chunkSummary

Yuwei Ni

5/29/2018

The lng package is used for observational study. Two new functions are added recently to divide the input data into four chunks and calculate the mean and standard deviation of each variable.

All changes are finished in the new branch: try. The purposes of two new functions are same, while one (chunkSummaryR) is a pure R funcion, and the other (chunkSummaryC) is a R function calls a C++ function.

Follows are the test part, it uses one sample dataset to check whether the two functions give the same resuls.

# Install required packages and select one data to test

devtools::install\_github("YuweiNi45/lng",ref = "try")  
library(devtools)  
library(Rcpp)  
library(MASS)  
library(lng)  
data<-birthwt  
summ<-head(data,10)  
knitr::kable(summ,caption = "The first ten rows of sample data")

The first ten rows of sample data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | low | age | lwt | race | smoke | ptl | ht | ui | ftv | bwt |
| 85 | 0 | 19 | 182 | 2 | 0 | 0 | 0 | 1 | 0 | 2523 |
| 86 | 0 | 33 | 155 | 3 | 0 | 0 | 0 | 0 | 3 | 2551 |
| 87 | 0 | 20 | 105 | 1 | 1 | 0 | 0 | 0 | 1 | 2557 |
| 88 | 0 | 21 | 108 | 1 | 1 | 0 | 0 | 1 | 2 | 2594 |
| 89 | 0 | 18 | 107 | 1 | 1 | 0 | 0 | 1 | 0 | 2600 |
| 91 | 0 | 21 | 124 | 3 | 0 | 0 | 0 | 0 | 0 | 2622 |
| 92 | 0 | 22 | 118 | 1 | 0 | 0 | 0 | 0 | 1 | 2637 |
| 93 | 0 | 17 | 103 | 3 | 0 | 0 | 0 | 0 | 1 | 2637 |
| 94 | 0 | 29 | 123 | 1 | 1 | 0 | 0 | 0 | 1 | 2663 |
| 95 | 0 | 26 | 113 | 1 | 1 | 0 | 0 | 0 | 0 | 2665 |

**The results of pure R function**

#chunkSummaryR  
result1<-chunkSummaryR(data)  
knitr::kable(result1,caption = "The results of pure R function")

The results of pure R function

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | Chunk1.mean1 | Chunk1.sd1 | Chunk2.mean2 | Chunk2.sd2 | Chunk3.mean3 | Chunk3.sd3 | Chunk4.mean4 | Chunk4.sd4 |
| low | 0.00 | 0.00 | 0.00 | 0.00 | 0.26 | 0.44 | 1.00 | 0.00 |
| age | 23.85 | 5.68 | 22.83 | 4.77 | 24.87 | 5.94 | 21.38 | 4.11 |
| lwt | 128.38 | 30.48 | 134.53 | 34.53 | 134.87 | 30.00 | 121.51 | 25.67 |
| race | 1.69 | 0.85 | 2.06 | 0.96 | 1.64 | 0.90 | 2.00 | 0.91 |
| smoke | 0.44 | 0.50 | 0.34 | 0.48 | 0.26 | 0.44 | 0.53 | 0.50 |
| ptl | 0.06 | 0.24 | 0.26 | 0.67 | 0.15 | 0.42 | 0.32 | 0.52 |
| ht | 0.02 | 0.14 | 0.04 | 0.20 | 0.13 | 0.34 | 0.06 | 0.25 |
| ui | 0.19 | 0.39 | 0.09 | 0.28 | 0.13 | 0.34 | 0.19 | 0.40 |
| ftv | 0.98 | 1.04 | 0.66 | 1.20 | 0.77 | 0.87 | 0.77 | 1.11 |
| bwt | 2848.46 | 162.50 | 3361.43 | 174.61 | 3315.72 | 1124.74 | 2254.79 | 191.50 |

**The results of R and C++ function**

#chunkSummaryC  
sourceCpp("/Users/summer/Downloads/package/lng/src/chunkSummaryC.cpp")  
result2<-chunkSummaryC(data)  
knitr::kable(result2,caption = "The results of R and C++ function")

The results of R and C++ function

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variables | Chunk1.mean1 | Chunk1.sd1 | Chunk2.mean2 | Chunk2.sd2 | Chunk3.mean3 | Chunk3.sd3 | Chunk4.mean4 | Chunk4.sd4 |
| low | 0.00 | 0.00 | 0.00 | 0.00 | 0.26 | 0.44 | 1.00 | 0.00 |
| age | 23.85 | 5.68 | 22.83 | 4.77 | 24.87 | 5.94 | 21.38 | 4.11 |
| lwt | 128.38 | 30.48 | 134.53 | 34.53 | 134.87 | 30.00 | 121.51 | 25.67 |
| race | 1.69 | 0.85 | 2.06 | 0.96 | 1.64 | 0.90 | 2.00 | 0.91 |
| smoke | 0.44 | 0.50 | 0.34 | 0.48 | 0.26 | 0.44 | 0.53 | 0.50 |
| ptl | 0.06 | 0.24 | 0.26 | 0.67 | 0.15 | 0.42 | 0.32 | 0.52 |
| ht | 0.02 | 0.14 | 0.04 | 0.20 | 0.13 | 0.34 | 0.06 | 0.25 |
| ui | 0.19 | 0.39 | 0.09 | 0.28 | 0.13 | 0.34 | 0.19 | 0.40 |
| ftv | 0.98 | 1.04 | 0.66 | 1.20 | 0.77 | 0.87 | 0.77 | 1.11 |
| bwt | 2848.46 | 162.50 | 3361.43 | 174.61 | 3315.72 | 1124.74 | 2254.79 | 191.50 |