

## 实验报告

成绩:

班级学号 202015116 姓名 梁祺若 实验日期 2022-9-14 仪器编号           

实验名称 实验 1 R 的数据组织和整理

### 1. 实验题目

R 的数据组织和整理

### 2. 实验步骤

本题目由两部分组成：（1）利用 R 软件做数据对象的创建和访问；（2）利用 R 软件自定义函数，对\*\*数据做\*\*\*分析。

（1）利用 R 软件做数据对象的创建和访问

- 读取数据

```
ConsumptData<-read.table("ConsumptData.txt",header=TRUE,sep="\t")
str(ConsumptData)
names(ConsumptData)
head(ConsumptData)
```

- 保存数据

```
write.table(ConsumptData,file="out.txt",sep="\t",row.names=FALSE,col.names=TRUE,quote=FALSE,ap
pend=FALSE,na="NA" )
```

- 数据框的创建与访问

```
ConsumptData<-read.table("ConsumptData.txt",header=TRUE, sep="\t")
str(ConsumptData)
ConsumptData$price
ConsumptData[["ID"]]
```

- 数组的创建与访问

```
a<-(1:120)
dim1<-c("row1","row2","row3")
dim2<-c("col1","col2","col3","col4","col5","col6","col7","col8","col9","col10")
dim3<-c("high1","high2","high3","high4")
a=array(a,c(3,10,4),dimnames = list(dim1,dim2,dim3))
is.array(a)
```

```

> a<-(1:120)
> dim1<-c("row1","row2","row3")
> dim2<-c("col1","col2","col3","col4","col5","col6","col7","col8","col9","col10")
> dim3<-c("high1","high2","high3","high4")
> a=array(a,c(3,10,4),dimnames = list(dim1,dim2,dim3))
> is.array(a)
[1] TRUE
> a
, , high1
      col1 col2 col3 col4 col5 col6 col7 col8 col9 col10
row1    1    4    7   10   13   16   19   22   25   28
row2    2    5    8   11   14   17   20   23   26   29
row3    3    6    9   12   15   18   21   24   27   30

, , high2
      col1 col2 col3 col4 col5 col6 col7 col8 col9 col10
row1   31   34   37   40   43   46   49   52   55   58
row2   32   35   38   41   44   47   50   53   56   59
row3   33   36   39   42   45   48   51   54   57   60

, , high3
      col1 col2 col3 col4 col5 col6 col7 col8 col9 col10
row1   61   64   67   70   73   76   79   82   85   88
row2   62   65   68   71   74   77   80   83   86   89
row3   63   66   69   72   75   78   81   84   87   90

, , high4
      col1 col2 col3 col4 col5 col6 col7 col8 col9 col10
row1   91   94   97  100  103  106  109  112  115  118
row2   92   95   98  101  104  107  110  113  116  119
row3   93   96   99  102  105  108  111  114  117  120
> |

```

- 矩阵的创建和访问

```
MatA<-(1:42)
```

```
myMat<-matrix(MatA,nrow = 6,ncol = 7,byrow = FALSE)
```

```
is.matrix(myMat)
```

```
str(myMat)
```

```
myMat[1, ]
```

```
print(myMat)
```

```

> MatA<-(1:42)
> myMat<-matrix(MatA,nrow = 6,ncol = 7,byrow = FALSE)
> is.matrix(myMat)
[1] TRUE
> str(myMat)
int [1:6, 1:7] 1 2 3 4 5 6 7 8 9 10 ...
> myMat[1, ]
[1] 1 7 13 19 25 31 37
> print(myMat)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    1    7   13   19   25   31   37
[2,]    2    8   14   20   26   32   38
[3,]    3    9   15   21   27   33   39
[4,]    4   10   16   22   28   34   40
[5,]    5   11   17   23   29   35   41
[6,]    6   12   18   24   30   36   42
> |

```

- 列表的创建和访问

```
myMatb<-matrix(nrow=5,ncol=2)
```

```

lista<-c(1:10)
listb<-myMatb
listb[,1]=seq(from=6,to=30,by=6)
listb[,2]=seq(from=20,to=0,by=-5)
listc<-array(1:20,c(2,5,2))
myList<-list(L1=lista,L2=listb,L3=listc)
names(myList)
is.list(myList)
str(myList)

> myMatb<-matrix(nrow=5,ncol=2)
> lista<-c(1:10)
> listb<-myMatb
> listb[,1]=seq(from=6,to=30,by=6)
> listb[,2]=seq(from=20,to=0,by=-5)
> listc<-array(1:20,c(2,5,2))
> myList<-list(L1=lista,L2=listb,L3=listc)
> names(myList)
[1] "L1" "L2" "L3"
>
> is.list(myList)
[1] TRUE
>
> str(myList)
list of 3
 $ L1: int [1:10] 1 2 3 4 5 6 7 8 9 10
 $ L2: num [1:5, 1:2] 6 12 18 24 30 20 15 10 5 0
 $ L3: int [1:2, 1:5, 1:2] 1 2 3 4 5 6 7 8 9 10 ...
>

```

(2) 利用 R 软件自定义\*\*函数，对\*\*数据做\*\*\*分析。

- 数据来源：CBA 联赛积分榜 中国篮球数据库 篮球-CBA 新浪体育 新浪网 (sina.com.cn)

分享到微博

sina 新浪竞技风暴

体育首页 | 新浪首页 | 新浪导航

## 中国篮球数据库 ★ 联赛、国家队一网打尽

THE DATABASE OF CHINESE BASKETBALL

首页 | 赛程 | 球队 | 比分 | 统计 | 积分 | 图片 | 微博

广东 八一 江苏 深圳 浙江 广厦 上海 福建 辽宁 新疆 山东 北京 山西 广州 吉林 天津 青岛 四川 同曦 北控

### CBA联赛常规赛排名/所有球队

联盟排名 | 胜率排名 | 主场排名 | 客场排名 | 连胜/连负

赛季 21-22 查询

排名	球队	胜	负	胜率	主场战绩	客场战绩	场均得分	场均失分	连胜/连负
1	辽宁	32	6	84.20%	16胜3负	16胜3负	106.39	93.74	7连胜
2	广厦	31	7	81.60%	16胜3负	15胜4负	107.03	91.03	3连胜
3	上海	28	10	73.70%	14胜5负	14胜5负	111.08	99.82	3连负
4	浙江	28	10	73.70%	16胜3负	12胜7负	104.58	92.16	2连胜
5	广东	26	12	68.40%	14胜5负	12胜7负	103.37	96.66	1连负
6	深圳	26	12	68.40%	13胜6负	13胜6负	106.71	100.21	1连胜
7	北京	24	14	63.20%	13胜6负	11胜8负	98.32	87.05	1连胜
8	广州	23	15	60.50%	11胜8负	12胜7负	100.66	99.50	1连胜
9	山西	23	15	60.50%	12胜7负	11胜8负	111.26	107.32	3连胜
10	吉林	22	16	57.90%	13胜6负	9胜10负	98.58	97.42	1连胜
11	山东	20	18	52.60%	9胜10负	11胜8负	103.53	103.53	3连胜
12	天津	19	19	50.00%	11胜8负	8胜11负	104.16	106.13	1连负
13	北控	18	20	47.40%	9胜10负	9胜10负	96.18	95.63	2连胜
14	新疆	18	20	47.40%	8胜11负	10胜9负	103.76	104.63	1连胜

- 从中国篮球数据库抓取数据。

```
library("XML")
library("RODBC")

sr1<-'http://cba.sports.sina.com.cn/cba/stats/teamrank/'
data1<-readHTMLTable(sr1)
print(data1)
```

```
> library("XML")
> library("RODBC")
>
> sr1<-'http://cba.sports.sina.com.cn/cba/stats/teamrank/'
> data1<-readHTMLTable(sr1)
> print(data1)
$ NULL
  排名 球队 胜 负 胜率 主场战绩 客场战绩 每场得分 每场丢分 连胜/连负
1    1  辽宁 32  6 84.20% 16胜3负 16胜3负 106.39  93.74  7连胜
2    2  广厦 31  7 81.60% 16胜3负 15胜4负 107.03  91.03  3连胜
3    3  上海 28 10 73.70% 14胜5负 14胜5负 111.08  99.82  3连负
4    4  浙江 28 10 73.70% 16胜3负 12胜7负 104.58  92.16  2连胜
5    5  广东 26 12 68.40% 14胜5负 12胜7负 103.37  96.66  1连负
6    6  深圳 26 12 68.40% 13胜6负 13胜6负 106.71 100.21  1连胜
7    7  北京 24 14 63.20% 13胜6负 11胜8负  98.32  87.05  1连胜
8    8  广州 23 15 60.50% 11胜8负 12胜7负 100.66  99.50  1连负
9    9  山西 23 15 60.50% 12胜7负 11胜8负 111.26 107.32  3连胜
10   10 吉林 22 16 57.90% 13胜6负 9胜10负  98.58  97.42  1连胜
11   11 山东 20 18 52.60% 9胜10负 11胜8负 103.53 103.53  3连胜
12   12 天津 19 19 50.00% 11胜8负 8胜11负 104.16 106.13  1连负
13   13 北控 18 20 47.40% 9胜10负 9胜10负  96.18  95.63  2连胜
14   14 新疆 18 20 47.40% 8胜11负 10胜9负 103.76 104.63  1连胜
15   15 青岛 13 25 34.20% 9胜10负 4胜15负  96.58  98.39  1连负
16   16 四川 12 26 31.60% 6胜13负 6胜13负  91.39 103.47  2连负
17   17 福建  7 31 18.40% 5胜14负 2胜17负  99.47 111.21  7连负
18   18 江苏  5 33 13.20% 4胜15负 1胜18负  84.92  99.18  3连负
19   19 宁波  3 35  7.90% 1胜18负 2胜17负  87.50 109.00  9连负
20   20 同曦  2 36  5.30% 1胜18负 1胜18负  91.89 111.29  7连负
```

- 由于原先抓取的数据有表头，且原先的表头是字符型数据，故我们需要先将原先的表头“胜”删除，并将数据转化为整数型。最后进行排序。

```
is.list(data1)
is.array(data1)
grades<-data1[[1]]$胜
grades<-grades[-1]
grades<-as.integer(grades)
```

```
> is.list(data1)
[1] TRUE
>
> is.array(data1)
[1] FALSE
> grades<-data1[[1]]$胜
> grades<-grades[-1]
> grades<-as.integer(grades)
> grades
[1] 31 28 28 26 26 24 23 23 22 20 19 18 18 13 12  7  5  3  2
> |
```

- 自定义冒泡排序函数

```
bubblesort=function(grades){
```

```
n=length(grades)
for(i in 1:(n-1)){
  for(j in (i+1):n){
    if(grades[i]<=grades[j]){
      t<-grades[i]
      grades[i]<-grades[j]
      grades[j]<-t
    }
  }
}
return(grades)
}
bubblesort(grades)

> bubblesort=function(grades){
+   n=length(grades)
+   for(i in 1:(n-1)){
+     for(j in (i+1):n){
+       if(grades[i]<=grades[j]){
+         t<-grades[i]
+         grades[i]<-grades[j]
+         grades[j]<-t
+       }
+     }
+   }
+   return(grades)
+ }
> bubblesort(grades)
[1] 31 28 28 26 26 24 23 23 22 20 19 18 18 13 12 7 5 3 2
> |
```

### 3. 实验结果

- 在实验的第一部分，我们对数据进行了读取和保存。然后我们分别创建并访问了我们自行设计的数据框、数组、矩阵和列表。
- 在实验的第二部分，我们从网页中抓取中国篮球数据库中“2021-2022 赛季 CBA 联赛常规赛排名/所有球队”中的数据，通过自定义冒泡排序函数对数据进行降序排序。