

## Responsi Analisis Biplot

### Pengantar:

Biplot adalah representasi grafis dari informasi dalam matriks data  $n \times p$ . *Bi* (pada Bi-plot) mengacu pada dua jenis informasi yang terkandung dalam matriks data. Informasi dalam baris berkaitan dengan sampel atau unit pengambilan sampel dan kolom berkaitan dengan variabel. Ketika hanya ada dua variabel, *scatter plots* dapat mewakili informasi pada unit sampling dan variabel dalam satu diagram. Gagasan dibalik biplots adalah untuk menambahkan informasi tentang variabel ke grafik komponen utama.

### Tahapan Manual Pembuatan Biplot:

1. Menghitung matriks  $\mathbf{X}'\mathbf{X}$
2. Menghitung nilai eigen ( $\lambda_i$ ) dari  $\mathbf{X}'\mathbf{X}$
3. Menghitung matriks  $\mathbf{U}$ ,  $\mathbf{L}$ , dan  $\mathbf{A}$  (rumus ada di buku)
4. Cari nilai  $\mathbf{G}$  dan  $\mathbf{H}$  (rumus ada di buku)
5. Menentukan  $\mathbf{G}^*$  : kolom ke-1 dan ke-2 dari matriks  $\mathbf{G}$  dan  $\mathbf{H}^*$  : kolom ke-1 dan ke-2 dari matriks  $\mathbf{H}$
6. Gambar biplotnya dengan koordinat ( $\mathbf{G}^*, \mathbf{H}^*$ )
7. Mengitung ukuran kebaikan biplot (besarnya keragaman yang dapat dijelaskan biplot)

### Contoh:

Table 12.9 gives the data on some universities for certain variables used to compare or rank maj or universities. These variables include:

X1 = average SAT score of new freshmen,

X2 = percentage of new freshmen in top 10% of high school class,

X3 = percentage of applicants accepted,

X4 = student – faculty ratio,

X5 = estimated annual expenses, and

X6 = graduation rate (%) .

University	SAT	Top10	Accept	SFRatio	Expenses	Grad
Harvard	14	91	14	11	39.525	97
Princeton	13.75	91	14	8	30.22	95
Yale	13.75	95	19	11	43.514	96
Stanford	13.6	90	20	12	36.45	93
MIT	13.8	94	30	10	34.87	91
Duke	13.15	90	30	12	31.585	95
Cal Tech	14.15	100	25	6	63.575	81
Dartmouth	13.4	89	23	10	32.162	95
Brown	13.1	89	22	13	22.704	94
JohnsHopkins	13.05	75	44	7	58.691	87
UChicago	12.9	75	50	13	38.38	87
UPenn	12.85	80	36	11	27.553	90
Cornell	12.8	83	33	13	21.864	90
Northwestern	12.6	85	39	11	28.052	89
Columbia	13.1	76	24	12	31.51	88
NotreDame	12.55	81	42	13	15.122	94
UVirginia	12.25	77	44	14	13.349	92

University	SAT	Top10	Accept	SFRatio	Expenses	Grad
Georgetown	12.55	74	24	12	20.126	92
CarnegieMellon	12.6	62	59	9	25.026	72
UMichigan	11.8	65	68	16	15.47	85
UCBerkeley	12.4	95	40	17	15.14	78
UWisconsin	10.85	40	69	15	11.857	71
PennState	10.81	38	54	18	10.185	80
Purdue	10.05	28	90	19	9.066	69
TexasA&M	10.75	49	67	25	8.704	67

## Penyelesaian Menggunakan Program R:

### Tahap 1. Input Data di R


```
> data<-read.delim("clipboard",header = T)
> head(data)

  University SAT Top10 Accept SFRatio Expenses Grad
1  Harvard  14.00   91    14     11   39.525   97
2 Princeton  13.75   91    14     8   30.220   95
3   Yale    13.75   95    19    11   43.514   96
4 Stanford  13.60   90    20    12   36.450   93
5    MIT    13.80   94    30    10   34.870   91
6    Duke    13.15   90    30    12   31.585   95

> rownames(data)=data$University
> data.new=data[,-1]
> head(data.new)

      SAT Top10 Accept SFRatio Expenses Grad
Harvard  14.00   91    14     11   39.525   97
Princeton 13.75   91    14     8   30.220   95
Yale     13.75   95    19    11   43.514   96
Stanford 13.60   90    20    12   36.450   93
MIT      13.80   94    30    10   34.870   91
Duke     13.15   90    30    12   31.585   95
```

### Keterangan:

- Copy data dari Excel, posisikan kursor pada *source code* `data<-read.delim()` kemudian pilih  atau Ctrl+Shift+Enter untuk *R Studio* (Ctrl+R untuk *R*). [opsional, dapat melakukan input data dengan cara lain]
- `head(data)` menampilkan 6 data baris pertama untuk memastikan struktur data yang diinginkan: baris untuk Universitas dan kolom untuk variabel.
- `rownames(data)=data$University` mengganti nama baris menjadi Universitas
- `data.new=data[,-1]` mendefinisikan data baru tanpa kolom Universitas
- `head(data.new)` memeriksa kembali struktur data

### Tahap 2. Eksplorasi Data

```
> summary(data.new)

      SAT      Top10      Accept
Min.   :10.05  Min.   : 28.00  Min.   :14.0
1st Qu.:12.40  1st Qu.: 74.00  1st Qu.:24.0
Median :12.85  Median : 81.00  Median :36.0
Mean   :12.66  Mean   : 76.48  Mean   :39.2
```

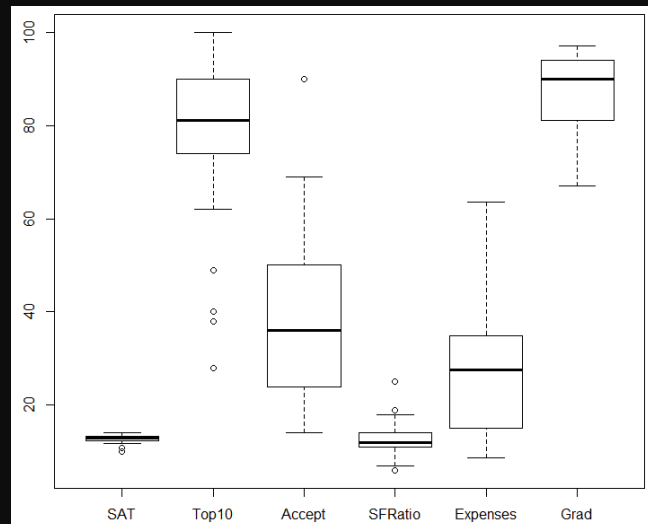
```
3rd Qu.:13.40 3rd Qu.: 90.00 3rd Qu.:50.0
Max.      :14.15 Max.      :100.00 Max.      :90.0
```

```

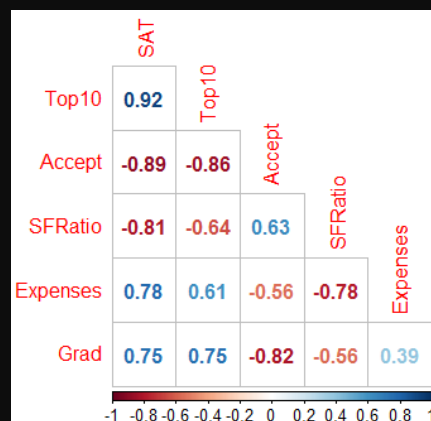
SFRatio      Expenses      Grad
Min.       : 6.00    Min.       : 8.704    Min.       :67.00
1st Qu.    :11.00    1st Qu.    :15.140    1st Qu.    :81.00
Median     :12.00    Median     :27.553    Median     :90.00
Mean       :12.72    Mean       :27.388    Mean       :86.72
3rd Qu.    :14.00    3rd Qu.    :34.870    3rd Qu.    :94.00
Max.       :25.00    Max.       :63.575    Max.       :97.00

```

```
> boxplot(data.new)
```



```
> library(corrplot)
> corrplot(cor(data.new), method = "number",is.cor = T,type = "lower",diag
=F)
```



```
> help("corrplot")
```

Keterangan:

- `summary(data.new)` melihat deksripsi data
- `boxplot(data.new)` melakukan pengecekan adanya pencilan pada variabel.  
[Peubah Expenses dan Grad tidak memiliki pencilan, peubah SAT dan Top10 hanya memiliki pencilan bawah artinya terdapat nilai yang jauh lebih kecil diantara nilai lainnya, peubah Accept memiliki satu pencilan atas artinya terdapat satu nilai yang

jauh lebih besar diantara nilai lainnya, sedangkan SfRatio memiliki pencilan atas dan bawah.]

- `library(corrplot)` digunakan untuk plot korelasi antar peubah, jika belum tersedia lakukan `install.packages(corrplot)`.
- `corrplot()` visualisasi nilai korelasi antar peubah dalam grafik.  
[korelasi antar peubah yang terbesar ditemukan pada peubah Top10 dan SAT yaitu sebesar 0.92, dsb]
- `help(corrplot)` untuk mengetahui keterangan fungsi `corrplot` pada R

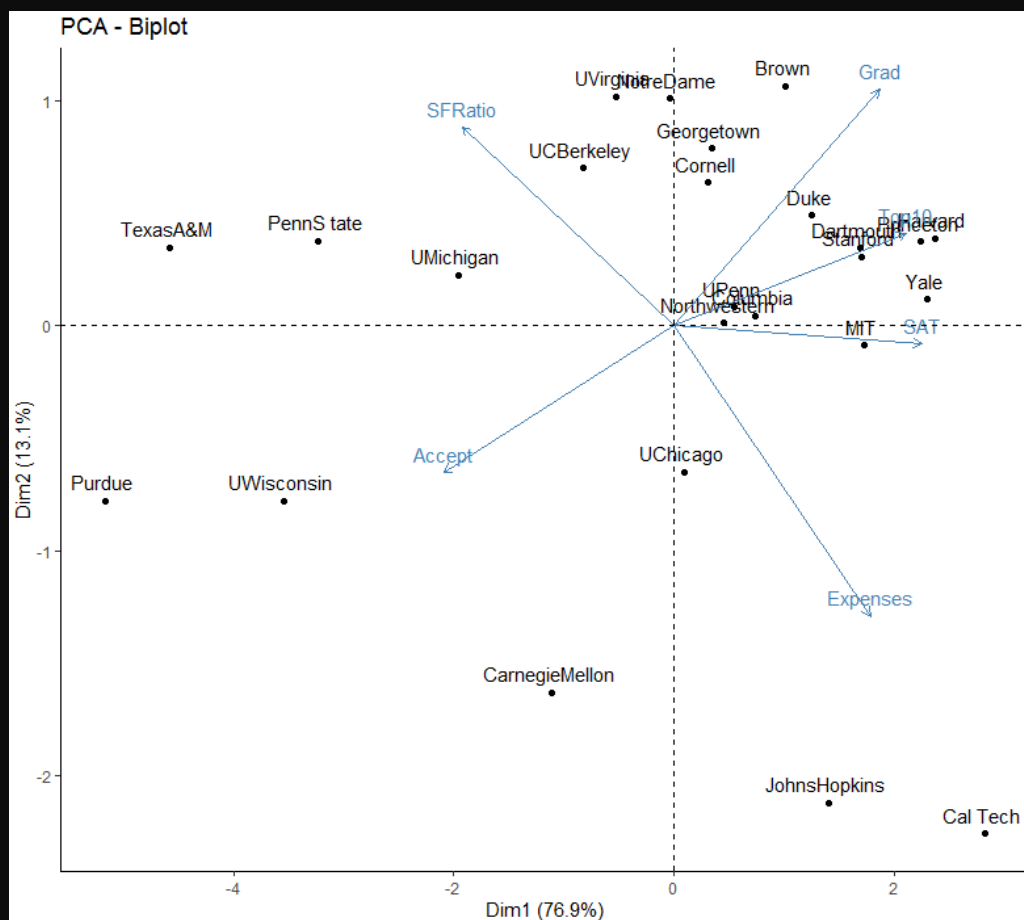
### Tahap 3. Analisis Biplot

```
> library(magrittr)
> df=data.new[,] %>% scale()
> head(df)
```

	SAT	Top10	Accept	SFRatio	Expenses	Grad
Harvard	1.2325607	0.7471478	-1.2774171	-0.4228798	0.8413933	1.1349362
Princeton	1.0018478	0.7471478	-1.2774171	-1.1604608	0.1963274	0.9141315
Yale	1.0018478	0.9529737	-1.0239613	-0.4228798	1.1179293	1.0245339
Stanford	0.8634201	0.6956914	-0.9732702	-0.1770194	0.6282200	0.6933268
MIT	1.0479904	0.9015172	-0.4663586	-0.6687401	0.5186870	0.4725221
Duke	0.4481368	0.6956914	-0.4663586	-0.1770194	0.2909556	0.9141315

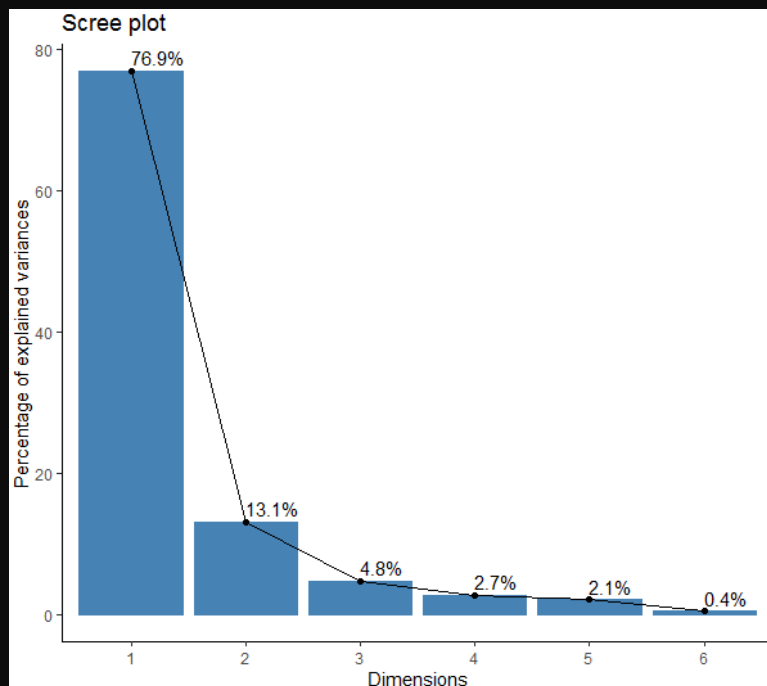
```
> library(FactoMineR)
> res.pca=PCA(df,graph=F)

> library(factoextra)
> fviz_pca_biplot(res.pca)+theme_classic()
```



```
> eig.val=get_eigenvalue(res.pca)
> eig.val
      eigenvalue variance.percent cumulative.variance.percent
Dim.1 4.61208506      76.8680843          76.86808
Dim.2 0.78681615      13.1136025          89.98169
Dim.3 0.28656186       4.7760310          94.75772
Dim.4 0.16378011       2.7296685          97.48739
Dim.5 0.12430620       2.0717700          99.55916
Dim.6 0.02645063       0.4408438         100.00000

> fviz_eig(res.pca,addlabels = T)+theme_classic()
```



Keterangan:

- `library(magrittr)` digunakan karena proses selanjutnya akan dilakukan transformasi data dan membutuhkan operator `%>%` jika belum tersedia lakukan `install.packages(magrittr)`.
- `df=data.new[,-1] %>% scale()` proses transformasi data, transformasi dilakukan karena peubah SAT dan Expenses memiliki satuan skala yang berbeda dengan peubah lainnya.
- `library(FactoMiner)` digunakan untuk fungsi komponen utama, jika belum tersedia lakukan `install.packages(FactoMiner)`.
- `res.pca=PCA(df,graph=F)` untuk memperoleh struktur komponen utama.
- `library(factoextra)` agar fungsi yang dapat memperoleh grafik Biplot dan persentase keragaman Biplot dapat digunakan, jika belum tersedia lakukan `install.packages(factoextra)`.
- `fviz_pca_biplot(res.pca)+theme_classic()` untuk grafik Biplot. [Interpretasi:
  1. Terdapat 5 Universitas dengan nilai SFRatio relatif tinggi yaitu TexasA&M, PennState, Umichigan, UC Berkeley, Uvirginia dan Notre Dame.
  2. Universitas Cal Tech dan Johns Hopkins memiliki karakteristik variabel yang mirip.
  3. Terdapat 3 Universitas dengan nilai Accept relatif tinggi yaitu Purdue dan Carnegie Mellon

4. Peubah TOP10 dan SAT keduanya berhubungan positif
  5. Peubah Expenses dan SFRatio keduanya berhubungan negatif
  6. dsb]
- `eig.val=get_eigenvalue(res.pca)` digunakan untuk memperoleh persentase keragaman berdasarkan nilai eigen. Kebaikan Biplot dilihat berdasarkan `cumulative.variance.percent` dari nilai eigen kedua. [Informasi keragaman yang mampu dijelaskan oleh Biplot adalah sebesar 89.98%. Artinya Biplot mampu memberikan informasi dari 6 peubah untuk menilai posisi relatif 25 Universitas.
  - `fviz_eig(res.pca,addlabels=T)+theme_classic()` grafik persentase keragaman yang dijelaskan setiap nilai eigen.

**Latihan:**

Lakukan analisis Biplot dengan pengambilan sampel 30% dari data `USArrests` pada program R!