

제1장 통계와 통계학



목차

1.1 통계와 통계학

1.1.1 통계와 통계학의 정의

1.1.2 통계학의 분류

1.2 통계학의 활용 및 통계의 오남용

1.2.1 통계학의 활용

1.2.2 통계의 오남용

1.3 통계학의 역사

1.3.1 통계의 탄생

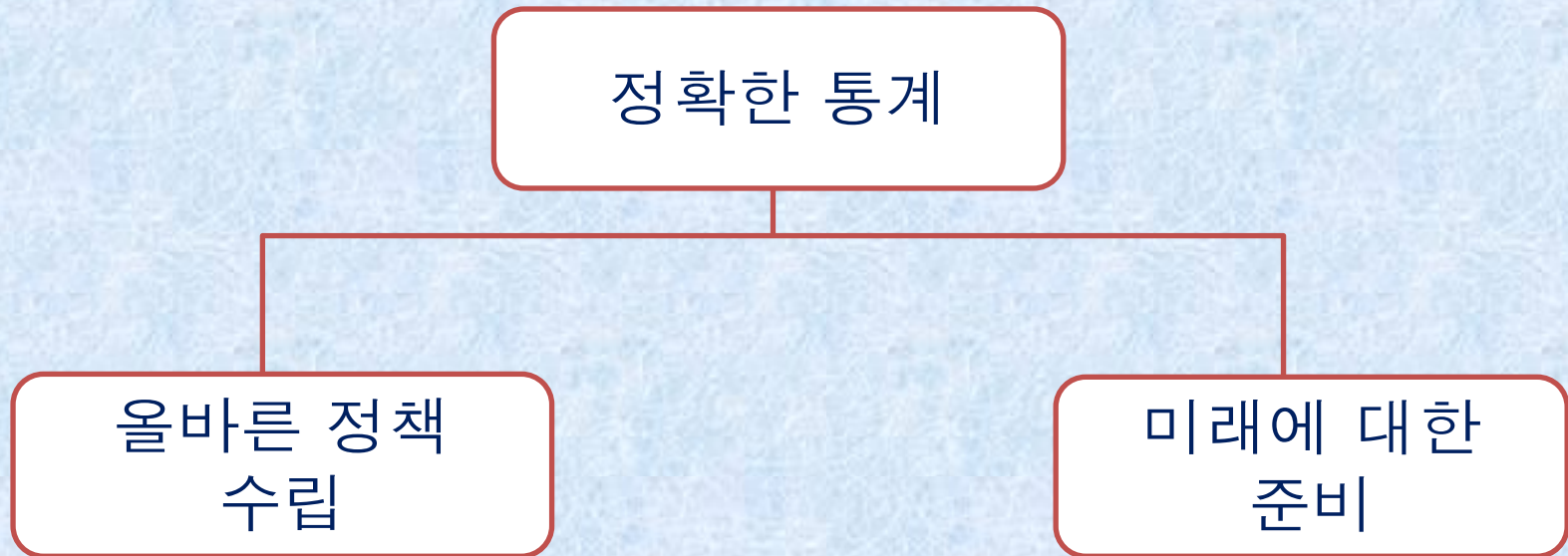
1.3.2 통계학의 발달

1.3.3 수리통계학의 정립

1.3.4 오늘날의 통계학

1.1 통계와 통계학

1.1.1 통계와 통계학의 정의



1.1 통계와 통계학

1.1.1 통계와 통계학의 정의

통계

사실이나 결과를 분석 처리한 수치적 정보

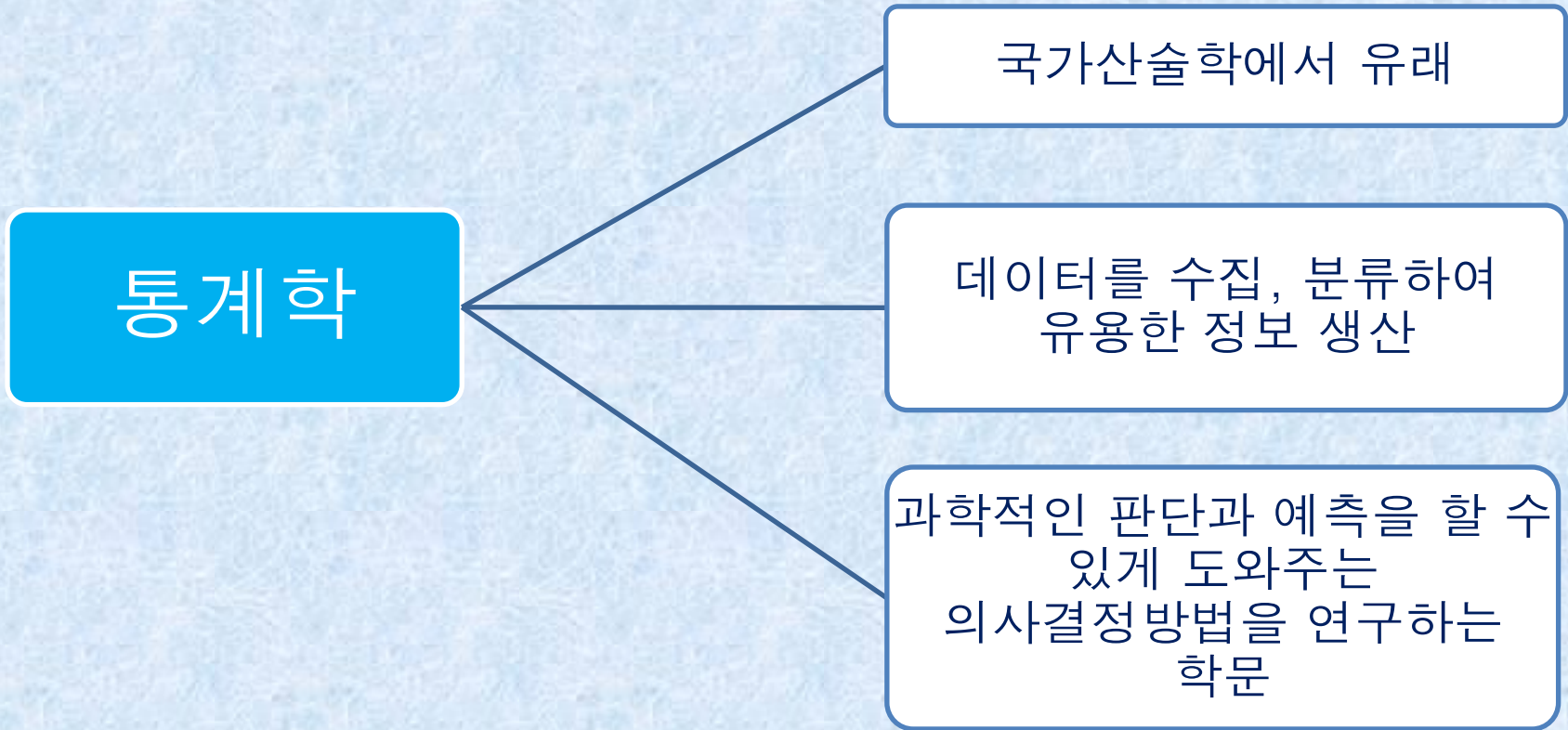
국가, 정부를 의미하는 라틴어의 '정부'에서 유래

자료에서 일차적 가공을 거쳐 생산되는 수량적인 요약 값들

소비자 물가지수, 주가지수, 실업률, 이혼율, 시청률, 지지율, 승률

1.1 통계와 통계학

1.1.1 통계와 통계학의 정의



1.1 통계와 통계학

1.1.1 통계와 통계학의 정의

통계



- 자료에서 일차적 분석, 요약을 통해 얻어지는 수량적 정보

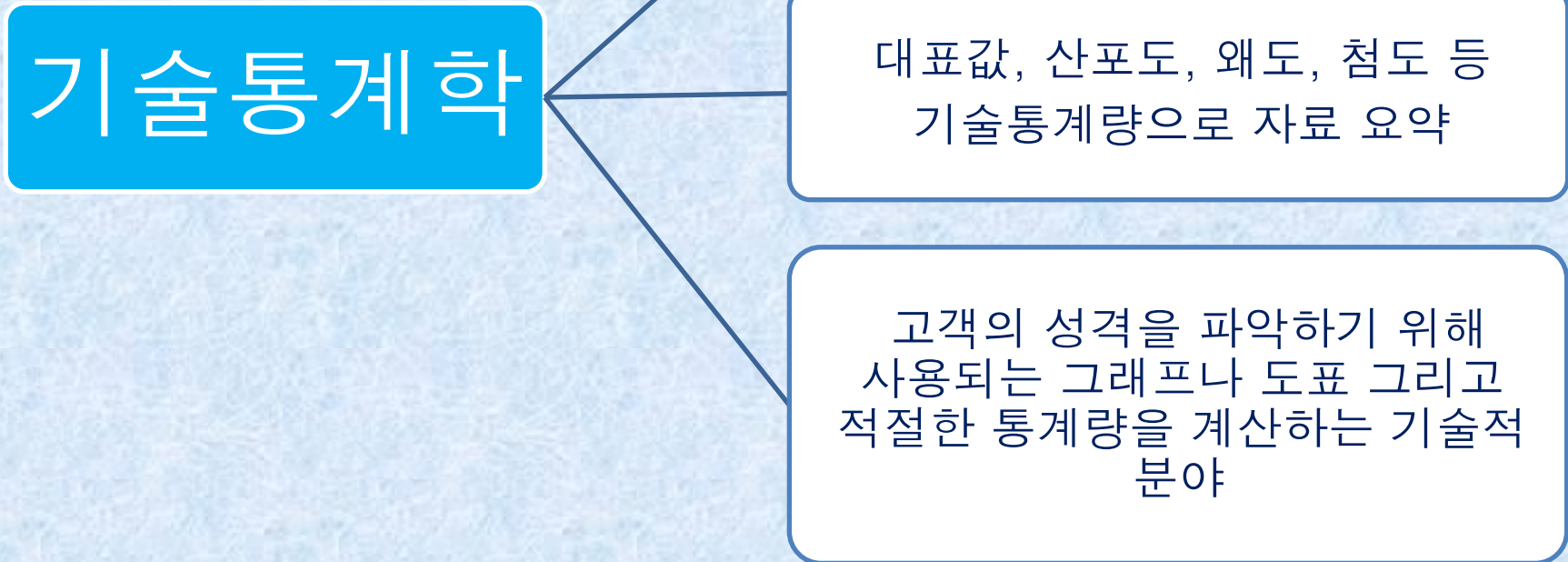
통계학



- 수집된 자료에 대한 해석과 분석을 위해 요구되는 수리과학적 방법

1.1 통계와 통계학

1.1.2 통계학의 분류



1.1 통계와 통계학

1.1.2 통계학의 분류

추측통계학

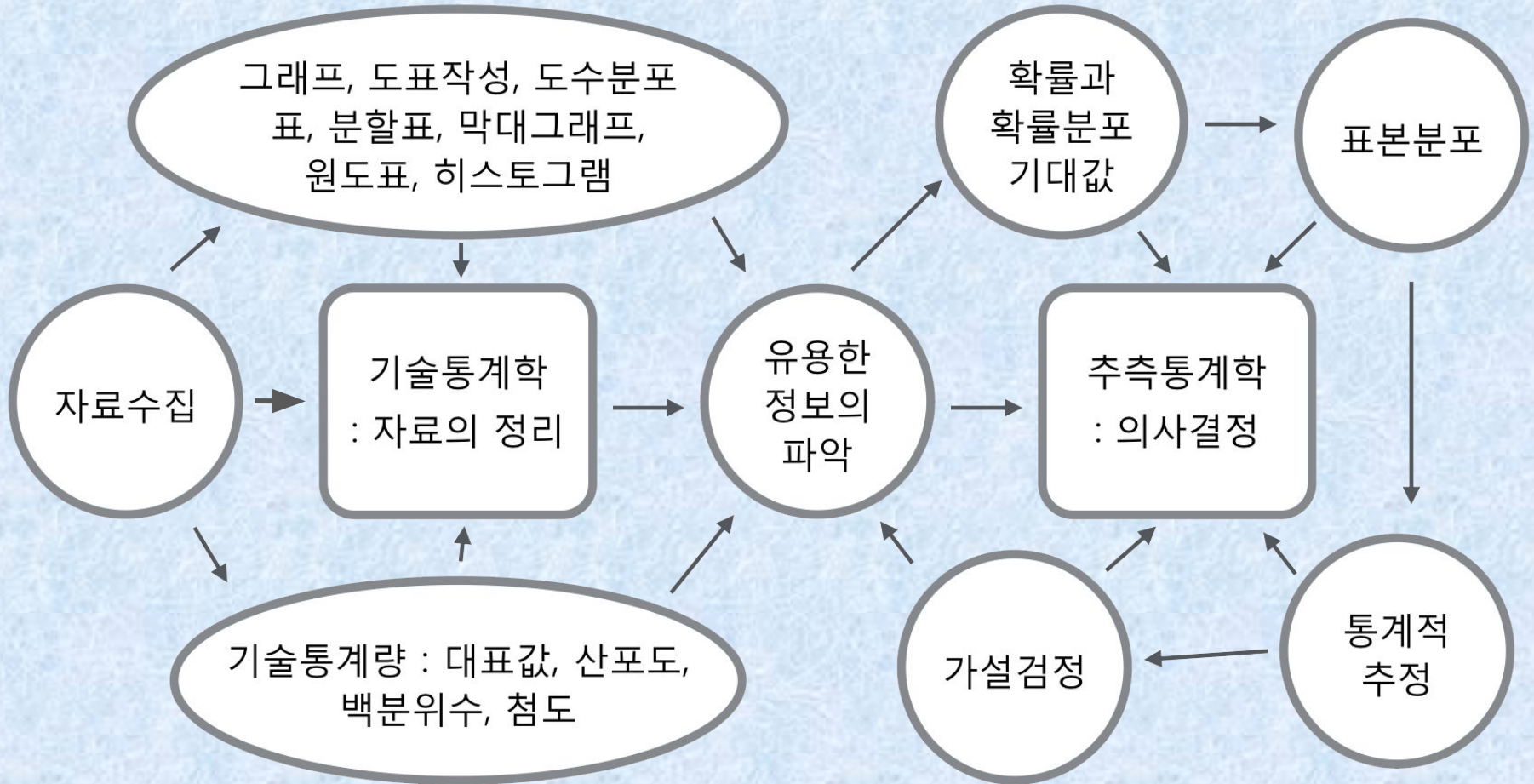
귀납적 추론 기법을 사용하여
미래 현상을 예측

새로운 정보의 진위여부를 확인하는
추정과 가설검정 관련 통계적 추론방법

제품의 평균수명을 추측

1.1 통계와 통계학

1.1.2 통계학의 분류



[그림 1.1] 기술통계학과 추측통계학의 관계

1.2 통계학의 활용 및 통계의 오남용

1.2.1 통계학의 활용

- 일기예보와 통계적 확률

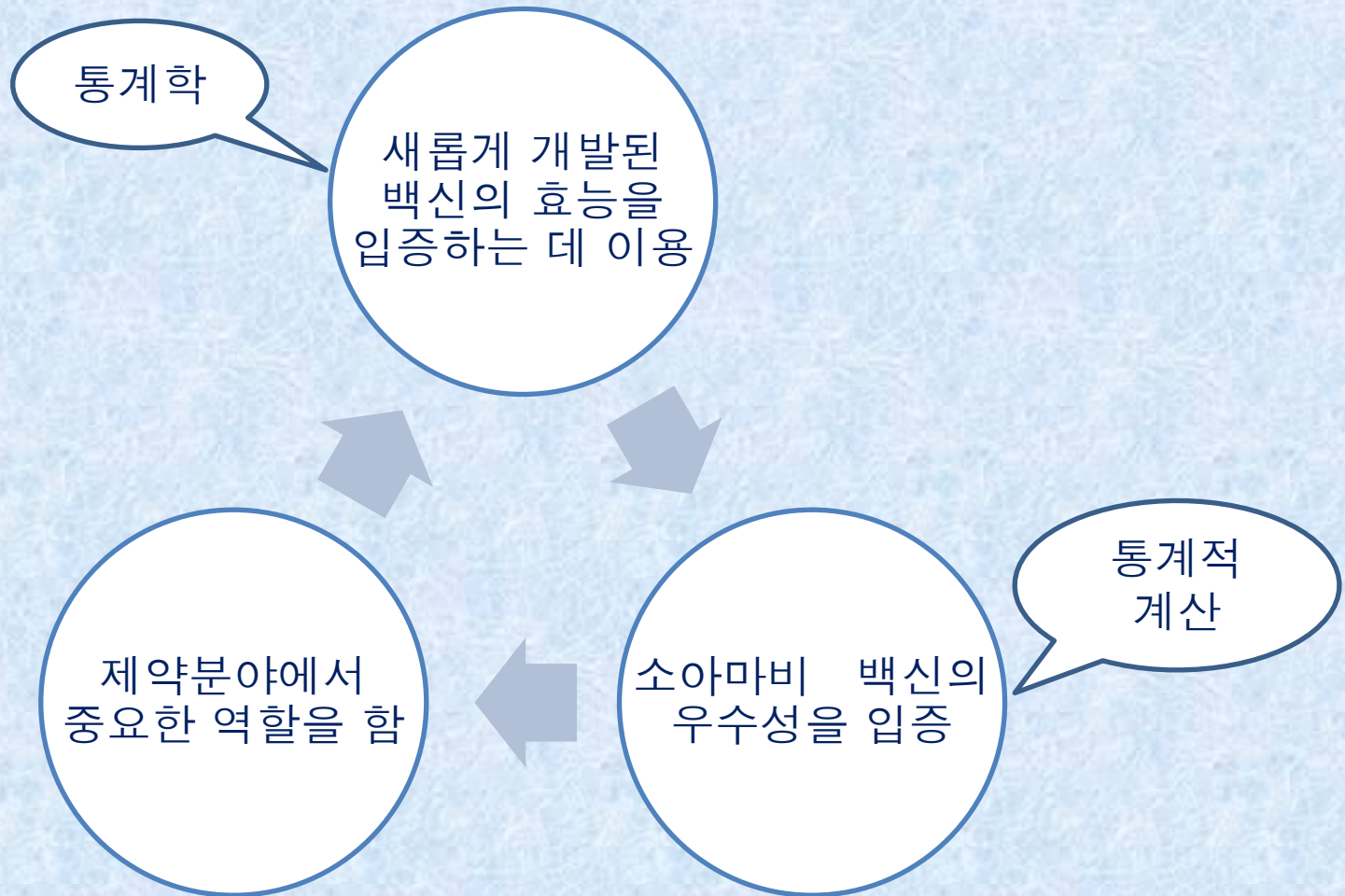


강수확률예보의 장점 - 필요한 정보 제공, 경제적 가치 생산
일기예보의 불확실성을 가능한 한 정확하게 표현

1.2 통계학의 활용 및 통계의 오남용

1.2.1 통계학의 활용

- 새 백신효능을 밝히는 통계



1.2 통계학의 활용 및 통계의 오남용

1.2.1 통계학의 활용

- 미래를 예측하는 통계



브라질에
비가
많이 옵니다

커피
생산량
증가

커피
가격은
급락

스타벅스
의 주가,
이윤증가

1.2 통계학의 활용 및 통계의 오남용

1.2.1 통계학의 활용

두 변량의 자료에 대한 상관성 분석 연구

상관관계

- 두 자료의 관련성 존재 유무 파악

회귀분석

- 미래의 사실을 예측

1.2 통계학의 활용 및 통계의 오남용

1.2.1 통계학의 활용

- 데이터마이닝 기법에 활용되는 통계

회사

새로운 판매 전략 수립
(데이터 마이닝)

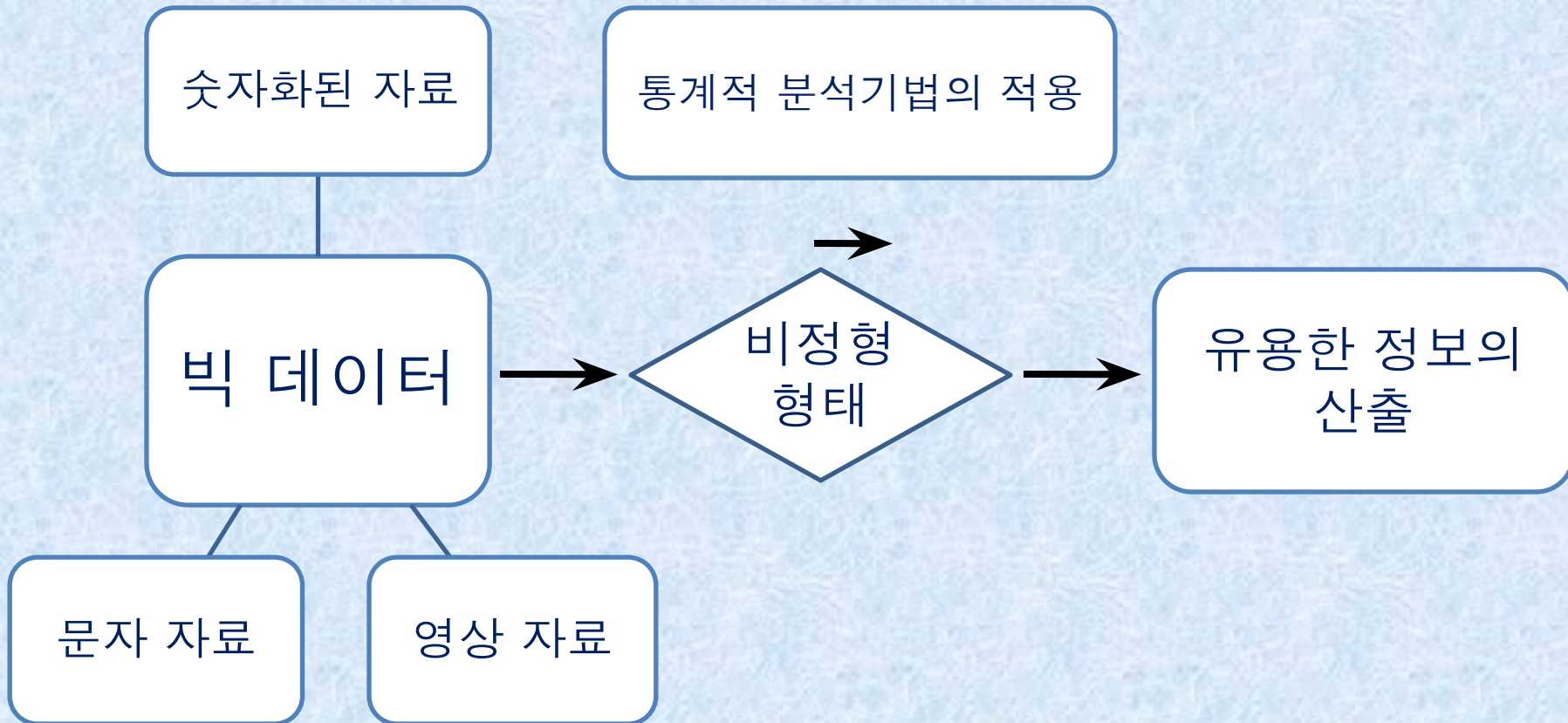
구입자들의
정보로
통계적 분석



1.2 통계학의 활용 및 통계의 오남용

1.2.1 통계학의 활용

- 빅 데이터 분석에 응용되는 통계



1.2 통계학의 활용 및 통계의 오남용

1.2.1 통계학의 활용

- 소비자들의 빅 데이터를 분석
 - 기업 이윤을 창출하는데 활용

필립스전자
회사의
이유식
제조기

매출 부진

빅 데이터
분석

판매시장
1위

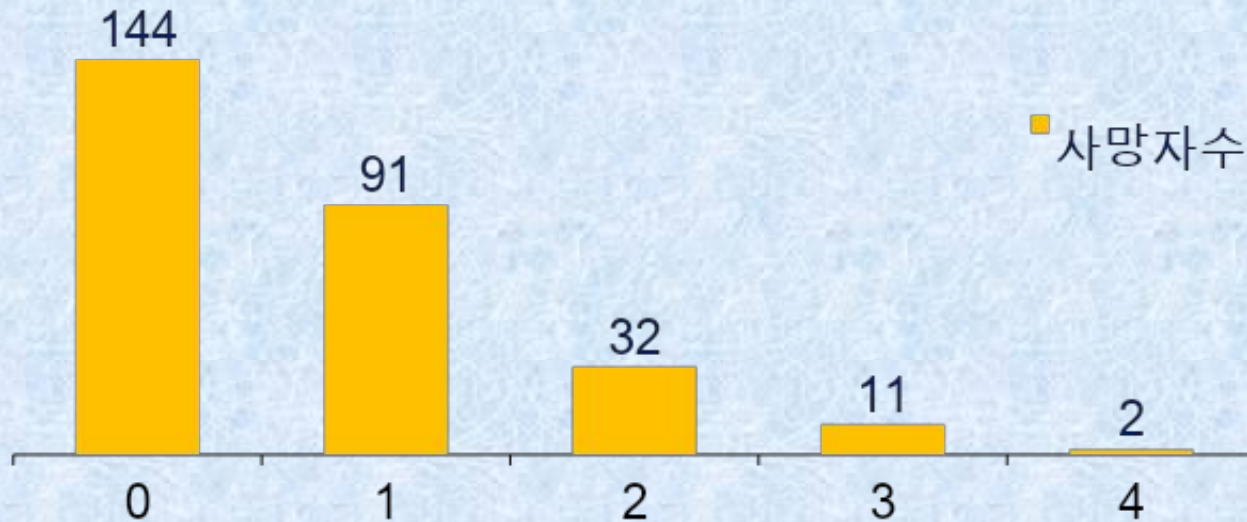
1.2 통계학의 활용 및 통계의 오남용

1.2.2 통계의 오남용

(1) 비대칭분포에서 산술평균값과 표준편차의 문제

사망자수	0	1	2	3	4	총
해당횟수	144	91	32	11	2	280

[표1.1] 사망자수의 분포



평균
0.7명

중앙값
0

대표값
사용시
부적절

[그림 1.2] 사망자수의 분포

1.2 통계학의 활용 및 통계의 오남용

1.2.2 통계의 오남용

산포도

- 대표값으로부터 퍼진 정도

사분위 범위

- 중앙값을 기준으로 위쪽 값들의 중앙값과 아래 쪽 값들의 중앙값의 차이 (0명, 1명)
- $\rightarrow 1\text{명} - 0\text{명} = 1\text{명}$

1.2 통계학의 활용 및 통계의 오남용

1.2.2 통계의 오남용

(2) 심슨의 역설



심슨의 역설

통계적으로 잘못
해석하는 문제

“부분이 작다고 해서 전체적인 것 또한 작지는 않고, 부분이 크다고 해서 전체적인 것 또한 크지는 않다.”

1.2 통계학의 활용 및 통계의 오남용

1.2.2 통계의 오남용

- 미국 버클리 대학 신입생 합격률 사례

성별	지원자수	합격자수	불합격자수	합격률
남성	2691	1400	1291	52.0%
여성	1835	772	1063	42.1%

(지원자수 가중치) 여자 합격률 < 남자 합격률

다른 변수에 대해서도 조사해 봐야함

분야	남자 지원자	남자 합격자	남자 합격률	여자 지원자	여자 합격자	여자 합격률	전체 지원자	전체 합격자	전체 합격률
남자	2691	1400	52.0%	1835	772	42.1%	4526	2172	48.0%
여자	1835	772	42.1%	2691	1400	52.0%	4526	2172	48.0%
남자 합격자	512	353	120	138	53	224	1400	1400	52.0%
여자 합격자	89	17	202	131	94	239	772	772	42.1%
남자 합격률	62.1%	63.0%	36.9%	33.1%	27.7%	60.1%	52.0%	52.0%	52.0%
여자 합격률	82.4%	68.0%	34.1%	34.9%	23.9%	70.1%	42.1%	42.1%	42.1%
전체 합격률	64.4%	63.2%	35.1%	34.0%	25.2%	64.8%	48.0%	48.0%	48.0%

[표 1.3] 남녀 성별과 지원분야별 대학원 합격률의 차이

1.2 통계학의 활용 및 통계의 오남용

1.2.2 통계의 오남용

(3) 이질적인 두 모집단의 비교

“미국 군인이 해외에서 죽는 사망률은 미국 뉴욕시민이 뉴욕에서 죽은 사망률보다 훨씬 작아서 생명의 위험률이 매우 낮다”



군대는 청년들만 있는 집단으로 특별한 사건이 없다면 사망률이 낮음



뉴욕시민은 늙은이, 어린이, 장애인 등 많은 종류의 사람이 살고 있으므로 특정한 사건이 없더라도 사망률이 높음



동질적이 아닌 두 모집단의 결과를 단순 비교하는 것은 부적절

1.2 통계학의 활용 및 통계의 오남용

1.2.2 통계의 오남용

(4) 부적절한 표본선택

존스 홉킨스 대학 여학생의 33.3%가 그 대학의 교수와 결혼한다는 뉴스



남녀 공학 첫해에 여학생 입학자 3명 중 1명이 교수와 결혼



이 경우 백분율을 계산하는 데 사용한 표본의 크기가 너무 작음



매우 작은 표본으로부터 얻은 결과로 모집단 전체를 언급하는 것은 부적절

1.3 통계의 역사

1.3.1 통계의 탄생



1.3 통계의 역사

1.3.1 통계의 탄생

<서양>

B.C. 3050
이집트 피라미드 건설
조사조직
최초의 통계

B.C. 2200
최초의 토지 조사

B.C. 1400
인구조사

B.C. 850
스파르타
기록부 토대로 토지
소유자 재분배

B.C. 1030
다윗 왕의 인구조사
구약성서 '역대상'

B.C. 1440
모세의 이스라엘 민족의
인구조사- '민수기'

B.C. 594
아테네 솔론왕
조세조사

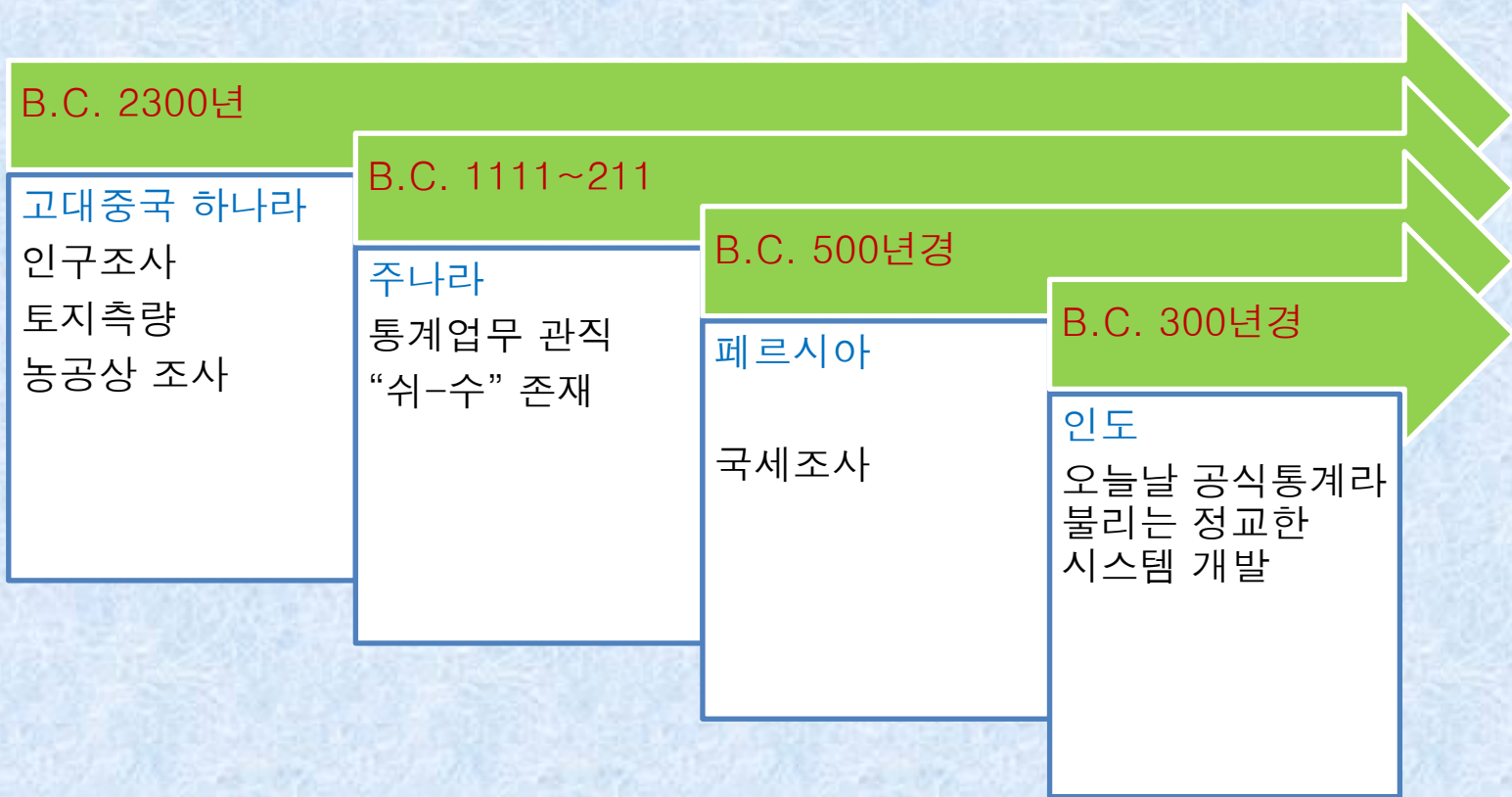
B.C. 435
로마 최초의 인구센서스
주기적 인구정태조사

A.D. 47 & 72
클라우디우스, 베스파시안
로마제국의 인구조사

1.3 통계의 역사

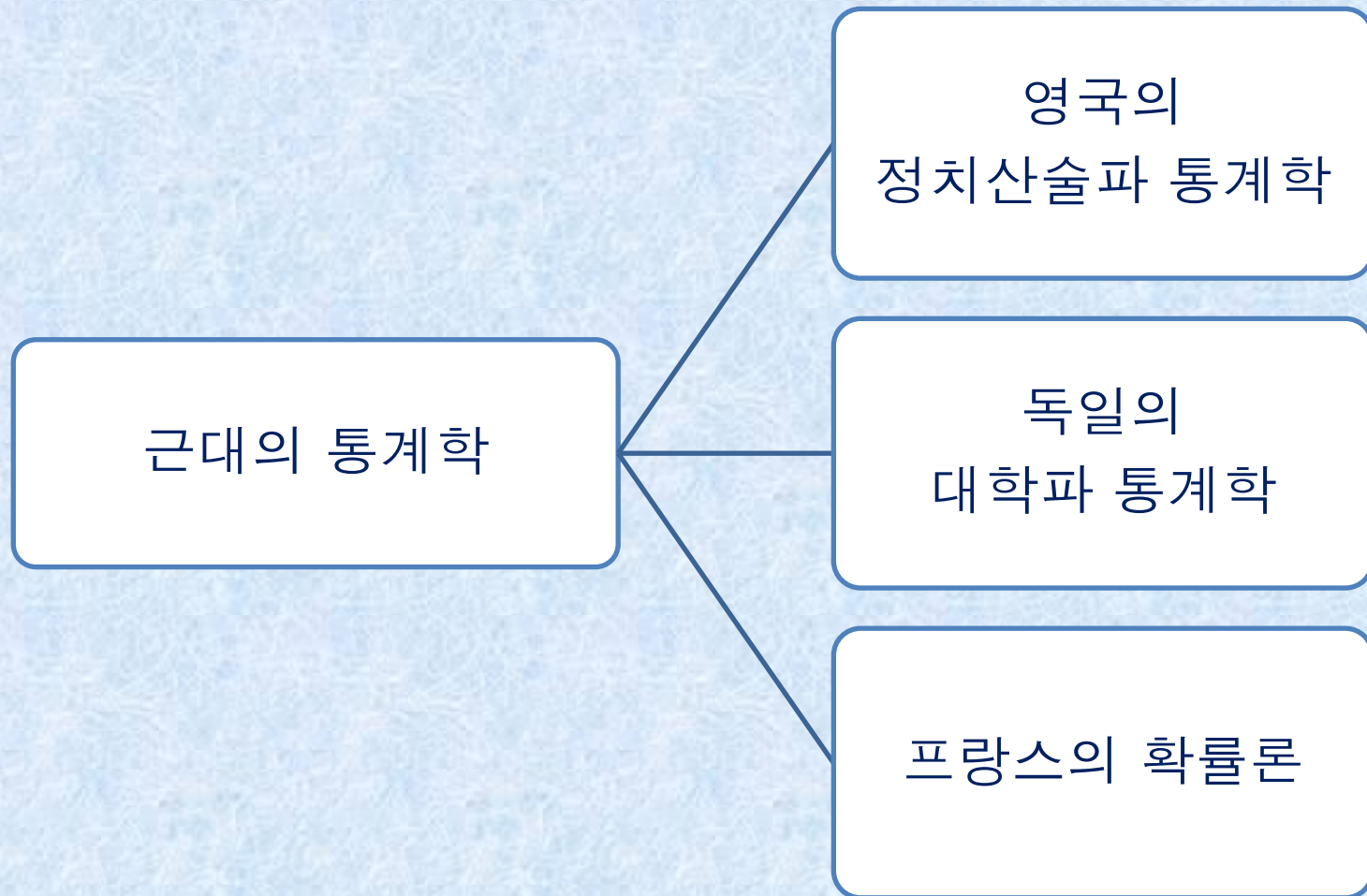
1.3.1 통계의 탄생

<동양>



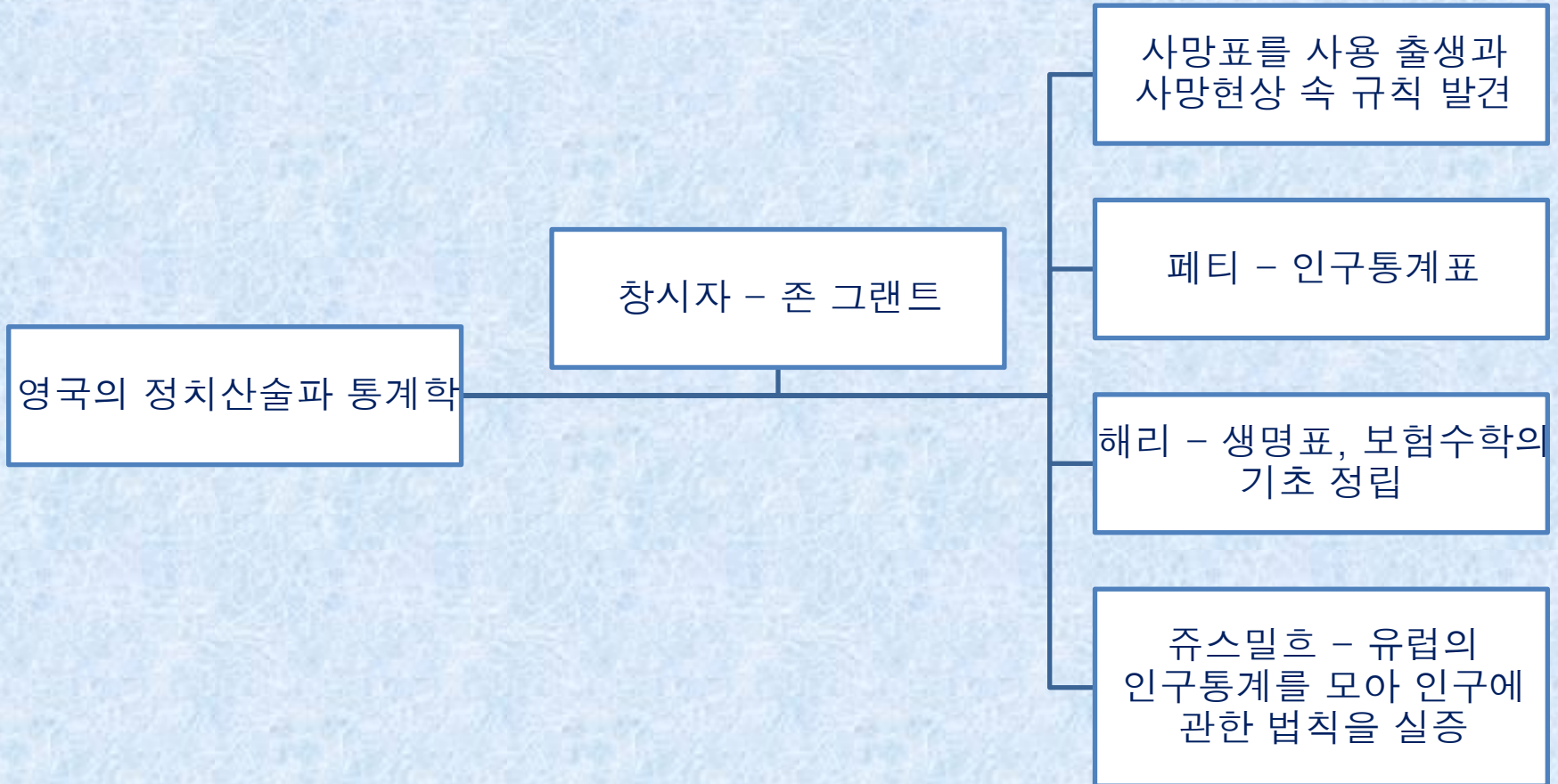
1.3 통계의 역사

1.3.2 통계학의 발달



1.3 통계의 역사

1.3.2 통계학의 발달



1.3 통계의 역사

1.3.2 통계학의 발달

독일의 대학과 통계학

- 콘링 - 17세기에 창시
- 국가의 토지와 인구를 명백히 하는데 목적

프랑스의 확률론

- 베르누이 - 순열과 조합을 사용해서 시행되는 사건들의 확률계산 기초 마련
- 드무아 - 정규분포 개념 발표
- 가우스 - 최소제곱법의 개념 고안, 정규분포를 유도
- 라플라스 - '확률해석론' 확률론을 체계화

1.3 통계의 역사

1.3.2 통계학의 발달



<케틀레>

- 세 가지 통계학의 흐름 종합
- 새로운 단계를 구축
- 확률론에 기초한 통계학 연구
- 통계조사법, 통계해석법, 통계사상사
- 통계제도의 정비
- 조사기술의 개선
- 통계지식 보급에 힘씀

1.3 통계의 역사

1.3.3 수리통계학의 정립

근원

근이론

이론

1.3 통계의 역사

1.3.4 오늘날의 통계학

