

**기초통계학**

**Basic Statistics**

기초통계학 I	기초통계학 II
<ul style="list-style-type: none"> <li>○ 통계학이란?</li> <li>○ 기술통계               <ul style="list-style-type: none"> <li>- 표와 그래프, 수치적 해석</li> </ul> </li> <li>○ 확률</li> <li>○ 확률변수와 확률분포</li> <li>○ 표집분포</li> </ul>	<ul style="list-style-type: none"> <li>○ 통계적 추론               <ul style="list-style-type: none"> <li>- 추정, 검정</li> </ul> </li> <li>○ 분산분석</li> <li>○ 회귀분석</li> <li>○ 범주형 자료분석</li> </ul>

## ■ 통계학이란?

- 포털사이트 검색창에서 “그냥도전 동전 돌리기”라고 검색
  - 500원짜리 동전을 돌렸을 때 학이 나올 확률이 70% 정도
  - 실제로 500원짜리 동전을 1000번 돌리는 실험
  - 실험 결과 1000번 중 학이 679번이 나옴
  - 이 결과를 토대로 학이 나올 가능성이 70%정도 된다는 것이 얼추 맞다고 주장

## ■ 모집단과 표본

◎ 조선(대한제국 포함)시대 **임금의 수명**은?

- 자료 수집 ⇨ 전체 자료 수집이 가능

【표 1】 조선 임금의 수명

임금	수명	임금	수명	임금	수명	임금	수명	임금	수명
태조	74	세조	52	명종	32	숙종	60	철종	33
정종	63	예종	20	선조	57	경종	37	고종	68
태종	56	성종	38	광해군	67	영조	83	순종	53
세종	54	연산군	31	인조	55	정조	49		
문종	39	중종	57	효종	41	순조	45		
단종	17	인종	31	현종	34	헌종	23		

- ◎ 궁에서 임금과 같이 생활했던 **내시의 수명**은?
  - 자료 수집 ⇨ 전체 자료 수집이 가능?
  - 이윤묵(1741~1816)의 양세계보(養世係譜)에는 81명의 내시 기록
  - 평균수명은 70세, 81명 중 100세 이상 생존한 사람은 3명
  - Q/P. **81명이 전체 내시를 대표할 수 있는가?**

## ● 모집단(population)

- 통계학에서는 잘 정의된 연구목적과 이에 대한 명확한 연구대상을 설정
  - 예) 임금의 평균수명, 내시의 평균수명
- 연구대상이 되는 모든 개체의 집합
  - 예) 전체 임금, 전체 내시
- 실제로 관심을 가지는 것은 대상 자체보다는 그 대상의 속성에 관심을 가지기 때문에 전체 대상의 속성이 모집단이 되기도 함
  - 예) 전체 임금의 수명, 전체 내시의 수명

- 대부분의 모집단은 매우 커 전체를 조사하기 힘들
  - 예) 대통령 선거에서 어떤 후보자의 지지율?
  - 대한민국 18대 대통령선거에서 유권자인 선거인 수는 40,507,842명
- 모집단을 명확하게 정의할 수 있는 경우도 있지만 애매한 경우도 있음
  - 예) “그냥도전 동전 돌리기”에서의 모집단은?
  - 통계분석을 할 때 이런 문제를 심심치 않게 만나는데 이 경우 모집단을 동전 돌리기 실험을 무한히 많이 반복수행하여 결과를 모아 놓은 것으로 이해

- 개념적으로 규정한 조사 대상 전체를 **목표모집단(target population)** 또는 대상모집단이라고 하고 실제로 표본을 추출하기 위해 규정한 조사 대상 전체를 **조사모집단(survey population)**
  - 예) 경제활동인구조사
  - 목표모집단: 군인 및 수감자 등을 제외한 대한민국 영토 내에 거주하는 15세 이상 모든 국민
  - 조사모집단: 조사의 편의나 여건을 고려해 도서지역, 기술시설 및 특수시설 거주자는 조사 대상에서 제외

## ○ 전수조사(census)

- 모집단 전체를 대상으로 조사하는 경우
- 센서스(census)는 추정하다(to estimate)라는 뜻의 라틴어 "censere"에서 유래 되었으며 센서스, 공공치안, 국가재정 등의 일을 담당하던 고대로마 관료를 censor라고 함
  - 아우구스투스(Augustus)가 센서스를 위해 출생 도시로 가서 호적 등록하라는 명령을 내렸으며 이 명령에 따라 예수의 부모는 고향을 가는 과정에서 예수를 출산
- 144년에 중국 한나라에서 실시한 조사에서는 994만 가구에 4973만 명이 사는 것으로 기록

- 우리나라의 경우
  - 1949년에 정부수립 후 대규모 조사를 시작하였으나 한국전쟁 중 자료가 모두 소실
  - 1955년에 간이총인구조사를 거쳐 1960년까지 국세(國勢)조사라는 이름으로 실시
  - 1963년 통계위원회에서 일본식 용어인 국세조사를 사용금지 하였으며 이후 '센서스'와 '총조사'라는 용어를 번갈아 사용하다가 1990년 이후 총조사로 통일
  - 2010년 11월에 실시한 '인구주택총조사'에서는 인터넷을 이용한 조사가 본격적으로 병행 실시했으며 2015년에는 전수조사 대신 행정자료 이용과 표본조사를 병행

## ● 표본(Sample)

- 모집단으로부터 선택된 일부의 개체
  - 예) 양세계보에 수록된 81명의 내시
  - 예) “그냥도전 동전 돌리기”에서 나온 1000번의 동전 결과
- Q/P. **추출된 표본이 모집단 특성을 대표할 수 있는가?**
  - 예) 양세계보에 기록된 내시가 모두 특정 시기의 내시라고 한다면 그 시기의 정치, 사회적 상황과 의학기술에 영향을 받음 ⇨ 조선시대의 전체 내시를 대표한다고 보기 어려움

- 1936년 미국대통령 선거
  - 공화당의 랜던과 민주당의 루즈벨트
  - 'Literary Digest'는 **전화기 및 자동차 보유자 대상**으로 엽서를 보내 회송된 236만여 명의 의견을 분석한 결과 랜던 57%, 루즈벨트 43%
  - Gallup은 수천 명의 표본조사를 토대로 루즈벨트 56%, 랜던 44%
  - 선거결과에서 루즈벨트 63%, 랜던 37%
  - Gallup이 루즈벨트의 당선을 예측했지만 예측한 득표율과 실제 득표율 간에 차이가 현재의 조사 결과들에 비해 큼

## ○ 확률추출(probability sampling)

- 연구목적에 필요한 자료와 정보를 여건이나 상황, 정확성 등을 고려하여 표본 수집 ⇨ 표본론, 실험계획법
- 모집단을 대표 할 수 있는 표본은 네이만이 제안한 **확률표본추출법**을 기반으로 얻을 수 있음
  - 단순확률추출
  - 계통표본추출
  - 층화확률추출
  - 집락표본추출

- 어떤 표본이 선택되는가에 따라 결과에 차이가 발생 ⇨ 변동성이 있는 표본의 정보를 이용하여 전체 모집단의 특성을 완벽하게 파악하는 것은 불가능
- 제한된 표본의 정보에 확률을 이용하여 모집단의 특성에 대해 추론 ⇐ **통계적 추론(statistical inference)**
- 확률추출방법에 의해 얻어진 표본을 이용하여 모집단에 대한 통계적 추론이 가능
- 많은 경우의 통계적 추론에서는 확률추출이 아닌 방법을 통해 얻어진 표본 또는 자료를 이용하는데 이런 추론은 모집단에 대해 심각하게 왜곡된 결론을 도출할 수 있음

## 통계학이란

- 관심 또는 연구의 대상이 되는 모집단의 특성을 파악하기 위해
- 모집단부터 일부의 자료(표본)를 수집하고
- 수집된 표본을 정리, 요약, 분석하여 표본의 특성을 파악한 후
- 표본의 특성을 이용하여 모집단의 특성에 대해 추론하는 원리와 방법을 제공하는 학문