

تحلیل مولفه اصلی برای داده ها فوتبال

ابتدا داده های خود را فراخوانی میکنیم :

```
> football<-read.table("C:/Users/12345/Desktop/football.txt",header=T)
```

۵ سطر ابتدایی از داده ها را نمایش می دهیم:

```
> head(football,5)
  WDIM CIRCUM FBEYE EYEHD EARHD JAW
1 13.5  57.15  19.5  12.5  14.0  11
2 15.5  58.42  21.0  12.0  16.0  12
3 14.5  55.88  19.0  10.0  13.0  12
4 15.5  58.42  20.0  13.5  15.0  12
5 14.5  58.42  20.0  13.0  15.5  12
```

تعریف متغیرها :

WDIM: عرض سر در وسیع ترین بخش

CIRCUM: دور سر

EYEHD: اندازه جلو تا پشت سر در سطح چشم

EARHD: اندازه چشم تا فرق سر

EARHD: اندازه گوش تا فرق سر

JAW: عرض فک

بردار میانگین را بدست می آوریم :

```
> colMeans(football)
  WDIM  CIRCUM  FBEYE  EYEHD  EARHD  JAW
15.50000 57.57483 19.80667 10.51333 13.57500 11.87333
```

تشکیل ماتریس واریانس_کوواریانس برای داده ها(ماتریس S)

```
(S<-cov(football))
      WDIM    CIRCUM    FBEYE    EYEHD    EARHD    JAW
WDIM    0.37016949 0.6020339 0.14881356 0.04440678 0.10711864 0.20932203
CIRCUM  0.60203390 2.6289847 0.80147571 0.66562938 0.10278390 0.37685989
FBEYE   0.14881356 0.8014757 0.45825989 0.01126554 -0.01322034 0.11984181
EYEHD   0.04440678 0.6656294 0.01126554 1.47371751 0.25220339 -0.05438418
EARHD   0.10711864 0.1027839 -0.01322034 0.25220339 0.48800847 -0.03559322
JAW     0.20932203 0.3768599 0.11984181 -0.05438418 -0.03559322 0.32368362
```

محاسبه مقادیر ویژه و بردارهای ویژه ماتریس واریانس_کوواریانس

```
> (eigen<-eigen(S))
eigen() decomposition
$values
[1] 3.32341443 1.37430806 0.47606880 0.32468424 0.15649723 0.08785095

$vectors
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	-0.20744390	0.1415256	-0.42155301	0.4425457	-0.1682617	0.7314861
[2,]	-0.87284535	0.2191281	0.08433843	-0.1309809	-0.3304096	-0.2380849
[3,]	-0.26126472	0.2314010	0.12087973	-0.3819134	0.7676277	0.3584302
[4,]	-0.32586218	-0.8911780	0.17308617	0.1733080	0.1641025	0.1126568
[5,]	-0.06563904	-0.2220298	-0.86746687	-0.3545187	0.1144323	-0.2347707
[6,]	-0.12788326	0.1868463	-0.13457558	0.6967211	0.4829502	-0.4603940

در اینجا به راحتی میتوان مقادیر ویژه را مشاهده کرد.

کاهش بعد به کمک مقادیر ویژه :

```
> labda<-eigen$values
> which(labda>mean(labda))
[1] 1 2
```

براساس مقادیر ویژه مولفه‌های با مقادیر ویژه بزرگتر از میانگین مقادیر ویژه را نگه میداریم که در اینجا مولفه‌های اصلی ۱ و ۲ را نگه میداریم.

تحلیل مولفه اصلی به کمک ماتریس واریانس-کوواریانس :

```
> football.pca<-princomp(football,cor=F,scores = T)
> summary(football.pca)
Importance of components:
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.8077677	1.1624986	0.68420342	0.56504233	0.39228682	0.293916
Proportion of Variance	0.5787074	0.2393088	0.08289803	0.05653739	0.02725092	0.015297
Cumulative Proportion	0.5787074	0.8180161	0.90091418	0.95745156	0.98470248	1.000000

در اینجا نیز میتوان نتیجه گرفت که میتوان کاهش بعد انجام داد چون که دو مولفه اصلی اول ۸۱ درصد کل واریانس را نمایش میدهند. همچنین میتوان انحراف استاندارد هر متغیر را در بخش (standard deviation) مشاهده کرد.

در اینجا نیز میتوانیم بردارهای ویژه را مشاهده کنیم:

```
> loadings(football.pca)

Loadings:
```

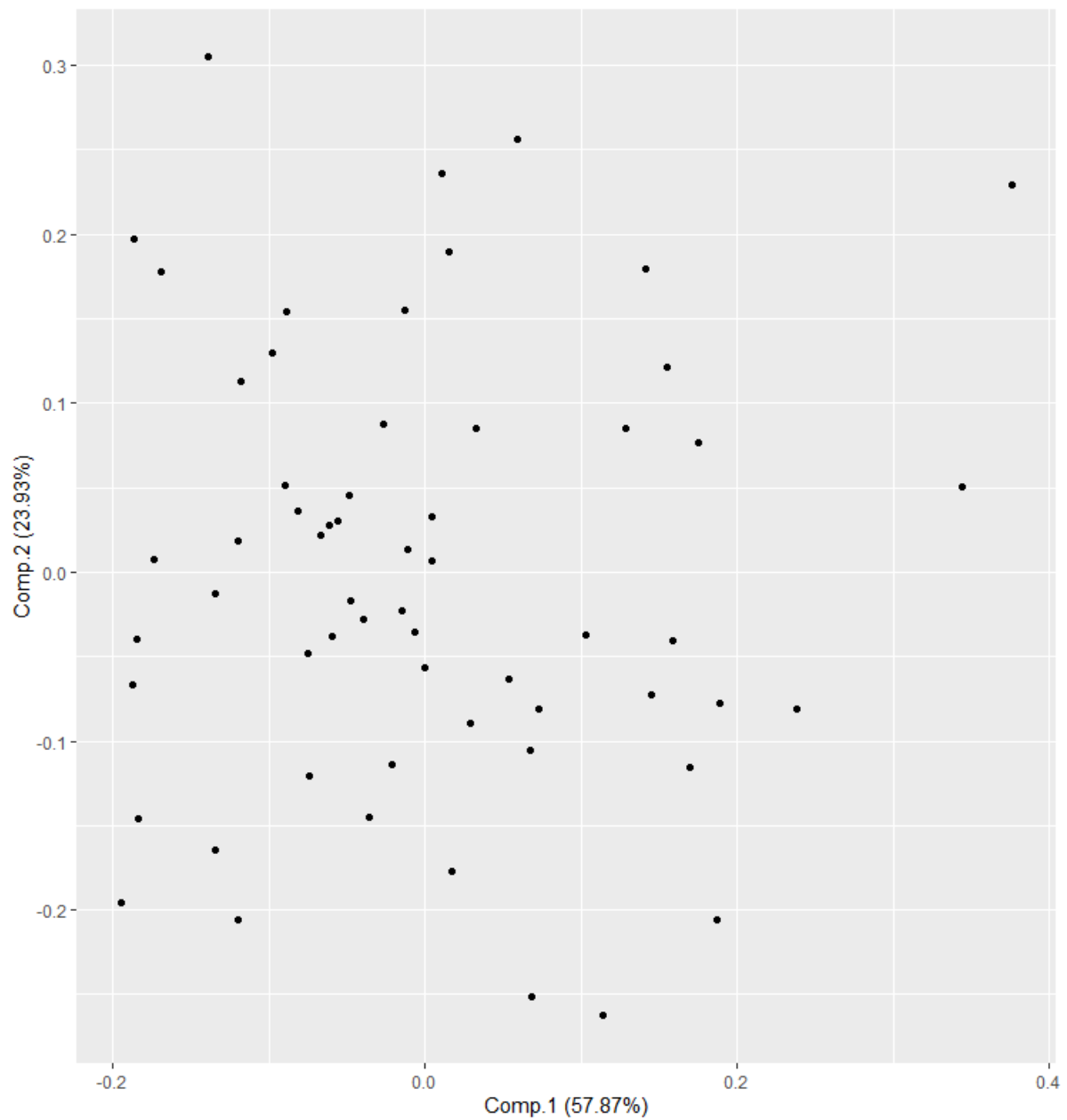
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
WDIM	0.207	0.142	0.422	0.443	0.168	0.731
CIRCUM	0.873	0.219		-0.131	0.330	-0.238
FBEYE	0.261	0.231	-0.121	-0.382	-0.768	0.358
EYEHD	0.326	-0.891	-0.173	0.173	-0.164	0.113
EARHD		-0.222	0.867	-0.355	-0.114	-0.235
JAW	0.128	0.187	0.135	0.697	-0.483	-0.460

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
SS loadings	1.000	1.000	1.000	1.000	1.000	1.000
Proportion Var	0.167	0.167	0.167	0.167	0.167	0.167
Cumulative Var	0.167	0.333	0.500	0.667	0.833	1.000

رسم نمودارها:

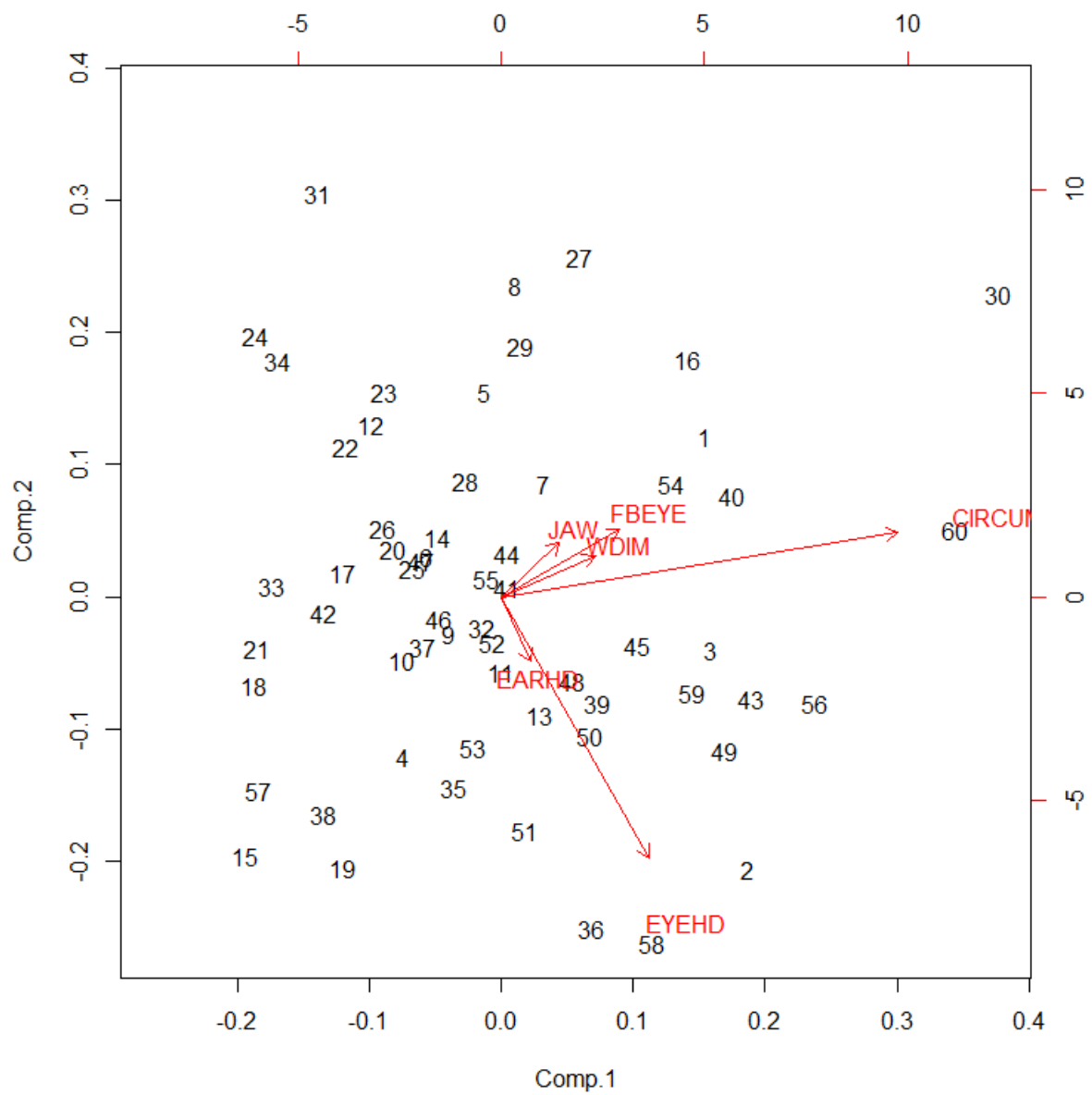
```
> require(ggfortify)
Loading required package: ggfortify
```

```
Loading required package: ggplot2
Warning message:
In (function (kind = NULL, normal.kind = NULL, sample.kind = NULL) :
  non-uniform 'Rounding' sampler used
> autoplot(football.pca)
```



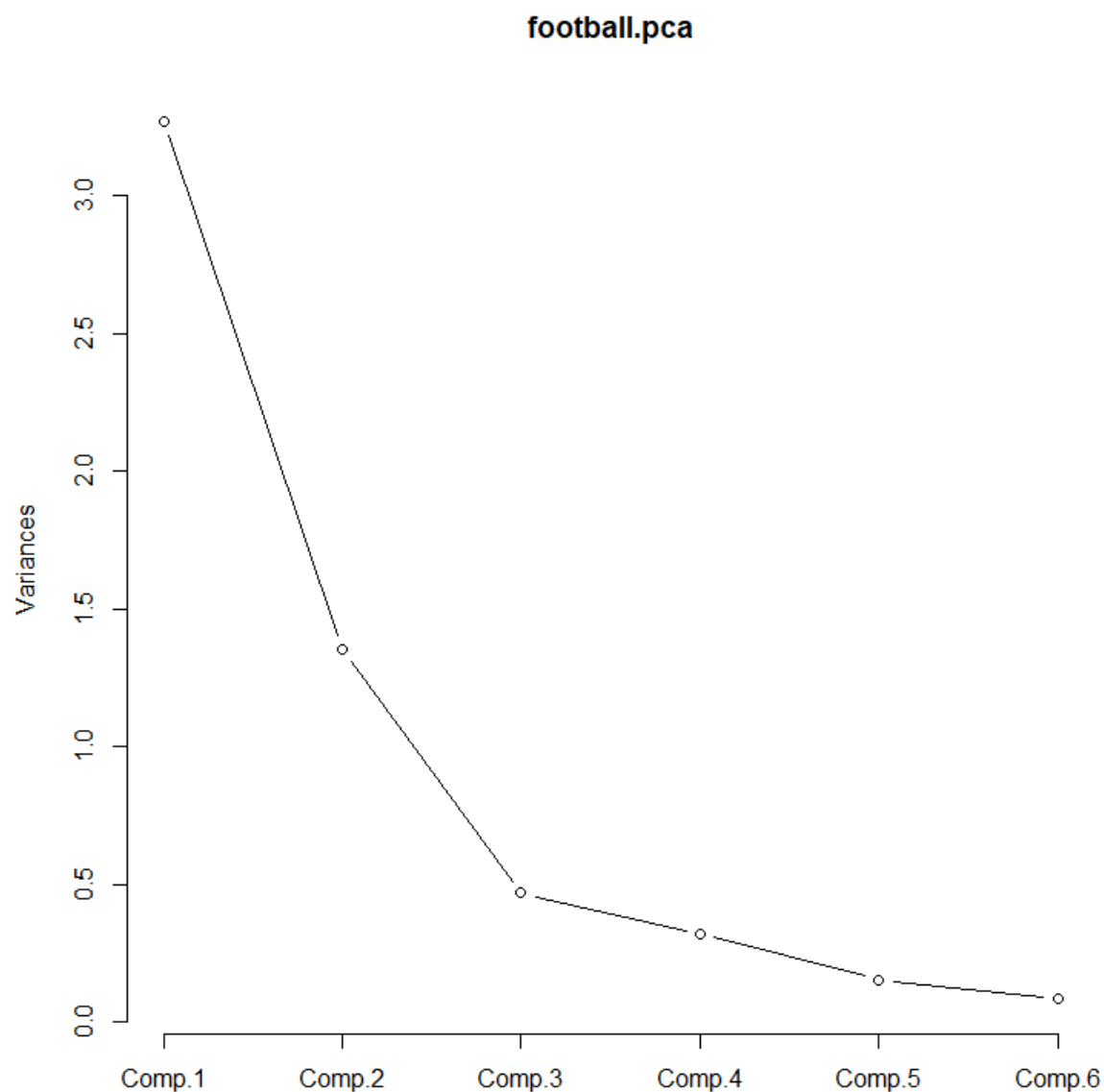
نمودار بالا اطلاعات خوبی در زمینه نرمال بودن داده‌ها، نقاط دور افتاده و همچنین میزان پراکنش داده‌ها نسبت به PCA نمایش می‌دهد. طبق نمودار بالا داده‌ها دارای توزیع نرمال هستند و نقطه دور افتاده هم نداریم.

```
> biplot(football.pca)
```



در نمودار بالا میتوان میزان تاثیرگذار بودن هر متغیر را بر اساس طول بردار مشخص نمود یعنی هرچه طول بردار بلندتر اثر بیشتر.

```
> plot(football.pca,type ="lines")
```



طبق نمودار بالا ما باید در کاهش بعد ۲ مولفه اصلی را نگه داریم.

در نمودار بالا باید نگاه به شیب خط کنیم یعنی تا جایی که شیب خط زیاده مولفه ها را نگه میداریم و از جایی که شیب خط کم شد مولفه ها را حذف می کنیم.

تحلیل مولفه اصلی با استفاده از ماتریس همبستگی (**R**)

ماتریس **R** را تعریف میکنیم :

```
> (R<-cor(football))
```

	WDIM	CIRCUM	FBEYE	EYEHD	EARHD	JAW
WDIM	1.00000000	0.61027644	0.36131523	0.06012317	0.25202932	0.60471992
CIRCUM	0.61027644	1.00000000	0.73019762	0.33816708	0.09074389	0.40853148
FBEYE	0.36131523	0.73019762	1.00000000	0.01370847	-0.02795587	0.31116595
EYEHD	0.06012317	0.33816708	0.01370847	1.00000000	0.29739257	-0.07874176
EARHD	0.25202932	0.09074389	-0.02795587	0.29739257	1.00000000	-0.08955570
JAW	0.60471992	0.40853148	0.31116595	-0.07874176	-0.08955570	1.00000000

مقادیر ویژه را بدست می آوریم:

```
> (eigen<-eigen(R))
eigen() decomposition
$values
[1] 2.5677926 1.3691056 0.9324399 0.6779565 0.3220909 0.1306146

$vectors
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	-0.5107369	0.008377599	0.4456068	0.03305842	-0.62058449	0.39280307
[2,]	-0.5613420	-0.086752441	-0.3196668	-0.02117623	-0.22757477	-0.72314844
[3,]	-0.4621079	0.146807619	-0.4753965	-0.47302264	0.31139558	0.46710208
[4,]	-0.1443386	-0.663982124	-0.3135873	0.59227988	0.09531954	0.28297679
[5,]	-0.1096575	-0.644048720	0.4703556	-0.48807667	0.31147128	-0.12926295
[6,]	-0.4214747	0.339382289	0.3920134	0.43092227	0.60001436	-0.08827771

کاهش بعد به کمک مقادیر ویژه :

```
> labda<-eigen$values
> which(labda>1)
[1] 1 2
```

در تحلیل مولفه اصلی با استفاده از ماتریس همبستگی مولفه‌هایی که مقادیر ویژه بزرگتر از ۱ دارند را نگه می‌داریم و بقیه حذف می‌کنیم پس در اینجا نیز به نتیجه قبل رسیدیم یعنی کاهش بعد به ۲ بعد.

تحلیل مولفه اصلی :

```
> football.pca.R<-princomp(covmat=R)
> summary(football.pca.R)
Importance of components:
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.6024333	1.1700878	0.9656293	0.8233811	0.56753050	0.36140646
Proportion of Variance	0.4279654	0.2281843	0.1554066	0.1129927	0.05368181	0.02176911
Cumulative Proportion	0.4279654	0.6561497	0.8115563	0.9245491	0.97823089	1.00000000

```
>
> loadings(football.pca.R)

Loadings:
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
WDIM	0.511	0.446	0.621	0.393		
CIRCUM	0.561	-0.320	0.228	-0.723		

FBEYE	0.462	0.147	-0.475	-0.473	-0.311	0.467	
EYEHD	0.144	-0.664	-0.314	0.592		0.283	
EARHD	0.110	-0.644	0.470	-0.488	-0.311	-0.129	
JAW	0.421	0.339	0.392	0.431	-0.600		
		Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
SS loadings		1.000	1.000	1.000	1.000	1.000	1.000
Proportion Var		0.167	0.167	0.167	0.167	0.167	0.167
Cumulative Var		0.167	0.333	0.500	0.667	0.833	1.000

برخلاف نتایج قبل در اینجا میبینیم که تحلیل مولفه اصلی به ما میگوید در کاهش بعد باید ۳ مولفه اصلی اول را نگه داریم و ۳ مولفه اصلی اول ۸۱.۱ درصد کل واریانس را نمایش میدهند.