Homework1

JinDian

2024-09-10

1

The Iowa data set iowa.csv is a toy example that summarises the yield of wheat (bushels per acre) for the state of Iowa between 1930-1962. In addition to yield, year, rainfall and temperature were recorded as the main predictors of yield.

a. First, we need to load the data set into R using the command read.csv(). Use the help function to learn what arguments this function takes. Once you have the necessary input, load the data set into R and make it a data frame called iowa.df.

b. How many rows and columns does iowa.df have?

c. What are the names of the columns of iowa.df?

d. What is the value of row 5, column 7 of iowa.df?

e. Display the second row of iowa.df in its entirety

```
iowa.df <- read.csv("data/iowa.csv",header=T,sep = ";")
iowa.df
```

```
##    Year Rain0 Temp1 Rain1 Temp2 Rain2 Temp3 Rain3 Temp4 Yield
## 1  1930 17.75  60.2  5.83  69.0  1.49  77.9  2.42  74.4  34.0
## 2  1931 14.76  57.5  3.83  75.0  2.72  77.2  3.30  72.6  32.9
## 3  1932 27.99  62.3  5.17  72.0  3.12  75.8  7.10  72.2  43.0
## 4  1933 16.76  60.5  1.64  77.8  3.45  76.4  3.01  70.5  40.0
## 5  1934 11.36  69.5  3.49  77.2  3.85  79.7  2.84  73.4  23.0
## 6  1935 22.71  55.0  7.00  65.9  3.35  79.4  2.42  73.6  38.4
## 7  1936 17.91  66.2  2.85  70.1  0.51  83.4  3.48  79.2  20.0
## 8  1937 23.31  61.8  3.80  69.0  2.63  75.9  3.99  77.8  44.6
## 9  1938 18.53  59.5  4.67  69.2  4.24  76.5  3.82  75.7  46.3
## 10 1939 18.56  66.4  5.32  71.4  3.15  76.2  4.72  70.7  52.2
## 11 1940 12.45  58.4  3.56  71.3  4.57  76.7  6.44  70.7  52.3
## 12 1941 16.05  66.0  6.20  70.0  2.24  75.1  1.94  75.1  51.0
## 13 1942 27.10  59.3  5.93  69.7  4.89  74.3  3.17  72.2  59.9
## 14 1943 19.05  57.5  6.16  71.6  4.56  75.4  5.07  74.0  54.7
## 15 1944 20.79  64.6  5.88  71.7  3.73  72.6  5.88  71.8  52.0
## 16 1945 21.88  55.1  4.70  64.1  2.96  72.1  3.43  72.5  43.5
## 17 1946 20.02  56.5  6.41  69.8  2.45  73.8  3.56  68.9  56.7
## 18 1947 23.17  55.6 10.39  66.3  1.72  72.8  1.49  80.6  30.5
## 19 1948 19.15  59.2  3.42  68.6  4.14  75.0  2.54  73.9  60.5
## 20 1949 18.28  63.5  5.51  72.4  3.47  76.2  2.34  73.0  46.1
## 21 1950 18.45  59.8  5.70  68.4  4.65  69.7  2.39  67.7  48.2
## 22 1951 22.00  62.2  6.11  65.2  4.45  72.1  6.21  70.5  43.1
## 23 1952 19.05  59.6  5.40  74.2  3.84  74.7  4.78  70.0  62.2
## 24 1953 15.67  60.0  5.31  73.2  3.28  74.6  2.33  73.2  52.9
## 25 1954 15.92  55.6  6.36  72.9  1.79  77.4  7.10  72.1  53.9
## 26 1955 16.75  63.6  3.07  67.2  3.29  79.8  1.79  77.2  48.4
## 27 1956 12.34  62.4  2.56  74.7  4.51  72.7  4.42  73.0  52.8
## 28 1957 15.82  59.0  4.84  68.9  3.54  77.9  3.76  72.9  62.1
## 29 1958 15.24  62.5  3.80  66.4  7.55  70.5  2.55  73.0  66.0
## 30 1959 21.72  62.8  4.11  71.5  2.29  72.3  4.92  76.3  64.2
## 31 1960 25.08  59.7  4.43  67.4  2.76  72.6  5.36  73.2  63.2
## 32 1961 17.79  57.4  3.36  69.4  5.51  72.6  3.04  72.4  75.4
## 33 1962 26.61  66.6  3.12  69.1  6.27  71.6  4.31  72.5  76.0
```

b.33 rows,10 columns.

c.Year,Rain0,Temp1,Rain1,Temp2,Rain2,Temp3,Rain3,Temp4,Yield.

d.the value of row 5,column 7 of iowa.df is 79.7.

e.

```
iowa.df[2,]
```

```
##   Year Rain0 Temp1 Rain1 Temp2 Rain2 Temp3 Rain3 Temp4 Yield
## 2 1931 14.76  57.5  3.83    75  2.72  77.2   3.3  72.6  32.9
```

2

a. For each of the following commands, either explain why they should be errors, or explain the non-erroneous result.

vector1 <- c(“5”, “12”, “7”, “32”) max(vector1) sort(vector1) sum(vector1)

1.vector1是一个由字符”5” “12” “7” “32”组成的向量

```
vector1 <- c("5", "12", "7", "32")
vector1
```

```
## [1] "5"  "12" "7"  "32"
```

2.max()函数输出最大值，由于vector1向量中都是字符，字符从左开始比较字符大小，分别比较5,1,7,3.显然最大的是”7”

```
max(vector1)
```

```
## [1] "7"
```

3.sort()函数默认对vector1向量进行递增排序，根据字符大小比较的规则，从小到大依次是”12” “32” “5” “7” 。

```
sort(vector1)
```

```
## [1] "12" "32" "5"  "7"
```

4.出现error,因为vector1向量全是字符类型，无法进行sum().

b. For the next series of commands, either explain their results, or why they should produce errors.

vector2 <- c(“5”,7,12) vector22 + vector2[3] dataframe3 <- data.frame(z1=“5”,z2=7,z3=12) dataframe3[1,2] + dataframe3[1,3] list4 <- list(z1=“6”, z2=42, z3=“49”, z4=126) list4[2]+list4[[4]] list42+list4[4]

1.向量必须包含同一种类型的元素，这里由于同时包含字符和数字类型数据，所以会把数字类型元素变成字符类型元素

```
vector2 <- c("5",7,12)
vector2
```

```
## [1] "5"  "7"  "12"
```

会出现error:二进列运算符中有非数值参数.原因是在给vector2赋值时，其第二个和第三个元素都变成了字符类型数据，所以不能进行加运算

2.

```
dataframe3 <- data.frame(z1="5",z2=7,z3=12)
dataframe3
```

```
##   z1 z2 z3
## 1  5  7 12
```

```
dataframe3[1,2] + dataframe3[1,3]
```

```
## [1] 19
```

dataframe3是dataframe类型的数据，它的nrow=1,dataframe3[1,2] =7,dataframe3[1,3]=12，两个数加起来是19.

3.

list4是lists类型数据，list4[2]和list4[[4]]分别表示list4的第2个和第4个元素，二者都是数字类型，可以进行相加计算.list42和list4[4]分别表示只包含第2个和第4个元素的新list，两个list不能直接进行相加运算。