
1

# 제2장
# 데이터의 정리와 요약

1. 도수분포표
2. 체크시트
3. 히스토그램
4. 각종 그래프
5. 중심위치의 척도
6. 산포의 척도

2/31

2

# 1. 도수분포표

[예 2-1] R 기본 패키지에 있는 'iris' 데이터의 두 번째 열 '꽃받침 너비'(Sepal.Width)의 도수분포표

| | 대표값 | 도수 | 누적도수 | 상대도수 | 상대누적도수 |
|---|---|---|---|---|---|
| (2, 2.2] | 2.1 | 4 | 4 | 0.026666667 | 0.02666667 |
| (2.2, 2.4] | 2.3 | 7 | 11 | 0.046666667 | 0.07333333 |
| (2.4, 2.6] | 2.5 | 13 | 24 | 0.086666667 | 0.16000000 |
| (2.6, 2.8] | 2.7 | 23 | 47 | 0.153333333 | 0.31333333 |
| (2.8, 3] | 2.9 | 36 | 83 | 0.240000000 | 0.55333333 |
| (3, 3.2] | 3.1 | 24 | 107 | 0.160000000 | 0.71333333 |
| (3.2, 3.4] | 3.3 | 18 | 125 | 0.120000000 | 0.83333333 |
| (3.4, 3.6] | 3.5 | 10 | 135 | 0.066666667 | 0.90000000 |
| (3.6, 3.8] | 3.7 | 9 | 144 | 0.060000000 | 0.96000000 |
| (3.8, 4] | 3.9 | 3 | 147 | 0.020000000 | 0.98000000 |
| (4, 4.2] | 4.1 | 2 | 149 | 0.013333333 | 0.99333333 |
| (4.2, 4.4] | 4.3 | 1 | 150 | 0.006666667 | 1.00000000 |

3/31

3

# 1. 도수분포표

- [예 2-2]* 규격 [5.00±1.00]Ω, 저항 데이터 100개 [tab2-1.csv] 도수분포표 (15개 구간)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4.91 | 5.03 | 5.07 | 5.21 | 4.74 | 5.03 | 5.08 | 4.95 | 4.89 | 4.65 |
| 4.79 | 5.01 | 4.77 | 4.95 | 4.59 | 5.07 | 4.97 | 5.19 | 5.05 | 5.27 |
| 4.77 | 4.76 | 5.11 | 5.17 | 4.94 | 4.69 | 5.01 | 5.11 | 4.75 | 5.05 |
| 5.01 | 4.93 | 5.01 | 5.08 | 4.69 | 4.89 | 5.23 | 4.99 | 5.14 | 4.95 |
| 4.91 | 4.81 | 4.99 | 4.89 | 4.79 | 4.74 | 5.09 | 5.07 | 5.25 | 5.28 |
| 4.87 | 4.88 | 4.87 | 4.75 | 4.99 | 4.59 | 5.07 | 4.99 | 4.99 | 5.07 |
| 4.94 | 5.29 | 4.97 | 4.99 | 4.95 | 4.65 | 4.77 | 4.83 | 4.95 | 5.05 |
| 5.02 | 4.97 | 5.07 | 4.89 | 4.77 | 4.88 | 5.08 | 4.78 | 5.23 | 5.18 |
| 4.66 | 4.95 | 5.19 | 4.84 | 4.93 | 4.98 | 5.08 | 4.85 | 5.04 | 4.89 |
| 4.79 | 5.09 | 4.98 | 4.94 | 5.17 | 4.88 | 4.96 | 4.92 | 4.79 | 5.10 |

4/31

4

[표 2-2] 저항 데이터([표 2-1])의 도수분포표

| 급 | 구간 | 대푯값 | 도수 | 누적도수 | 상대도수 | 상대누적도수 |
|---|---|---|---|---|---|---|
| 1 | [4.575, 4.625] | 4.60 | 2 | 2 | 0.02 | 0.02 |
| 2 | [4.625, 4.675] | 4.65 | 3 | 5 | 0.03 | 0.05 |
| 3 | [4.675, 4.725] | 4.70 | 2 | 7 | 0.02 | 0.07 |
| 4 | [4.725, 4.775] | 4.75 | 9 | 16 | 0.09 | 0.16 |
| 5 | [4.775, 4.825] | 4.80 | 6 | 22 | 0.06 | 0.22 |
| 6 | [4.825, 4.875] | 4.85 | 5 | 27 | 0.05 | 0.27 |
| 7 | [4.875, 4.925] | 4.90 | 11 | 38 | 0.11 | 0.38 |
| 8 | [4.925, 4.975] | 4.95 | 15 | 53 | 0.15 | 0.53 |
| 9 | [4.975, 5.025] | 5.00 | 13 | 66 | 0.13 | 0.66 |
| 10 | [5.025, 5.075] | 5.05 | 12 | 78 | 0.12 | 0.78 |
| 11 | [5.075, 5.125] | 5.10 | 9 | 87 | 0.09 | 0.87 |
| 12 | [5.125, 5.175] | 5.15 | 3 | 90 | 0.03 | 0.9 |
| 13 | [5.175, 5.225] | 5.20 | 4 | 94 | 0.04 | 0.94 |
| 14 | [5.225, 5.275] | 5.25 | 4 | 98 | 0.04 | 0.98 |
| 15 | [5.275, 5.325] | 5.30 | 2 | 100 | 0.02 | 1.00 |
| | | | 100 | | 1.00 | |

5/31

5

# 2. 체크시트

- 2.2.1 계수표(tally sheet)

| 결점유형 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 도수 |
|---|---|---|---|---|---|---|---|---|
| 색상 | 𝍸 | 𝍸 | 𝍸 | | | | | 12 |
| 마무리 | 𝍸 | | | | | | | 1 |
| 인쇄 | 𝍸 | | | | | | | 2 |
| 복원력 | 𝍸 | 𝍸 | 𝍸 | 𝍸 | | | | 16 |
| 얼룩 | 𝍸 | | | | | | | 1 |
| 선명도 | 𝍸 | 𝍸 | | | | | | 8 |
| 흠집 | 𝍸 | 𝍸 | | | | | | 10 |

6/31

6

# 2. 체크시트

- 2.2.2 분할표(contingency table)

| | 교 | 대 | | |
|---|---|---|---|---|
| 결점의 유형 | 낮 | 저녁 | 밤 | 합계 |
| 복원력 | \ | \\ | \\ | 16 |
| 색상 | \ | \\ | \ | 12 |
| 흠집 | \ | \\ | \ | 10 |
| 선명도 | | \ | \\ | 8 |
| | 5 | 23 | 18 | 46 |

7/31

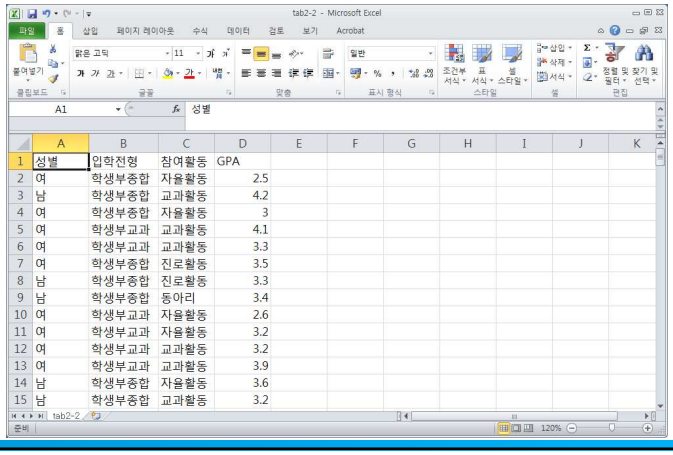
7

[예 2-3] S대학 A학부 신입생 210명에 대해 조사한 데이터 [tab2-2.csv]

(1) 이들 신입생의 입학전형 분포표

(2) 이들 신입생이 중점적으로 참여한 활동 분포표

(3) 이들 신입생의 입학전형과 참여활동 결합분포표



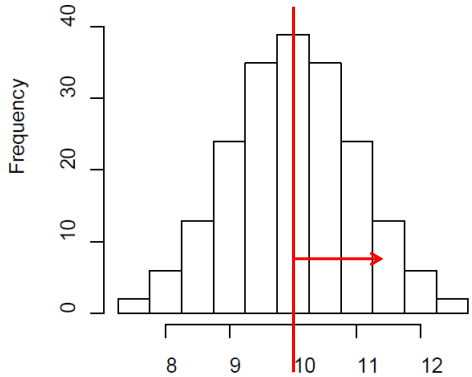
| | A | B | C | D |
|---|---|---|---|---|
| 1 | 성별 | 입학전형 | 참여활동 | GPA |
| 2 | 여 | 학생부종합 | 자율활동 | 2.5 |
| 3 | 남 | 학생부종합 | 교과활동 | 4.2 |
| 4 | 여 | 학생부종합 | 자율활동 | 3 |
| 5 | 여 | 학생부교과 | 교과활동 | 4.1 |
| 6 | 여 | 학생부교과 | 교과활동 | 3.3 |
| 7 | 여 | 학생부종합 | 진로활동 | 3.5 |
| 8 | 남 | 학생부종합 | 진로활동 | 3.3 |
| 9 | 남 | 학생부종합 | 동아리 | 3.4 |
| 10 | 여 | 학생부교과 | 자율활동 | 2.6 |
| 11 | 여 | 학생부교과 | 자율활동 | 3.2 |
| 12 | 여 | 학생부교과 | 교과활동 | 3.2 |
| 13 | 여 | 학생부교과 | 교과활동 | 3.9 |
| 14 | 남 | 학생부종합 | 자율활동 | 3.6 |
| 15 | 남 | 학생부종합 | 교과활동 | 3.2 |

8/31

8

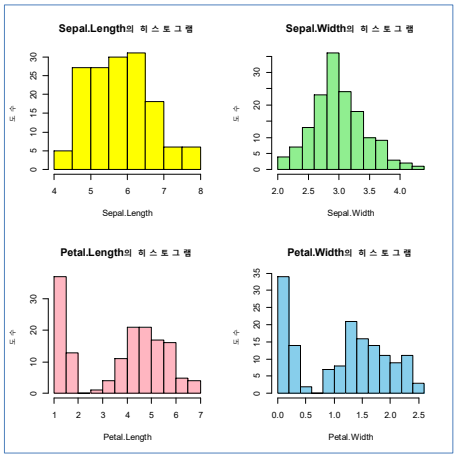
# 3. 히스토그램(histogram)

- 표본 데이터로부터 모집단 분포의 특성을 추측
  ① 모집단 분포의 형태(shape)
  ② 모집단 분포의 중심위치(location)
  ③ 모집단 분포의 산포(spread)




9

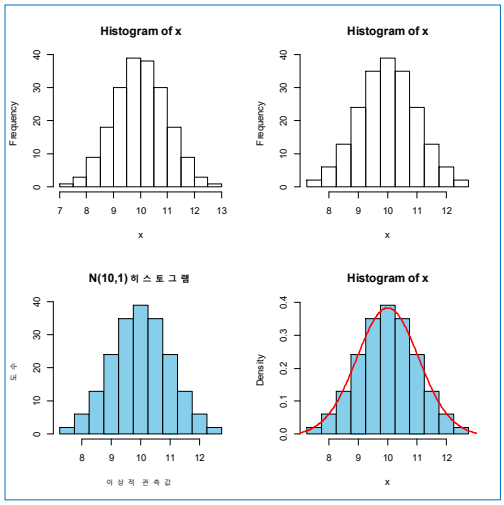
- [예 2-4] R 기본 패키지에 있는 'iris' 데이터의 1열에서 4열까지 변수의 히스토그램을 한 화면에 작성




10/31

10

- [예 2-5]* 안정된 프로세스 N(10,1)의 분포 형태




11/31

11

# 3. 히스토그램(histogram)

- 불안정(이상) 프로세스

1. 낙도형

   프로세스가 불안정하여 오염된 분포가 소량 혼합된 경우

2. 쌍봉우리형

   프로세스가 두 가지 특성을 갖는 하부프로세스로 분리된 경우

3. 이빠진형

   계측기에 문제가 있어 특정 영역의 값이 측정되지 않는 경우

4. 절벽형

   전수검사 후 어떤 경계치 이하(이상)의 제품을 제외한 경우

12/31

12

# 3. 히스토그램(histogram)

[예 2-6]* 불안정 프로세스의 히스토그램 정상 프로세스 N(10, 1)

(a) 낙도형 : 90%의 N(10, 1)과 10%의 오염된 $N(6, 0.5^2)$

(b) 쌍봉우리형 : 50%의 N(10, 1)과 50%의 N(6, 1)

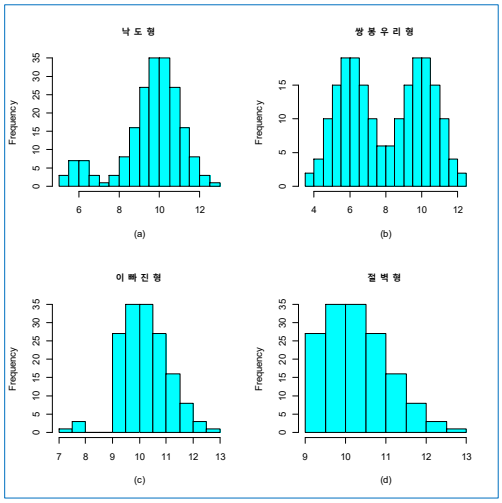
(c) 이빠진형 : 계측기에 문제가 있어 [8,9]영역의 값이 측정되지 않는 경우

(d) 절벽형 : 전수검사 후 9.0 미만의 데이터를 제외한 경우



13

[예 2-6]* 불안정 프로세스의 분포 형태

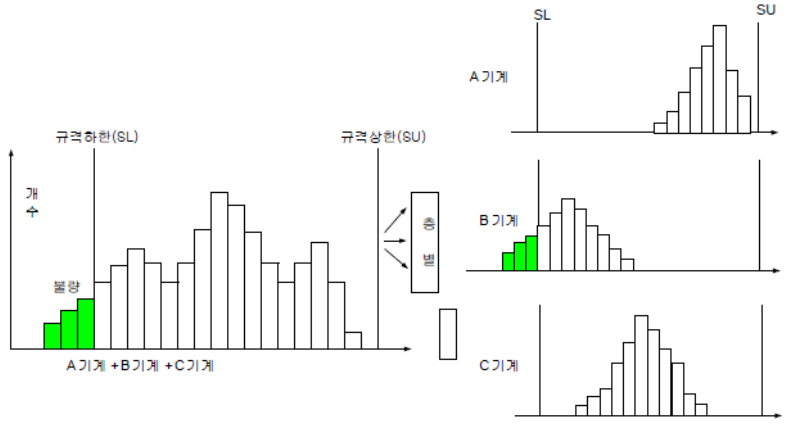





14

# 3. 히스토그램(histogram)

- 층화(stratified) 히스토그램

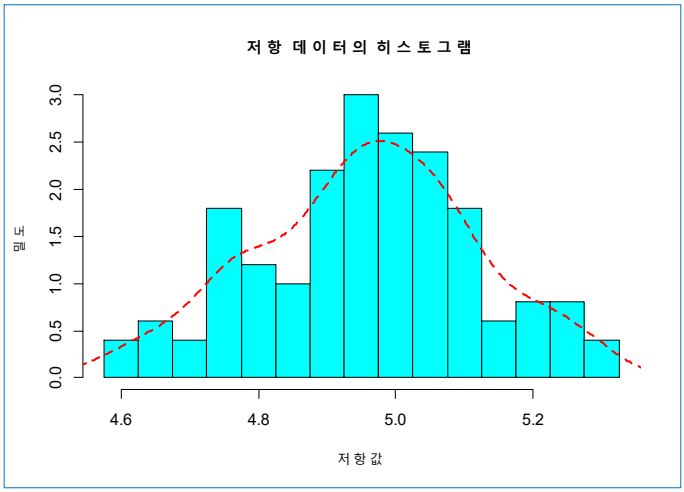



15/31

15

# 3. 히스토그램(histogram)

[예 2-7] 저항 데이터 히스토그램




16/31

16

# 4.1 줄기-잎 그림 (stem-and-leaf plot)

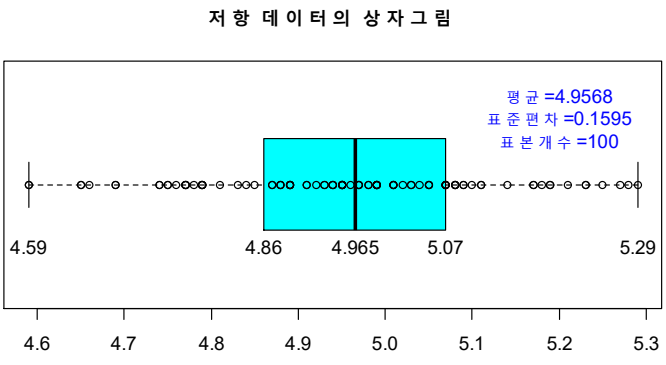
- [예 2-8] [표 2-1]의 저항 데이터 → 줄기-잎 그림

```
# 저항 데이터 줄기-잎 그림 ⇒ stem( ) 함수
stem(x)
 45  99
 46  55699
 47  44556777789999
 48  13457788899999
 49  1123344455555677788999999
 50  11112334555777777888899
 51  011477899
 52  1335789
```

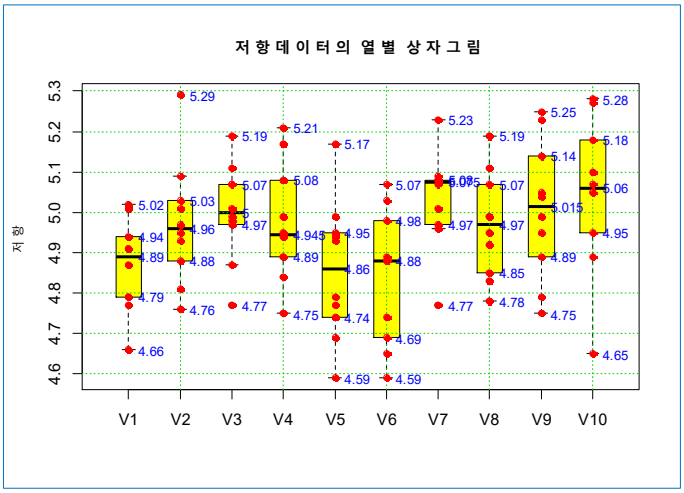




- [예 2-9] 저항 데이터 상자그림



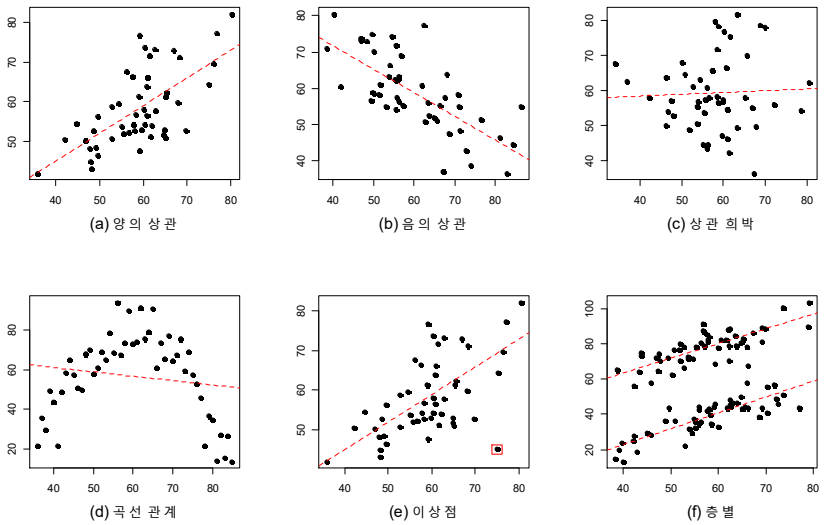
- [예 2-10] 저항 데이터 열별 상자그림




19/31

19

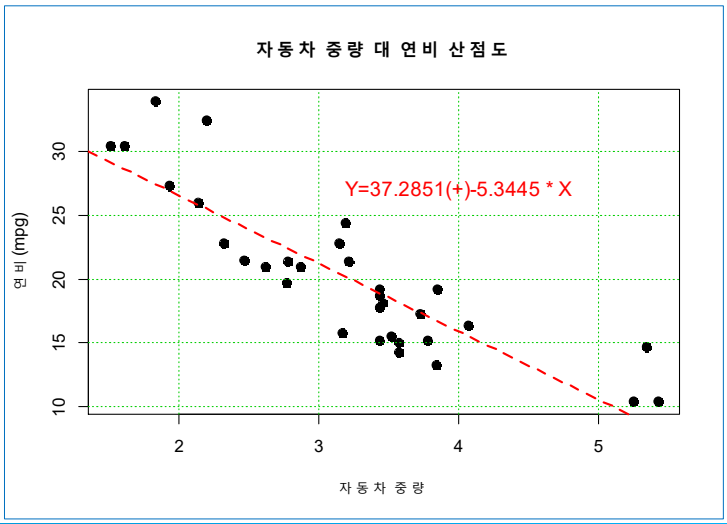
# 4.3 산점도(scatter diagram)




20/31

20

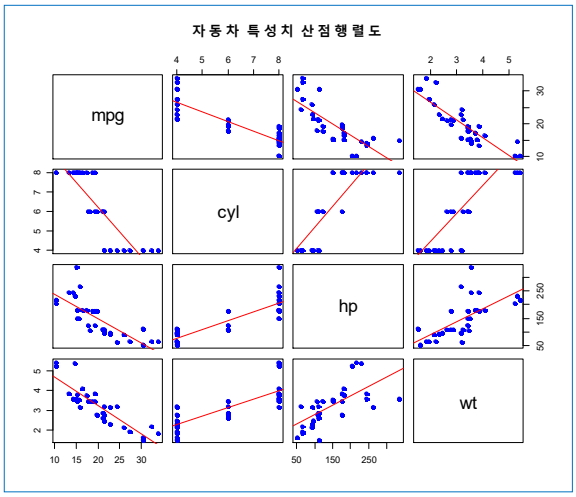
- [예 2-11] 'mtcars' 데이터의 중량(wt) 대 연비(mpg)산점도




21/31

21

[예 2-12] 'mtcars' 데이터의 산점행렬도




22/31

22

# 5. 중심위치의 척도

[예 2-13] 10개의 표본 데이터에 대한 중심위치의 척도

| 5 | 4 | 6 | 3 | 5 | 4 | 3 | 9 | 5 | 10 |
|---|---|---|---|---|---|---|---|---|---|

**(1) 평균** $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{54}{10} = 5.4$

**(2) 중앙값**

| 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|

$\Rightarrow \frac{5+5}{2} = 5$

**(3) 최빈값**

| 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|

$\Rightarrow 5$

**(4) 기하평균** $\bar{x}_g = \left(\prod_{i=1}^{n} x_i\right)^{1/n} = (9,720,000)^{1/10} \approx 4.998$

**(5) 조화평균** $\bar{x}_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \approx \frac{10}{2.1444} \approx 4.663$

**(6) 절사평균**

| 3 | 4 | 4 | 5 | 5 | 5 | 6 | 9 |
|---|---|---|---|---|---|---|---|

$\Rightarrow \bar{x}_{0.1} = \frac{1}{8}\sum_{i=2}^{9} x_i = \frac{41}{8} = 5.125$





# 5. 중심위치의 척도

- 중심위치의 대표값을 선정하는 기준

  ① 명목척도로 측정된 데이터는 최빈값 사용

  ② 분포가 대칭이고 이상점이 존재하지 않으면 표본평균 사용

  ③ 비대칭이거나 이상치가 존재하면 중앙값을 사용하고, 표본평균을 참고 값으로 비교

  ④ 순위 척도로 측정된 데이터는 중앙값 사용

[예 2-14] 저항 데이터의 중심위치 척도

# 6. 산포의 척도

① 표본분산 $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2 / n}{n-1}$

② 표본표준편차 $s = \sqrt{s^2}$

③ 데이터의 범위 $R = x_{\max} - x_{\min}$

④ 사분위수 범위 $IQR = Q_3 - Q_1$

⑤ 변동계수 $CV = s / \bar{x}$





# 6. 산포의 척도

[예 2-15] 10개의 표본 데이터에 대한 산포의 척도

| 5 | 4 | 6 | 3 | 5 | 4 | 3 | 9 | 5 | 10 |
|---|---|---|---|---|---|---|---|---|---|

① 표본분산 $s^2 = \left\{ \sum_{i=1}^{10} x_i^2 - \frac{(\sum_{i=1}^{10} x_i)^2}{10} \right\} / (10-1) = \left( \frac{342 - 54^2/10}{9} \right) = 5.60$

② 표본표준편차 $\Rightarrow s = \sqrt{5.60} \approx 2.366$

③ 데이터의 범위 $R = x_{\max} - x_{\min} = 10 - 3 = 7$

| 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|

④ 사분위수 범위 $1 + 0.25 \times (10-1) = 3.25 \Rightarrow Q_1 = 4$

$1 + 0.75 \times 9 = 7.75 \Rightarrow Q_3 = 5 + (6-5) \times 0.75 = 5.75$ $\Rightarrow Q_3 - Q_1 = 1.75$

⑤ 변동계수 $\Rightarrow CV = \frac{s}{\bar{x}} \approx \frac{2.366}{5.4} \approx 0.438 \ (43.8\%)$

# 6. 산포의 척도

[예 2-16] 저항 데이터 산포의 척도

① 표본분산

$$s^2 = \left\{ \sum_{i=1}^{100} x_i^2 - \frac{(\sum_{i=1}^{100} x_i)^2}{100} \right\} / (100-1) = \left( \frac{2459.5046 - 495.68^2/100}{99} \right) \approx 0.0254$$

② 표본표준편차 $\Rightarrow s = \sqrt{s^2} \approx 0.1595$

③ 데이터의 범위 $R = x_{\max} - x_{\min} = 5.29 - 4.59 = 0.70$

④ 사분위수 범위

$1 + 0.25 \times (100-1) = 25.75,\ x_{(25)} = 4.85,\ x_{(26)} = 4.87$

$\Rightarrow Q_1 = 4.85 + 0.75 \times (4.87 - 4.85) = 4.865$

$1 + 0.75 \times 99 = 75.25,\ x_{(75)} = x_{(76)} = 5.07$

$\Rightarrow Q_3 = 5.07 \quad \Rightarrow Q_3 - Q_1 = 5.07 - 4.865 = 0.205$

⑤ 변동계수 $\Rightarrow CV = \frac{s}{\bar{x}} \approx \frac{0.1595}{4.9568} \approx 0.0322\ (3.22\%)$

27/31

27