

## 第2章 数据操作

1. 将表2-1中的领导行为数据存储为数据框 leadership, 各个变量名称为 manager、data、country、gender、age、q1、q2、q3、q4、q5;

表2-1 各变量数据表

经理人	日期	国籍	性别	年龄/岁	q1	q2	q3	q4	q5
1	10/24/08	US	M	32	5	4	5	5	5
2	10/28/08	US	F	45	3	5	2	5	5
3	10/01/08	UK	F	25	3	5	5	5	2
4	10/12/08	UK	M	39	3	3	4		
5	05/01/09	UK	F	99	2	2	1	2	1

【参考答案】根据题意可知,我们要在这里创建一个名叫 leadership 的数据框。参考代码如下:

```
>manager<-c(1, 2, 3, 4, 5)
>date<-c("10/24/08", "10/28/08", "10/01/08", "10/12/08", "05/01/09")
>country<-c("US", "US", "UK", "UK", "UK")
>gender<-c("M", "F", "F", "M", "F")
>age<-c(32, 45, 25, 39, 99)
>q1<-c(5, 3, 3, 3, 2)
>q2<-c(4, 5, 5, 3, 2)
>q3<-c(5, 2, 5, 4, 1)
>q4<-c(5, 5, 5, NA, 2)
>q5<-c(5, 5, 2, NA, 1)
>leadership<-data.frame(manager, date, country, gender, age, q1, q2, q3, q4, q5)
>leadership
```

参考结果:

```
  manager    date country gender age q1 q2 q3 q4 q5
1      1 10/24/08     US      M  32  5  4  5  5  5
2      2 10/28/08     US      F  45  3  5  2  5  5
3      3 10/01/08     UK      F  25  3  5  5  5  2
4      4 10/12/08     UK      M  39  3  3  4 NA NA
5      5 05/01/09     UK      F  99  2  2  1  2  1
> |
```

2. 假设现有一个名为 mydata 的数据框, 其中的变量是  $x_1$  和  $x_2$ , 现在需要创建两个新变量 sum、mean 分别存储二者的和、均值。

【参考答案】

参考代码如下:

```
>mydata<-data.frame(x1=c(2, 2, 6, 4), x2=c(3, 4, 2, 8))
```

```

>attach(mydata)
>mydata $ sum<-x1+x2
>mydata $ mean<-(x1+x2)/2
>detach(mydata)
>mydata

```

参考结果:

	x1	x2	sum	mean
1	2	3	5	2.5
2	2	4	6	3.0
3	6	2	8	4.0
4	4	8	12	6.0

3. 将变量名 manger 修改为 managerID, 并将 date 修改为 testDate。

【参考答案】

参考代码如下:

```

>names(leadership)[1]<-"mangerID"
>names(leadership)[2]<-"testDate"
>leadership

```

>

参考结果:

	mangerID	testDate	country	gender	age	q1	q2	q3	q4	q5	agecat
1	1	10/24/08	US	M	32	5	4	5	5	5	Young
2	2	10/28/08	US	F	45	3	5	2	5	5	Young
3	3	10/01/08	UK	F	25	3	5	5	5	2	Young
4	4	10/12/08	UK	M	39	3	3	4	NA	NA	Young
5	5	05/01/09	UK	F	NA	2	2	1	2	1	<NA>

4. 综合编程.: 根据习题 1, 移除 leadership 数据集中含有缺失数据的行; 将 leadership 数据集中变量名 date 的字符串形式转换为日期格式; 将两个数据框按照 ID 和 Country 进行合并。

【参考答案】

(1) 参考代码:

```

>newdata<-na.omit(leadership)
>newdata

```

参考结果:

	manager	date	country	gender	age	q1	q2	q3	q4	q5	agecat
1	1	10/24/08	US	M	32	5	4	5	5	5	Young
2	2	10/28/08	US	F	45	3	5	2	5	5	Young
3	3	10/01/08	UK	F	25	3	5	5	5	2	Young

(2) 参考代码:

```

>dates<-as.Date(leadership[, 2], "% m/% d/% y")
>dates

```

参考结果:

```

[1] "2008-10-24" "2008-10-28" "2008-10-01" "2008-10-12"
[5] "2009-05-01"

```

(3) 参考代码:

```

> ID = c(1, 2, 3, 4, 5)

```

```
>country = c("china", "UK", "France", "Germany", "Australia")
>person = c(200, 400, 300, 800, 450)
>dataframeA = data.frame(ID, country)
>dataframeB = data.frame(ID, country, person)
>total = merge(dataframeA, dataframeB, by = c("ID", "country"))
```