3. 도수분포표와 상자그림

도수분포표

자료가 도수분포표로 요약되고 원 자료는 주어지지 않을 경우

↓

계급구간의 모든 관측값이

계급의 중간값을 갖는다고 가정하여 평균과 분산을 계산

↓

원 자료를 그룹화에 의해 정보가 상실되기 때문에

가능하다면 원 자료를 이용

도수분포표에서의 평균

계급의 개수 : k

각 계급의 도수 : $\mathrm{f_i}$,

각 계급의 중간값 : $\mathrm{m_i}$

자료의 개수 : n($= \sum_{i=1}^{k} f_i$)

$$\bar{x} = \frac{1}{n}(m_1\mathrm{f}_1 + \mathrm{m}_2\mathrm{f}_2 + \cdots + \mathrm{m_k f_k})$$

$$= \sum_{i=1}^{k} m_i \left(\frac{f_i}{n}\right)$$

Σ(각 계급의 중간값 × 각 계급의 상대도수)

도수분포표에서의 분산, 표준편차

계급의 개수 : k

각 계급의 도수 : f_i,

각 계급의 중간값 : m_i

자료의 개수 : n(= $\sum_{i=1}^{k} f_i$)

분산

$$s_g^2 = \frac{1}{n-1}\sum_{i=1}^{k}(m_i \,, \overline{x_g})^2 f_i$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{k} m_i^2\, f_i - n\overline{x_g}^2\right)$$

표준편차

$$s_g = \sqrt{s_g^2}$$

상자 그림

```
plt.boxplot()
```

다섯 가지 요약 수치(최솟값, Q1, Q2, Q3, 최댓값)를 그림으로 표현

일반적 그래프에선 드러나지 않는 수치를 함께 제공

제 1사분위수에서 제 3사분위수까지 상자로 그림

좌우에 선을 그어 최솟값, 최댓값을 나타냄

상자 그림

- 상자-수염그림(box-whisker plot) 이라고도 함
- 봉우리가 하나 있는 분포의 특징을 나타내는데 적절
- 봉우리가 여러 개 있는 분포에서는 효과적인 분석 어려움
- 대략적인 자료의 분포를 먼저 파악 후 상자 그림 작성