

*Saluton
Denove!*

확률 및 통계학

동국대학교

2025년 2학기

문동휘

월 12 ~ 13:30

수 11 ~ 12:30

자료의 요약

- ▶ 통계학을 하기 위한 기본:
 1. 통계학을 적용할 자료 구하기
 2. 구한 자료를 목적에 따라 분류/정리
- ▶ 질적 자료
 - ▶ 범주나 속성으로 자료표현
 - ▶ 범주형 자료
 - ▶ 성별 (남, 여), 학년 (1학년, 2학년, 3학년, 4학년), 나라(미국, 한국, 필리핀, 러시아)

자료의 요약

- ▶ 양적 자료
 - ▶ 관측결과 숫자로 표현
 - ▶ 무게, 연령, 길이 등
- ▶ 양적 자료와 질적 자료의 경계
 - ▶ 무게를 **50kg**으로 측정하면 양적자료, **50**키로 미만, **50**부터 **70**키로 이하, **70**키로 이상 등 **범주**로 나누면 질적자료다.

자료의 요약

- ▶ 양적 자료의 구분
 - ▶ 이산형 자료: 가족 구성원의 수, 한 개인이 소유한 주식의 개수, 한 나무가 가진 나뭇잎 개수 등은 이산형 자료이다.
 - ▶ 연속형 자료: 시간, 속도, 무게 등
 - ▶ 핵심: 셀 수 있으면 이산형 자료, 셀 수 없으면 연속형 자료

자료 구분법: 측정척도

- ▶ 명목척도(Nominal scale), 순위척도(Ordinal scale), 구간척도(Interval scale), 비율척도(Ratio scale)
- ▶ 명목척도
 - ▶ 범주로 구분되는 자료를 측정할 때 사용
 - ▶ 범주 사이에 순서나 본질적으로 순위가 있지 않음 (예: 성별, 혈액형, 나라 등)
 - ▶ 각 범주를 숫자로 나타내기도 하나 그 숫자는 의미없음.
 - ▶ 예: 여자를 0으로, 남자를 1로 나타낼 수도, 여자를 -100으로 남자를 +100으로 나타내거나 그 반대가 되게 할 수도 있음. 여기서 이 숫자들은 여자, 남자라는 범주를 상징하는 거 외에 다른 그 어떤 의미도 부여되어 있지 않음.
 - ▶ 따라서 그 어떤 연산(덧셈, 뺄셈, 곱셈, 나눗셈 등)도 불가능.

자료 구분법: 측정척도

- ▶ 순위척도
 - ▶ 범주 + 순위/순서
 - ▶ 예) [매우 불만족, 불만족, 보통, 만족, 매우 만족], 학점 [A, B, C, D, F]
 - ▶ 명목척도와 마찬가지로 4칙 연산 불가능
- ▶ 구간척도
 - ▶ 측정된 관측치 간의 차이가 양적으로 계산될 수 있는 순위척도
 - ▶ 크기의 순서, 변수 간의 차이에 대한 분석은 의미가 있음.
 - ▶ 그러나 곰셈이나 나눗셈 같은 비율에 대한 분석은 의미 없음
 - ▶ 예) 섭씨 40도가 섭씨 20도 보다 2배 더 덥다는 의미는 부여할 수 없음. 20도 더 덥다고 할 수는 있음.

자료 구분법: 측정척도

- ▶ 비율척도
 - ▶ 절대 0도, 절대 기준점이 있는 척도. 즉 0의 의미가 없다는 개념이 포함된 자료.
 - ▶ 예) 켈빈 온도, 킬로그램, 센치미터 등
 - ▶ 4칙 연산 다 가능.
- ▶ 결론: 자료를 보고 그것이 어떤 척도에 해당하는지 파악, 그 후에 분석방법을 적용하기.
- ▶ 때로는 어떤 분석방법을 염두해두고 자료를 구할 때 그 분석방법에 맞는 척도로 계산된 자료를 구한다.

자료의 정리

- ▶ 정리되지 않은 자료 = 원자료 또는 Raw Data
- ▶ 이런 원자료를 정리하고 파악하기 위하여 여러 표나 그래프 등을 이용한다.
- ▶ 자료의 정리에 사용되는 표: 도수분포표
- ▶ 예) 관심 집단의 나이

45	46	64	57	85
92	51	71	54	48
27	66	76	55	69
54	44	54	75	46
61	68	78	61	83
88	45	89	67	56
81	58	55	62	38
55	56	64	81	38
49	68	91	56	68
46	47	83	71	62

자료의 정리

- ▶ 도수분포표로 정리한 자료

Class limits	Tally	Frequency
27–35	/	1
36–44	///	3
45–53		9
54–62		15
63–71		10
72–80	///	3
81–89	//	7
90–98	//	2
		50

- ▶ 자료의 특성파악을 용이하게 한다
- ▶ 만약 연속형 자료이거나 수많은 관측값들이 있는 이산형 자료라면 자료를 몇 개의 그룹으로 나누어 도수분포표를 만든다.

자료의 정리

- ▶ 자료를 계급으로 나누는 원칙
 1. 동일한 간격을 가진 계급들
 2. 계급이 겹치면 안 됨.
 3. 모든 자료는 하나의 계급에 속해야 함
- ▶ 급하한 = 각 계급의 아래에 있는 값 (계급 구간에서 가장 작은 값) $\Rightarrow 36 - 44 : 36$
- ▶ 급상한 = 각 계급의 위에 있는 값 (계급 구간에서 가장 큰 값) $\Rightarrow 36 - 44 : 44$
- ▶ 급간격 = n 번째 계급 급상한 - ($n - 1$) 번째 계급 급상한
- ▶ 예) 옆 표에서 급간격은 $44 - 35 = 80 - 71 = 9$ 이다.

Class limits	Tally	Frequency
27–35	/	1
36–44	///	3
45–53		9
54–62		15
63–71		10
72–80	///	3
81–89	//	7
90–98	//	2
		50

자료의 정리

- ▶ Sturges 법칙
 - ▶ 근사적으로 필요한 계급의 수를 결정하는데 사용하는 공식.
 - ▶ 원래는 Histogram의 Bin(계급)수를 결정하기 위해 나옴.
 - ▶ Binomial Approximation to Normal Distribution 사용하기 때문에 만능은 아님.
 - ▶ 늘 자료를 요약하는 목적을 먼저 생각하고, 그것에 맞게 자료를 나누는 계급을 설정하는 것이 바람직함.

자료의 정리

- ▶ Sturges 법칙
 - ▶ $C = \underline{3.3(\log_{10} n) + 1}$
 - ▶ 여기서 C 는 계급의 수, n 은 자료의 수이다.
 - ▶ C 는 가장 가까운 정수로 반올림 한다.
- ▶ 범위 = $R = (\text{자료의}) \underline{\text{최댓값} - \text{최소값}}$
- ▶ 계급 간격 = $Width = W = \underline{R/C} = \underline{\text{범위}/\text{계급수}}$
- ▶ W 는 R/C 보다 큰 최소의 정수를 취한다.
- ▶ 자료의 최솟값이 첫 계급에 포함되도록 첫 계급의 급하한을 정하고 동일한 급간격으로 C 개의 계급을 구하면서 도수분포표 작성.
- ▶ 계급 값 = 계급의 맷표값 = (각 계급의) [급하한 + 급상한] / 2

자료의 정리

- ▶ 상대도수 = 도수/자료개수 = f/n

Class boundaries	Midpoints	Relative frequency
5.5–10.5	8	0.05
10.5–15.5	13	0.10
15.5–20.5	18	0.15
20.5–25.5	23	0.25
25.5–30.5	28	0.20
30.5–35.5	33	0.15
35.5–40.5	38	0.10
		Total 1.00

상대도수

1/20

- ▶ 누적도수 = 각 계급의 급상한보다 작은 값을 갖는 자료의 수

특정 구간 값의 합계

	Cumulative frequency	Cumulative relative frequency
Less than 5.5	0	0.00
Less than 10.5	1	0.05
Less than 15.5	3	0.15
Less than 20.5	6	0.30
Less than 25.5	11	0.55
Less than 30.5	15	0.75
Less than 35.5	18	0.90
Less than 40.5	20	1.00

누적도수

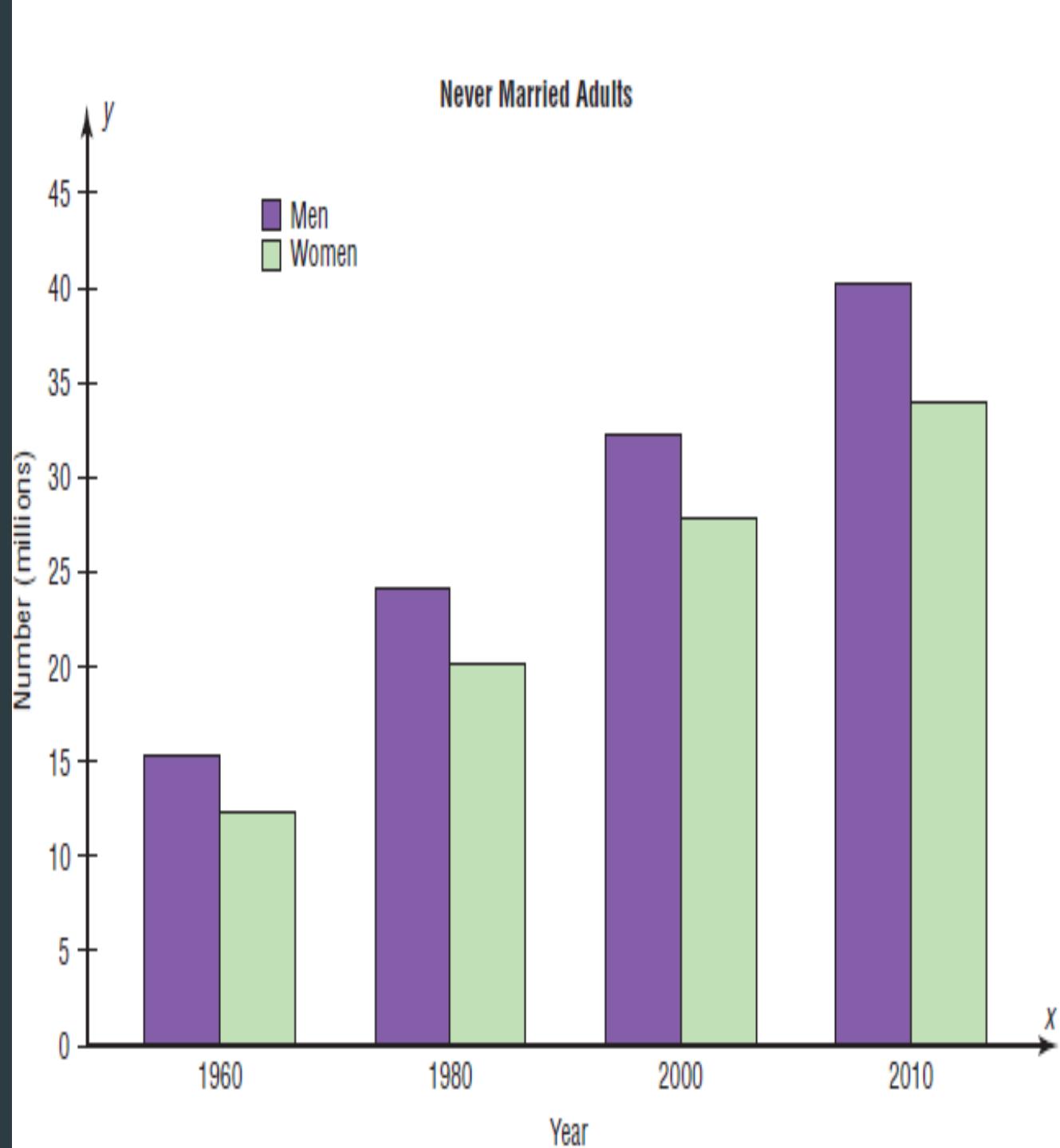
누적 상대도수

Class boundaries	Frequency
5.5–10.5	1
10.5–15.5	2
15.5–20.5	3
20.5–25.5	5
25.5–30.5	4
30.5–35.5	3
35.5–40.5	2
Total 20	

18/20

자료의 정리

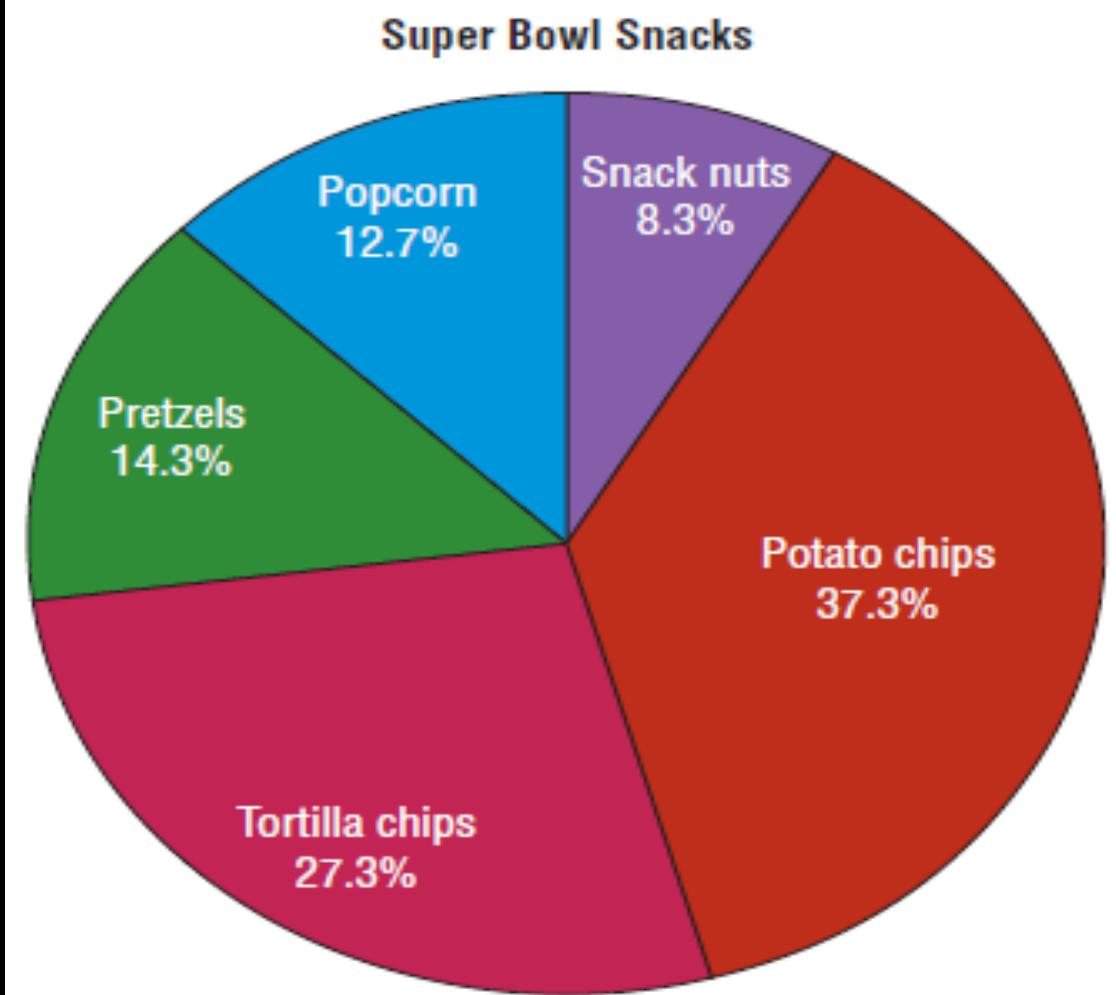
- ▶ 막대 그래프 (Bar Graph)
- ▶ 범주형 자료에 주로 사용
- ▶ X축에는 각각의 범주를, Y축에는 각 범주에 대응되는 도수(때에 따라 상대도수나 누적도수 사용가능)를 그 크기에 비례하도록 막대로 그린다.



막대 그래프 그려보기

Class boundaries	Frequency
5.5–10.5	1
10.5–15.5	2
15.5–20.5	3
20.5–25.5	5
25.5–30.5	4
30.5–35.5	3
35.5–40.5	2
Total	20

FIGURE 2–14 Pie Graph for Example 2–11



자료의 정리

- ▶ 원 그래프 (Pie Chart)
- ▶ 범주형 자료에 주로 사용
- ▶ 전체에 대하여 각 부분들이 차지하는 비중을 보고싶을 때 사용
- ▶ 먼저 원을 그리고 원의 중심각을 각 범주의 도수에 비례하도록 나누어 부채꼴로 그리기
- ▶ 주로 상대도수나 상대도수에 100을 곱한 퍼센트로 많이 나타낸다.

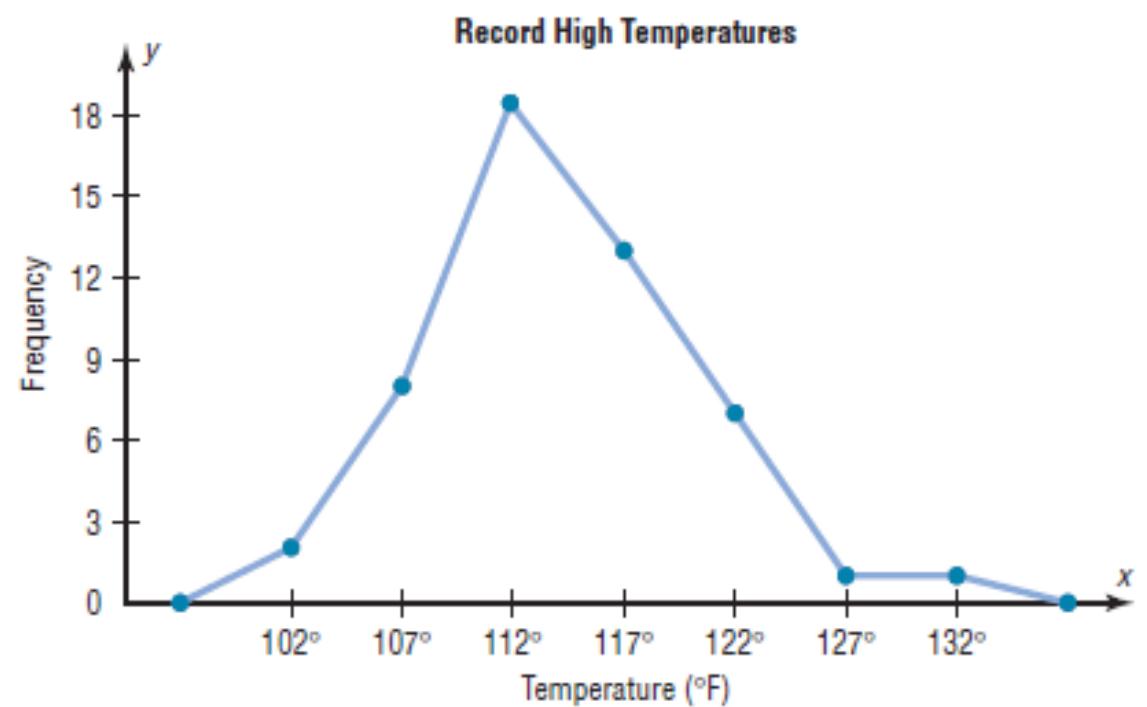
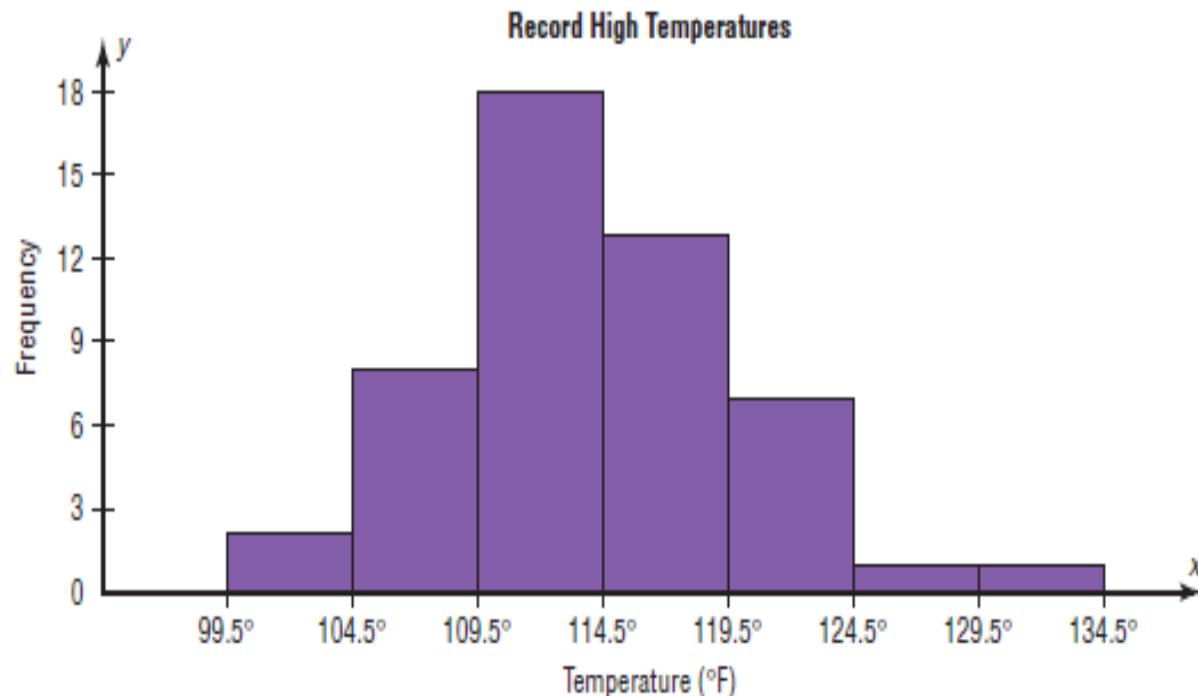
원 그래프 그려보기

Class boundaries	Midpoints	Relative frequency
5.5–10.5	8	0.05
10.5–15.5	13	0.10
15.5–20.5	18	0.15
20.5–25.5	23	0.25
25.5–30.5	28	0.20
30.5–35.5	33	0.15
35.5–40.5	38	0.10
Total		1.00

Class boundaries	Frequency	
5.5–10.5	1	
10.5–15.5	2	
15.5–20.5	3	
20.5–25.5	5	
25.5–30.5	4	
30.5–35.5	3	
35.5–40.5	2	
Total		20

자료의 정리

- ▶ 히스토그램
- ▶ 막대그래프의 일종.
- ▶ 작성 시 주의 사항: 막대의 넓이가 도수에 비례해야 하므로 막대의 폭이 일정해야 한다.
- ▶ 각 계급은 그 계급의 계급값으로 표현하기도 한다.
- ▶ 상대도수 다각형 (Frequency Polygon): 히스토그램에서 각 막대의 중심점 (계급값)을 직선으로 연결하여 히스토그램에 중첩하여 그리거나 따로 그림.



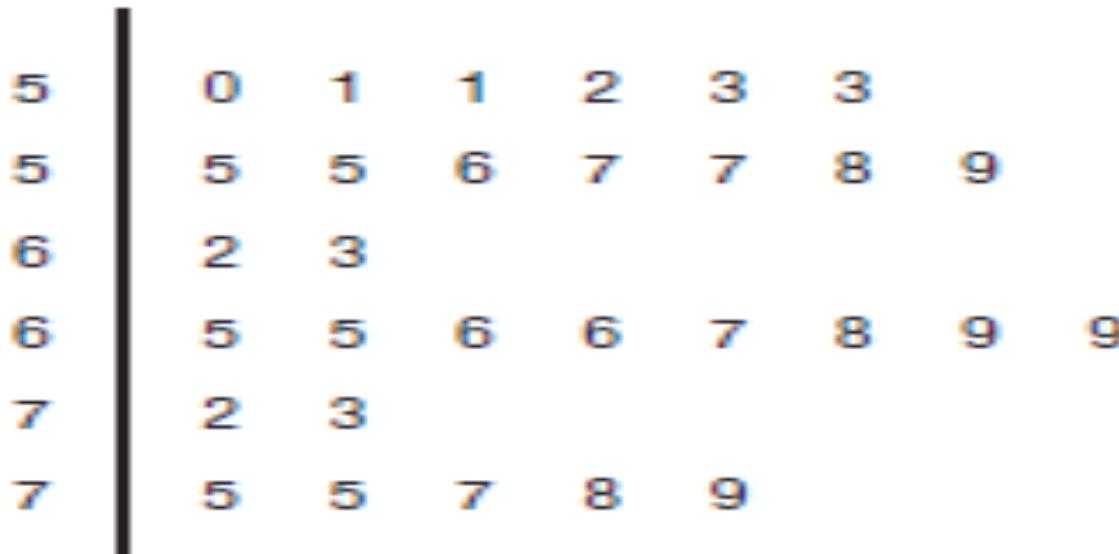
자료의 정리

- ▶ 줄기-잎 그림 (Stem and Leaf Plot)
- ▶ 도수분포표는 각 계급에 몇 개의 자료가 있다면 알려줄 뿐, 정확히 자료가 무엇인지 알려주지 않음.
- ▶ 이 단점을 보완하기 위해 나온 줄기-잎 그림.
- ▶ 작성 순서
 1. 모든 자료에서 줄기 부분을 결정하고 나머지는 잎으로 결정한다.
 2. 줄기를 세로축에 크기순으로 나열하고 그 옆에 수직선을 긋는다. 각 줄기에 해당되는 자료의 잎 부분을 그 줄기의 오른쪽에 가로로 나열한다.
 3. 각 줄기에서 잎의 갚을 크기순으로 나열한다.

자료의 정리

▶ 줄기-잎 그림

FIGURE 2–18
Stem and Leaf Plot for
Example 2–15



EXAMPLE 2–15 Number of Car Thefts in a Large City

An insurance company researcher conducted a survey on the number of car thefts in a large city for a period of 30 days last summer. The raw data are shown. Construct a stem and leaf plot by using classes 50–54, 55–59, 60–64, 65–69, 70–74, and 75–79.

52	62	51	50	69
58	77	66	53	57
75	56	55	67	73
79	59	68	65	72
57	51	63	69	75
65	53	78	66	55

SOLUTION

Step 1 Arrange the data in order.

50, 51, 51, 52, 53, 53, 55, 55, 55, 56, 57, 57, 58, 59, 62, 63,
65, 65, 66, 66, 67, 68, 69, 69, 72, 73, 75, 75, 77, 78, 79

Step 2 Separate the data according to the classes.

50, 51, 51, 52, 53, 53 55, 55, 56, 57, 57, 58, 59
62, 63 65, 65, 66, 66, 67, 68, 69, 69 72, 73
75, 75, 77, 78, 79

Step 3 Plot the data as shown here.

Leading digit (stem)	Trailing digit (leaf)
5	0 1 1 2 3 3
5	5 5 6 7 7 8 9
6	2 3
6	5 5 6 6 7 8 9 9
7	2 3
7	5 5 7 8 9

The graph for this plot is shown in Figure 2–18.

점검

- ▶ 1. 학점에 대한 표본자료
 - ▶ A C D B C C C D F F D F A D C B C D D B
 - ▶ 막대그래프와 원그래프를 그려보시오.
- ▶ 2. 아래 자료에 대하여 줄기 - 잎 그림과 상대도수 분포표를 작성해 보시오.

45	46	64	57	85
92	51	71	54	48
27	66	76	55	69
54	44	54	75	46
61	68	78	61	83
88	45	89	67	56
81	58	55	62	38
55	56	64	81	38
49	68	91	56	68
46	47	83	71	62

Gis revido!