peyvasteproject

kiyana mirbaghestan

2023-01-29

در [کورتز و مورایس، 2007]، «مساحت» خروجی برای اولین بار با یک تابع ln(x+1) تبدیل شد.

سپس چندین روش داده کاوی به کار گرفته شد.

1. X - مختصات فضایی محور x در نقشه پارک مونتهسینیو: 1 تا 9

2. مختصات فضایی محور y در نقشه پارک مونتهسینیو: 2 تا 9

3. ماه - ماه سال: 'ژانویه' تا 'دسامبر'

4. روز - روز هفته: "از دوشنبه" تا "یکشنبه"

5. FFMC - شاخص FFMC از سیستم FWI: 18.7 تا 96.20

6. شاخص DMC - DMC از سیستم FWI: 1.1 تا 291.3

7. شاخص DC - DC از سیستم FWI: 7.9 تا 860.6

8. ISI - شاخص ISI از سیستم FWI: 0.0 تا 56.10

9. دما - درجه حرارت بر حسب درجه سانتیگراد: 2.2 تا 33.30

10. RH - رطوبت نسبی بر حسب درصد: 15.0 تا 100

11. باد - سرعت باد بر حسب کیلومتر بر ساعت: 0.40 تا 9.40

12. باران - باران بیرونی در mm/m2: 0.0 تا 6.4

13. مساحت(پاسخ) - مساحت سوخته جنگل (در هکتار): 0.00 تا 1090.84

(این متغیر خروجی بسیار به سمت 0.0 منحرف است، بنابراین ممکن است باعث حس مدل سازی با تبدیل لگاریتمی شود).

##preprocessing

دیتاست را فراخوانی می کنیم.کارهایی که قبل از شروع کار لازم است انجام می دهیم.

fire<-read.csv("C:/Users/asus/Documents/forestfires.csv")  
fired<-as.data.frame(fire)

نوع داده ها:

str(fired)

## 'data.frame': 517 obs. of 13 variables:  
## $ X : int 7 7 7 8 8 8 8 8 8 7 ...  
## $ Y : int 5 4 4 6 6 6 6 6 6 5 ...  
## $ month: chr "mar" "oct" "oct" "mar" ...  
## $ day : chr "fri" "tue" "sat" "fri" ...  
## $ FFMC : num 86.2 90.6 90.6 91.7 89.3 92.3 92.3 91.5 91 92.5 ...  
## $ DMC : num 26.2 35.4 43.7 33.3 51.3 ...  
## $ DC : num 94.3 669.1 686.9 77.5 102.2 ...  
## $ ISI : num 5.1 6.7 6.7 9 9.6 14.7 8.5 10.7 7 7.1 ...  
## $ temp : num 8.2 18 14.6 8.3 11.4 22.2 24.1 8 13.1 22.8 ...  
## $ RH : int 51 33 33 97 99 29 27 86 63 40 ...  
## $ wind : num 6.7 0.9 1.3 4 1.8 5.4 3.1 2.2 5.4 4 ...  
## $ rain : num 0 0 0 0.2 0 0 0 0 0 0 ...  
## $ area : num 0 0 0 0 0 0 0 0 0 0 ...

summary(fired)

## X Y month day   
## Min. :1.000 Min. :2.0 Length:517 Length:517   
## 1st Qu.:3.000 1st Qu.:4.0 Class :character Class :character   
## Median :4.000 Median :4.0 Mode :character Mode :character   
## Mean :4.669 Mean :4.3   
## 3rd Qu.:7.000 3rd Qu.:5.0   
## Max. :9.000 Max. :9.0   
## FFMC DMC DC ISI   
## Min. :18.70 Min. : 1.1 Min. : 7.9 Min. : 0.000   
## 1st Qu.:90.20 1st Qu.: 68.6 1st Qu.:437.7 1st Qu.: 6.500   
## Median :91.60 Median :108.3 Median :664.2 Median : 8.400   
## Mean :90.64 Mean :110.9 Mean :547.9 Mean : 9.022   
## 3rd Qu.:92.90 3rd Qu.:142.4 3rd Qu.:713.9 3rd Qu.:10.800   
## Max. :96.20 Max. :291.3 Max. :860.6 Max. :56.100   
## temp RH wind rain   
## Min. : 2.20 Min. : 15.00 Min. :0.400 Min. :0.00000   
## 1st Qu.:15.50 1st Qu.: 33.00 1st Qu.:2.700 1st Qu.:0.00000   
## Median :19.30 Median : 42.00 Median :4.000 Median :0.00000   
## Mean :18.89 Mean : 44.29 Mean :4.018 Mean :0.02166   
## 3rd Qu.:22.80 3rd Qu.: 53.00 3rd Qu.:4.900 3rd Qu.:0.00000   
## Max. :33.30 Max. :100.00 Max. :9.400 Max. :6.40000   
## area   
## Min. : 0.00   
## 1st Qu.: 0.00   
## Median : 0.52   
## Mean : 12.85   
## 3rd Qu.: 6.57   
## Max. :1090.84

دراین قسمت مدل را برازش داده و ضرایب را برآورد میکنیم،پایین تر با استاندارد کردن این کار را انجام خواهیم داد زیرا با تفاوت مقیاس ها ممکن است به یک داده بیش از حد اهمیت داده شود.

model<-lm(area~X+Y+factor(month)+factor(day)+FFMC+DC+ISI+temp+RH+wind+rain,data=fired)  
model

##   
## Call:  
## lm(formula = area ~ X + Y + factor(month) + factor(day) + FFMC +   
## DC + ISI + temp + RH + wind + rain, data = fired)  
##   
## Coefficients:  
## (Intercept) X Y factor(month)aug   
## -62.726777 1.971740 0.751741 19.703569   
## factor(month)dec factor(month)feb factor(month)jan factor(month)jul   
## 26.243535 4.298081 18.089812 9.202270   
## factor(month)jun factor(month)mar factor(month)may factor(month)nov   
## -2.192176 -4.052049 7.046492 -5.620819   
## factor(month)oct factor(month)sep factor(day)mon factor(day)sat   
## 19.332580 33.552575 5.708702 20.017819   
## factor(day)sun factor(day)thu factor(day)tue factor(day)wed   
## 5.278037 10.493833 7.479151 4.538615   
## FFMC DC ISI temp   
## 0.303603 -0.044808 -0.559326 1.568709   
## RH wind rain   
## -0.002969 2.138627 -3.519250

خوشبختانه تعداد داده ای گمشده صفر است.

length(which(!complete.cases(fired)))

## [1] 0

متغیرهای زیر که کاراکتر بودند فاکتور می کنیم.

fire$month<-as.factor(fire$month)  
length(fire$month)

## [1] 517

str(fire$month)

## Factor w/ 12 levels "apr","aug","dec",..: 8 11 11 8 8 2 2 2 12 12 ...

fire$day<-as.factor(fire$day)  
length(fire$day)

## [1] 517

str(fire$day)

## Factor w/ 7 levels "fri","mon","sat",..: 1 6 3 1 4 4 2 2 6 3 ...

ماه و روز کاراکتری را حذف می کنیم.

time<-data.frame(fire$month,fire$day)  
firee<-data.frame(fire,time)  
firee<-firee[,-c(3,4)]  
str(firee)

## 'data.frame': 517 obs. of 13 variables:  
## $ X : int 7 7 7 8 8 8 8 8 8 7 ...  
## $ Y : int 5 4 4 6 6 6 6 6 6 5 ...  
## $ FFMC : num 86.2 90.6 90.6 91.7 89.3 92.3 92.3 91.5 91 92.5 ...  
## $ DMC : num 26.2 35.4 43.7 33.3 51.3 ...  
## $ DC : num 94.3 669.1 686.9 77.5 102.2 ...  
## $ ISI : num 5.1 6.7 6.7 9 9.6 14.7 8.5 10.7 7 7.1 ...  
## $ temp : num 8.2 18 14.6 8.3 11.4 22.2 24.1 8 13.1 22.8 ...  
## $ RH : int 51 33 33 97 99 29 27 86 63 40 ...  
## $ wind : num 6.7 0.9 1.3 4 1.8 5.4 3.1 2.2 5.4 4 ...  
## $ rain : num 0 0 0 0.2 0 0 0 0 0 0 ...  
## $ area : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ fire.month: Factor w/ 12 levels "apr","aug","dec",..: 8 11 11 8 8 2 2 2 12 12 ...  
## $ fire.day : Factor w/ 7 levels "fri","mon","sat",..: 1 6 3 1 4 4 2 2 6 3 ...

a<-as.numeric(firee$fire.month)  
b<-as.numeric(firee$fire.day)  
fireee<-data.frame(firee,a,b)  
fireeee<-fireee[,-c(12,13)]

#####Normalization(Min-Max) #####principal component

در این قسمت چون مولفه ها می توانند از واحد های مختلف باشند ، خوب است که از ابتدا آن ها را استاندارد کنیم.

DATA = (fireeee - min(fireeee)) / (max(fireeee) - min(fireeee))

m2<-lm(area~X+Y++FFMC+DC+ISI+temp+RH+wind+rain,data=DATA)  
m2

##   
## Call:  
## lm(formula = area ~ X + Y + +FFMC + DC + ISI + temp + RH + wind +   
## rain, data = DATA)  
##   
## Coefficients:  
## (Intercept) X Y FFMC DC ISI   
## -0.026096 1.777507 0.923195 0.117641 0.006429 -0.701600   
## temp RH wind rain   
## 1.078861 -0.134648 1.704794 -3.567796

str(DATA)

## 'data.frame': 517 obs. of 13 variables:  
## $ X : num 0.00642 0.00642 0.00642 0.00733 0.00733 ...  
## $ Y : num 0.00458 0.00367 0.00367 0.0055 0.0055 ...  
## $ FFMC: num 0.079 0.0831 0.0831 0.0841 0.0819 ...  
## $ DMC : num 0.024 0.0325 0.0401 0.0305 0.047 ...  
## $ DC : num 0.0864 0.6134 0.6297 0.071 0.0937 ...  
## $ ISI : num 0.00468 0.00614 0.00614 0.00825 0.0088 ...  
## $ temp: num 0.00752 0.0165 0.01338 0.00761 0.01045 ...  
## $ RH : num 0.0468 0.0303 0.0303 0.0889 0.0908 ...  
## $ wind: num 0.006142 0.000825 0.001192 0.003667 0.00165 ...  
## $ rain: num 0 0 0 0.000183 0 ...  
## $ area: num 0 0 0 0 0 0 0 0 0 0 ...  
## $ a : num 0.00733 0.01008 0.01008 0.00733 0.00733 ...  
## $ b : num 0.000917 0.0055 0.00275 0.000917 0.003667 ...

summary(DATA)

## X Y FFMC DMC   
## Min. :0.0009167 Min. :0.001833 Min. :0.01714 Min. :0.001008   
## 1st Qu.:0.0027502 1st Qu.:0.003667 1st Qu.:0.08269 1st Qu.:0.062887   
## Median :0.0036669 Median :0.003667 Median :0.08397 Median :0.099281   
## Mean :0.0042804 Mean :0.003942 Mean :0.08310 Mean :0.101639   
## 3rd Qu.:0.0064171 3rd Qu.:0.004584 3rd Qu.:0.08516 3rd Qu.:0.130542   
## Max. :0.0082505 Max. :0.008251 Max. :0.08819 Max. :0.267042   
## DC ISI temp RH   
## Min. :0.007242 Min. :0.000000 Min. :0.002017 Min. :0.01375   
## 1st Qu.:0.401250 1st Qu.:0.005959 1st Qu.:0.014209 1st Qu.:0.03025   
## Median :0.608889 Median :0.007700 Median :0.017693 Median :0.03850   
## Mean :0.502310 Mean :0.008270 Mean :0.017316 Mean :0.04060   
## 3rd Qu.:0.654450 3rd Qu.:0.009901 3rd Qu.:0.020901 3rd Qu.:0.04859   
## Max. :0.788933 Max. :0.051428 Max. :0.030527 Max. :0.09167   
## wind rain area   
## Min. :0.0003667 Min. :0.000e+00 Min. :0.0000000   
## 1st Qu.:0.0024752 1st Qu.:0.000e+00 1st Qu.:0.0000000   
## Median :0.0036669 Median :0.000e+00 Median :0.0004767   
## Mean :0.0036830 Mean :1.986e-05 Mean :0.0117774   
## 3rd Qu.:0.0044920 3rd Qu.:0.000e+00 3rd Qu.:0.0060229   
## Max. :0.0086172 Max. :5.867e-03 Max. :1.0000000   
## a b   
## Min. :0.0009167 Min. :0.0009167   
## 1st Qu.:0.0018334 1st Qu.:0.0018334   
## Median :0.0064171 Median :0.0036669   
## Mean :0.0061954 Mean :0.0034257   
## 3rd Qu.:0.0110007 3rd Qu.:0.0045836   
## Max. :0.0110007 Max. :0.0064171

نوع داده:

str(DATA)

## 'data.frame': 517 obs. of 13 variables:  
## $ X : num 0.00642 0.00642 0.00642 0.00733 0.00733 ...  
## $ Y : num 0.00458 0.00367 0.00367 0.0055 0.0055 ...  
## $ FFMC: num 0.079 0.0831 0.0831 0.0841 0.0819 ...  
## $ DMC : num 0.024 0.0325 0.0401 0.0305 0.047 ...  
## $ DC : num 0.0864 0.6134 0.6297 0.071 0.0937 ...  
## $ ISI : num 0.00468 0.00614 0.00614 0.00825 0.0088 ...  
## $ temp: num 0.00752 0.0165 0.01338 0.00761 0.01045 ...  
## $ RH : num 0.0468 0.0303 0.0303 0.0889 0.0908 ...  
## $ wind: num 0.006142 0.000825 0.001192 0.003667 0.00165 ...  
## $ rain: num 0 0 0 0.000183 0 ...  
## $ area: num 0 0 0 0 0 0 0 0 0 0 ...  
## $ a : num 0.00733 0.01008 0.01008 0.00733 0.00733 ...  
## $ b : num 0.000917 0.0055 0.00275 0.000917 0.003667 ...

سامری می گیریم:

summary(DATA)

## X Y FFMC DMC   
## Min. :0.0009167 Min. :0.001833 Min. :0.01714 Min. :0.001008   
## 1st Qu.:0.0027502 1st Qu.:0.003667 1st Qu.:0.08269 1st Qu.:0.062887   
## Median :0.0036669 Median :0.003667 Median :0.08397 Median :0.099281   
## Mean :0.0042804 Mean :0.003942 Mean :0.08310 Mean :0.101639   
## 3rd Qu.:0.0064171 3rd Qu.:0.004584 3rd Qu.:0.08516 3rd Qu.:0.130542   
## Max. :0.0082505 Max. :0.008251 Max. :0.08819 Max. :0.267042   
## DC ISI temp RH   
## Min. :0.007242 Min. :0.000000 Min. :0.002017 Min. :0.01375   
## 1st Qu.:0.401250 1st Qu.:0.005959 1st Qu.:0.014209 1st Qu.:0.03025   
## Median :0.608889 Median :0.007700 Median :0.017693 Median :0.03850   
## Mean :0.502310 Mean :0.008270 Mean :0.017316 Mean :0.04060   
## 3rd Qu.:0.654450 3rd Qu.:0.009901 3rd Qu.:0.020901 3rd Qu.:0.04859   
## Max. :0.788933 Max. :0.051428 Max. :0.030527 Max. :0.09167   
## wind rain area   
## Min. :0.0003667 Min. :0.000e+00 Min. :0.0000000   
## 1st Qu.:0.0024752 1st Qu.:0.000e+00 1st Qu.:0.0000000   
## Median :0.0036669 Median :0.000e+00 Median :0.0004767   
## Mean :0.0036830 Mean :1.986e-05 Mean :0.0117774   
## 3rd Qu.:0.0044920 3rd Qu.:0.000e+00 3rd Qu.:0.0060229   
## Max. :0.0086172 Max. :5.867e-03 Max. :1.0000000   
## a b   
## Min. :0.0009167 Min. :0.0009167   
## 1st Qu.:0.0018334 1st Qu.:0.0018334   
## Median :0.0064171 Median :0.0036669   
## Mean :0.0061954 Mean :0.0034257   
## 3rd Qu.:0.0110007 3rd Qu.:0.0045836   
## Max. :0.0110007 Max. :0.0064171

library(devtools)

## Warning: package 'devtools' was built under R version 4.2.2

## Loading required package: usethis

library(ggbiplot)

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.2.2

## Loading required package: plyr

## Loading required package: scales

## Loading required package: grid

تحلیل ممیزی مولفه اصلی را انجام می دهیم:

library(graphics)  
library(psych)

## Warning: package 'psych' was built under R version 4.2.2

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:scales':  
##   
## alpha, rescale

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

DATA.pca<-princomp(DATA,cor=TRUE)   
summary(DATA.pca(

1.7 رادیکال لانداحت1،.......0.44رادیکال لاندا حت13 است.

برای اینکه بفهمیم چند مولفه اصلی لازم است سطر سوم را نگاه میکنیم که در ایجا بنظر اگر معیار بالای0.7 باشد 6متغیر لازم داریم.

## Importance of components:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5  
## Standard deviation 1.7025252 1.2705254 1.1770316 1.1469103 1.02550756  
## Proportion of Variance 0.2229686 0.1241719 0.1065695 0.1011849 0.08089737  
## Cumulative Proportion 0.2229686 0.3471405 0.4537100 0.5548949 0.63579223  
## Comp.6 Comp.7 Comp.8 Comp.9 Comp.10  
## Standard deviation 0.99216661 0.97095511 0.89912230 0.82184562 0.68064484  
## Proportion of Variance 0.07572266 0.07251952 0.06218622 0.05195617 0.03563672  
## Cumulative Proportion 0.71151489 0.78403441 0.84622064 0.89817681 0.93381353  
## Comp.11 Comp.12 Comp.13  
## Standard deviation 0.66629955 0.46352961 0.44900924  
## Proportion of Variance 0.03415039 0.01652767 0.01550841  
## Cumulative Proportion 0.96796392 0.98449159 1.00000000

DATA.pca$loadings

با ای ها مولفه ها را میسازیم:مثلا مولفه اصلی ائل ضرایب -0.419،-0.425،-0.423،-0.36و-0.484و0.233و0.11و0.1249است.

## Loadings:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10  
## X 0.600 0.317 0.113 0.113 0.461   
## Y 0.601 0.324 0.117 0.131 -0.457   
## FFMC -0.419 -0.106 0.144 0.283 0.166 0.101 -0.363 -0.553   
## DMC -0.425 0.112 -0.446 -0.199 0.218 -0.156   
## DC -0.423 0.252 -0.387 0.208 0.248 0.119   
## ISI -0.360 0.180 -0.321 0.256 0.156 0.127 -0.421 0.411   
## temp -0.484 0.201 -0.121 -0.251 0.262 0.246   
## RH 0.233 0.169 -0.682 -0.129 0.153 -0.326   
## wind 0.119 0.148 -0.489 0.481 -0.146 0.335 0.581   
## rain 0.221 -0.179 -0.211 0.905 -0.190   
## area 0.170 0.134 0.186 -0.922 -0.220   
## a -0.312 0.519 0.326 0.318 0.515   
## b -0.124 0.127 -0.174 0.182 -0.671 -0.169 0.645   
## Comp.11 Comp.12 Comp.13  
## X 0.527   
## Y -0.513   
## FFMC 0.425 -0.215   
## DMC 0.393 0.566   
## DC -0.690   
## ISI -0.464 0.227   
## temp -0.135 -0.660 0.229   
## RH -0.530   
## wind -0.132   
## rain   
## area   
## a -0.166 0.349   
## b   
##   
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9  
## SS loadings 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000  
## Proportion Var 0.077 0.077 0.077 0.077 0.077 0.077 0.077 0.077 0.077  
## Cumulative Var 0.077 0.154 0.231 0.308 0.385 0.462 0.538 0.615 0.692  
## Comp.10 Comp.11 Comp.12 Comp.13  
## SS loadings 1.000 1.000 1.000 1.000  
## Proportion Var 0.077 0.077 0.077 0.077  
## Cumulative Var 0.769 0.846 0.923 1.000

loadings(DATA.pca)

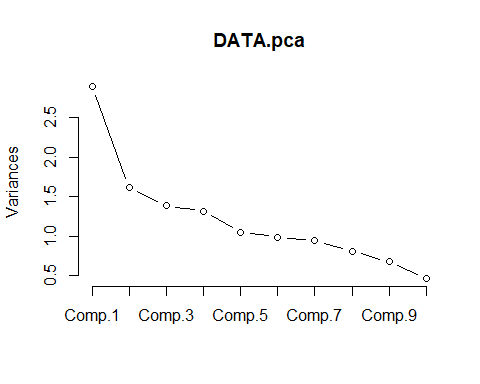
ضرایب ایکس برای ایگرگ ها:

##   
## Loadings:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10  
## X 0.600 0.317 0.113 0.113 0.461   
## Y 0.601 0.324 0.117 0.131 -0.457   
## FFMC -0.419 -0.106 0.144 0.283 0.166 0.101 -0.363 -0.553   
## DMC -0.425 0.112 -0.446 -0.199 0.218 -0.156   
## DC -0.423 0.252 -0.387 0.208 0.248 0.119   
## ISI -0.360 0.180 -0.321 0.256 0.156 0.127 -0.421 0.411   
## temp -0.484 0.201 -0.121 -0.251 0.262 0.246   
## RH 0.233 0.169 -0.682 -0.129 0.153 -0.326   
## wind 0.119 0.148 -0.489 0.481 -0.146 0.335 0.581   
## rain 0.221 -0.179 -0.211 0.905 -0.190   
## area 0.170 0.134 0.186 -0.922 -0.220   
## a -0.312 0.519 0.326 0.318 0.515   
## b -0.124 0.127 -0.174 0.182 -0.671 -0.169 0.645   
## Comp.11 Comp.12 Comp.13  
## X 0.527   
## Y -0.513   
## FFMC 0.425 -0.215   
## DMC 0.393 0.566   
## DC -0.690   
## ISI -0.464 0.227   
## temp -0.135 -0.660 0.229   
## RH -0.530   
## wind -0.132   
## rain   
## area   
## a -0.166 0.349   
## b   
##   
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9  
## SS loadings 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000  
## Proportion Var 0.077 0.077 0.077 0.077 0.077 0.077 0.077 0.077 0.077  
## Cumulative Var 0.077 0.154 0.231 0.308 0.385 0.462 0.538 0.615 0.692  
## Comp.10 Comp.11 Comp.12 Comp.13  
## SS loadings 1.000 1.000 1.000 1.000  
## Proportion Var 0.077 0.077 0.077 0.077  
## Cumulative Var 0.769 0.846 0.923 1.000

plot(DATA.pca, type="lines")

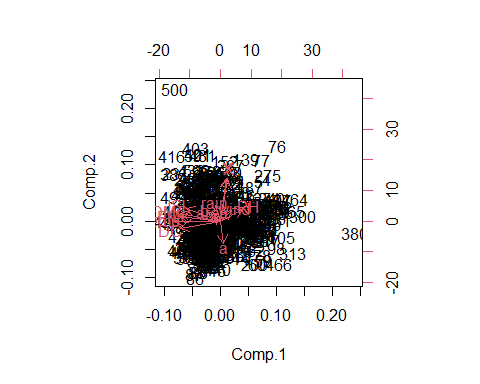
شکست در متغیر6 رخ داده به این معنی که در بعد ازآن کاهش واریانس شدید داریم

elbow



biplot(DATA.pca)

شکل صفحه بعد ارتفاع هرمیله برابر واریانس یکی از مولفه ها استدر زیر اگر هر یک از خطوط را ادامه دهیم متغیر ویژگی مورد نظر را کمتر دارا است.



yDATA<-DATA.pca$scores  
head(yDATA)#e'i

## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6  
## [1,] 3.4619055 0.7659244 -0.1844547 0.4108252 1.60984899 0.2991678  
## [2,] 0.1442968 -0.3761890 1.6893250 0.9113803 -1.28563942 0.4628570  
## [3,] 0.5627201 -0.5334295 1.8168833 0.4375973 -0.06372262 0.6765000  
## [4,] 3.2335547 2.1509317 0.1751244 -1.4207120 1.38489671 0.7731040  
## [5,] 2.6720484 2.0564197 0.6975824 -1.2563941 -0.38342534 0.6422584  
## [6,] -0.5447141 2.1983903 -0.4396010 1.4877467 0.35679389 0.6263357  
## Comp.7 Comp.8 Comp.9 Comp.10 Comp.11 Comp.12  
## [1,] 0.05730138 -0.13463143 0.3255265 -0.14412971 0.30405046 0.2106985  
## [2,] 0.71080077 0.81388972 -0.8193792 0.57294447 0.73645251 0.1478299  
## [3,] 0.55064841 0.03797649 -0.7399018 0.45013713 0.75257811 0.5791841  
## [4,] 0.17527998 -0.11595249 -2.0842475 -0.38561339 -0.07178193 -1.0167090  
## [5,] -0.39418654 0.46950390 -2.5555831 0.03632114 -0.28580699 -1.1122840  
## [6,] -0.37360892 -0.19539290 0.3979147 0.47320257 -0.24895848 0.1258685  
## Comp.13  
## [1,] 0.2380181  
## [2,] -0.6566165  
## [3,] -0.8411546  
## [4,] 0.4200231  
## [5,] 0.7479538  
## [6,] -0.4607539

#Factor analysis  
#m=6

#Factor analysis  
#m=6  
DATA.fac<-factanal(DATA,6,rotation="varimax")  
DATA.fac

دراین قسمت تعداد مهمترین مولفه های اصلی را در نطر میگیریم.

##   
## Call:  
## factanal(x = DATA, factors = 6, rotation = "varimax")  
##   
## Uniquenesses:  
## X Y FFMC DMC DC ISI temp RH wind rain area a b   
## 0.629 0.197 0.585 0.005 0.343 0.005 0.124 0.005 0.851 0.963 0.981 0.005 0.949   
##   
## Loadings:  
## Factor1 Factor2 Factor3 Factor4 Factor5 Factor6  
## X 0.604   
## Y 0.893   
## FFMC 0.532 0.297 -0.174   
## DMC 0.188 0.946 -0.177 0.125 0.133   
## DC 0.223 0.638 -0.113 0.155 0.396   
## ISI 0.973 0.105 -0.141   
## temp 0.425 0.291 -0.281 -0.227 0.691   
## RH -0.239 0.122 0.910 -0.294   
## wind -0.364   
## rain 0.150   
## area   
## a 0.956 -0.162 0.234   
## b 0.104 -0.182   
##   
## Factor1 Factor2 Factor3 Factor4 Factor5 Factor6  
## SS loadings 1.572 1.507 1.186 1.131 1.000 0.961  
## Proportion Var 0.121 0.116 0.091 0.087 0.077 0.074  
## Cumulative Var 0.121 0.237 0.328 0.415 0.492 0.566  
##   
## Test of the hypothesis that 6 factors are sufficient.  
## The chi square statistic is 16.3 on 15 degrees of freedom.  
## The p-value is 0.363

متغیر سوم ،چهارم،پنجم،ششم،هفتم،هشتم و سیزدهم با داشتن مقادیر زیاد روی عامل اول بارگذاری شده اند.

متغیر اسوم،چهارم،پنجم،ششم و هفتم روی عامل دوم هم بار دارند

DATA.load<-DATA.fac$loadings  
DATA.load

##   
## Loadings:  
## Factor1 Factor2 Factor3 Factor4 Factor5 Factor6  
## X 0.604   
## Y 0.893   
## FFMC 0.532 0.297 -0.174   
## DMC 0.188 0.946 -0.177 0.125 0.133   
## DC 0.223 0.638 -0.113 0.155 0.396   
## ISI 0.973 0.105 -0.141   
## temp 0.425 0.291 -0.281 -0.227 0.691   
## RH -0.239 0.122 0.910 -0.294   
## wind -0.364   
## rain 0.150   
## area   
## a 0.956 -0.162 0.234   
## b 0.104 -0.182   
##   
## Factor1 Factor2 Factor3 Factor4 Factor5 Factor6  
## SS loadings 1.572 1.507 1.186 1.131 1.000 0.961  
## Proportion Var 0.121 0.116 0.091 0.087 0.077 0.074  
## Cumulative Var 0.121 0.237 0.328 0.415 0.492 0.566

L<-DATA.load[1:2,]#L matrix  
S<-cov(DATA)  
si<-c()  
LLP=L%\*%t(L)  
for(i in 1:2){si[i]=S[i,i]-sum(L[i,]^2)}  
si

## [1] -0.3711376 -0.8030417

R<-cor(DATA)  
si\_mat<-diag(c(si[1],si[2]))  
si\_mat

## [,1] [,2]  
## [1,] -0.3711376 0.0000000  
## [2,] 0.0000000 -0.8030417

#Hambastegi kanooni

همبستگی را با تعریف کردن دو بردار دلخواه بدست می اوریم.

x<-DATA[,c(1,2,3,4,5,6,7,8,9,10,12,13)]  
y<-DATA[,c(11)]  
corL<-cancor(x,y)  
corL

## $cor

در مورد این عدد یعنی همبستگی بسیار کمی بین متغیرهای پیشگو وپاسخ داریم.

## [1] 0.1572381  
##   
## $xcoef

ها نمایش داده میشوند. e`i در زیر

## [,1] [,2] [,3] [,4] [,5]  
## X 9.02579036 -15.011368414 5.78864362 -6.98734562 -0.68557878  
## Y 2.52745620 46.229084023 -1.44596832 -3.56749749 5.63995962  
## FFMC -0.51526216 0.029607867 -8.80697480 -3.47620185 -0.90068936  
## DMC 0.52677317 -0.030269310 0.10034721 0.65982206 -0.69155197  
## DC -0.07017924 0.004032622 -0.01336873 0.02051735 0.27021107  
## ISI -2.93055216 0.168394667 -0.55825306 0.85676568 0.05632112  
## temp 4.70247148 -0.270212260 0.89579334 -1.37479764 -0.09037493  
## RH -0.88719847 0.050979980 -0.16900612 0.25937815 0.01705071  
## wind 8.55105876 -0.491358836 1.62892672 -2.49995677 -0.16433939  
## rain -15.60098928 0.896460209 -2.97189728 4.56104908 0.29982920  
## a 4.66960549 -0.268323723 0.88953256 -1.36518906 -0.08974329  
## b 2.39661983 -0.137713979 0.45654207 -0.70066715 -0.04605968  
## [,6] [,7] [,8] [,9] [,10]  
## X -2.569936899 10.04837453 -4.74342101 5.94279070 0.909134265  
## Y -1.397364888 2.81140332 -0.12707219 -1.59766053 0.391259977  
## FFMC 5.108846605 1.78376811 -2.86100240 -0.41371664 -0.606897381  
## DMC 0.063031429 0.32160192 0.28524195 0.18235814 0.053843257  
## DC 0.002159086 0.04751666 0.04175184 -0.04708894 0.004152039  
## ISI -12.218055209 1.41913138 2.58669088 2.22221330 -0.024746819  
## temp -0.537816695 -8.65286636 -8.95554769 -2.03322868 -1.546354083  
## RH 0.101467952 -0.32795422 -3.80736010 -0.45737591 -0.734284087  
## wind -0.977975557 3.16091145 -1.89930670 -27.00858383 -1.781868245  
## rain 1.784268663 -5.76692863 3.46519235 -2.81853079 164.996425357  
## a -0.534057848 1.72612654 -1.03718302 0.84362771 0.327248879  
## b -0.274098879 0.88591405 -0.53232193 0.43298195 0.167956620  
## [,11] [,12]  
## X 7.7596450 -1.41933032  
## Y 0.6648735 -2.15922666  
## FFMC 0.8789632 0.19290948  
## DMC -0.3156677 0.01484535  
## DC 0.1367590 -0.01660873  
## ISI -2.3676045 -0.55828959  
## temp -1.4616887 -0.56458647  
## RH -0.7797161 0.34995013  
## wind -0.7814174 0.56223677  
## rain -8.0561226 -6.70914993  
## a -12.0148819 1.73487168  
## b 0.5818040 25.51315482  
##   
## $ycoef  
## [,1]  
## [1,] 0.7543938  
##   
## $xcenter  
## X Y FFMC DMC DC ISI   
## 4.280413e-03 3.941739e-03 8.309622e-02 1.016394e-01 5.023102e-01 8.270382e-03   
## temp RH wind rain a b   
## 1.731617e-02 4.060009e-02 3.683035e-03 1.985941e-05 6.195428e-03 3.425749e-03   
##   
## $ycenter  
## [1] 0.01177743

A<-corL$xcoef  
A

## [,1] [,2] [,3] [,4] [,5]  
## X 9.02579036 -15.011368414 5.78864362 -6.98734562 -0.68557878  
## Y 2.52745620 46.229084023 -1.44596832 -3.56749749 5.63995962  
## FFMC -0.51526216 0.029607867 -8.80697480 -3.47620185 -0.90068936  
## DMC 0.52677317 -0.030269310 0.10034721 0.65982206 -0.69155197  
## DC -0.07017924 0.004032622 -0.01336873 0.02051735 0.27021107  
## ISI -2.93055216 0.168394667 -0.55825306 0.85676568 0.05632112  
## temp 4.70247148 -0.270212260 0.89579334 -1.37479764 -0.09037493  
## RH -0.88719847 0.050979980 -0.16900612 0.25937815 0.01705071  
## wind 8.55105876 -0.491358836 1.62892672 -2.49995677 -0.16433939  
## rain -15.60098928 0.896460209 -2.97189728 4.56104908 0.29982920  
## a 4.66960549 -0.268323723 0.88953256 -1.36518906 -0.08974329  
## b 2.39661983 -0.137713979 0.45654207 -0.70066715 -0.04605968  
## [,6] [,7] [,8] [,9] [,10]  
## X -2.569936899 10.04837453 -4.74342101 5.94279070 0.909134265  
## Y -1.397364888 2.81140332 -0.12707219 -1.59766053 0.391259977  
## FFMC 5.108846605 1.78376811 -2.86100240 -0.41371664 -0.606897381  
## DMC 0.063031429 0.32160192 0.28524195 0.18235814 0.053843257  
## DC 0.002159086 0.04751666 0.04175184 -0.04708894 0.004152039  
## ISI -12.218055209 1.41913138 2.58669088 2.22221330 -0.024746819  
## temp -0.537816695 -8.65286636 -8.95554769 -2.03322868 -1.546354083  
## RH 0.101467952 -0.32795422 -3.80736010 -0.45737591 -0.734284087  
## wind -0.977975557 3.16091145 -1.89930670 -27.00858383 -1.781868245  
## rain 1.784268663 -5.76692863 3.46519235 -2.81853079 164.996425357  
## a -0.534057848 1.72612654 -1.03718302 0.84362771 0.327248879  
## b -0.274098879 0.88591405 -0.53232193 0.43298195 0.167956620  
## [,11] [,12]  
## X 7.7596450 -1.41933032  
## Y 0.6648735 -2.15922666  
## FFMC 0.8789632 0.19290948  
## DMC -0.3156677 0.01484535  
## DC 0.1367590 -0.01660873  
## ISI -2.3676045 -0.55828959  
## temp -1.4616887 -0.56458647  
## RH -0.7797161 0.34995013  
## wind -0.7814174 0.56223677  
## rain -8.0561226 -6.70914993  
## a -12.0148819 1.73487168  
## b 0.5818040 25.51315482

B<-corL$ycoef  
B

## [,1]  
## [1,] 0.7543938

cancorL<-corL$cor  
cancorL

## [1] 0.1572381

library("MASS")

در حالت اول با فرض نرمال بودن دو جامعه و برابری واریانس دو جامعه با احتمال های نابرابرولی تعداد بسیار بالا بود.

xlda<-lda(area~.,proir=c(1,1,1,1,1,1,1,1,1,1,1,1)/12,data=fireeee)  
xlda$scaling[,1]

## X Y FFMC DMC DC   
## -0.0555795382 0.0163906943 -0.0150016045 0.0006422652 0.0001813784   
## ISI temp RH wind rain   
## -0.0070097181 -0.0425261909 -0.0207219054 -0.0840492198 11.1918471603   
## a b   
## 0.0111943938 -0.0549010690

predict(xlda,fireeee[,1:13])$class

## [1] 0 0 0 0 2.21 0 0 0 0 0   
## [11] 0 0 11.16 0 0 0 0 0 0 0   
## [21] 0 0 0.17 0 0 0 0 0 0 0   
## [31] 0 6.1 0 0 0 0 0 0 0 0   
## [41] 0 0 0 0 0 0 0 0 0 0   
## [51] 0 0 0 0 0 0 0 0 0 0   
## [61] 0 0 0 0 0 0 0 0 0 0   
## [71] 0 0 0 0 0 0.36 1.36 0 0 0   
## [81] 0 0 0 0 0 0 0 0 0 0   
## [91] 0 0 0 0 0 0 0 0 0 0   
## [101] 0 0 0 0 0 0 0 0 0 0   
## [111] 0 0 0 0 0 0 0 0 0 0   
## [121] 0 0 0 0 0 0 0 0 0 0   
## [131] 0 0 0 0 0 0 0 0 0.36 0   
## [141] 0 0 0.61 0 0 0 0 0.96 1.07 0   
## [151] 0 1.36 0 0 95.18 1.56 0 0 0 0   
## [161] 0 0 0 0 0 0 2.29 0 2.53 0   
## [171] 0 2.69 0 0 3.5 0 10.93 0 4.88 0   
## [181] 0 5.44 0 0 0 0 0 0 0 0   
## [191] 0 0 0 0 0 10.02 4.61 0 0 11.32   
## [201] 0 0 13.05 13.7 0 0 15.45 0 0 23.41   
## [211] 0 26 0 27.35 0 0 0 30.32 31.72 0   
## [221] 0 0 0 37.02 0 0 0 0 0 0   
## [231] 0 95.18 0 105.66 0 0 0 0 0 0   
## [241] 13.05 0 0 2.17 0 0 0 0.75 0 0   
## [251] 0 0 0 0 0 0 0 0 0 0   
## [261] 8.85 3.3 0 0 0 0.79 0.17 0 0 0   
## [271] 0 0 0 8.98 11.19 5.38 17.85 10.73 22.03 17.85   
## [281] 5.38 24.77 10.93 1.1 24.24 0 0 0 0 0   
## [291] 0 8 0 86.45 0 0 0 0 0 11.32   
## [301] 0 0 0 0 13.99 0 0 0 0 0   
## [311] 0 0 11.32 0 0 0 0 0 0 0   
## [321] 34.36 0 0 2.18 0 0 0 0 0 0   
## [331] 0 0 0 0 0 0 0 0 0 0   
## [341] 0 0 0 0 0 0 0 0 0 0   
## [351] 0 0 0 0 0 0 0 0 0 0   
## [361] 0 0 0 0 0 0 0 0 0 3.71   
## [371] 3.95 0 0 0 18.3 0 0 0 0 11.32   
## [381] 0 0 5.86 42.87 12.18 0 0 0 0 28.74   
## [391] 0 0 70.76 0 0 0 0 0 0 8.16   
## [401] 8.16 0 4.96 0 0 0 3.95 0 7.8 0   
## [411] 0 4.62 0 0 0 0 0 0 0 0   
## [421] 185.76 0 6.3 0 0 0 0 0 0 0   
## [431] 11.06 0 0 0 0 0 0.54 0 0 0   
## [441] 7.8 1.23 0 0.71 18.3 2.21 0 5.55 11.06 0   
## [451] 0 9.71 0 0 0 0 0 0 0 1.94   
## [461] 0 0 3.71 5.39 2.14 6.84 0 0 0 61.13   
## [471] 0 0 1.94 70.32 10.08 3.19 0 0 2.21 0   
## [481] 2.75 0 0 0 26.43 2.07 0 0 46.7 0   
## [491] 0 0 43.32 0 0 0 2.77 0 40.54 10.82   
## [501] 0 0 0 0 49.59 0 3.3 0 0 2.17   
## [511] 0 6.44 6.44 54.29 11.16 0 0   
## 251 Levels: 0 0.09 0.17 0.21 0.24 0.33 0.36 0.41 0.43 0.47 0.52 0.54 ... 1090.84

خوشه بندی

10 سطر از داده ها را به تصادف انتخاب کرده و خوشه بندی را روی آن ها به روش های مختلف انجام می دهیم.

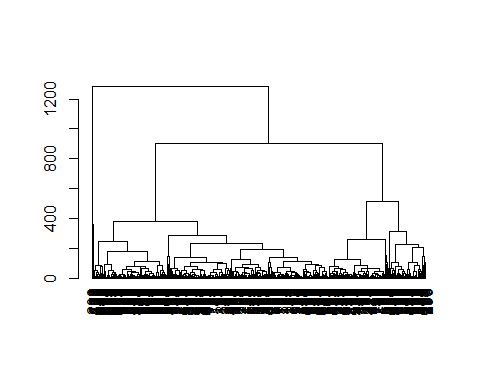
set.seed(2020)  
sample.ran = sample(1:nrow(fireeee),size= 10,replace = FALSE)  
  
s1 = fireeee[412,]  
s2 = fireeee[236,]  
s3 = fireeee[400,]  
s4 = fireeee[220,]  
s5 = fireeee[111,]  
s6 = fireeee[321,]  
s7 = fireeee[273,]  
s8 = fireeee[164,]  
s9 = fireeee[170,]  
s10 = fireeee[71,]

ماتریس فاصله ها به روش اقلیدسی:

d = rbind(s1,s2,s3,s4,s5,s6,s7,s8,s9,s10)  
  
D<-dist(fireeee,method="euclidean")

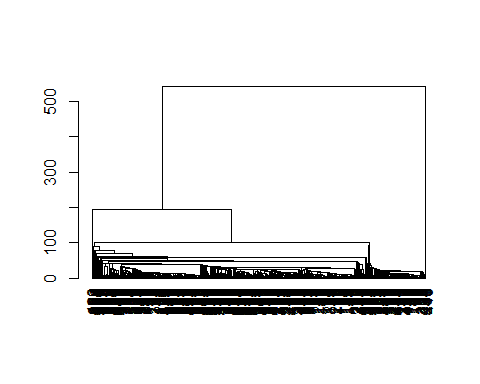
خوشه بندی(پیوند کامل)

h1<-hclust(D,method="complete")  
dend1<-as.dendrogram(h1)  
plot(dend1)



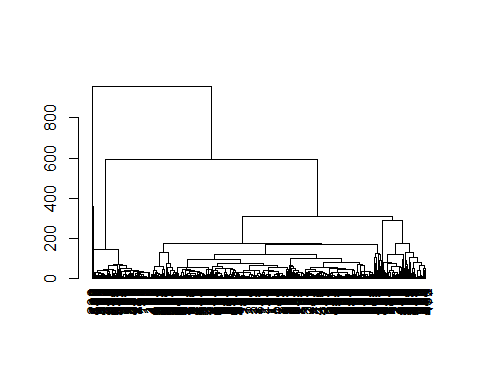
خوشه بندی(پیوند منفرد)

h2<-hclust(D,method="single")  
dend2<-as.dendrogram(h2)  
plot(dend2)



خوشه بندی(پیوند میانگین)

h3<-hclust(D,method="average")  
dend3<-as.dendrogram(h3)  
plot(dend3)

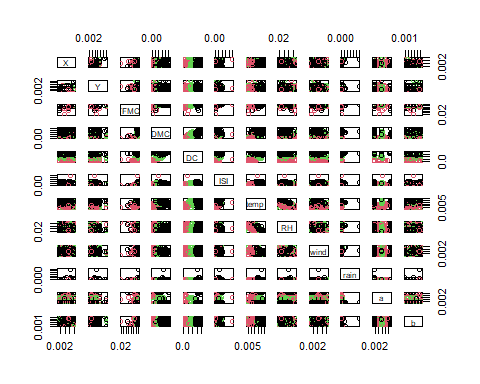


خوشه بندی غیر سلسله مراتبی:

kc<-kmeans(x,3)  
kc

## K-means clustering with 3 clusters of sizes 360, 93, 64  
##   
## Cluster means:  
## X Y FFMC DMC DC ISI  
## 1 0.004125261 0.003855337 0.08408658 0.12498727 0.63638419 0.008875933  
## 2 0.004514623 0.004080904 0.07927599 0.02692312 0.06880758 0.005866053  
## 3 0.004812805 0.004225528 0.08307675 0.07887987 0.37807733 0.008357951  
## temp RH wind rain a b  
## 1 0.01887154 0.04015000 0.003457071 2.750174e-05 0.006608057 0.003511565  
## 2 0.01140090 0.04294806 0.004174548 1.971451e-06 0.005835495 0.002976891  
## 3 0.01716281 0.03971996 0.004239852 2.864765e-06 0.004397414 0.003595280  
##   
## Clustering vector:  
## [1] 2 1 1 2 2 3 3 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 2 3 1 1 1 1 1 1 1 1 1 1 1 1 1  
## [38] 1 1 2 3 1 1 1 1 1 1 3 2 2 1 1 1 1 1 1 1 1 2 2 2 2 3 1 1 1 1 1 1 2 2 1 2 1  
## [75] 1 2 2 2 3 1 1 1 1 1 1 1 1 1 1 2 1 2 3 1 1 1 2 2 1 1 1 1 1 1 2 2 2 1 1 1 2  
## [112] 2 1 1 2 2 2 2 2 3 1 1 1 1 1 1 2 1 1 1 2 2 2 1 2 1 1 1 3 1 1 3 3 3 3 1 2 1  
## [149] 1 1 3 3 3 1 1 1 1 1 1 1 2 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1  
## [186] 1 2 1 2 2 2 1 1 1 1 1 2 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1 2 2 2 1 1 1 2 1 1  
## [223] 2 3 1 1 1 3 1 3 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1  
## [260] 1 3 1 1 1 1 3 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3 3 3 2 2 3 3 3 3 3 3 3 3 3 3 3  
## [297] 3 3 3 2 3 3 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
## [334] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
## [371] 1 3 1 1 1 1 1 1 2 2 3 1 1 1 1 1 1 2 1 1 2 1 1 2 2 2 1 1 1 3 3 1 1 1 1 1 1  
## [408] 2 1 3 2 2 3 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 2 3  
## [445] 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 1  
## [482] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2  
##   
## Within cluster sum of squares by cluster:  
## [1] 2.2611685 0.2523843 0.4243913  
## (between\_SS / total\_SS = 89.7 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
## [6] "betweenss" "size" "iter" "ifault"

plot(x,col=kc$cluster)



نیجه گیری:

بیشترین مناطق آسیب دیده با یک استثنا بیشترین اتش سوزی را دارند.

حتی اگر مناطقی که بیشترین تعداد آتش سوزی را دارند بیشترین خسارت وارد شده را دارند، خسارت بیشتر به دلیل یک آتش سوزی بزرگ است تا بسیاری از آتش سوزی های متوسط. در واقع بیشتر آتش سوزی ها بسیار ناچیز هستند.

با شک و تردید (اما جالب) این ممکن است دلیلی بر کنترل آتش به جای خاموش کردن هر آتش کوچک باشد. آتش‌سوزی‌های جزئی کنترل‌شده باعث می‌شود تا زیر خراش‌ها سوزانده شوند و از آتش‌سوزی‌های شدید جنگل‌ها جلوگیری کنند.

محدودیت های مجموعه داده: ما از میزان دخالت انسان در اندازه آتش اطلاعی نداریم. جایی که آتش سوزی توسط گشت های جنگلی خاموش می شود. سپس، این یک سطح اضافی از پیچیدگی را اضافه می کند زیرا ما اطلاعاتی در مورد زمان تشخیص و سپس زمان پاسخ نداریم. این یک عامل بسیار پیچیده در استنتاج می افزاید زیرا در مجموعه داده به حساب نمی آید.لذا شاید نتیجه چه در تحلیل رگرسیونی چه عاملی باب میل نباشد.

منبع: https://archive.ics.uci.edu/ml/datasets/forest+fires