



# R软件在CGED-Q JSL中的运用

## (三) 制表

作者：陈俊（硕士研究生，华中师范大学）；康文林（教授，香港科技大学；华中师范大学）





01

简单制表

PART one

02

整理变量

PART two

03

制作可以导出的表格

PART three

04

指定条件制表

PART four

01

# 简单制表



## table()函数

table()函数的格式和逻辑:

table(x,	————→	变量名, 可以两个输入两个变量, 用逗号 (英文) 并列
exclude = ,	————→	排除某一个类型的值, 比如空白、缺失值、“涂黑”等特定字符; 如果等于NULL, 就是不排除任何值; 如果是NA, 就是排除NA的值; 如果等于特定字符, 需要加上引号再输入特定字符。
useNA = ,	————→	选择是否统计NA值得频数, 如果等于“no”则为不统计; 如果等于“ifany”或者
)		“always”则为统计

## 代码

一个变量:

```
table(JSL1900_1912_delete_kongbaiming$阳历年份)
```

两个变量:

```
table(JSL1900_1912_delete_kongbaiming$籍贯省,  
JSL1900_1912_delete_kongbaiming$qiren)
```

两个变量加上参数:

```
table(JSL1900_1912_delete_kongbaiming$籍贯省,  
JSL1900_1912_delete_kongbaiming$qiren,  
exclude= NULL,  
useNA= "always")
```

02

# 整理变量



## factor()函数

格式与逻辑:

`factor(x, levels = c(...), labels = c(...))`

- `x` → 需要整理的变量
- `levels = c(...)` → 该变量中需要整理的字符，可以同时整理多个字符。用“”引起，用逗号分隔
- `labels = c(...)` → 根据levels整理出的变量，需要与levels中的变量一一对应，labels当中的内容即为整理出来后的标签，用“”引起来，逗号分隔

## 实例代码：整理籍贯省

```
JSL1900_1912_delete_kongbaiming$jiguansheng_sort <-  
  factor(JSL1900_1912_delete_kongbaiming$籍贯省,  
        levels = c("","? ? ","? 西","奧國","比國","丹國","德國","俄國",  
                  "法國","韓國","和國","美國","腦威國","葡國","日本國","瑞典國",  
                  "義國","英國","安徽","奉天","福","福建","甘肅","广西",  
                  "廣","廣? ","廣東","廣東駐防","廣西","貴州","漢軍","河南",  
                  "黑龍江","湖北","湖南","吉林","江","江南","江蘇","江西",  
                  "滿洲","蒙古","南","山东","山東","山西","陝西","顺天",  
                  "順天","順天府","四川","天津","西","新疆","雲","雲南",  
                  "浙江","直隸"),  
        labels = c("京师","不详","不详","外国","外国","外国","外国","外国",  
                  "外国","外国","外国","外国","外国","外国","外国","外国",  
                  "外国","外国","安徽","奉天","福建","福建","甘肃","广西",  
                  "不详","不详","广东","广东","广西","贵州","不详","河南",  
                  "黑龙江","湖北","湖南","吉林","不详","江南","江苏","江西",  
                  "不详","不详","不详","山东","山东","山西","陕西","顺天",  
                  "顺天","顺天","四川","天津","不详","新疆","云南","云南",  
                  "浙江","直隶")  
  )
```



进行变量整理之后，即可进行简单制表：

```
table(JSL1900_1912_delete_kongbaiming$jiguansheng_sort,  
      JSL1900_1912_delete_kongbaiming$qiren,  
      exclude= NULL,  
      useNA= "always")
```

03

# 制作可以导出的表格



## table1()函数

### 格式与逻辑:

`table1(x,` → 需要制表的变量, 如果是多个变量, 用+号连接。|号用来分层, 即|号之后通常是列变量。\*号用于嵌套, 嵌套两个逻辑值。比较复杂, 但是结合代码理解, 要稍微好一点。

`data,` → 所引用的数据集

`overall = ,` → 总计。“”内可以输入任意字符, 比如“overall”、“total”等。如果令其等于F, 则不显示总计的数值。

`rowlabelhead = ,` → 抬头行标题, 对应第一列变量, 可以等于任意字符, 记得用“”

)

## 实例代码

利用整理好的籍贯省变量和旗人变量制表（籍贯省在纵轴，旗人在横轴）：

```
table1(~jiguansheng_sort|qiren, data = JSL1900_1912_delete_kongbaiming, overall = "total")
```

籍贯省和贡生在纵轴，旗人在横轴：

```
table1(~jiguansheng_sort+gongsheng|qiren, data = JSL1900_1912_delete_kongbaiming, overall = "total")
```

籍贯省在纵轴，旗人在横轴，贡生嵌套与旗人之上：

```
table1(~jiguansheng_sort|qiren*gongsheng, data = JSL1900_1912_delete_kongbaiming, overall = "total", rowlabelhead = "")
```

观察三条代码的符号，有相同之处，也有不同之处。~号不可省略，|号用来区分行变量和列变量，通常在|号之前，是表格中的行，|号之后是表格中的列。+号用来连接多个行变量；\*号用来嵌套列变量。请注意符号的运用。

03

# 指定条件制表



## subset()函数

用于从某一个数据集中选择出符合某条件的数据或是相关的列格式与逻辑:

subset(x,            →     数据集

subset,            →     逻辑条件, 即从数据集中选取数据的条件, 必须是逻辑表达式,  
满足符号要求, 可以用&、|号进行多条件筛选

select =c(...), →     指定从相关的列中选择数据  
)

## 实例代码

指定一个条件:

```
JSL1900_1912_only1910 <- subset(JSL1900_1912_delete_kongbaiming,阳历年份numeric == 1910)
```

指定两个或多个条件:

```
JSL1900_1912_京师1910 <- subset(JSL1900_1912_delete_kongbaiming,jiguansheng_sort == "京师"  
                                & 阳历年份numeric == 1910)
```

```
JSL1900_1912_京师1910候补 <- subset(JSL1900_1912_delete_kongbaiming,jiguansheng_sort == "京师"  
                                     & 阳历年份numeric == 1910 & houbu == 1)
```

如果有多个变量同时满足一个条件时:

```
JSL1900_1912_西南after1908 <- subset(JSL1900_1912_delete_kongbaiming, (jiguansheng_sort == "四川" |  
                                                                           jiguansheng_sort == "贵州" |  
                                                                           jiguansheng_sort == "云南" )  
                                     & (阳历年份numeric > 1908)  
                                     )
```

谢谢

