

Finance Analytics

Chapter 2. 자료분석 시작하기

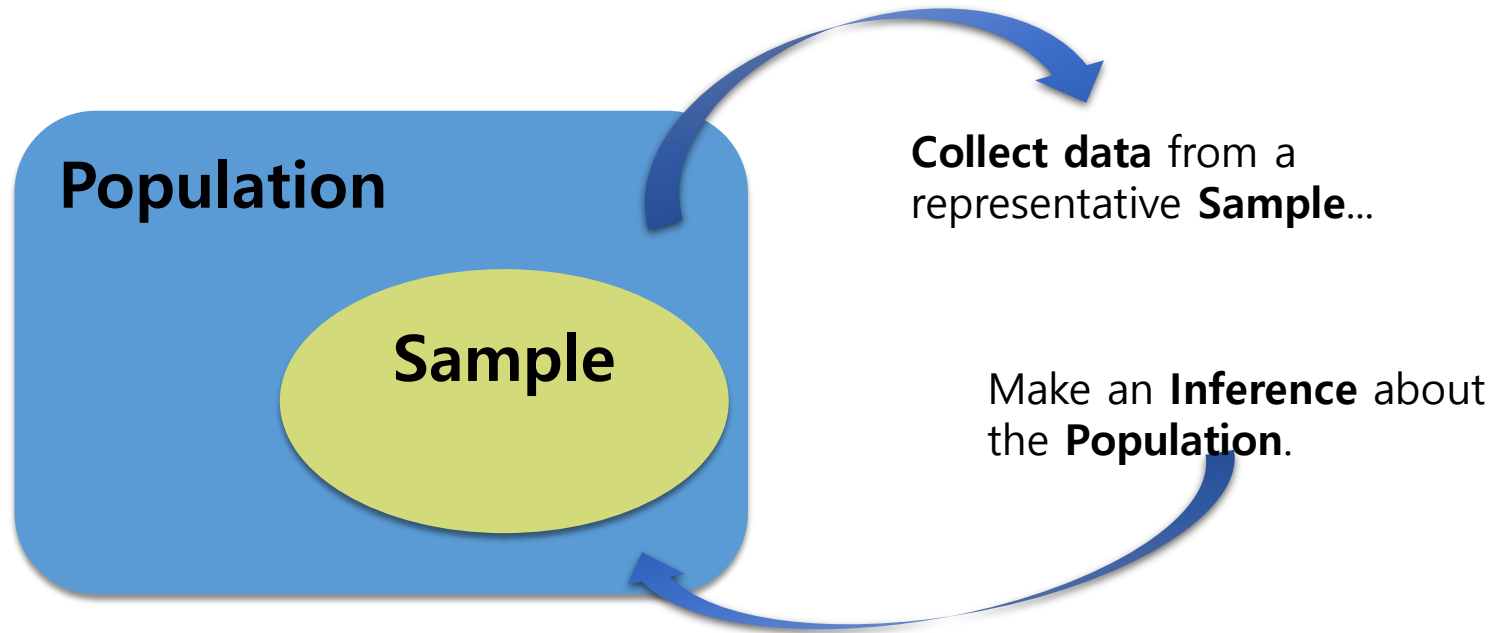
권태연

한국외대 국제금융학과

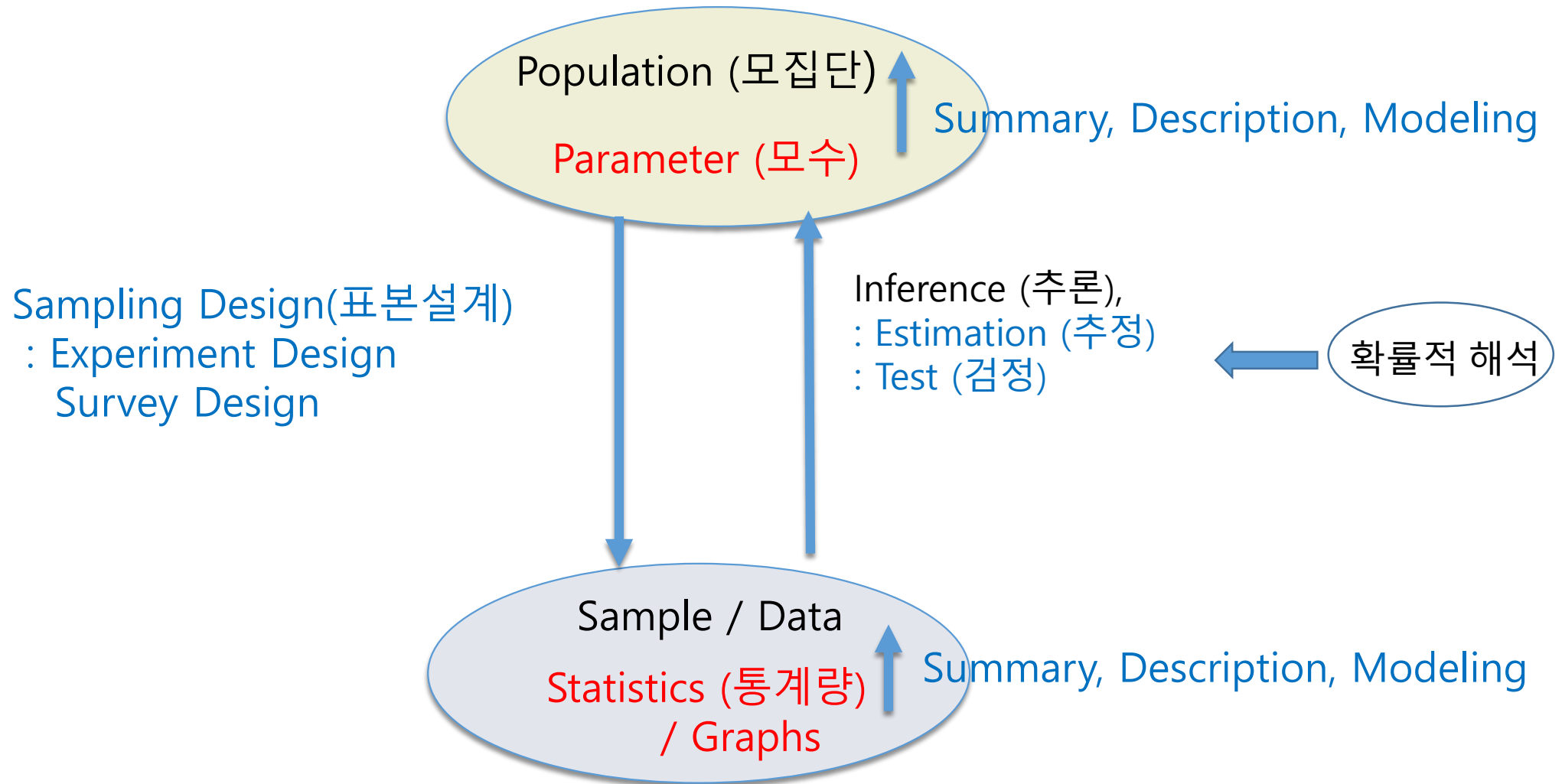
1. 용어

모집단과 표본 (Population and Sample)

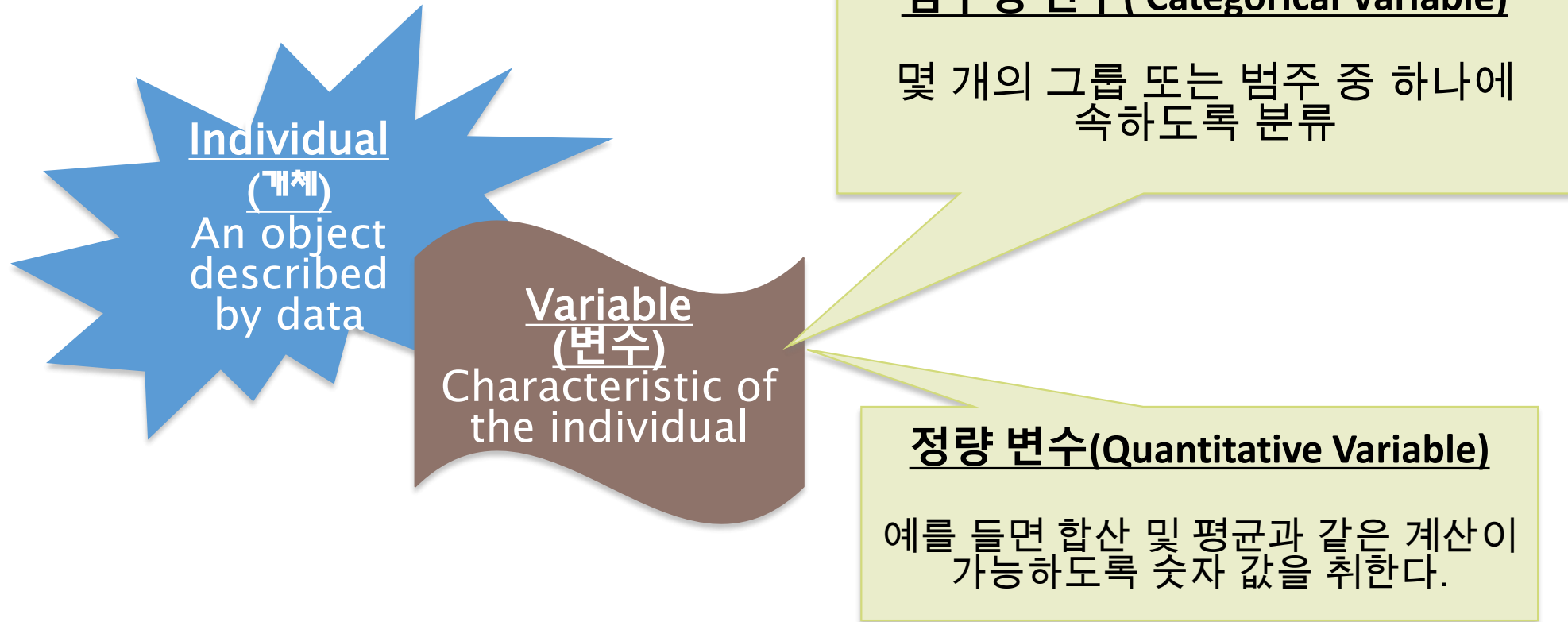
- 모집단(population): 정보를 원하는 개체들(individuals)의 전체집단
- 표본(sample): 실제로 정보를 수집한 모집단의 일부분
- 표본을 사용하여 전체 모집단에 관한 결론을 도출한다



모집단과 표본 (Population and Sample)



자료 (Data)의 구성



- 개체 (individual)는 자료 세트로 설명하는 목적물,
개체는 사람일 수도 있고 동물이거나 사물일 수도 있다.
- 변수 (variable)는 개체의 특성
변수는 상이한 개체에 대해 상이한 값을 취할 수 있다.

자료 예제 : GDP와 기대수명

**Per Capita Gross Domestic Product and Average Life Expectancy
for Countries in Western Europe**

Country	Per Capita GDP (x)	Life Expectancy (y)
Austria	21.4	77.48
Belgium	23.2	77.53
Finland	20.0	77.32
France	22.7	78.63
Germany	20.8	77.17
Ireland	18.6	76.39
Italy	21.5	78.51
Netherlands	22.0	78.15
Switzerland	23.8	78.99
United Kingdom	21.2	77.37

자료 예제 : 주식가격과 시험점수

	A	B	C	D	E	F
1	DATE	Kospi200	SP500	Nikkei225	SAMSUNG	SHINHANF
2	2009. 04	176	872.81	8828.26	592000	31700
3	2009. 05	178.7	919.14	9522.5	558000	31450
4	2009. 06	178.99	919.32	9958.44	592000	32250
5	2009. 07	202.96	987.48	10356.83	724000	41600
6	2009. 08	207.44	1020.62	10492.53	771000	41000
7	2009. 09	219.75	1057.08	10133.23	815000	47000
8	2009. 10	206.81	1036.19	10034.74	723000	46150
9	2009. 11	204.75	1095.63	9345.55	720000	45550
10	2009. 12	221.86	1115.1	10546.44	799000	43200
11	2010. 01	210.34	1073.87	10198.04	784000	41000
12	2010. 02	208.36	1104.49	10126.03	744000	41450
13	2010. 03	221.58	1169.43	11089.94	818000	44450
14	2010. 04	227.95	1186.69	11057.4	849000	47550
15	2010. 05	214.34	1089.41	9768.7	776000	43400
16	2010. 06	220.85	1030.71	9382.64	774000	45650
17	2010. 07	229.25	1101.6	9537.3	810000	48500
18	2010. 08	226.81	1049.33	8824.06	756000	45900
19	2010. 09	242.95	1141.2	9369.35	777000	43650
20	2010. 10	242.00	1102.26	9202.45	745000	42550

개체 ?
변수?
차이점?

	A	B	C
1	Student	exam	exam2
2	1	80	30
3	2	73	80
4	3	92	73
5	4	85	92
6	5	75	85
7	6	98	75
8	7	93	98
9	8	55	93
10	9	80	55
11	10	90	80
12	11	92	90
13	12	80	92
14	13	87	80
15	14	90	87
16	15	72	90
17	16	65	72
18	17	70	65

자료분석 (Data Analysis)

자료 분석 과정

0. 자료구조 파악: 개체 및 변수에 대한 파악
1. 각각의 개별 변수를 검토해 보자.
2. 변수들 간의 관계를 살펴보도록 하자 : 시각화, 수치화 할 수 있다.
3. 나아가 변수들 간의 어떠한 관계가 있는지 수량화 하자. : 계량모형

예제 :

1. GDP가 높은 국가의 국민들이 평균적으로 더 오래 사는가?
2. 흡연은 건강에 유해 한가?
3. 아버지의 키는 아들의 키에 영향을 주는가?
4. 삼성전자의 주식 수익률은 SP500지수의 수익률에 영향을 받는가?

계량모형

얼마나 영향을 주는(받는) 가?

계량모형

: 어떠한 변수가 어떠한 변수에 얼마나 영향을 주는가? 를 수량화, 수치화 함

- ✓ 변수간의 관계에 주로 방향성이 있다.
- ✓ 선/후관계 나아가 인과관계가 있다.
- ✓ 계량모형을 만드는 찾아내는 이유?
 - ✓ 현상의 설명
 - ✓ 예측하여 대비

두 변수간의 관계의 방향성에 따른 변수구분

- 반응변수 및 설명변수(Response Variables, Explanatory Variable)
 - **반응변수**(response variable)는 연구의 결과를 측정한다.
 - **종속변수** (dependent variable), **y변수**
 - **설명변수**(explanatory variable)는 반응변수의 변화를 설명하거나 또는 영향을 미친다.
 - **독립변수**(independent variable), **x변수**

예제 :

1. GDP가 높은 국가의 국민들이 평균적으로 더 오래 사는가?
2. 흡연은 건강에 유해 한가?
3. 아버지의 키는 아들의 키에 영향을 주는가?
4. 삼성전자의 주식 수익률은 SP500지수의 수익률에 영향을 받는가?

자료에서 변수의 종류

반응변수 or 종속변수 or Y	설명변수 or 독립변수 or X
정량 (& 연속형)	정량
정량 (& 연속형)	범주
범주	정량
범주	범주