

Практика №5. Многомерные методы

Долаева А. Р., г.20.М04-мм

13/05/2021

Вариант №3.

X1	X2	X3	X4	X5	X6
insomia.1	anoreksia.1	SBP.1	SV.1	CO.1	TPR.1

Данные на первый день исследования:

insomia - бессонница, категориальный признак (0,1,2)

anoreksia - анорексия, категориальный (0,1,2)

SBP - систолическое артериальное давление, метрический

SV - ударный объем крови, метрический

CO - сердечный выброс, метрический

TPR - общее периферическое (сосудистое) сопротивление, метрический.

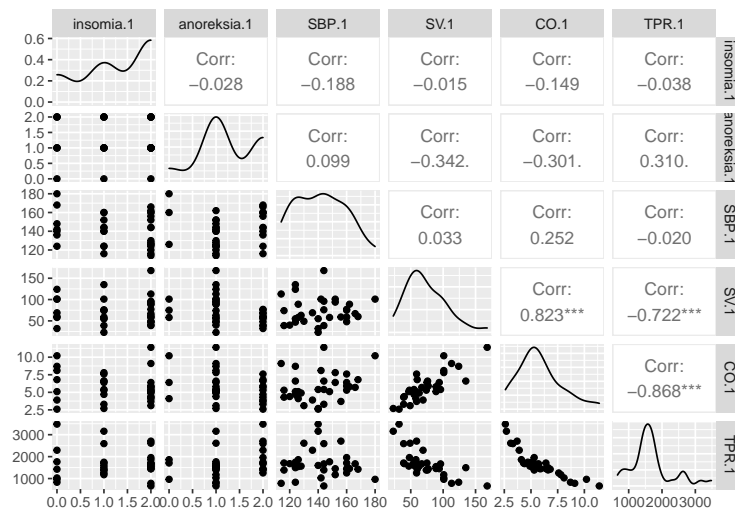
1. Факторный анализ. Найти наиболее значимые главные компоненты, построить факторные нагрузки, проинтерпретировать факторы, построить двумерные диаграммы факторов.

Читаем данные:

```
data <- read.table(file = "data.csv", header = TRUE, sep = ",")
X<-na.omit(data[,c("insomia.1", "anoreksia.1", "SBP.1", "SV.1", "CO.1", "TPR.1")])
```

Построим попарные диаграммы рассеяния:

```
ggpairs(X)
```



Из графиков можем предположить, что SV.1 и CO.1 положительно скоррелированы между собой, а TPR.1 отрицательно с SV.1 и CO.1.

Очевидно, что ударный объем крови (SV), сердечный выброс (CO) и общее периферическое сопротивление (TPR) связаны друг с другом.

Применяем метод главных компонент, и выводим вклады компонент:

```
pc<-princomp(scale(X))
pc$sdev/sum(pc$sdev)
```

```
##      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6
## 0.31176675 0.20982379 0.17359559 0.15952175 0.09846574 0.04682637
```

Так 31% информации содержит первая компонента, 21% вторая, 17% третья.
Сумма первых трех компонент 72%

```
sum(pc$loading[,seq(3)])
```

```
## [1] 0.7521103
```

Можно выбрать первые 2 компонента, которые дают более 50% информации. Мы отберем первые 3, выведем матрицу, показывающую, как связаны параметры с главными компонентами:

```
pc$loading[,seq(3)]
```

```
##              Comp.1      Comp.2      Comp.3
## insomia.1    0.04491319 0.61734859 0.71429965
## anoreksia.1  0.28631691 -0.32855307 0.56391127
## SBP.1        -0.08472262 -0.68998048 0.37202303
## SV.1         -0.53680536 0.06185078 0.05122519
## CO.1         -0.57177003 -0.15361282 0.09641548
## TPR.1        0.54198391 -0.08624449 -0.14648887
```

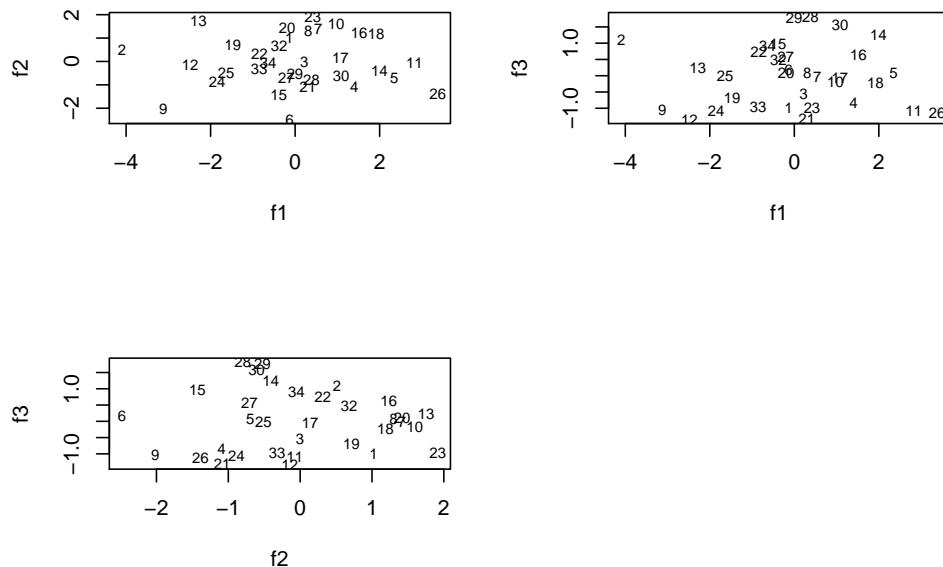
Первая компонента отвечает за SV, CO, TPR и в целом несет информацию о состоянии сердца: низкие ударный объем крови и сердечный выброс, высокое общее периферическое сопротивление.

Вторая связана с высокой бессонницей (insomia) и пониженным систолическим артериальным давлением (SBP), Эта компонента характеризуется состоянием нервной системы.

Третья с высокой бессонницей (insomia) и аноксией (anoreksia), и сообщает о состоянии здоровья со стороны пищеварительной системы.

Строим графики рассеивания и оцениваем разброс по компонентам:

```
op<-par(mfrow=c(2,2))
plot(pc$scores[,1],pc$scores[,2],type="n",xlab="f1", ylab="f2")
text(pc$scores[,1],pc$scores[,2],row.names(X),cex=0.7)
plot(pc$scores[,1],pc$scores[,3],type="n", xlab="f1",ylab="f3")
text(pc$scores[,1],pc$scores[,3],row.names(X),cex=0.7)
plot(pc$scores[,2],pc$scores[,3],type="n",xlab="f2",ylab="f3")
text(pc$scores[,2],pc$scores[,3],row.names(X),cex=0.7)
par(op)
```



Из первого графика по первой и второй компонентам: индивиды, находящиеся близко к левой границе (2), характеризуются высокими ударным объемом крови и сердечным выбросом, низким общим периферическим сопротивлением. У 26 индивида противоположные показатели. Индивидам, близким к нижней границе (6), свойственны низкий уровень бессоницы и повышенное систолическое артериальное давление. И наоборот, для тех, ко находится ближе к верхней границе.

Из второго графика по первой и третьей компонентам: индивиды, находящиеся близко к левой границе (2), характеризуются высокими ударным объемом крови и сердечным выбросом, низким общим периферическим сопротивлением. У 26 индивида противоположные показатели. Индивидов, близкие к нижней границе, отличается низкие уровни бессоницы и аноксии. И наоборот, для тех, ко находится ближе к верхней границе.

Из третьего графика по второй и третьей компонентам: индивиды, находящиеся близко к левой границе (6), характеризуются низким уровнем бессоницы и повышенным систолическим артериальным давлением. У 23 индивида противоположные показатели. Индивидам, близким к нижней границе (6), свойственны низкие уровни бессоницы и аноксии. И наоборот, для тех, ко находится ближе к верхней границе.

```
op<-par(mfrow=c(1,2))

plot(pc$scores[,2],pc$scores[,3],type="n",xlab="f1", ylab="f3")

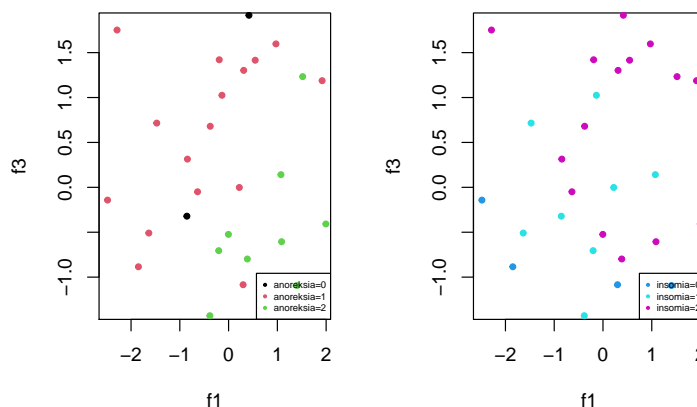
for(i in c(0,1,2))
  lines(pc$scores[X$anoreksia.1==i,1],pc$scores[X$anoreksia.1==i,2],
        type="p",pch=20,col=i+1)
legend('bottomright',c("anoreksia=0", "anoreksia=1", "anoreksia=2"),
      cex=0.5,pch=20,col=seq(3))

plot(pc$scores[,2],pc$scores[,3],type="n",xlab="f1", ylab="f3")

for(i in c(0,1,2))
  lines(pc$scores[X$insomnia.1==i,1],pc$scores[X$insomnia.1==i,2],
        type="p",pch=20,col=i+4)

legend('bottomright',c("insomnia=0", "insomnia=1", "insomnia=2"),
```

```
cex=0.5,pch=20,col=seq(3)+3)
```



2. Дискриминантный анализ. Найти дискриминантные функции, их веса, построить матрицу классификации (оценить чувствительность, специфичность, вероятность правильной классификации).

Объединяем признаки 0 и 1 в таблицах:

```
table(X$insomia.1)
```

```
##
##  0  1  2
##  7 10 16
```

```
table(X$anoreksia.1)
```

```
##
##  0  1  2
##  3 18 12
```

```
Group<-ifelse(X$insomia.1==2,1,0)
Group.<-ifelse(X$anoreksia.1==2,1,0)
```

Применяем линейный дискриминантный анализ и выводим коэффициенты (веса) для метрических переменных:

```
LDA<-lda(Group~.,subset(cbind(X,Group=Group),select=-c(insomia.1,anoreksia.1)))
LDA$scaling
```

```
##          LD1
## SBP.1  0.015411183
## SV.1   0.034057801
## CO.1  -1.369653549
## TPR.1 -0.002916678
```

Матрица классификации:

```
tab<-table(Group,predict(LDA,X)$class);tab
```

```
##
```

```
## Group  0  1
##      0 11  6
##      1  5 11
```

Оценим чувствительность и специфичность для предсказанных категорий фактора insomnia.1:

```
round(tab[1,1]/sum(tab[,1])*100)
```

```
## [1] 69
```

```
round(tab[2,2]/sum(tab[,2])*100)
```

```
## [1] 65
```

чувствительность равна 69%, процент предсказанных нулей. специфичность равна 65%, процент предсказанных единиц.

Оценим вероятность правильной классификации для предсказанных категорий фактора insomnia.1:

```
Prob<-sum(diag(tab))/sum(sum(tab));Prob
```

```
## [1] 0.6666667
```

Высокий показатель: 67% правильной классификации.

Значения дискриминантной функции:

```
P1<-predict(LDA,X)$x
LDA<-lda(Group~.,subset(cbind(X,Group=Group.),select=-c(insomnia.1,anoreksia.1)))
LDA$scaling
```

```
##          LD1
## SBP.1  0.033757358
## SV.1   -0.028807454
## CO.1   0.144888639
## TPR.1  0.000599413
```

```
tab<-table(Group,predict(LDA,X)$class);tab
```

```
##
## Group  0  1
##      0 10  7
##      1 13  3
```

Оценим чувствительность и специфичность для предсказанных категорий фактора anoreksia.1:

```
round(tab[1,1]/sum(tab[,1])*100)
```

```
## [1] 43
```

```
round(tab[2,2]/sum(tab[,2])*100)
```

```
## [1] 30
```

```
Prob<-sum(diag(tab))/sum(sum(tab));Prob
```

```
## [1] 0.3939394
```

Низкий показатель: 40% правильной классификации.

Значения дискриминантной функции:

```
P2<-predict(LDA,X)$x
cbind(P1=P1,P2=P2,Group=Group,Group.=Group.)
```

##		LD1	LD1	Group	Group.
## 1	1.17784100	-1.121433835	0	0	
## 2	-1.43242403	-2.434282765	1	0	
## 3	0.14822667	0.377049369	0	0	
## 4	0.59369659	0.037375217	0	1	
## 5	-0.35795938	1.264548397	0	1	
## 6	-0.66581746	1.277963179	0	1	
## 7	0.13049843	0.005342884	1	0	
## 8	1.01391536	-0.350124087	1	0	
## 9	-2.39966340	0.718837276	0	0	
## 10	0.46253859	0.037677834	1	0	
## 11	-1.79357683	1.842772777	0	0	
## 12	-0.03404501	-2.131800484	0	0	
## 13	-0.96773682	-2.123907848	1	0	
## 14	-0.37514862	1.220375557	1	1	
## 15	-1.42013555	1.050123068	0	1	
## 16	0.41338437	-0.064451205	1	1	
## 17	0.42873680	-0.175777283	0	1	
## 18	-0.01754080	0.486881383	1	0	
## 19	1.28310541	-2.355537828	0	0	
## 20	1.65008199	-1.149043721	1	0	
## 21	0.29750029	0.617321240	0	0	
## 22	0.72634664	-0.285776931	1	0	
## 23	1.04518999	-0.212086210	1	0	
## 24	-0.29276665	-0.830445439	0	0	
## 25	-0.15359382	-0.448824163	0	0	
## 26	-2.29509510	1.763226626	0	1	
## 27	-0.11766327	-0.042490434	0	1	
## 28	0.32796409	0.997473785	1	1	
## 29	0.96714698	0.399166674	1	1	
## 30	-0.69499317	1.455148498	1	1	
## 32	1.39624037	-0.625869075	1	0	
## 33	-0.55288184	0.686871509	0	0	
## 34	1.50862817	0.113696035	1	0	

Классификация по двум факторам:

красная точка - предсказанные значения низкой бессоницы индивидам;

зеленая точка - предсказанные значения высокой бессоницы индивидам;

черная окружность - предсказанные значения низкой анорексии индивидам;

голубая окружность - предсказанные значения высокой анорексии индивидам.

```
plot(P1,P2,type="n")
lines(P1[Group==0],P2[Group==0],type="p",pch=20,col=2)
lines(P1[Group==1],P2[Group==1],type="p",pch=20,col=3)
legend('topright',c("insomx<2","insom=2"),pch=1,col=c(2,3))
lines(P1[Group.==0],P2[Group.==0],type="p",pch=1,col=1,cex=2)
lines(P1[Group.==1],P2[Group.==1],type="p",pch=1,col=5,cex=2)
legend('bottomleft',c("anorex<2","anorex=2"),pch=1,col=c(1,5))
```

