

집단별 데이터 조작

국민대학교 경영학부 빅데이터경영통계학과

2018. 4. 2(월)

1. 주어진 벡터에 그룹별 함수 적용하기 : `tapply`

`tapply(v, g, func)`

내용 벡터 v	2	2	6	8	9	1	2	7	5
집단 벡터 g	1	3	1	2	3	2	2	1	2

집단 벡터 g는 집단을 효과적으로 알려주기만 하면 된다.

따라서 `c(1,3,1,2,3,2,2,1,2)`,

`c(0,7,0,5,7,5,5,0,5)`,

`c('A','C','A','B','C','B','B','A','B')`,

`c("paul","jack","paul","susan","jack","susan","susan","paul","susan")`은

모두 동일한 결과를 산출한다.

집단	1	내용	2 6 7	함수 <code>mean(c(2,6,7))</code>	= 5
집단	2	내용	8 1 2 5	함수 <code>mean(c(8,1,2,5))</code>	= 4
집단	3	내용	2 9	함수 <code>mean(c(2,9))</code>	= 5.5

```
> v=c(2,2,6,8,9,1,2,7,5)
> g=c(1,3,1,2,3,2,2,1,2)
> tapply(v, g, FUN=mean)
 1  2  3
5.0 4.0 5.5
```

2. 그 밖에도 적어하는 데이터 구조에 따라 `tapply`, `apply`, `lapply(sapply)`, `aggregate` 등을 쓸 수 있다.

`tapply(tag-apply)`
`apply(array-apply?)`
`lapply(list-apply)`
`sapply(simplified-list-apply)`

벡터	tapply	메트릭스	apply																																																							
	<p>tapply(v, g, func)</p> <p>v <table><tr><td>2</td><td>2</td><td>6</td><td>8</td><td>9</td><td>1</td><td>2</td><td>7</td></tr></table></p> <p>g <table><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table></p>	2	2	6	8	9	1	2	7									<p>apply(, 1, f) apply(, 2, f)</p> <table><tr><td>8</td><td>2</td><td>2</td><td>3</td></tr><tr><td>1</td><td>2</td><td>4</td><td>2</td></tr><tr><td>5</td><td>9</td><td>4</td><td>1</td></tr><tr><td>6</td><td>8</td><td>5</td><td>4</td></tr></table> <table><tr><td>8</td><td>2</td><td>2</td><td>3</td></tr><tr><td>1</td><td>2</td><td>4</td><td>2</td></tr><tr><td>5</td><td>9</td><td>4</td><td>1</td></tr><tr><td>6</td><td>8</td><td>5</td><td>4</td></tr><tr><td>7</td><td>3</td><td>6</td><td>5</td></tr></table>	8	2	2	3	1	2	4	2	5	9	4	1	6	8	5	4	8	2	2	3	1	2	4	2	5	9	4	1	6	8	5	4	7	3	6	5				
2	2	6	8	9	1	2	7																																																			
8	2	2	3																																																							
1	2	4	2																																																							
5	9	4	1																																																							
6	8	5	4																																																							
8	2	2	3																																																							
1	2	4	2																																																							
5	9	4	1																																																							
6	8	5	4																																																							
7	3	6	5																																																							
결과	<table><tr><td></td><td></td><td></td></tr></table>				결과	<table><tr><td></td><td></td><td></td><td></td></tr></table>																																																				
리스트	lapply sapply	데이터프레임	aggregate																																																							
	<p>lapply(lst, func)</p> <p>[[1]] <table><tr><td>2</td><td>2</td><td>6</td><td>8</td></tr></table></p> <p>[[2]] <table><tr><td>2</td><td>2</td><td>6</td><td>8</td><td>9</td></tr></table></p> <p>[[3]] <table><tr><td>2</td><td>2</td><td>6</td></tr></table></p> <p>[[4]] <table><tr><td>2</td><td>2</td><td>6</td></tr></table></p>	2	2	6	8	2	2	6	8	9	2	2	6	2	2	6	<p>aggregate(df, by=list(), func)</p> <table><tr><td>8</td><td>2</td><td>2</td><td>3</td></tr><tr><td>1</td><td>2</td><td>4</td><td>2</td></tr><tr><td>5</td><td>9</td><td>4</td><td>1</td></tr><tr><td>6</td><td>8</td><td>5</td><td>4</td></tr><tr><td>7</td><td>3</td><td>6</td><td>5</td></tr><tr><td>8</td><td>9</td><td>4</td><td>3</td></tr><tr><td>2</td><td>4</td><td>3</td><td>7</td></tr><tr><td>9</td><td>5</td><td>6</td><td>3</td></tr></table> <table><tr><td></td></tr><tr><td></td></tr><tr><td></td></tr><tr><td></td></tr><tr><td></td></tr><tr><td></td></tr><tr><td></td></tr><tr><td></td></tr></table>	8	2	2	3	1	2	4	2	5	9	4	1	6	8	5	4	7	3	6	5	8	9	4	3	2	4	3	7	9	5	6	3									<p>aggregate(y1+y2~x1+x2, df, func)</p>
2	2	6	8																																																							
2	2	6	8	9																																																						
2	2	6																																																								
2	2	6																																																								
8	2	2	3																																																							
1	2	4	2																																																							
5	9	4	1																																																							
6	8	5	4																																																							
7	3	6	5																																																							
8	9	4	3																																																							
2	4	3	7																																																							
9	5	6	3																																																							
결과	<table><tr><td></td><td></td><td></td><td></td></tr></table>					결과	<table><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td><td></td></tr></table>																																																			