Nonparametric_code H24105323

Liau Zhi Xuan

2024-11-11

目录

Correlation

```
library(Hmisc)
```

```
Attaching package: 'Hmisc'

The following objects are masked from 'package:base':

    format.pval, units
```

```
data(airquality)
str(airquality)
```

```
'data.frame':   153 obs. of  6 variables:
 $ Ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...
 $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
 $ Wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
 $ Temp   : int  67 72 74 62 56 66 65 59 61 69 ...
 $ Month  : int  5 5 5 5 5 5 5 5 5 5 ...
 $ Day    : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
describe(airquality)
```

```
airquality

 6  Variables      153  Observations
--------------------------------------------------------------------------------
Ozone
       n  missing distinct     Info     Mean      Gmd      .05      .10
     116       37       67    0.999    42.13    35.28     7.75    11.00
     .25      .50      .75      .90      .95
   18.00    31.50    63.25    87.00   108.50

lowest :   1   4   6   7   8, highest: 115 118 122 135 168
--------------------------------------------------------------------------------
Solar.R
       n  missing distinct     Info     Mean      Gmd      .05      .10
     146        7      117        1    185.9    102.7    24.25    47.50
     .25      .50      .75      .90      .95
  115.75   205.00   258.75   288.50   311.50

lowest :   7   8  13  14  19, highest: 320 322 323 332 334
--------------------------------------------------------------------------------
Wind
       n  missing distinct     Info     Mean      Gmd      .05      .10
     153        0       31    0.997    9.958    3.964     4.60     5.82
     .25      .50      .75      .90      .95
    7.40     9.70    11.50    14.90    15.50

lowest : 1.7  2.3  2.8  3.4  4   , highest: 16.1 16.6 18.4 20.1 20.7
--------------------------------------------------------------------------------
Temp
       n  missing distinct     Info     Mean      Gmd      .05      .10
     153        0       40    0.999    77.88    10.74     60.2     64.2
     .25      .50      .75      .90      .95
    72.0     79.0     85.0     90.0     92.0

lowest : 56 57 58 59 61, highest: 92 93 94 96 97
--------------------------------------------------------------------------------
Month
       n  missing distinct     Info     Mean      Gmd
     153        0        5     0.96    6.993    1.608

Value          5     6     7     8     9
Frequency     31    30    31    31    30
Proportion 0.203 0.196 0.203 0.203 0.196

For the frequency table, variable is rounded to the nearest 0
--------------------------------------------------------------------------------
Day
       n  missing distinct     Info     Mean      Gmd      .05      .10
     153        0       31    0.999     15.8    10.26      2.0      4.0
     .25      .50      .75      .90      .95
     8.0     16.0     23.0     28.0     29.4

lowest :  1  2  3  4  5, highest: 27 28 29 30 31
--------------------------------------------------------------------------------
```

```r
# Spearman
spearman_test <- cor.test(airquality$Ozone, airquality$Temp, method = "spearman")
```

```
Warning in cor.test.default(airquality$Ozone, airquality$Temp, method =
"spearman"): Cannot compute exact p-value with ties
```

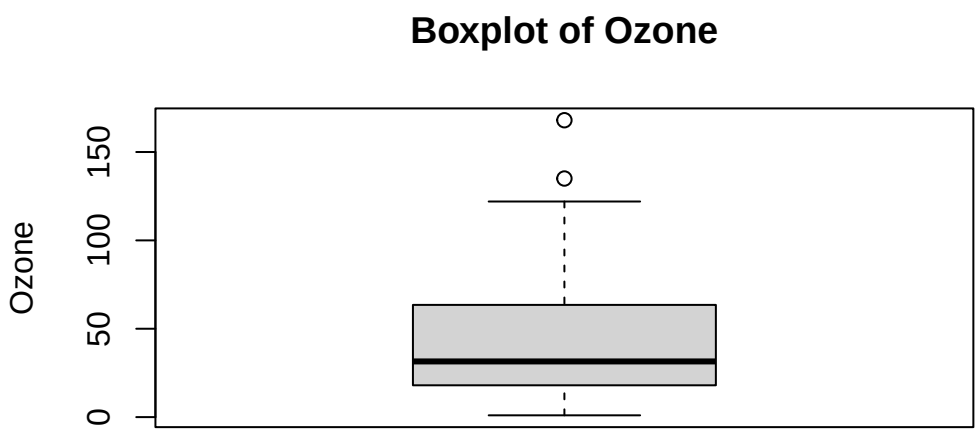
```r
spearman_test
```

```

    Spearman's rank correlation rho

data:  airquality$Ozone and airquality$Temp
S = 58778, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
     rho
0.774043
```

```r
# Kendall
kendall_test <- cor.test(airquality$Ozone, airquality$Temp, method = "kendall")
kendall_test
```

```

    Kendall's rank correlation tau

data:  airquality$Ozone and airquality$Temp
z = 9.1599, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.5862988
```

```r
#
boxplot(airquality$Ozone, main = "Boxplot of Ozone", ylab = "Ozone")
```

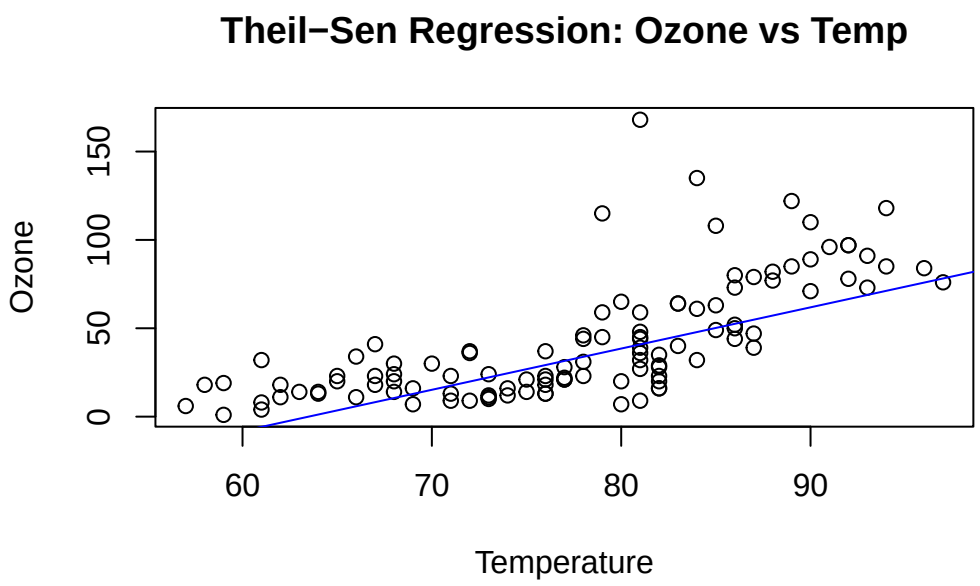


Theil-Sen

```r
library(mblm)
data("airquality")

#  NA Theil-Sen
airquality_clean <- na.omit(airquality)
theilsen_model <- mblm(Ozone ~ Temp, data = airquality_clean)
summary(theilsen_model)
```

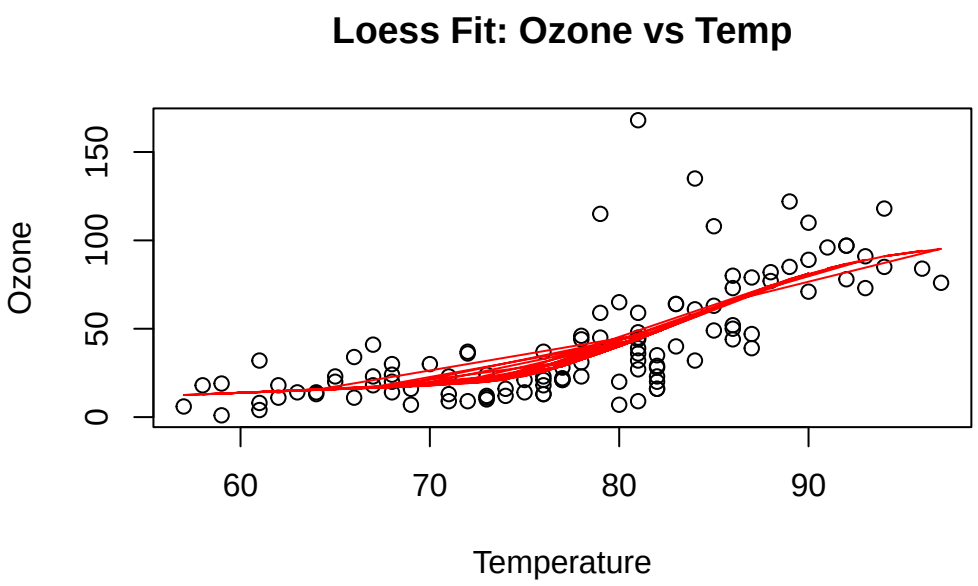
```
Call:
mblm(formula = Ozone ~ Temp, dataframe = airquality_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-31.846  -8.513   8.154  19.487 127.154

Coefficients:
            Estimate      MAD V value Pr(>|V|)
(Intercept) -148.154   86.590      19   <2e-16 ***
Temp           2.333    1.050    6214   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25.51 on 109 degrees of freedom
```

```r
plot(airquality_clean$Temp, airquality_clean$Ozone, main = "Theil-Sen Regression: Ozone vs Temp",
     xlab = "Temperature", ylab = "Ozone")
abline(theilsen_model, col = "blue")
```



LOESS

```r
data("airquality")

#  NA Loess
airquality_clean <- na.omit(airquality)

loess_model <- loess(Ozone ~ Temp, data = airquality_clean)

plot(airquality_clean$Temp, airquality_clean$Ozone, main = "Loess Fit: Ozone vs Temp",
     xlab = "Temperature", ylab = "Ozone")
lines(airquality_clean$Temp, predict(loess_model, newdata = airquality_clean), col = "red")
```



KDE

```
library(Hmisc)
data("faithful") #      Old Faithful
str(faithful)  #eruptions        waiting
```

```
'data.frame':   272 obs. of  2 variables:
 $ eruptions: num  3.6 1.8 3.33 2.28 4.53 ...
 $ waiting  : num  79 54 74 62 85 55 88 85 51 85 ...
```

```
describe(faithful)
```

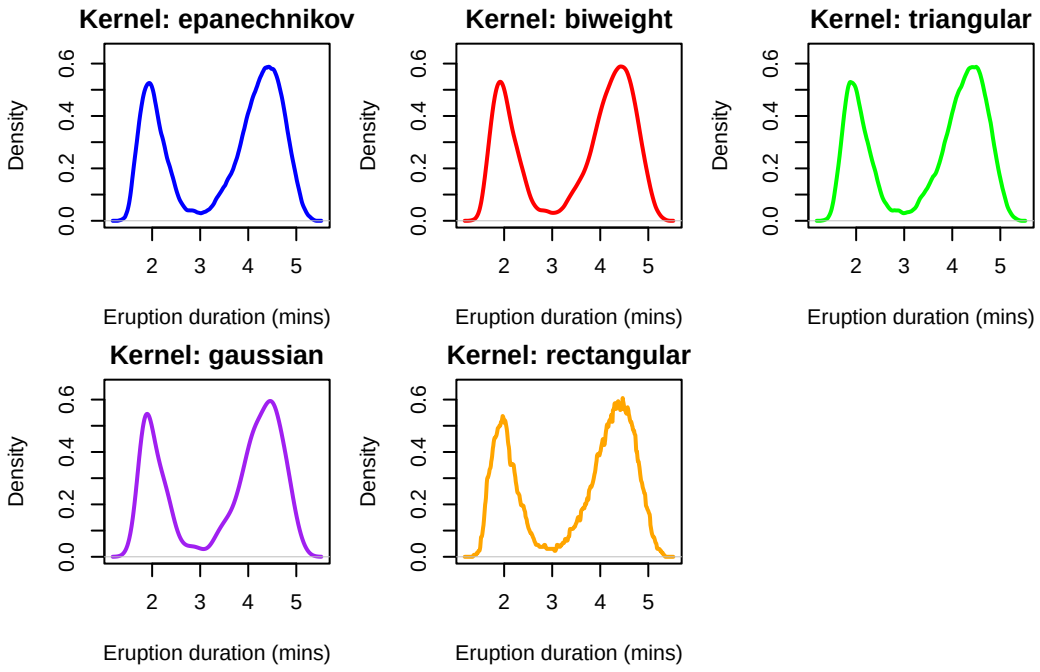
```
faithful

 2  Variables      272  Observations
---------------------------------------------------------------------------
eruptions
       n  missing distinct     Info     Mean      Gmd      .05      .10
     272        0      126        1    3.488    1.266    1.800    1.852
     .25      .50      .75      .90      .95
   2.163    4.000    4.454    4.700    4.817

lowest : 1.6   1.667 1.7   1.733 1.75 , highest: 4.933 5     5.033 5.067 5.1
---------------------------------------------------------------------------
waiting
       n  missing distinct     Info     Mean      Gmd      .05      .10
     272        0       51    0.999     70.9    15.37       48       51
```

```
     .25      .50      .75      .90      .95
      58       76       82       86       89

lowest : 43 45 46 47 48, highest: 91 92 93 94 96
------------------------------------------------------------------------------
```

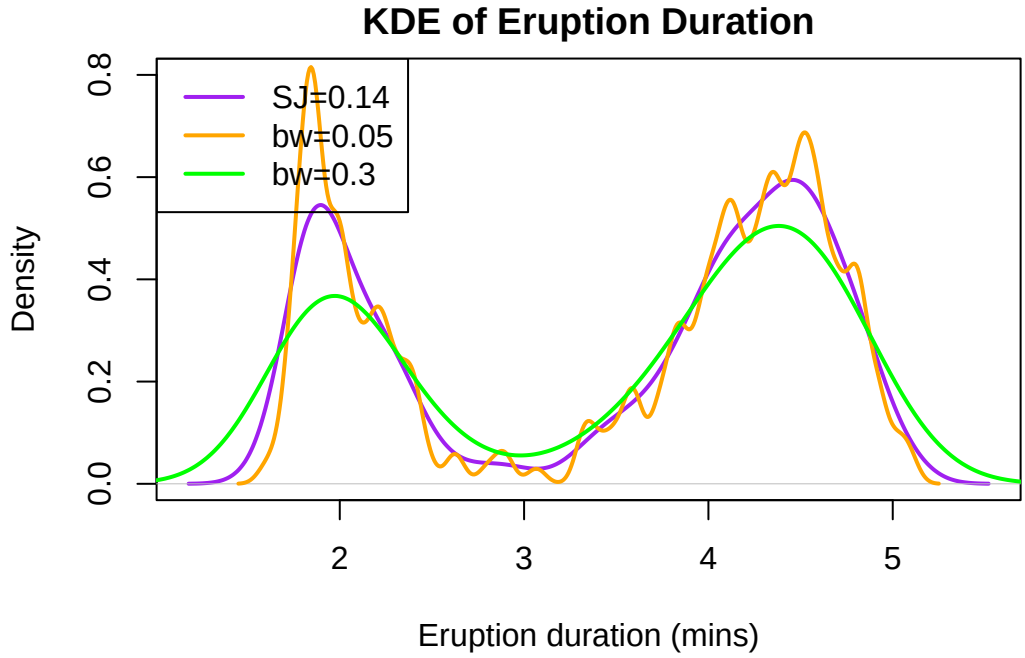
```r
#--------------------------

kernels <- c("epanechnikov","biweight","triangular", "gaussian","rectangular")
colors <- c("blue", "red", "green", "purple", "orange")
par(mfrow = c(2, 3), mar = c(4, 4, 2, 1))
# Loop through each kernel and plot KDE individually
for (i in 1:length(kernels)) {
  kde <- density(faithful$eruptions, kernel = kernels[i],bw="SJ")
  plot(kde, main = paste("Kernel:", kernels[i]), xlab = "Eruption duration (mins)",
       ylab = "Density", col = colors[i], lwd = 2, ylim = c(0, 0.65))
}

#--------------------------
# Gaussian & SJ
kde_choosen<- density(faithful$eruptions,kernel="gaussian",bw="SJ")
kde_choosen
```

```
Call:
    density.default(x = faithful$eruptions, bw = "SJ", kernel = "gaussian")

Data: faithful$eruptions (272 obs.);    Bandwidth 'bw' = 0.14

       x                y
 Min.   :1.180   Min.   :0.0001834
 1st Qu.:2.265   1st Qu.:0.0422638
 Median :3.350   Median :0.1709243
 Mean   :3.350   Mean   :0.2301726
 3rd Qu.:4.435   3rd Qu.:0.4134348
 Max.   :5.520   Max.   :0.5945634
```

```r
par(mfrow = c(1,1))
```




```
plot(kde_choosen, main = "KDE of Eruption Duration",xlab = "Eruption duration (mins)", ylab = "Densit
#      KDE
lines(density(faithful$eruptions, bw = 0.05), col = "orange", lwd = 2)  #
lines(density(faithful$eruptions, bw = 0.3), col = "green", lwd = 2)  #

legend("topleft", legend = c("SJ=0.14", "bw=0.05", "bw=0.3"),
       col = c("purple", "orange", "green"), lwd = 2)
```



```r
# kde_default$x   x
# kde_default$y     PDF
pdf_values <- data.frame(Duration = kde_choosen$x, Density = kde_choosen$y)
head(pdf_values)
```

```
  Duration      Density
1 1.179869 0.0001833851
2 1.188363 0.0002229529
3 1.196857 0.0002696114
4 1.205350 0.0003278130
5 1.213844 0.0003968097
6 1.222338 0.0004779552
```