

PCA 실습 1

차준범 강사

데이터 파악

1. 데이터 확인

```
head(iris)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1         3.5         1.4         0.2   setosa
## 2           4.9         3.0         1.4         0.2   setosa
## 3           4.7         3.2         1.3         0.2   setosa
## 4           4.6         3.1         1.5         0.2   setosa
## 5           5.0         3.6         1.4         0.2   setosa
## 6           5.4         3.9         1.7         0.4   setosa
```

2. 결측치 확인

```
colSums(is.na(iris))
```

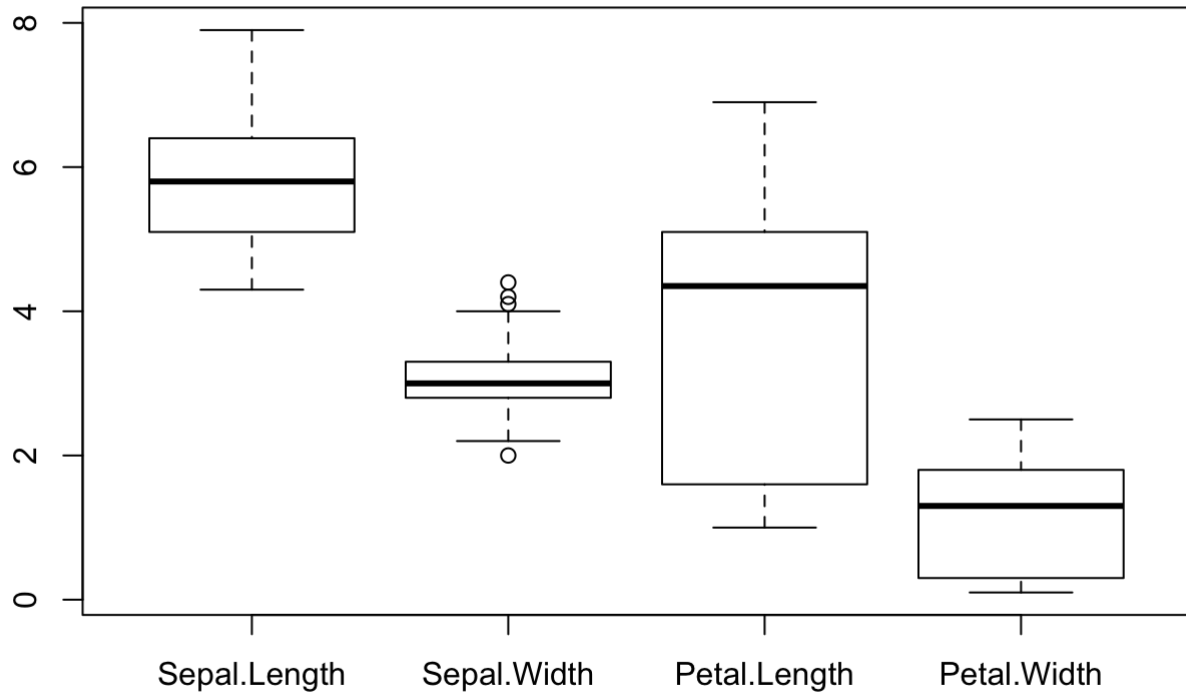
```
## Sepal.Length Sepal.Width Petal.Length Petal.Width      Species
##           0           0           0           0           0
```

3. 변수별 기술통계 및 분포 확인

```
summary(iris)
```

```
##      Sepal.Length      Sepal.Width      Petal.Length      Petal.Width
## Min.       :4.300   Min.       :2.000   Min.       :1.000   Min.       :0.100
## 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
## Median :5.800   Median :3.000   Median :4.350   Median :1.300
## Mean     :5.843   Mean     :3.057   Mean     :3.758   Mean     :1.199
## 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
## Max.     :7.900   Max.     :4.400   Max.     :6.900   Max.     :2.500
##      Species
## setosa      :50
## versicolor:50
## virginica   :50
##
##
##
```

```
boxplot(iris[,1:4])
```



분석, 결과치 확인 및 해석

4. pca 함수 적용 및 요약 결과 확인

```
iris.pca <- prcomp(iris[1:4], center = T, scale. = T) # pca 함수
```

Name	Type	Value
iris.pca	list [5] (S3: prcomp)	List of length 5
sdev	double [4]	1.708 0.956 0.383 0.144
rotation	double [4 x 4]	0.5211 -0.2693 0.5804 0.5649 -0.3774 -0.9233 -
center	double [4]	5.84 3.06 3.76 1.20
scale	double [4]	0.828 0.436 1.765 0.762
x	double [150 x 4]	-2.25714 -2.07401 -2.35634 -2.29171 -2.38186 -

```
summary(iris.pca) # pca 요약정보. standard deviation 제공 = 분산 = eigenvalue
```

```
## Importance of components:
##              PC1      PC2      PC3      PC4
## Standard deviation   1.7084 0.9560 0.38309 0.14393
## Proportion of Variance 0.7296 0.2285 0.03669 0.00518
## Cumulative Proportion 0.7296 0.9581 0.99482 1.00000
```

```
iris.pca$rotation # 각 주성분의 eigenvector
```

```
##              PC1          PC2          PC3          PC4
## Sepal.Length  0.5210659 -0.37741762  0.7195664  0.2612863
## Sepal.Width   -0.2693474 -0.92329566 -0.2443818 -0.1235096
## Petal.Length  0.5804131 -0.02449161 -0.1421264 -0.8014492
## Petal.Width   0.5648565 -0.06694199 -0.6342727  0.5235971
```

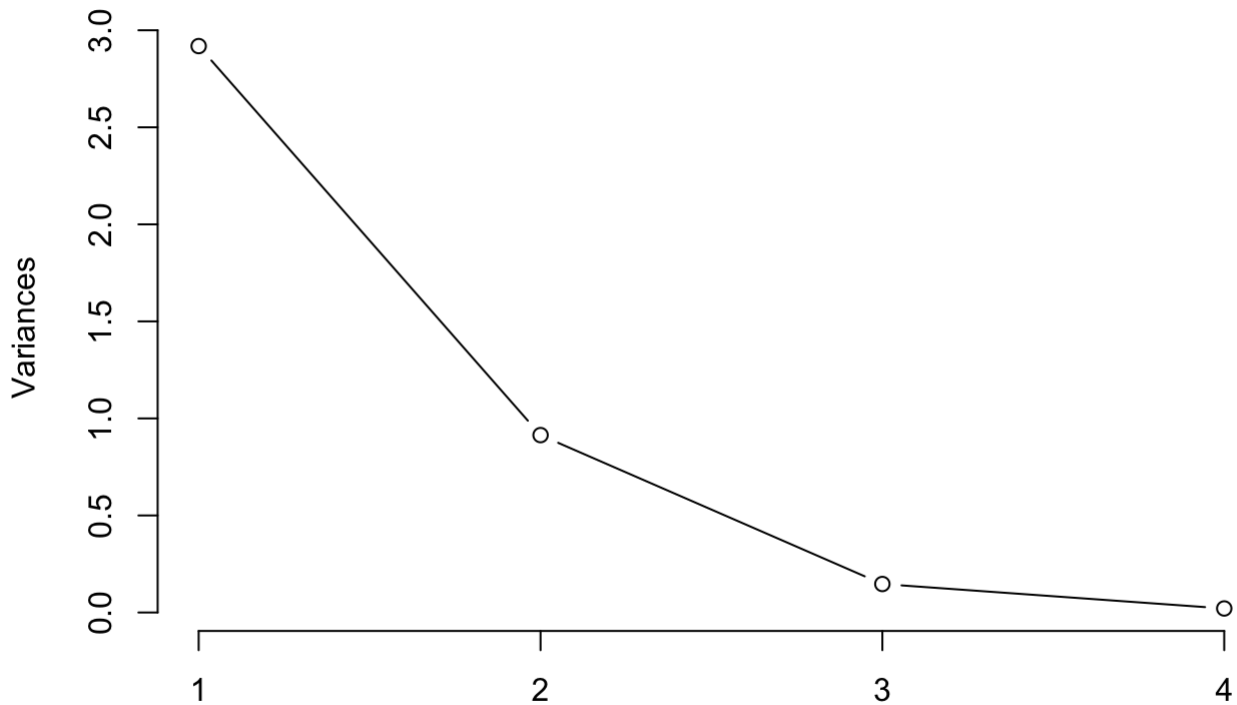
```
head(iris.pca$x, 10) # 각 주성분의 값
```

```
##              PC1          PC2          PC3          PC4
## [1,] -2.257141 -0.47842383  0.12727962  0.024087508
## [2,] -2.074013  0.67188269  0.23382552  0.102662845
## [3,] -2.356335  0.34076642 -0.04405390  0.028282305
## [4,] -2.291707  0.59539986 -0.09098530 -0.065735340
## [5,] -2.381863 -0.64467566 -0.01568565 -0.035802870
## [6,] -2.068701 -1.48420530 -0.02687825  0.006586116
## [7,] -2.435868 -0.04748512 -0.33435030 -0.036652767
## [8,] -2.225392 -0.22240300  0.08839935 -0.024529919
## [9,] -2.326845  1.11160370 -0.14459247 -0.026769540
## [10,] -2.177035  0.46744757  0.25291827 -0.039766068
```

5. scree plot 확인

```
plot(iris.pca, type = 'l', main = 'Scree Plot') # PC의 분산을 y축으로 scree plot 생성
```

Scree Plot



6. 차원축소

```
head(iris.pca$x[,1:2], 10) # 2개의 차원으로 축소
```

```
##           PC1           PC2
## [1,] -2.257141 -0.47842383
## [2,] -2.074013  0.67188269
## [3,] -2.356335  0.34076642
## [4,] -2.291707  0.59539986
## [5,] -2.381863 -0.64467566
## [6,] -2.068701 -1.48420530
## [7,] -2.435868 -0.04748512
## [8,] -2.225392 -0.22240300
## [9,] -2.326845  1.11160370
## [10,] -2.177035  0.46744757
```

7. 2차원 시각화

```
library(ggfortify)
autoplot(iris.pca, data = iris, colour = 'Species') # 2차원으로 축소된 데이터 시각화
```

