HW5-577

LianZuo

2023-12-04

1. Read the description of the data from Help page. Name the variables that are discrete, continuous and categorical.

discrete variables: cyl,carb

continuous variables:pmg,disp,hp,drat,wt,qsec

categorical variables:vs,am,gear

1. Decide how many factors you may consider for factor analysis by
2. finding number of eigenvalues that are greater than the average eigenvalue. (You should report all the eigenvalues in your answer).

We may consider 2 factors in our study.

(ev.mtcars <- eigen(cor(mtcars))$values)

[1] 6.60840025 2.65046789 0.62719727 0.26959744 0.22345110 0.21159612  
 [7] 0.13526199 0.12290143 0.07704665 0.05203544 0.02204441

mean(ev.mtcars)

[1] 1

which(ev.mtcars>mean(ev.mtcars))

[1] 1 2

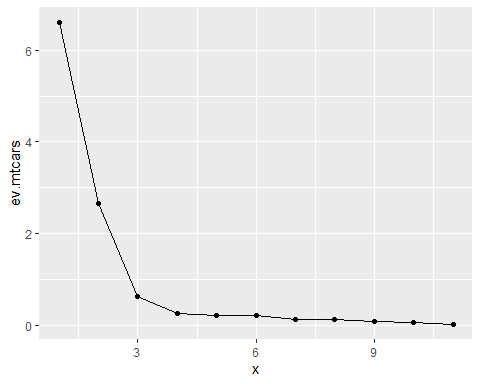
1. drawing a plot of eigenvalues and checking the point beyond which eigenvalues are relatively small and the line graph connecting has a sharp drop. (You should provide your graph here.)

We see that a sharp fall occurs at the second eigenvalue and afterwards the line has a very small slope and becomes more flat.

library(ggplot2)

Warning: package 'ggplot2' was built under R version 4.3.2

x <- 1:11  
ev.mtcars.data <- data.frame(cbind(x,ev.mtcars))  
ggplot(ev.mtcars.data,mapping = aes(x=x,y=ev.mtcars))+geom\_point()+geom\_line()



Now, perform a factor analysis using 3 factors with “varimax” rotation and answer the following questions.

mtcars.fac.vm <- factanal(mtcars,factors = 3,rotation = "varimax")  
mtcars.fac.vm

Call:  
factanal(x = mtcars, factors = 3, rotation = "varimax")  
  
Uniquenesses:  
 mpg cyl disp hp drat wt qsec vs am gear carb   
0.135 0.055 0.090 0.127 0.290 0.060 0.051 0.223 0.208 0.125 0.158   
  
Loadings:  
 Factor1 Factor2 Factor3  
mpg 0.643 -0.478 -0.473   
cyl -0.618 0.703 0.261   
disp -0.719 0.537 0.323   
hp -0.291 0.725 0.513   
drat 0.804 -0.241   
wt -0.778 0.248 0.524   
qsec -0.177 -0.946 -0.151   
vs 0.295 -0.805 -0.204   
am 0.880   
gear 0.908 0.224   
carb 0.114 0.559 0.719   
  
 Factor1 Factor2 Factor3  
SS loadings 4.380 3.520 1.578  
Proportion Var 0.398 0.320 0.143  
Cumulative Var 0.398 0.718 0.862  
  
Test of the hypothesis that 3 factors are sufficient.  
The chi square statistic is 30.53 on 25 degrees of freedom.  
The p-value is 0.205

1. Is it sufficient to have 3 factors in this study? Justify your answer by providing the test statistic value, and p-value.

Since p-value is 0.205, The chi square statistic is 30.53 on 25 degrees of freedom. we conclude that 3 factors considered are sufficient to the fit.

1. Using a cutoff value of 0.6, write down the variables that constitute factors 1, 2 and 3.

Variables mpg,cyl,disp,drat,wt,am and gear can be associated with Factor 1.

Variables cyl, hp, qsec and vs can be descibed by Factor 2.

Factor 3 can be considered as a measure of carb variable.

1. What are the estimated error variances associated with drat and am variables?

mtcars.fac.vm$uniquenesses[c(5,9)]

drat am   
0.2899972 0.2083812

The estimated error variances associated with drat and am variables is 0.2899972 and 0.2083812 .

1. Show the calculation how the sum of squared loadings for factor 1 is 4.38.

loads <- mtcars.fac.vm$loadings  
apply(loads^2,2,sum)[1]

Factor1   
4.380409

1. How much variance among the observed variables is explained by factor 2?

The variance among the observed variables is explained by factor 2 is 3.520 .

1. Show the calculation how the proportion of variance explained by factor 2 is found to be 0.320.

round(apply(loads^2,2,sum)[2]/11,2)

Factor2   
 0.32

1. [2 points] Write down the variable that is highly correlated with factor 1. What is this correlation value?

The gear variable that is highly correlated with factor 1 and its correlation value is 0.908 .

1. [2 points] Write down the variable that is highly correlated with factor 2. What is this correlation value?

The qsec variable that is highly correlated with factor 2 and its correlation value is -0.946 .

1. [2 points] How much variance in variable hp is explained by all the factors in the model? What do we call this variance?

community <- apply(loads^2,1,sum)  
community[4]

hp   
0.8732196

The variance in variable hp is explained by all the factors in the model is 0.8732196 , we call this variance as communality or the common variance.