

Chapter 2

Descriptive Statistics I:

Graph and Table

경영대학 재무금융학과
윤선중

0

Objectives

- 기술통계방법 (Descriptive Statistical Method)
 - 그래프 기법 (Graphical Technique)
 - 수치 방법 (Numerical Technique)
- 범주데이터
 - 막대그래프 (Bar chart), 파이차트(Pie chart), 도수분포 (Frequency distribution)
- 구간데이터
 - 히스토그램 (Histogram)
 - 대칭성(symmetry), 최빈 계급의 수(number of modes), 종모양 (Bell shape)
- 횡단면 데이터와 시계열 데이터
- 교차분류표, 막대 그래프, 산포도 등의 사용

1

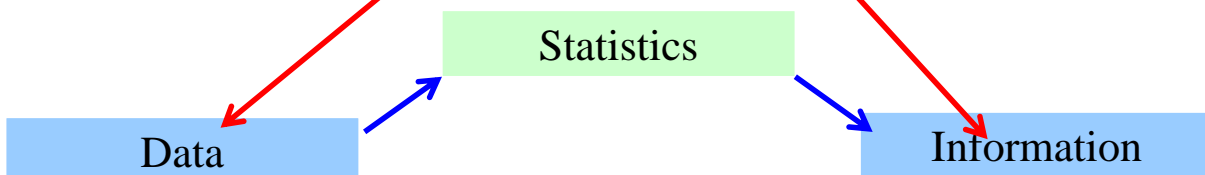
I. Basic Concept

2

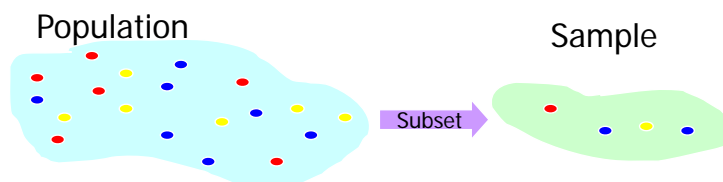
Definition

■ 기술통계학

- 기술통계학이란 Data Set으로부터 의미있는 정보를 추출하기 위하여 데이터를 정리하고, 통합하고, 나타내는 일련의 과정



- 이 과정에서 Graphical technique이나 numerical measure가 사용됨
- 본 강의자료는 Graphical Technique에 대해 살펴봄
- 모집단과 표본집단 모두에서 사용될 수 있음.



3

Terminology

■ 변수 (Variable)

- 모집단 또는 표본집단의 특성

■ 관측치 (Observation)

- 변수가 가지는 하나의 측정값

■ 데이터(Data)

- 한 변수에 대한 관측치들의 집합

■ 예제

- 재무계량분석 I을 수강하는 세 학생의 시험점수는 70, 90, 30점
- 과거 3일간 KOSPI 지수는 1100, 1040, 1045 포인트

4

Type of Data

■ 구간 데이터 (Interval Data)

- 실수 값들로 이루어진 관측치들의 집합
- 정량 데이터(Quantitative data), 수치 데이터 (Numerical data)
- 높이, 무게, 소득, 거리, 점수 등

■ 범주 데이터 (Nominal Data)

- 범주를 나타내는 값들로 이루어진 관측치들의 집합
- 정성 데이터 (Qualitative data), 카테고리 데이터 (Categorical data)
- 결혼상태 (기혼/미혼), 취업분야 (금융/마케팅/영업/생산 등)

■ 서열 데이터 (Ordinal Data)

- 범주를 나타내지만 순서를 가지는 값들로 이루어진 관측치들의 집합
- 강의평가 (매우우수/우수/보통/부실/매우부실), 거래량 상위 3개 종목 (삼성/포스코/국민은행)

5

Calculation of Data

■ 구간데이터 (Interval Data)

- 값들의 차이에 일관성이 유지됨
- 모든 산술 계산이 가능 (평균/분산 등)
- 범주 데이터 및 서열 데이터로 전환 ____

■ 범주데이터(Nominal Data)

- 발생도수(빈도)에 기초한 계산만이 가능
- 구간데이터 또는 서열데이터로 전환 ____

■ 서열데이터(Ordinal Data)

- 값들의 차이에 일관성이 유지되지는 않음: A, B, C 학생의 성적이 각각 우수, 보통, 미흡이라고 하면?
- 순위를 유지하는 계산만 가능
- 범주데이터로는 전환 ____; 구간데이터로는 전환 ____

6

II. Graph and Table: Univariate

7

Using Graph & Table for Nominal Data-도수분포표

■ 기본 용어

- 도수 (frequency): 각 범주에 속하는 데이터의 개수
- 도수분포(frequency distribution): 분류된 범주 별로 도수를 대응시킨 것
- 상대도수분포(relative frequency distribution): 분류된 범주 별로 도수의 비율을 대응시킨 것

■ 예제 2.1: 라이트 비어 선호도 서베이

- Xm02-01 참조
- 맥주회사 담당자는 라이트비어를 마시는 전문대학과 대학 학생들에 대한 라이트비어 판매량을 분석하였다
 1. Budweiser Light
 2. Busch Light
 3. Coors Light
 4. Michelob Light
 5. Miller Lite
 6. Natural Light
 7. Other brand

	A	B	C	D
1	Number	Brand	Gender	
2	1	1	2	
3	2	1	1	
4	3	5	2	
5	4	1	1	
6	5	3	1	
7	6	3	1	
8	7	3	2	
9	8	7	1	
10	9	2	1	
11	10	6	1	
12	11	1	2	
13	12	6	1	

8

Using Graph & Table for Nominal Data-도수분포표

Light Beer Brand	Frequency	Relative Frequency
Budweiser Light	90	31.6%
Busch Light	19	6.7
Coors Light	62	21.8
Michelob Light	13	4.6
Miller Lite	59	20.7
Natural Light	25	8.8
Other brands	17	6.0
Total	285	100

9

Using Graph & Table for Nominal Data-도수분포표

■ Excel 이용

● COUNTIF([Input Range], [Criteria])이용

- Input Range:: B1:B286
- COUNT(B1:B286, 1)

1)

F	G	H
	도수	상대도수
1. Budweiser Light	=countif(b2:b286,1)	
2. Busch Light		
3. Coors Light		
4. Michelob Light		
5. Miller Lite		
6. Natural Light		
7. Other brand		

2)

F	G	H	I
	도수	상대도수	
1. Budweiser Light	90	=G3/sum(G3:G9)	
2. Busch Light	19		
3. Coors Light	62		
4. Michelob Light	13		
5. Miller Lite	59		
6. Natural Light	25		
7. Other brand	17		

F	G	H	I
	도수	상대도수	
1. Budweiser Light	90	0.32	
2. Busch Light	19	0.07	
3. Coors Light	62	0.22	
4. Michelob Light	13	0.05	
5. Miller Lite	59	0.21	
6. Natural Light	25	0.09	
7. Other brand	17	0.06	
		1.00	

10

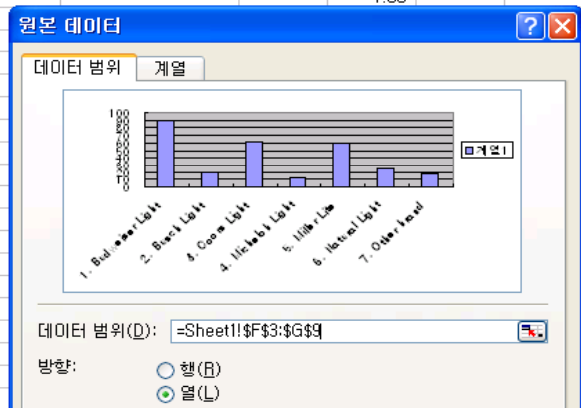
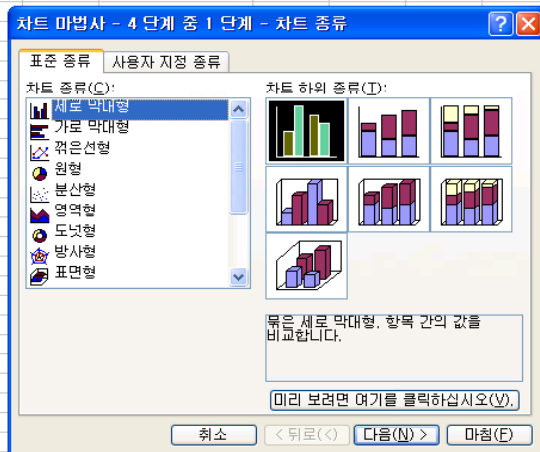
Using Graph & Table for Nominal Data-막대그래프 (Frequency)

■ 막대그래프

● 막대모양의 Bar를 이용해 빈도를 표시하는 그래프

E	F	G	H	I	J
		도수	상대도수		
	1. Budweiser Light	90	0.32		
	2. Busch Light	19	0.07		
	3. Coors Light	62	0.22		
	4. Michelob Light	13	0.05		
	5. Miller Lite	59	0.21		
	6. Natural Light	25	0.09		
	7. Other brand	17	0.06		

E	F	G	H	I	J
		도수	상대도수		
	1. Budweiser Light	90	0.32		
	2. Busch Light	19	0.07		
	3. Coors Light	62	0.22		
	4. Michelob Light	13	0.05		
	5. Miller Lite	59	0.21		
	6. Natural Light	25	0.09		
	7. Other brand	17	0.06		
			1.00		



11

Using Graph & Table for Nominal Data- 막대그래프

	E	F	G	H	I	J
			도수	상대도수		
		1. Budweiser Light	90	0.32		
		2. Busch Light	19	0.07		
		3. Coors Light	62	0.22		
		4. Michelob Light	13	0.05		
		5. Miller Lite	59	0.21		
		6. Natural Light	25	0.09		
		7. Other brand	17	0.06		
				1.00		

차트 마법사 - 4 단계 중 3 단계 - 차트 옵션

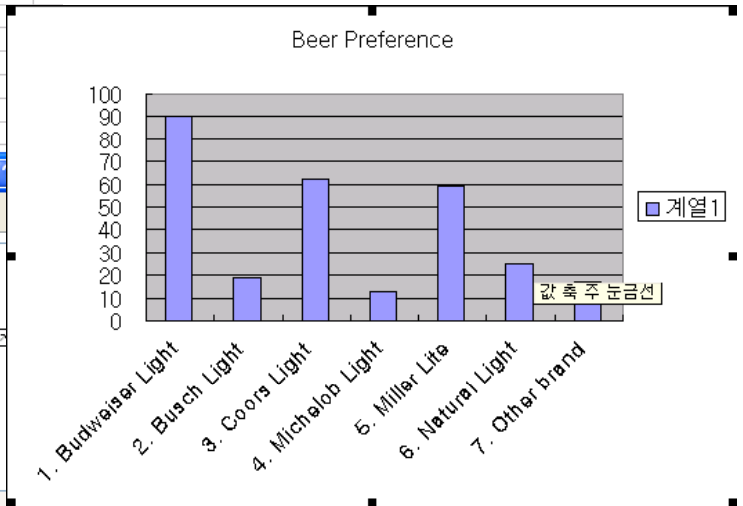
제목: Beer Preference

X(항목) 축(C):

Y(값) 축(V):

보조 X(항목) 축(X):

보조 Y(값) 축(Y):



Using Graph & Table for Nominal Data- 원그래프 (Relative Frequency)

원그래프

- 원형 모양으로 상대빈도를 나타내기 위해 유용한 그래프

	E	F	G	H	I	J
			도수	상대도수		
		1. Budweiser Light	90	0.32		
		2. Busch Light	19	0.07		
		3. Coors Light	62	0.22		
		4. Michelob Light	13	0.05		
		5. Miller Lite	59	0.21		
		6. Natural Light	25	0.09		
		7. Other brand	17	0.06		
				1.00		

	도수	상대도수
1. Budweiser Light	90	0.32
2. Busch Light	19	0.07
3. Coors Light	62	0.22
4. Michelob Light	13	0.05
5. Miller Lite	59	0.21
6. Natural Light	25	0.09
7. Other brand	17	0.06
		1.00

차트 마법사 - 4 단계 중 1 단계 - 차트 종류

표준 종류 사용자 지정 종류

차트 종류(C):

차트 하위 종류(I):

원형 전체에 대한 각 값의 기여도를 보여줍니다.

미리 보려면 여기를 클릭하십시오(V).

취소 < 뒤로(B) 다음(N) > 마침(E)

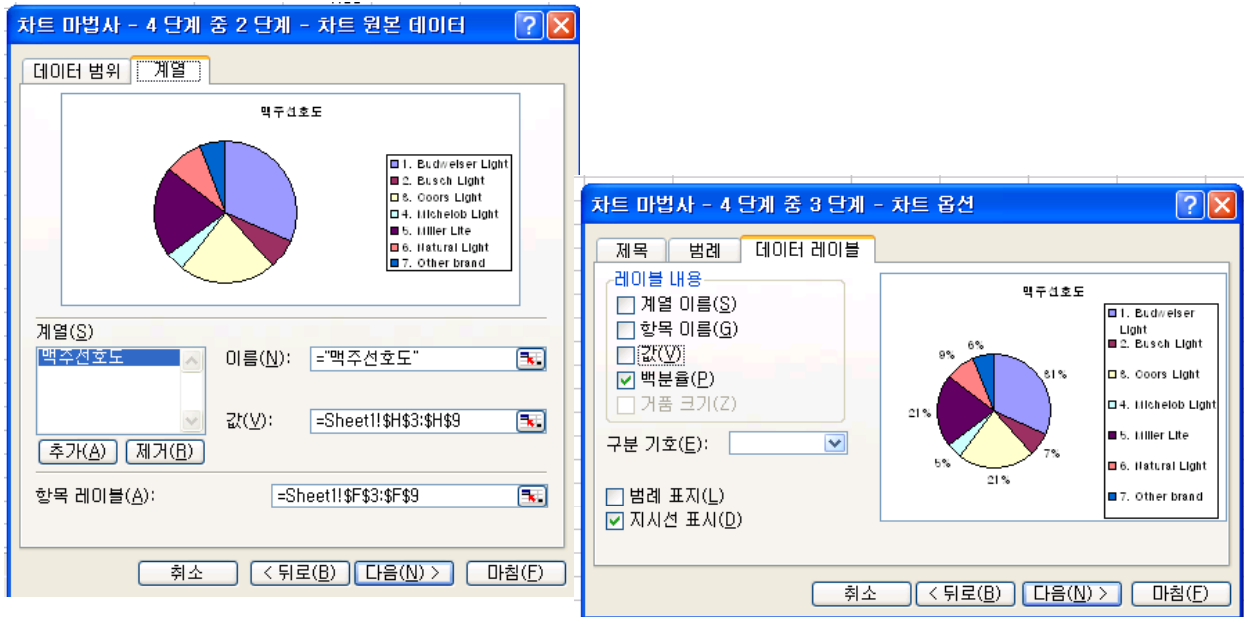
원본 데이터

데이터 범위: =Sheet1!\$H\$3:\$H\$9

방향: 행(R) 열(L)

취소 < 뒤로(B) 다음(N) > 마침(E)

Using Graph & Table for Nominal Data- 원그래프



14

Using Graph & Table for Interval Data - 도수분포표

- 구간데이터의 도수분포 작성
 - Step 1: 최대값과 최소값을 확인
 - Step 2: 최대값과 최소값의 차를 구함
 - Step 3: 전체 계급의 개수를 결정
 - Step 4: 범위를 계급의 개수로 나누어서 계급 구간 계산
 - Step 5: 각 계급의 도수 및 상대도수 계산

■ 예제 2.4: 장거리 전화비용의 분석

- Xm02-04 참조
- 장거리 전화회사가 신규계약자의 첫 달 장거리 전화비용에 관한 정보를 추출

	A	B	C
1	Bills		
2	42.19		
3	38.45		
4	29.23		
5	89.35		
6	118.04		
7	110.46		
8	0.00		
9	72.88		
10	83.05		
11	95.73		
12	103.15		
13	94.52		

15

도수분포표

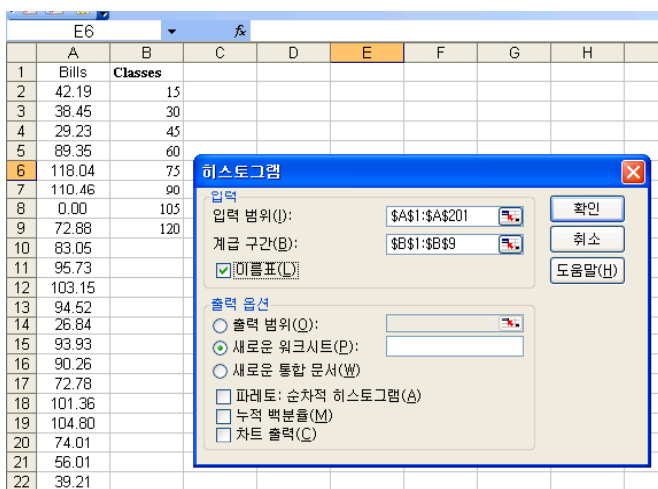
Classes

0~15
15~30
30~45
45~60
60~75
75~90
90~105
105~120

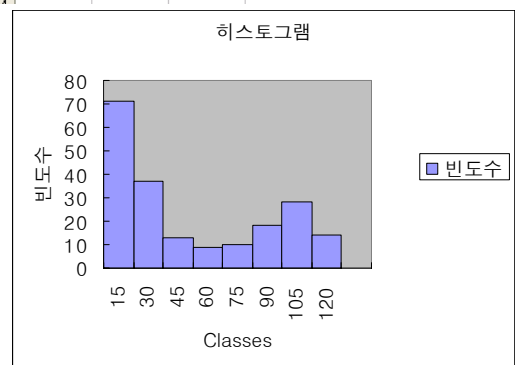
■ Excel 이용

- 각 계급의 상한값 (계급구간; bins)을 입력
- 도구 (데이터) - 데이터분석 - 히스토그램 클릭
- 입력범위 A1-A201
- 계급구간 B1-B9
- 차트출력

도수분포표 & 히스토그램



	A	B	C	D
1	Classes	빈도수		
2	15	71		
3	30	37		
4	45	13		
5	60	9		
6	75	10		
7	90	18		
8	105	20		
9	120	14		
10	기타	0		
11				
12				
13				



도수분포표

■ 계급구간

- 계급 구간의 개수에 대한 엄밀한 규정은 없음
- **Sturges' Formula**
 - 계급구간의 개수 = $1 + 3.3 \log(n)$
 - 단, \log 는 밑이 10인 상용로그

■ 앞의 예제

- 범위
 - 최대값 119.63 ; 최소값 0 : 범위=119.63
- 계급구간의 개수
 - $1+3.3\log(200)=8.59 \approx 9$

18

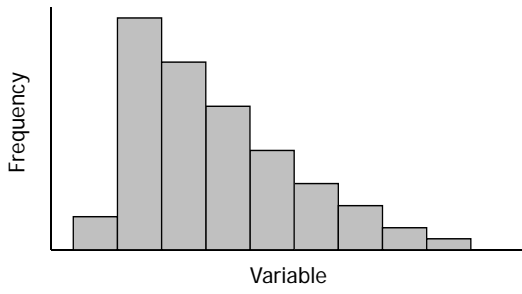
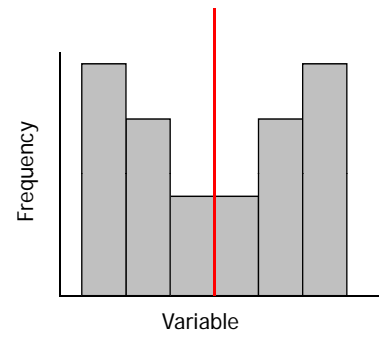
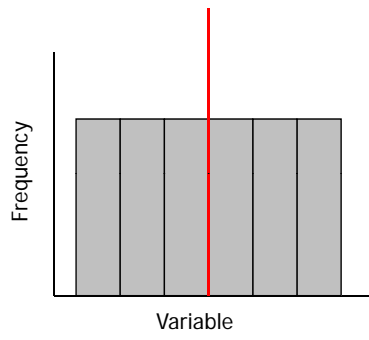
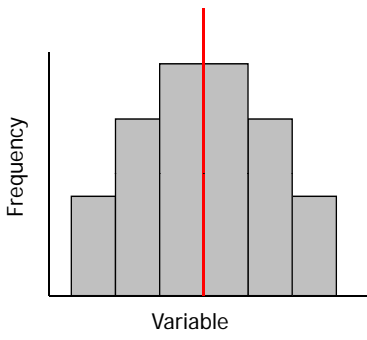
히스토그램

■ 히스토그램의 유형

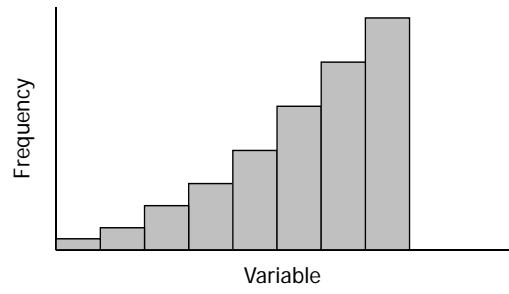
- **대칭 (Symmetric) vs. 비대칭 (Non-Symmetric)**
 - 비대칭의 척도: 왜도 (Skewness)
 - 양의 왜도 (Positively Skewed): _____으로 긴 꼬리
 - 음의 왜도 (Negatively Skewed): _____으로 긴 꼬리
- **종모양 (Bell shaped) vs. 비 종모양**
 - 첨도 (Kurtosis): 특히 꼬리의 두께를 판단하는데 이용
- **단봉 (Unimodal) vs. 양봉 (Bimodal)**
 - 봉우리를 나타내는 계급의 수에 따른 분류
- 예제
 - Xm 02-05: 투자수익률의 비교
 - Xm 02-06: 경영통계학 점수
 - Xm 02-07: 수리통계학 점수

19

히스토그램



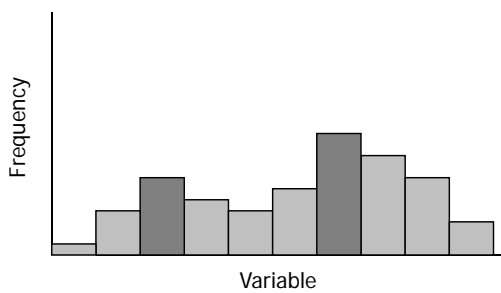
Positively Skewed



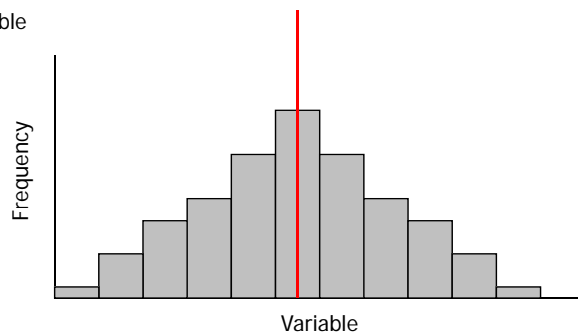
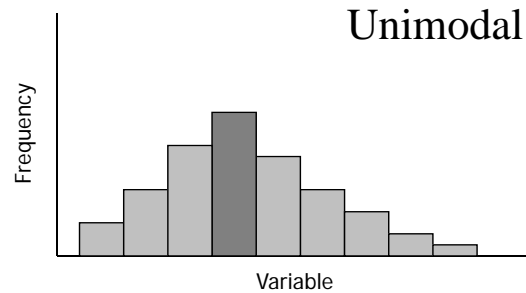
Negatively Skewed

히스토그램

Bimodal



Unimodal



Bell Shaped

줄기-잎 그림

■ 줄기-잎 그림 (Stem-and-leaf display)

- 자료의 줄기와 잎 두부분으로 분리하여 이를 모두 표현하는 방법
- 자료의 수가 적은 경우에 효율적이며, 막대 그래프보다 많은 정보를 담고 있음
- 즉, 해당 구간의 관측치 개수 뿐만 아니라 개별 관측치들이 얼마인가에 관한 정보도 포함하고 있음

■ 예시

- 데이터: 11,13,25,28,29,26,32,14,35
- 줄기: 10의 자리수, 잎: 1의 자리수
- 그래프
 - 1:134
 - 2:5689
 - 3:25

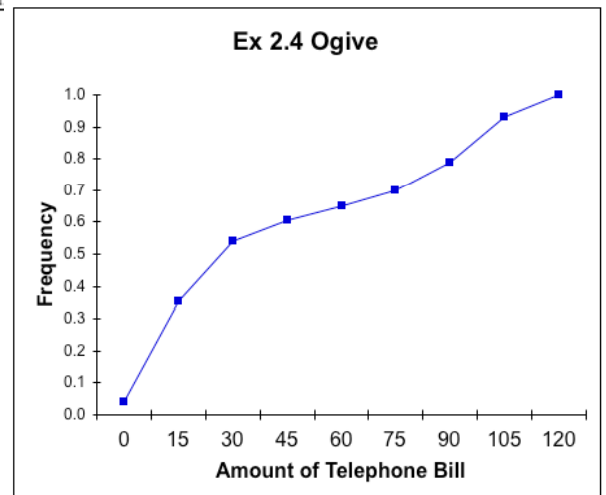
누적 백분율 곡선

■ 상대도수를 이용하여 누적되는 백분율 곡선을 표현

■ 몇 번째 사용자의 Phone Bill에 대한 정보가 한눈에 들어옴

Table 2.8 Cumulative Relative Frequency Distribution for Example 2.4

Class Limits	Relative Frequency	Cumulative Relative Frequency
0 to 15	$71/200 = .355$	$71/200 = .355$
15 to 30	$37/200 = .185$	$108/200 = .540$
30 to 45	$13/200 = .065$	$121/200 = .605$
45 to 60	$9/200 = .045$	$130/200 = .650$
60 to 75	$10/200 = .05$	$140/200 = .700$
75 to 90	$18/200 = .09$	$158/200 = .790$
90 to 105	$28/200 = .14$	$186/200 = .930$
105 to 120	$14/200 = .07$	$200/200 = 1.00$



III. Time Series Data

24

Overview

■ 관측시점에 따른 데이터의 분류

- 횡단면 데이터 (cross-sectional data)
 - 동일한 시점에서 측정된 데이터
- 시계열 데이터 (time-series data)
 - 연속적인 시점들에서 측정된 데이터

■ 예제

- 오늘 종목별 주식가격: ____ 데이터
- 과거 2년간 삼성전자의 주식가격: _____ 데이터

■ 통계적 검증과 데이터

- 검증하고자 하는 내용에 적합한 데이터를 사용하여야 함
- 예 1) “몸무게가 많이 나갈수록 술을 잘 마신다” 는 명제를 검증하기 위해서는 ____ 데이터 사용
- 예 2) “이자율은 평균으로 수렴하는 경향이 있다” 는 명제를 검증하기 위해서는 ____ 데이터 사용

25

Line Chart for Time Series Data

Ex 2.8: 가솔린 가격의 변화; Xm02-08

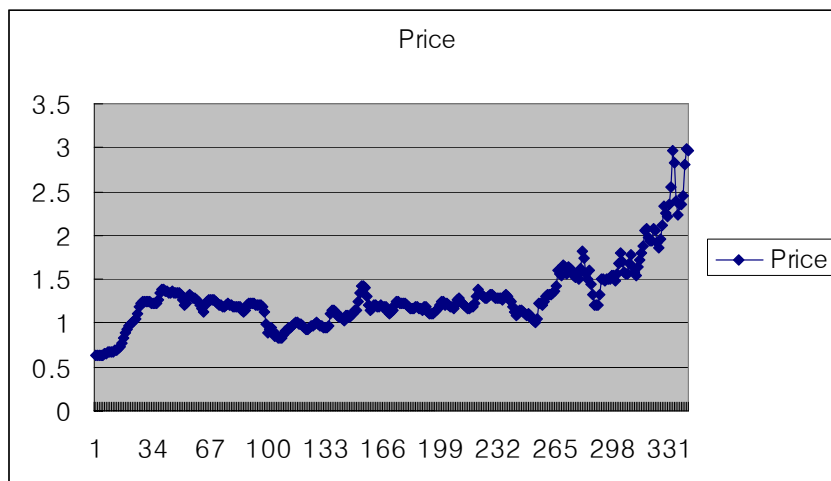
The screenshot shows Microsoft Excel with a data table and two dialog boxes. The data table is as follows:

Year	Month	Price
1978	1	0.631
1978	2	0.629
1978	3	0.629
1978	4	0.631
1978	5	0.637
1978	6	0.645
1978	7	0.655
1978	8	0.663
1978	9	0.669
1978	10	0.671
1978	11	0.676
1978	12	0.685
1979	1	0.695
1979	2	0.707
1979	3	0.733
1979	4	0.78
1979	5	0.823
1979	6	0.88
1979	7	0.93
1979	8	0.967
1979	9	0.998
1979	10	1.006
1979	11	1.019

The '차트 마법사' dialog box shows '표준 종류' (Standard Types) with '선' (Line) selected. The '원본 데이터' dialog box shows the data range as '=Sheet1!\$C\$1:\$C\$343' and '방향' (Orientation) set to '열(L)' (Columns).

26

Line Chart for Time Series Data



27

IV. Graph and Table: Bivariate

Using Graph & Table for Nominal Data - 분할표

■ 분할표(contingency table)

- 두 종류의 범주 데이터를 행과 열로 교차시켜 표를 만들고, 표의 각 칸에 도수(빈도)를 기입한 표

■ 예제2.10: 신문독자 서베이

- Xm 02-10
- 경쟁상태에 있는 4개의 신문
Globe and Mail(1), Post (2), Star (3), Sun(4)
구독신문과 직업과의 관계를 분석
블루칼러 (1), 화이트칼러 (2)
전문직종사자 (3)

1	Reader	Occupation	Newspaper
2	1	2	2
3	2	1	4
4	3	2	1
5	4	3	2
6	5	1	3
7	6	3	3
8	7	2	1
9	8	1	3
10	9	3	3
11	10	3	1
12	11	3	2
13	12	1	3
14	13	3	2
15	14	1	3
16	15	1	3
17	16	1	1
18	17	3	2
19	18	3	3
20	19	2	2
21	20	1	2
22	21	1	4

Using Graph & Table for Nominal Data - 분할표

■ 피벗 테이블 이용

	A	B	C	D	E
1	Reader	Occupation	Newspaper		
2	1	2	2		
3	2	1	4		
4	3	2	1		
5	4	3	2		
6	5	1	3		
7	6	3	3		
8	7	2	1		
9	8	1	3		
10	9	3	3		
11	10	3	1		

Using Graph & Table for Nominal Data - 분할표

피벗 테이블/피벗 차트 마법사 - 3단계 중 2단계

사용할 데이터가 있는 영역을 입력하거나 선택합니다.
범위(B): Sheet1!\$A\$1:\$C\$355

피벗 테이블/피벗 차트 마법사 - 3단계 중 3단계

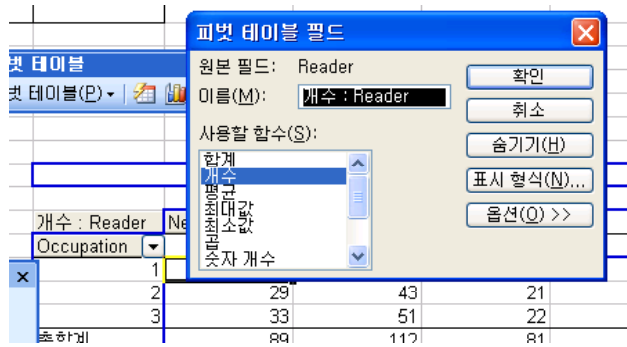
피벗 테이블 보고서 작성 위치:
 새 워크시트(N)
 기존 워크시트(E)
 Sheet1!\$F\$4

[마침]을 클릭하면 피벗 테이블 보고서가 작성됩니다.

피벗 테이블 필드 목록

- 피벗 테이블 보고서로 항목
- 피벗 테이블
- Reader
- Occupation
- Newspaper

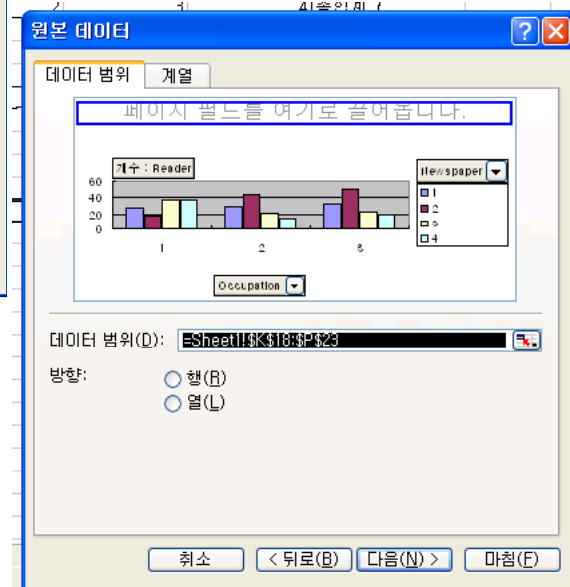
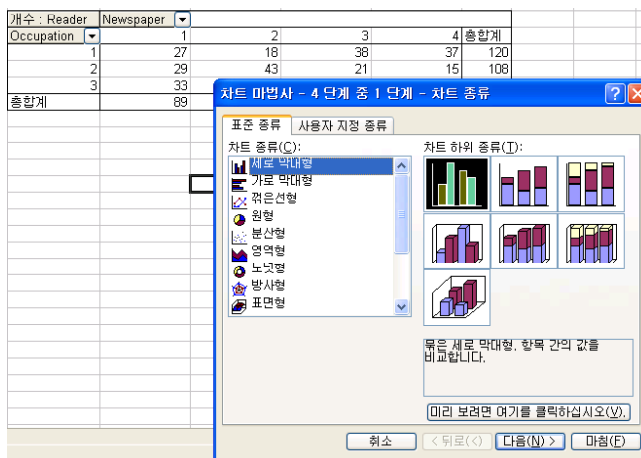
Using Graph & Table for Nominal Data - 분할표



개수 : Reader	Newspaper	1	2	3	4	총합계
Occupation	1	27	18	38	37	120
	2	29	43	21	15	108
	3	33	51	22	20	126
총합계		89	112	81	72	354

32

Using Graph & Table for Nominal Data - 막대그래프



33

Using Graph & Table for Interval Data - 산점도

■ 산점도(Scatter Diagram)

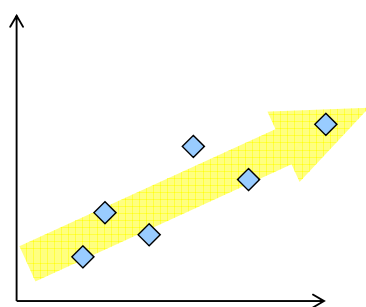
- 상관관계를 나타내기 위한 도표의 하나
- 가로축과 세로축에 서로 다른 두 변수를 설정하고, 각각 (x, y)에 해당하는 점을 도표상에 표현하여 상관관계를 나타냄

■ 상관관계의 종류

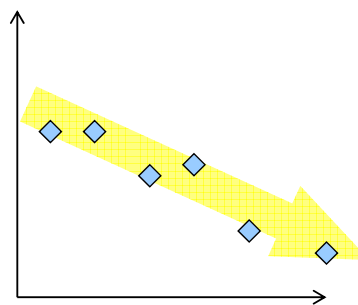
- 선형관계 (linear relationship)
 - 양 (positive)
 - 음 (negative)
- 비 선형관계
 - 이차 관계(quadratic relationship)
 - 지수관계(exponential relationship)

34

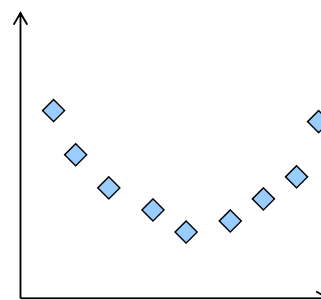
Using Graph & Table for Interval Data - 산점도



Positive Linear Relationship



Negative Linear Relationship



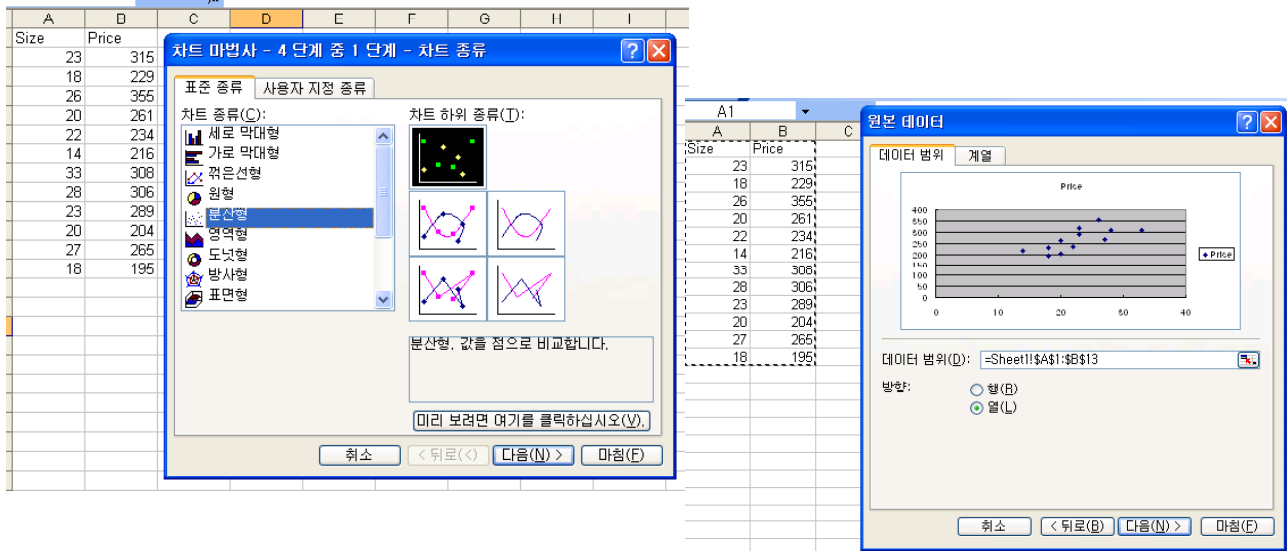
Weak or Non-Linear Relationship

35

Using Graph & Table for Interval Data - 산점도

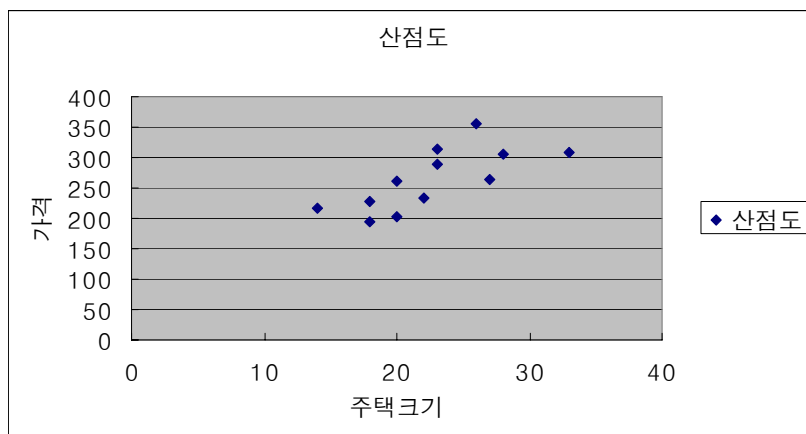
■ 예제 2.12: 주택가격과 주택크기의 관계 분석

● Xm02-12



36

Using Graph & Table for Interval Data - 산점도



37

Summary

	구간 데이터	범주데이터
하나의 데이터 셋	히스토그램 (Histogram)	도수분포표(Frequency and Relative Frequency Tables), 막대 및 원 그래프(Bar and Pie Charts)
두개의 데이터 셋	산점도 (Scatter Diagram)	분할표(Cross-classification Table) 막대그래프(Bar Charts)