



탐색적 데이터 분석 2강

- 데이터분석의 모든 것 -

강사명_이정인 저자

Contents

학습 목표와 내용

탐색적 데이터 분석

2. 데이터 분산도 탐색

2.1 최소값, 최대값으로 범위 탐색

2.2 분산과 표준편차

Summary

[02강] 학습 목표와 내용

학습 목표	탐색적 데이터 분석을 이해한다. 데이터 분산도 탐색을 학습하고, R을 통해 실습한다.
학습 내용	var(), sd()
학습 자료	-
핵심 키워드	var(), sd()
참고 자료	데이터분석의 모든 것(출판 : 아이리포, 저자 : 이정인/장원중)



2. 데이터 분산도 탐색

2.1 최소값, 최대값으로 범위 탐색

2.2 분산과 표준편차

2. 1 최소값, 최대값으로 범위 탐색

- ◆ 분산도(degree of dispersion) : 관측된 데이터가 흩어져 있는 정도
 - 범위, 분산, 표준편차 등
- ◆ 범위(range) : 관측된 값들 중에서의 최대값과 최소값의 차이로 분산도를 측정하는 간단한 방법
 - 이해가 쉽고 계산도 편리하지만 두 극단적인 수치의 차이만 나타냄
 - 극단적인 수치들 사이에서 의 분포 양상은 설명하지 못함

2. 1 최소값, 최대값으로 범위 탐색

- ◆ 최소값 - min(), 최대값 - max() 구하기,
- ◆ A기업과 B기업의 연봉의 범위

범위 : 최대값 - 최소값

```
A_salary.max() - A_salary.min()
```

75

```
B_salary.max() - B_salary.min()
```

290

- B기업의 최소값, 최대값의 범위가 더 넓음을 알 수 있음
- 극단값만 차이가 나는 것일 수도 있음

2. 2 분산과 표준편차

- ◆ 객관적인 편차 확인 위해 분산과 표준편차를 구해 봐야함
- ◆ 분산 - `var()`, 표준편차 - `sd()`

분산 `var()`, 표준편차 `std()`

```
A_salary.std()
```

```
21.556468140243165
```

```
A_salary.var()
```

```
464.68131868131866
```

```
B_salary.std()
```

```
77.53056823594814
```

```
B_salary.var()
```

```
6010.989010989011
```



Summary

boraborafeel@gmail.com



수고하셨습니다.