

تحلیل مولفه اصلی برای داده های اندازه سرپسرن

ابتدا برای فراخوانی داده ها پکیج زیر را نصب می کنیم :

```
> install.packages("calibrate")
```

بعد از نصب پکیج از دستور زیر آن را اجرا میکنیم. دستور زیر به گونه ای کار میکند که در صورت نصب نبودن پکیج آن را نصب میکند :

```
> require(calibrate)
```

اکنون داده های خود را بصورت زیر تعریف میکنیم :

```
> data("heads")
> Head_Lenght<-heads$X1
> Head_Breadth<-heads$X2
> head_data=cbind(Head_Lenght,Head_Breadth)
```

۵ سطر ابتدایی از داده ها را نمایش می دهیم:

```
> head(head_data,5)
      Head_Lenght Head_Breadth
[1,]          191          155
[2,]          195          149
[3,]          181          148
[4,]          183          153
[5,]          176          144
```

میانگین اندازه دور سر(Head\_Breadth) و اندازه طول سر(Head\_Lenght) را بدست می آوریم :

```
> colMeans(head_data)
      Head_Lenght Head_Breadth
          185.72          151.12
```

تشکیل ماتریس واریانس\_کوواریانس برای داده ها(ماتریس S)

```
> (S<-cov(head_data))
      Head_Lenght Head_Breadth
Head_Lenght    95.29333    52.86833
Head_Breadth   52.86833    54.36000
```

محاسبه مقادیر ویژه و بردارهای ویژه ماتریس واریانس\_کوواریانس

```
> eigen(S)
eigen() decomposition
$values
[1] 131.5183  18.1350

$vectors
      [,1]      [,2]
[1,] -0.8249295  0.5652357
[2,] -0.5652357 -0.8249295
```

مشاهده می‌شود که  $\lambda_1 = 131.51$ ,  $\lambda_2 = 18.13$  پس میتوانیم سهم هر متغیر را از واریانس کل بدست آوریم:

$$\text{سهم } Z_1 \text{ از واریانس کل} = \frac{\lambda_1}{\text{sum}(\lambda)} = \frac{131.52}{131.52 + 18.135} = 0.87$$

$$\text{سهم } Z_2 \text{ از واریانس کل} = \frac{\lambda_2}{\text{sum}(\lambda)} = \frac{18.135}{131.52 + 18.135} = 0.13$$

نتیجه میگیریم که میتوان کاهش بعد انجام داد زیرا  $Z_1$  ۸۷ درصد از واریانس کل داده‌ها را نمایش میدهد.

تحلیل مولفه اصلی به کمک ماتریس واریانس\_کوواریانس

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

	Comp.1	Comp.2
Standard deviation	11.2364407	4.1724812
Proportion of Variance	0.8788199	0.1211801
Cumulative Proportion	0.8788199	1.0000000

در اینجا نیز میتوان نتیجه گرفته شده در بالا را مشاهده کرد.

همچنین میتوان انحراف استاندارد هر متغیر را در بخش (standard deviation) مشاهده کرد.

در اینجا نیز میتوانیم بردارهای ویژه را مشاهده کنیم:

```
> loadings(head.pca)

Loadings:
```

	Comp.1	Comp.2
Head_Lenght	0.825	0.565
Head_Breadth	0.565	-0.825

	Comp.1	Comp.2
SS loadings	1.0	1.0
Proportion Var	0.5	0.5
Cumulative Var	0.5	1.0

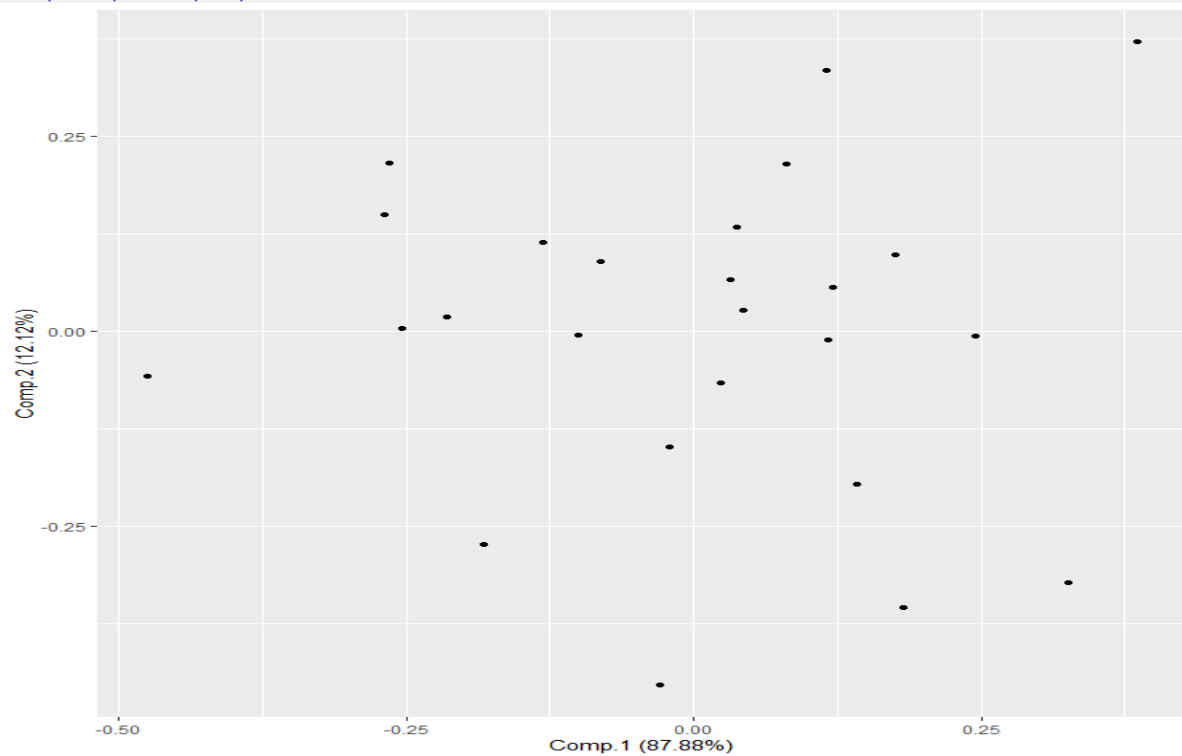
رسم نمودارها:

```
> 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(head.pca)

```



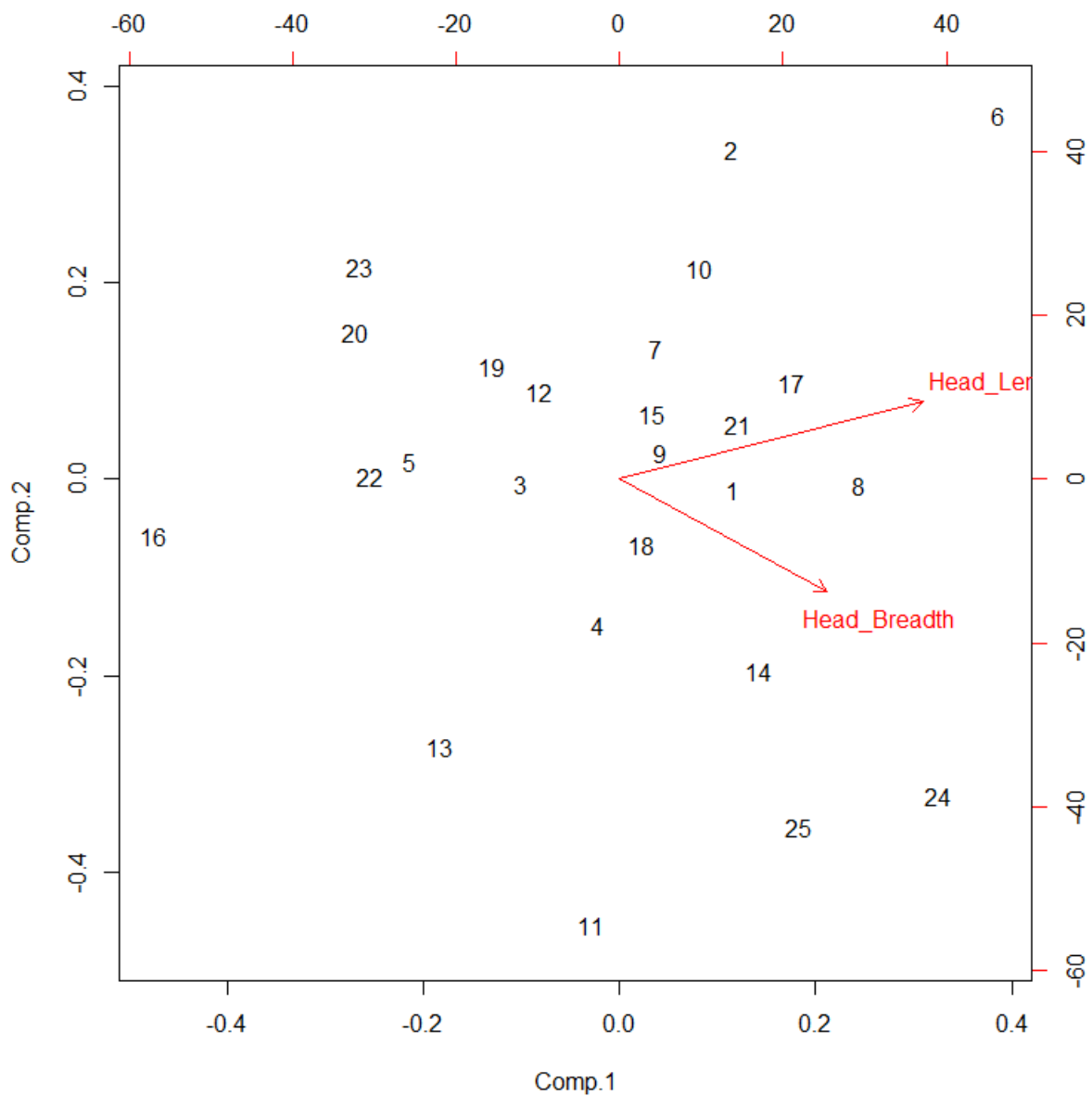
این نمودار اطلاعات خوبی در زمینه نرمال بودن داده‌ها، نقاط دور افتاده و همچنین میزان پراکنش داده‌ها نسبت به PCA نمایش می‌دهد. طبق نمودار بالا داده‌ها دارای توزیع نرمال چند متغیره هستند چون داده‌ها حول محور ۰ بدون الگوی خاصی پراکنده شده‌اند. طبق نمودار بالا نقطه دور افتاده نداریم. طبق نمودار بالا داده‌ها دارای خوشه بندی خاصی نیستند.

```

> biplot(head.pca)

```

در نمودار زیر میتوان میزان تاثیرگذار بودن هر متغیر را بر اساس طول بردار مشخص نمود



اکنون با استفاده از ماتریس همبستگی (**R**) تحلیل مولفه اصلی را انجام می‌دهیم :

مقادیر ویژه ماتریس **R** را بدست می‌آوریم

```
> (R<-cor(head_data))
      Head_Lenght Head_Breadth
Head_Lenght  1.0000000  0.7345555
Head_Breadth  0.7345555  1.0000000
>
> eigen(R)
eigen() decomposition
```

```
$values
[1] 1.7345555 0.2654445

$vectors
      [,1]      [,2]
[1,] 0.7071068 -0.7071068
[2,] 0.7071068  0.7071068
```

میبینیم که لاتدها نسبت به ماتریس S تغییر کرده‌اند و لاتدهای جدید ما برابر ۱.۷۳ و ۰.۲۶ می‌باشد.

```
> head.pca.R<-princomp(head_data,cor=F,scores = T,covmat =R)
Warning message:
In princomp.default(head_data, cor = F, scores = T, covmat = R) :
  both 'x' and 'covmat' were supplied: 'x' will be ignored
> summary(head.pca.R)
Importance of components:
              Comp.1      Comp.2
Standard deviation    1.3170253 0.5152130
Proportion of Variance 0.8672778 0.1327222
Cumulative Proportion 0.8672778 1.0000000
>
>
> loadings(head.pca.R)

Loadings:
              Comp.1 Comp.2
Head_Lenght    0.707  0.707
Head_Breadth    0.707 -0.707

              Comp.1 Comp.2
SS loadings      1.0    1.0
Proportion Var    0.5    0.5
Cumulative Var    0.5    1.0
```

طبق خروجی بالا می‌بینیم که مولفه اصلی اول ۸۶ درصد واریانس کل را بیان میکند که این مقدار با استفاده از ماتریس S برابر ۸۷ درصد بود.