

- 변수(자료) 형태

경우	X	Y
①	범주형 자료	범주형 자료
②	범주형 자료	수치자료
③	수치자료	범주형 자료
④	수치자료	수치자료

- 분석목적

- 비교: X에 따라 Y에 차이가 있는가?
- 관계: X와 Y가 관련이 있는가? X가 Y에 영향을 주는가?
- 분류: X에 따라 Y를 분류할 수 있는가?
 - X, Y가 비슷한 것끼리 묶을 수 있는가?

□ 분할표(cotingency table)

- 두 개 이상의 변수를 동시에 고려하여 각각의 범주에 관측개체의 빈도를 정리한 교차표(cross **tabulation**)
- 예】 성별(남녀)과 혈액형(A,B,AB,O) 간에 관계

성별	혈액형			
	A	B	AB	O
남자	Ⓐ			
여자				

- 칸(cell) : 각 범주에 교차되는 부분

- ◎ 세 가지 스마트폰모델에 대한 남녀별로 선호도 비교
- 남자 76명과 여자 70명을 대상으로 세 가지 모델(A,B,C) 중 가장 마음에 드는 모델을 선택
 - 데이터세트 구조

번호	성별	선호모델
1	남	A
2	남	C
3	여	B
⋮	⋮	⋮

- 남자 중 모델 A는 35, B는 23, C는 18명이 선택하고
여자 중 A는 17명, B는 33, C는 20명이 선택

성별	스마트폰모델			합계
	A	B	C	
남자	35	23	18	76
여자	17	33	20	70
합계	52	56	38	146

Q. 비율 표시는?

- 비율(상대도수)표시
 - 분모
 - 성별 합계(76, 70)
 - 스마트폰 모델별 합계(56, 56, 38)
 - 전체 합(146)
 - 비율은 분석 목적 또는 자료가 어떻게 수집되었는지에 따라 다르게 표시
 - 분석목적은 선호도에서 남녀 간 차이여부
 - ⇒ 남자 중 각각의 모델을 선호한 비율과 여자 중 각각의 모델을 선호한 비율을 비교

○ 비교를 위한 비율표시

성별	스마트폰모델			합계
	A	B	C	
남자	35 (46.0%)	23 (30.3%)	18 (23.7%)	76 (100%)
여자	17 (24.3%)	33 (47.1%)	20 (28.6%)	70 (100%)
합계	52 (35.6%)	56 (38.4%)	38 (26.0%)	146 (100%)

- 만약 146명을 무작위로 조사한 경우

성별	스마트폰모델			합계
	A	B	C	
남자	35 (24.0%)	23 (15.8%)	18 (12.3%)	76 (52.1%)
여자	17 (11.6%)	33 (22.6%)	20 (13.7%)	70 (47.9%)
합계	52 (35.6%)	56 (38.4%)	38 (26.0%)	146 (100%)

- 성별과 스마트폰 선호도 간 관련성 여부
- 차이 또는 관련 여부 확인 \Rightarrow “통계학의 이해 II”

- 수치자료를 범주화하고 분할표 작성
- ◎ 부자의 신장: 165, 170, 175, 180을 기준으로 범주화

부 자	165 미만	[165, 170)	[170, 175)	[175, 180)	180 이상	합계
180 이상	0	0	2	1	2	5
[175, 180)	1	3	6	4	1	15
[170, 175)	0	6	8	3	0	18
[165, 170)	2	4	3	1	0	10
165 미만	1	1	0	0	0	2
합계	4	14	19	9	4	50

○ k 차원 분할표(k -dimensional contingency table)

- k 개 범주형 변수에 대한 분할표

◎ 타이타닉(RMS Titanic)호 생존자와 사망자수

- 변수: 생존여부, 성별(어린이), 객실등급(승무원)

번호	성별	객실등급	생존여부
1	남	1등실	생존
2	여	승무원	사망
3	어린이	3등실	사망
⋮	⋮	⋮	⋮

그룹		남자		여자		어린이		전체		
생존여부		생존	사망	생존	사망	생존	사망	생존	사망	합
등급	1등실	57	118	140	4	5	1	202	123	325
	2등실	14	154	80	13	24	0	118	167	285
	3등실	75	387	76	89	27	52	178	528	706
	승무원	192	693	20	3	-	-	212	696	908
합계		338	1,352	316	109	56	53	710	1,514	2,224

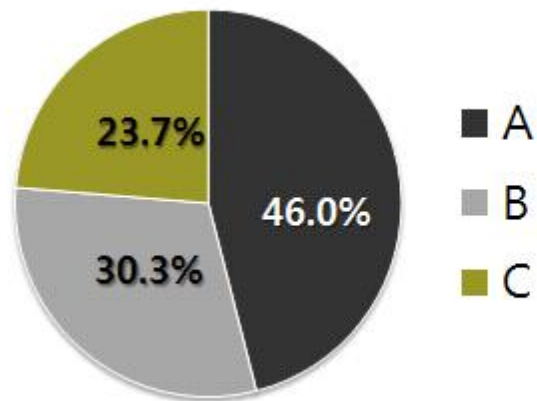
○ 생존율

- 1등실:62.2%, 2등실:41.4%, 3등실:25.2%,
승무원: 23.3%, 전체:31.9%
- 남자: 20.0%, 여자 74.4%, 어린이: 51.4%

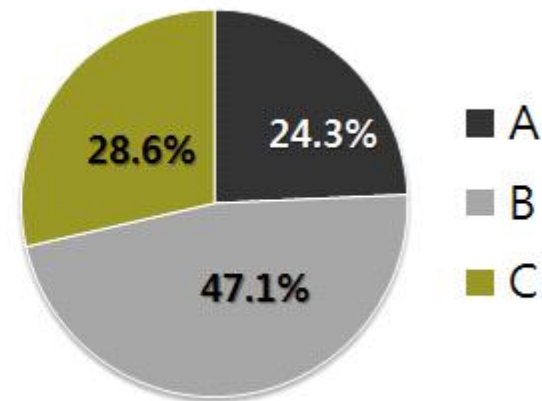
■ 그래프를 이용한 자료정리

○ 원도표(pie chart)

- 비교 그룹별로 파이차트 작성



남자



여자

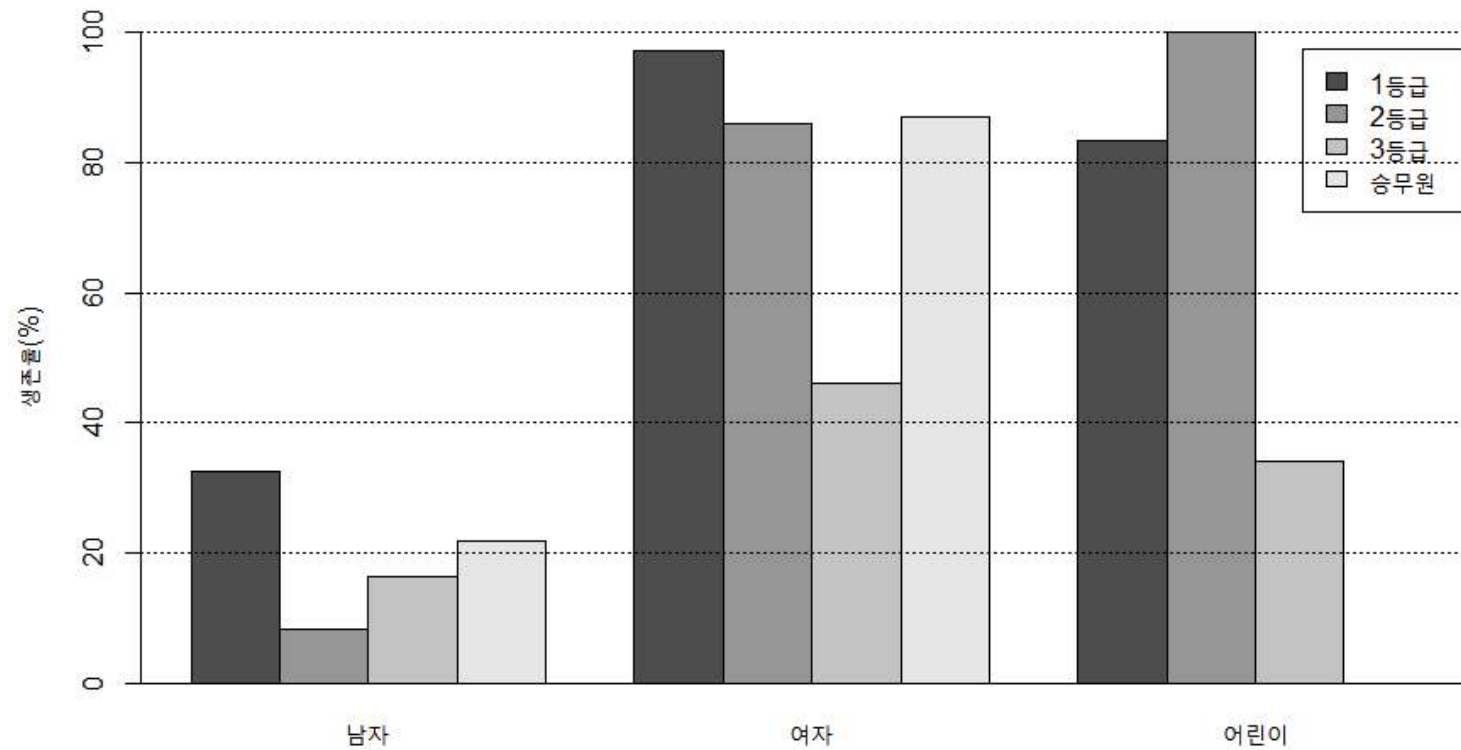
○ 막대그래프(Bar chart)

- 분석목적과 자료형태에 맞게 빈도/비율 중 선택

● 타이타닉호 생존율

- 등실과 성별에 따른 생존율(%)

	남자	여자	어린이
1등실	32.6	97.2	83.3
2등실	8.3	86.0	100
3등실	16.2	46.1	34.2
승무원	21.7	87.0	-



- 생존율
 - 1등실 62.2%, 2등실 41.2%, 3등실 25.2%, 승무원 23.3%
 - 남자: 20.0%, 여자 74.4%, 어린이: 51.4%

- 남녀로 구분했을 때, 승무원의 생존율은 1등실을 제외한 나머지 등실의 승객보다 생존율이 높음

- 남성비율
 - 1등실 54.9%, 2등실 64.4%, 3등실 73.7%, 승무원 97.5%

- 정리
 - 분할표: 비율표시
 - k 차원 분할표
 - 분류 변수에 따른 원도표, 막대그래프