

# 변수측정과 자료수집

# 사례: 체성분 분석

◆ 체성분 분석: 인체를 구성하는 성분들 사이의 관계를 분석

체성분분석 Body Composition Analysis	측정치 Value	체수분 Total Body Water	근육량 Soft Lean Mass	체지방량 Fat Free Mass	체중 Weight	표준범위
세포내수분 (ℓ) Intracellular Water	15.7	25.1	32.3	34.3	46.9	14.9~18.3
세포외수분 (ℓ) Extracellular Water	9.4					9.2~11.2
단백질 (kg) Protein	6.8	non-osseous		6.5~7.9		
무기질 (kg) Mineral	2.41	osseous : 1.99		2.22~2.72		
체지방 (kg) Body Fat Mass	12.6			9.4~15.1		

▶ 유기질은 추정치입니다.

골격근-지방 Muscle-Fat Analysis	표준이하	표준	표준이상	UNIT: %	표준범위							
체중 (kg) Weight	55	70	85	100	115	130	145	160	175	190	205	40.2~54.4
골격근량 (kg) Skeletal Muscle Mass	70	80	90	100	110	120	130	140	150	160	170	17.7~21.6
체지방량 (kg) Body Fat Mass	40	60	80	100	120	140	160	180	200	220	240	9.4~15.1

비만진단 Obesity Diagnosis	표준이하	표준	표준이상	UNIT: %	표준범위							
BMI (kg/m <sup>2</sup> ) Body Mass Index	10	15	18.5	21	26	30	35	40	45	50	55	18.5~23.0
체지방률 (%) Percent Body Fat	8	13	18	23	28	33	38	43	48	53	58	18.0~28.0
복부지방률 Waist-Hip Ratio	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	1.05	1.10	0.70~0.80

# 사례: 체성분 분석

## ◆ BMI(kg/m<sup>2</sup>)

- 체질량 지수(Body Mass Index)라고 하며 신장과 체중을 이용하여 비만 여부를 판정하기 때문에 겉보기 비만지수라고 함
- BMI = 체중(kg) ÷ 신장<sup>2</sup>(m<sup>2</sup>)

## ◆ 체지방률(%)

- 체중에서 체지방이 차지하는 비율
- 체지방률 = 체지방(kg) ÷ 체중(kg) × 100

## ◆ 복부지방률

- 허리와 엉덩이의 둘레비로 WHR(Waist-Hip Ratio)이라고도 함

# 자료와 변수

◆ 개체(individual): 자료 수집을 위한 개별 조사대상, 객체(subject)라고 포함

● 체성분 분석과 관련된 자료를 획득하기 위한 개체는 사람

◆ 표본(sample): 조사대상으로 선정된 개체의 집단

◆ 변수(variable): 각 개체가 갖는 특성

◆ 체성분 분석 변수

변수명	변수의 의미	변수 설명
Gender	성별 구분	남자=0, 여자=1
Age	연령	단위: 세, 만 나이
Hight	신장	단위: cm
Weight	체중	단위: kg
BFM	체지방량	단위: kg
FFM	제지방량	단위: kg, 체중 - 체지방량
SLM	근육량	단위: kg
MIN	빠가 된 무기질	단위: kg, 제지방량 - 근육량
SMM	골격근량	단위: kg
VFA	내장지방	단위: cm <sup>2</sup>
WHR	복부지방율	단위: cm/cm

# 변수와 관측치

- ◆ 관측치(observation): 표본을 구성하는 각 개체는 변수들에 대해 특정한 값을 갖는데, 자료는 이러한 변수들의 값

ID	Gender	Age	Hight	Weight	BFM	FFM	SLM	MIN	SMM	VFA	WHR
1	1	22	163	52.10	15.40	36.70	34.50	2.20	19.55	46.48	0.79
2	1	23	157	61.90	19.10	42.80	40.40	2.40	23.70	43.20	0.78
3	1	25	161	44.40	8.80	35.60	33.60	2.00	19.30	21.70	0.74
4	1	26	151	46.10	14.90	31.20	29.30	1.90	16.56	52.05	0.81
5	1	27	168	54.40	13.60	40.80	38.40	2.40	22.00	57.90	0.82
6	1	27	163	57.00	17.60	39.40	37.10	2.30	21.20	69.40	0.82
7	1	27	152	62.30	28.40	33.90	31.90	2.00	17.59	100.91	0.92
8	1	28	157	43.70	8.60	35.10	33.10	2.00	18.89	17.05	0.74
9	1	29	158	59.80	16.80	43.00	40.50	2.50	23.85	40.34	0.80
10	1	29	163	48.50	10.80	37.70	35.60	2.10	20.26	45.87	0.78

- ◆ 자료 (Data) : 표본을 구성하는 각 개체의 특성을 숫자로 표현한 ‘변수’들의 관측치

# 자료의 측정수준

주점에 들어가서 자리에 앉고 나면, 아주머니가 술은 무엇으로 하겠냐고 묻는다. 소주인지, 맥주인지, 아니면 막걸리인지를 말하라는 것이다. 소주라고 대답하면, 어떤 소주를 마실지에 대해 물으신다. 참이슬이라고 말씀드리면, 비로소 마실 술의 범주(category)가 결정된 것이다. 그러면 아주머니는 다시 물으신다. “몇 병?” 5병을 주문하면, 계산서에 그 양을 “正”라고 표시하신다. 참이슬은 정성적인 범주형 자료이고, 5병은 정량적인 수치형 자료이다.

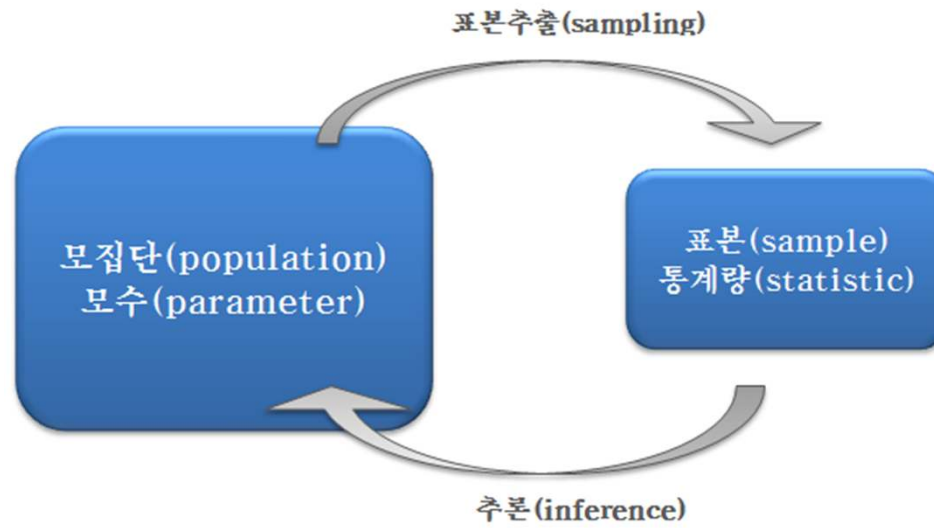
## ◆ 자료의 측정 수준에 따른 분류(비공식)

- 범주형
- 수치형
- 박쥐형

## ◆ 자료의 측정 수준에 따른 분류(공식)

범주형 변수	(명목)	범주		
박쥐형 변수	(서열)	범주	크기	
수치형 변수	(등간, 비율)	범주	크기	간격

# 표본과 통계량



# 횡단면 자료 VS. 종단면 자료

- ◆ 횡단면 자료: 어떤 동일 시점에서 여러 개체들로부터 수집된 자료
  - 예) 대학생의 흡연율
- ◆ 종단면 자료: 동일한 개체들을 여러 시점에서 관측하여 수집된 자료를 시계열 자료(time-series data)
  - 예) 다이어트 전후 체중변화



# 기술통계량과 추론통계량

- ◆ 기술통계량(descriptive statistic): 자료를 정리 및 요약하기 위해 사용되는 통계적인 값
  - 예) 한림대 경영학부 신입생의 내신성적 평균
- ◆ 추론통계량(inferential statistic): 모수의 추론을 위해 사용되는 통계적인 값
  - 추정통계량(Estimator) : 모수가 갖는 값을 추정하기 위해 사용되는 통계적인 값
    - ❖ 점 추정량(point estimator)
    - ❖ 구간 추정량(interval estimator)
  - 검정통계량(test statistic): 가설의 검정을 위해 사용되는 통계적인 값
  - 참고 : 추정치(estimate)는 자료를 통해 계산된 추정량의 값을 말함

# 표본추출의 방법

- ◆ **단순임의표본추출(simple random sampling):** 표본을 무작위 비복원 추출하는 방법

## 엑셀 활용

임의표본을 추출하는 경우처럼, 특정 범위의 수들에서 n개의 정수를 무작위로 추출하는 방법이다.

예를 들어, 1과 821 사이에서 10개의 정수를 얻고자 한다면,

□ =RANDBETWEEN(1, 821)을 10개의 셀에 입력한다. 새로운 정수들을 추출하려면 F9를 누른다.

□ =INT(1+821\*RAND())을 10개의 셀에 입력한다. 새로운 정수들을 추출하려면 F9를 누른다.

- ◆ **층화임의표본추출(stratified random sampling):** 모집단을 몇 개의 이질적인 소집단 혹은 계층으로 구분할 수 있을 때, 각 계층으로부터 단순임의 표본을 추출하는 방법
- ◆ **군집표본추출(cluster sampling):** 특정 군집들을 선정하고 이 집단들에서 단순임의표본을 추출하는 방법