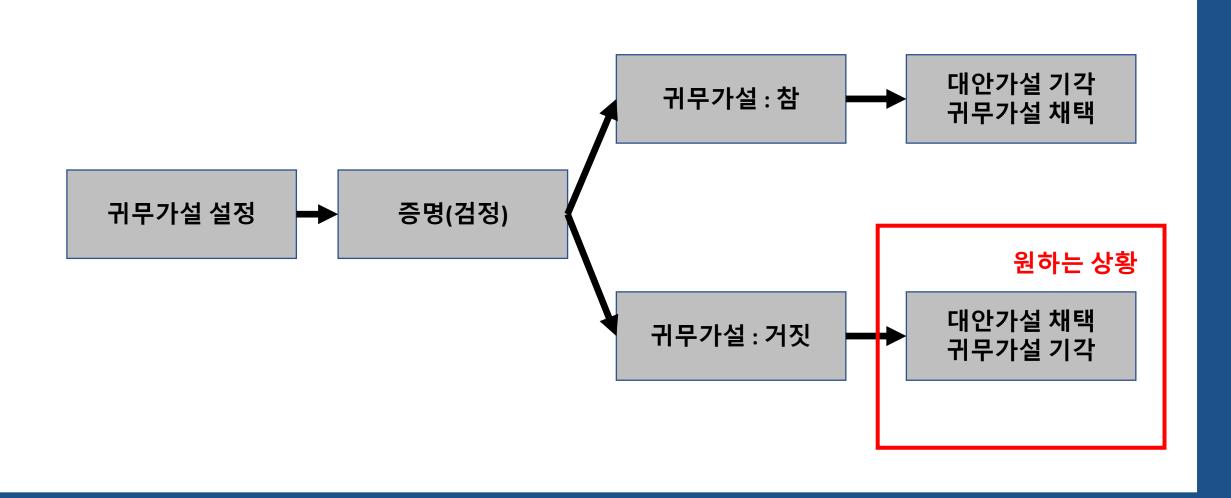
가설과 검정

• 모든 것은 참 아니면 거짓이라는 전제

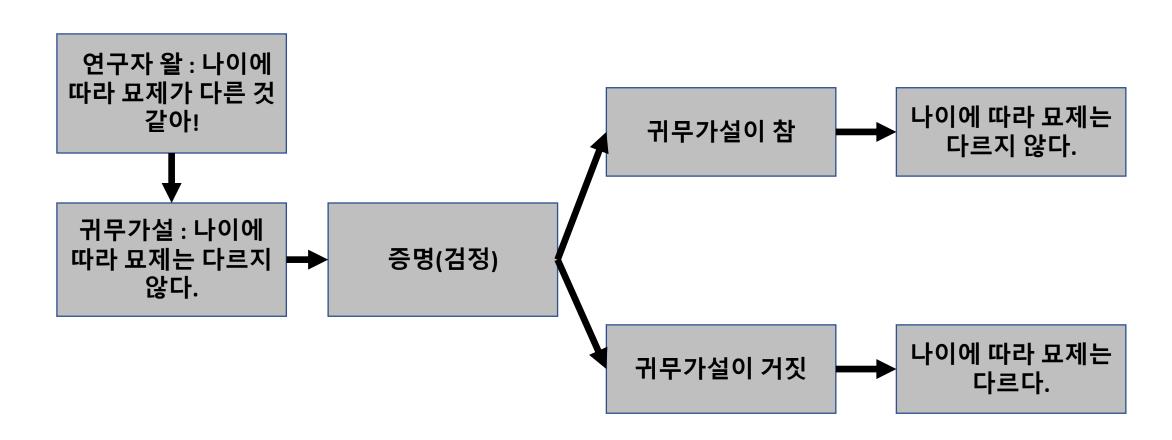
• 한 명제에 대한 가설을 세우고, 이 가설이 거짓임을 증명하여 명제가 참임을 증명(검정) -> 다양한 검정 방법은 6주차에...

• 여기서 처음 세워지는 가설이 귀무가설(영가설), 이에 반대되는 가설이 대립가설(대안가설)

가설과 검정



가설과 검정



오류

• 어느 가설 검정도 100% 확실하지 않음에 따라 오류가 발생

• 1종 오류: 귀무가설이 참인데 기각

• 2종 오류: 귀무가설이 거짓인데 채택

오류

*1종 오류와 2종 오류의 가능성은 항상 존재함 어떤 오류를 줄일지는 상황에 따라 판단

연구 주제 : 나이에 따라 묘제가 다르다.

- 귀무가설 : 나이에 따라 묘제가 다르지 않다.

- 대립가설 : 나이에 따라 묘제가 다르다.

1종 오류

- 실제 : 나이에 따라 묘제가 다르지 않음

- 검정결과 : 나이에 따라 묘제가 다름

-> 없는 현상을 있다고 하는 것

2종 오류

- 실제 : 나이에 따라 묘제가 다름

- 검정결과 : 나이에 따라 묘제가 다르지 않음

-> 있는 현상을 없다고 하는 것

오류

출처: https://support.minitab.com/

한 의료 분야 연구자가 두 약품의 효과를 비교하려고 합니다. 귀무 가설과 대립 가설은 다음과 같습니다.

- 귀무 가설(H₀): μ₁= μ₂
 두 약품의 효과가 동일합니다.
- 대립 가설(H₁): µ₁≠ µ₂
 두 약품의 효과가 동일하지 않습니다.

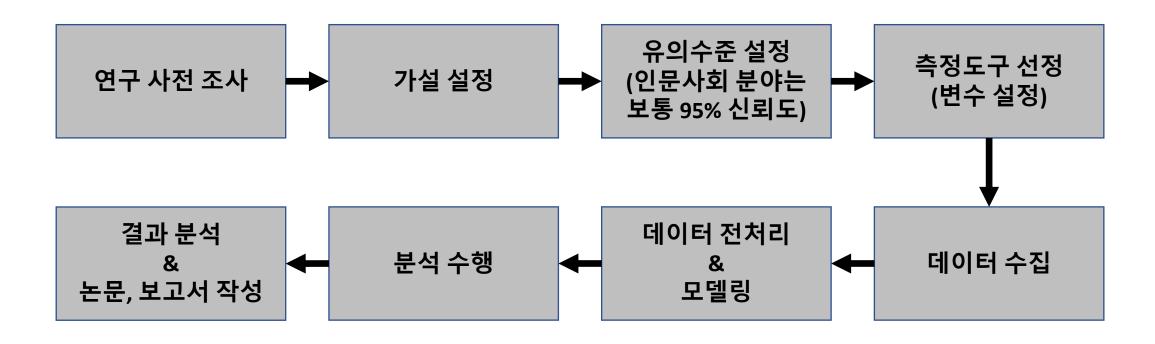
두 약품이 다르지 않지만 연구자가 귀무 가설을 기각하고 두 약품이 다르다는 결론을 내리는 경우 제1종 오류가 발생합니다. 약품의 효과가 동일한 경우, 환자는 어느 약품을 복용하거나 관계 없이 동일한 수준의 효과를 얻기 때문에 연구자는 이 오류가 매우 심각한 것으로 생각하지 않을 수도 있습니다. 그러나 제2종 오류가 발생하면 연구자는 귀무 가설을 기각해야 하지만 기각하지 못합니다. 즉, 연구자는 두 약품이 실제로는 다르지만 같다는 결론을 내리는 것입니다. 이 오류는 효과가 더 낮은 약품이 효과가 더 높은 약품 대신 판매될 경우 잠재적으로 생명을 위협할 수도 있습니다. *각 오류의 위험성을 인지하고 삼황에 맞게 적절히 조점해야함

분석 절차

- 확증적 데이터 분석
 - 추론통계 : 검점, 신뢰구간 추정 등...

- 탐색적 데이터 분석
 - 기술통계: 기술통계량, 데이터 분포 등...

확증적 데이터 분석(Confirmatory Data Analysis, CDA)



탐색적 데이터 분석(Exploratory Data Analysis, CDA)

