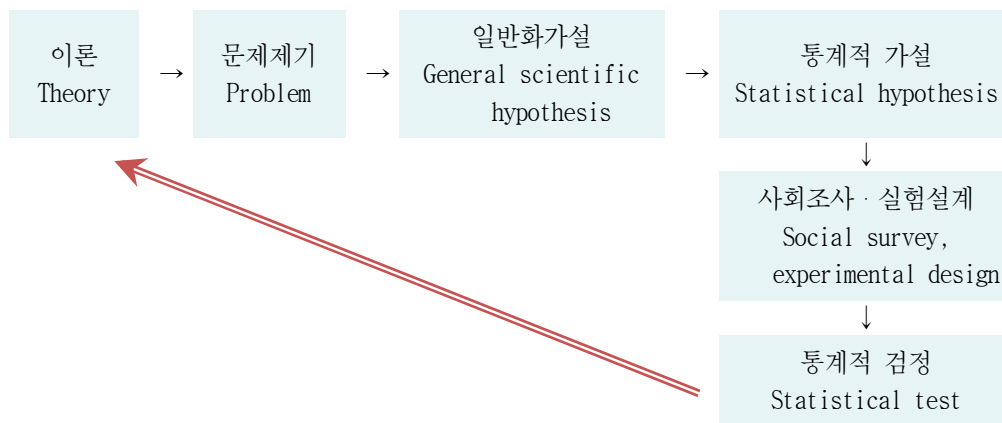


I. 과학과 통계

1. 통계의 역할과 필요성

1-1. 통계란

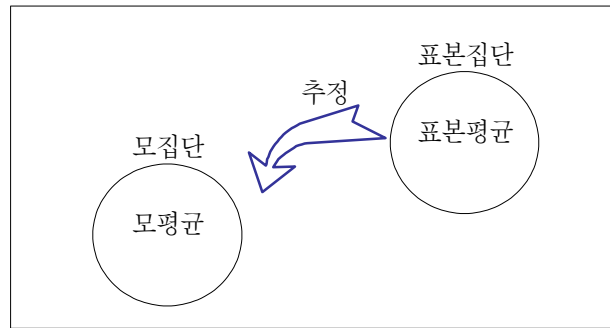
- ① 새로운 이론을 도출하거나 기존에 존재하는 이론을 지지, 거부 혹은 수정하기 위하여 경험적 연구방법을 사용할 때 자료 수집을 거쳐 모은 자료를 검정함으로써 결론에 도달하는 절차를 거치게 된다.



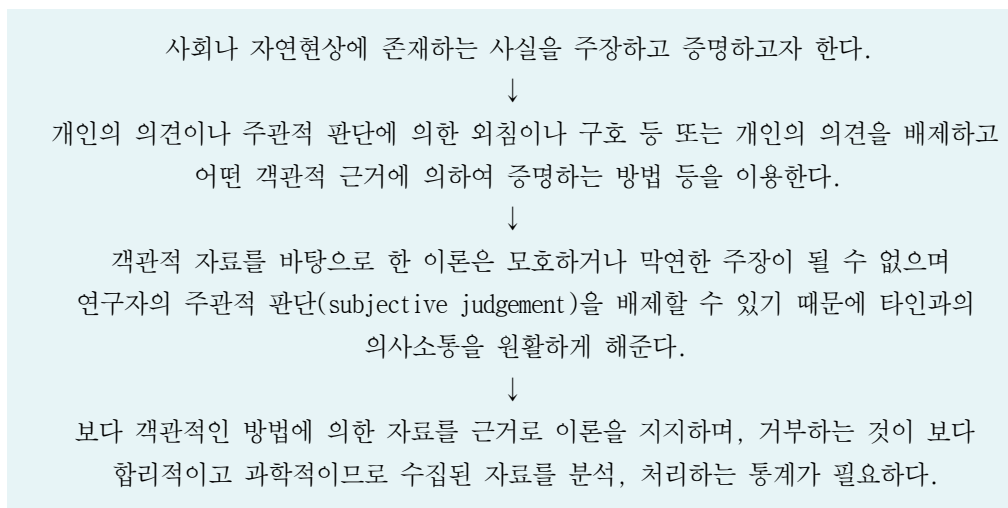
- ② 통계란 이론을 도출, 지지, 거부, 수정하기 위하여 수집한 자료를 가지고 가설을 검정하는, 즉 확률적으로 판정하는 수리적 논리라 정의할 수 있다(성태제).
- ③ 통계학이란 연구목적에 필요한 자료 및 정보를 최적의 방법으로 수집하고, 수집된 자료를 연구목적에 알맞게 정리하고 분석하여 얻은 결과로 미래를 예측하는 학문이다.

1-2. 통계의 역할과 필요성

- ① 통계는 얻어진 자료를 요약하고 설명하는 역할과 얻어진 자료의 결과를 일반화하는 역할을 한다.
- ㉠ 기술통계학(descriptive statistics) : 관찰한 자료의 중요한 특성을 쉽고 간단하게 요약하거나 서술하는 통계
- 자료를 도표나 그림으로 요약하여 대푯값이나 변동의 크기 등을 구하는 방법을 다루는 분야로 F. Galton, K. person 등에 의하여 발전하였다.
 - 수집된 자료의 특성을 쉽게 파악하기 위하여 도표, 그림, 평균, 최빈값, 표준편차 등을 이용하여 분포의 형태를 기술한다.
 - 관심대상의 자료를 어떻게 하면 효율적으로 정리하고 요약할 수 있는가가 주목적이다.
- ㉡ 추측통계학(inferential statistics) : 수집된 자료로부터 그 자료가 추출된 전체집단의 현상이나 특징을 추정, 예견, 일반화하는 통계
- 모집단에서 뽑은 표본을 분석하여 이 결과로부터 모집단의 특성을 밝히고 규명하는 분야이다.



② 통계학의 필요성은 다음의 흐름과 같다.



2. 통계학의 기본 개념

2-1. 모집단

- ① 통계적인 관찰의 대상이 되는 집단 전체이다.
- ② 모집단의 평균, 분산을 각각 모평균, 모분산이라 한다.

2-2. 표본

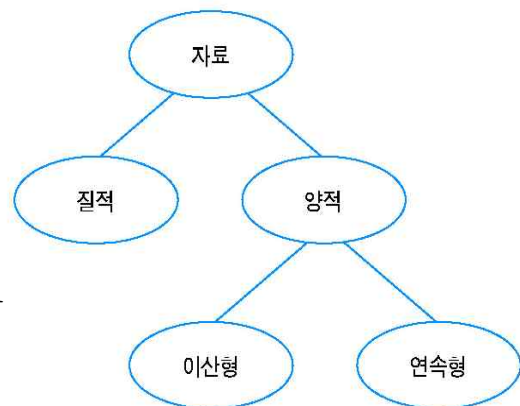
- ① 모집단을 대표하는 추출된 대상의 군집이다.
- ② 표본의 평균, 분산, 표준편차를 각각 표본평균, 표본분산, 표본표준편차라고 한다.
- ③ 표집(sampling)이란 모집단에서 표본을 추출하는 과정이다.
표집방법에는 단순무작위표집, 층화표집, 군집표집 등이 있다.

2-3. 변수

- ① 변수란 사람의 키, 체중 등과 같이 변하는 자료이다.
- ② 인과관계에 따라 독립변수와 종속변수로 나뉜다.
 - ㉠ 독립(independent)변수란 다른 변수에 영향을 주는 변수이다.
 - ㉡ 종속(dependent)변수란 다른 변수로부터 영향을 받는 변수이다.

【예】 통제된 실험상황에서 알코올의 섭취량에 따른 반응속도를 연구한다면 알코올 섭취량은 독립변수이고 반응속도는 종속변수이다.

 - ㉢ 매개(extraneous)변수란 독립변수와 종속변수 사이의 매개자 역할을 하는 변수이다. 독립변수의 결과인 동시에 종속변수의 원인이 된다.
- ③ 질적변수와 양적변수
 - ㉠ 질적(Qualitative=categorical)변수란 관측된 값이 몇 개의 범주(category)를 나타내는 문자나 숫자로 표시되는 변수이며, 수치로 나타낼 수 없는 변수이다. 범주형 자료라고도 한다.
질적자료는 명목척도, 서열척도로 구성된다.
 - 【예】 성별(남, 여), 직업, 학력, 종교, 경제상태(상, 중, 하)
 - ㉡ 양적(Quantitative=numerical)변수란 크기, 길이, 무게, 개수 등과 같은 양을 나타내기 위하여 수치로 나타낼 수 있는 변수이다. 양적자료는 등간척도, 비율척도로 구성된다. 그리고 양적변수는 연속변수와 불연속변수로 구별된다.
 - i) 연속(continuous)변수 : 어떤 구간 내에서 취할 수 있는 값이 무한히 많은 변수로, 값과 값 사이가 서로 연결되어 있어 그 사이의 값이 의미를 가진다. 측정도구의 정밀성이 측정값의 정확성에 크게 영향을 준다.
【예】 신장, 체중, 거리, 수질오염도, 시력 등
 - ii) 불연속(discrete)변수 : 어떤 구간 내에서 취할 수 있는 값이 한정되어 있는 변수로, 정수로 측정되므로 값과 값 사이가 서로 분리되어 있어 그 사이의 값이 아무런 의미를 가지지 않는다.
【예】 가구당 자녀 수, 입원 환자 수, 학생 수



【참고】 양적자료와 질적자료의 구분은 칼로 자르듯이 명확하게 구분되지 않는 경우가 있다. 예를 들어, 시험 성적을 점수 그대로 사용하는 경우는 양적자료에 해당하지만, 상위권, 중위권, 하위권으로 구분하여 사용하는 경우는 양적자료와 질적자료의 성질이 모두 해당된다.

【예】 다음 변수들을 질적인지 양적인지 구분하시오.

1. 가장 빈도가 높은 전자렌용도: 조리, 해동, 데우기, 기타
2. 전화 조사에 대한 대답을 거절한 고객의 수
3. 쥐 실험에서 쥐가 선택한 문; A, B, C
4. A지역에서 경주마의 우승시간
5. 어느 대상에서 영자 신문을 구독하고 있는 학생 수

<답> 1과 3은 질적변수이고, 2, 4, 5는 양적변수이다. 또한 2와 5는 이산형이고, 4는 연속형이다.

2-4. 측정과 척도

종류		적용 예
질적 자료	명명척도	성별, 결혼유무, 종교, 인종, 직업유형, 장애유형, 지역, 계절 등
	서열척도	사회계층, 선호도, 석차, 소득수준, 수여 받은 학위, 자격등급, 장애등급, 변화에 대한 평가, 서비스효율성 평가 등
양적 자료	등간척도	IQ, EQ, 온도, 학력, 시험점수, 물가지수, 경제성장률, 사회지표 등
	비율척도	연령, 무게, 신장, 수입, 출생률, 사망률, 이혼율, 가족 수, 졸업생 수, 서비스 대기인 수, 서비스 수혜기간 등

- ① 측정(measurement)이란 사물을 구분하기 위하여 이름을 부여하거나 사물의 속성을 구체화하기 위하여 수를 부여하는 절차이다.
- ② 수를 부여하기 위하여 단위와 수를 부여하는 규칙이 필요하며 이는 척도로 해결된다.
- ③ 척도(scale)는 사물의 속성을 구체화하기 위한 측정의 단위이다. 질적변수에 명명척도와 서열척도가, 양적변수에 등간척도와 비율척도가 있다.
 - ㉠ 명명척도(nominal scale) : 성별(남자 1, 여자 2), 인종, 색깔 등
 - 사물을 구분하기 위하여 이름을 부여하는 척도이다.
 - 일대일 변환(one to one transformation)으로 하나의 사물에 하나의 이름을 부여하여 사물을 구분한다.
 - 사용할 수 있는 통계기법으로는 최빈수, 도수, 상관관계계수 등이 있다.
 - ㉡ 서열척도(ordinal scale) : 수학 등수, 가장좋아한다/보통이다/매우싫어한다 등
 - 측정대상의 특징 및 속성에 따라 일정한 범주로 분류하여 이들에 대해 상대적인 순서, 서열상의 관계를 나타내는 척도이다.
 - 사물이나 현상을 분류하고 명칭을 부여할 뿐만 아니라 순서 또는 서열까지 부여한다.
 - ㉢ 등간척도(interval scale) : IQ 점수, 온도 등
 - 측정대상을 특징 및 속성에 따라 서열화하는 것은 물론 서열 간의 간격이 일정하도록 연속선상에 수치를 부여하는 것이다.
 - 낮은 부분에서 높은 부분으로 서열을 정할 수 있을 뿐만 아니라 부분과 부분 간의 간격(거리)의 차이를 측정할 수 있는 자료이다.

- 진정한 그리고 절대적인 '0' 을 가지지 않고 단지 임의적인 '0' 만을 가지므로 '0' 은 '아무것도 없는 상태' 가 아니라 '0만큼 있는' 을 의미한다.
- 가감(+, -)과 같은 수학적 조작이 가능하다.
- ㉔ 비율척도(ration scale) : 공식적인 교육, 소득, 백분율, 연령, 가족구성원의 수 등
 - 측정대상의 특징 및 속성에 절대적인 0(절대영점)과 임의단위(임의로 협약한 단위, 1cm, 1foot 등)를 가지고 수치를 부여하는 것이다.
 - 사물이나 현상을 분류하고 명칭을 부여하며, 서열을 정하고 가감과 같은 수학적 조작을 할 수 있을 뿐만 아니라, 가감과 승제와 같은 수학적 조작이 모두 적용되는 가장 고차원적인 측정이다.
- ㉕ 절대척도(absolute scael) : 사람 수, 자동차 수, 공의 수
 - 절대영점, 절대단위를 가지고 있는 척도이다. 이때, 절대영점(아무것도 없음)을 의미, 절대단위는 협약에 의한 것이 아닌 절대적인 것이다.
 - 가감승제 모두 적용할 수 있다.

2-5. 가설

- ① 가설이란 둘 이상의 변수 또는 현상 간의 관계를 설명하는 검증되지 않은 명제이다.
- ② 가설은 연구 문제에 관해 검증할 수 있도록 기술된 잠정적인 응답으로, 독립변수와 종속변수와의 관계의 형태로 기록된다.
- ③ 가설에는 크게 연구가설(=대립가설)과 귀무가설(=귀무가설)로 구분된다.
- ④ 새롭게 주장하고자 하는 가설을 '연구가설' 이라고 하고, 이러한 새로운 주장이 타당한 것으로 볼 수 없을 때는 저절로 원상이나 현재 믿어지는 가설로 돌아가게 되는데 이 가설을 '귀무가설' 이라 한다.
- ⑤ 연구가설(research hypothesis)은 연구문제에 대한 잠정적 대답으로, 연구자가 주장하고자 하는 가설이다. 연구가설에는 " \neq ", " $>$ ", " $<$ ", " \sim 같지 않다", " \sim 차이가 있다", "모두 같은 것은 아니다", "유의하다", "독립이 아니다", "연관성이 있다" 라는 말이 들어간다.
- ⑥ 귀무가설(null hypothesis)은 연구가설과 논리적으로 반대의 입장을 취하는 가설로, 처음부터 버릴 것을 예상하는 가설이다. 귀무가설에는 " $=$ ", " \sim 같다", " \sim 차이가 없다", "유의하지 않다", "독립이다", "연관이 없다" 라는 말이 들어간다.

영가설	연구가설
<ul style="list-style-type: none"> - 새로운 주장이 타당한 것으로 볼 수 없을 때 저절로 원상이나 현재 믿어지는 가설로 돌아가게 되는 가설 - 연구가설과 논리적으로 반대의 입장을 취하는 가설 - 처음부터 버릴 것을 예상하는 가설 	<ul style="list-style-type: none"> - 새롭게 주장하고자 하는 가설 - 연구문제에 대한 잠정적 대답 - 연구자가 주장하고자 하는 가설
<p>"$=$", "\sim같다", "\sim차이가 없다", "유의하지 않다", "독립이다", "연관이 없다"</p>	

- ⑦ 분석하는 사람의 입장에서는 귀무가설을 채택하게 되면 현재 믿어지는 가설이 옳은 것이 되므로 귀무가설을 기각하기를 원한다.
- ⑧ 연구가설은 귀무가설이 직접 채택될 수 없을 때 자동적으로 받아들여지는 가설로서 직접 검증할 필요가 없는 반면 귀무가설은 직접 검증을 거쳐야 하는 가설이다.
- 【예1】 문과와 이과에 따라 수능점수의 평균에 차이가 있는지를 검정하고자 한다.
- 귀무가설: 문과와 이과에 따라 수능점수의 모평균에 차이가 없다.
 - 연구가설: 문과와 이과에 따라 수능점수의 모평균에 차이가 있다.
- 【예2】 성별(남자, 여자)에 따라 결혼만족도에 차이가 있는지를 검정하고자 한다.
- 귀무가설: 성별에 따라 결혼만족도에 차이가 없다.
 - 연구가설: 성별에 따라 결혼만족도에 차이가 있다.
- 【예3】 성별과 거래횟수가 관련이 있는지를 검정하고자 한다.
- 귀무가설: 성별과 거래횟수는 서로 독립이다(연관성이 없다).
 - 연구가설: 성별과 거래횟수는 서로 독립이 아니다(연관성이 있다).