

统软 06-1400012141

邵智轩 1400012141 物理学院

第 25 章练习

1

```
library(readr)
d.patients<-read_csv('patients.csv',locale = locale(encoding='GBK'),
                      col_types = cols(
                        `出生日期`=col_date(format = "%Y/%m/%d"),
                        `发病日期`=col_date(format = "%Y/%m/%d")
                      ))
d.patients[[' 发病年龄（带小数年）']] <- as.numeric(
  difftime(d.patients[[" 发病日期"]], d.patients[[" 出生日期"]], units='days')/365.25)
d.patients[[' 发病年龄（周岁）']]<-floor(d.patients$`发病年龄（带小数年）`)
knitr::kable(head(d.patients[,c(' 出生日期',
                                ' 发病日期', ' 发病年龄（带小数年）', ' 发病年龄（周岁）')]))
```

出生日期	发病日期	发病年龄（带小数年）	发病年龄（周岁）
1941-03-08	2007-01-01	65.81793	65
1972-01-24	2007-01-01	34.93771	34
1932-06-01	2007-01-01	74.58453	74
1947-05-17	2007-01-01	59.62765	59
1943-03-10	2007-01-01	63.81383	63
1940-01-08	2007-01-01	66.98152	66

2

```
d.cancer<-read.csv('cancer.csv', encoding="GBK")
d.cancer[, " 肿瘤缩减率"]<-with(d.cancer, (v0-v1)/v0)
knitr::kable(head(d.cancer))
```

id	age	sex	type	v0	v1	肿瘤缩减率
1	70	F	腺癌	26.51	2.91	0.8902301
2	70	F	腺癌	135.48	35.08	0.7410688
3	69	F	腺癌	209.74	74.44	0.6450844
4	68	M	腺癌	61.00	34.97	0.4267213
5	67	M	鳞癌	237.75	128.34	0.4601893
6	75	F	腺癌	330.24	112.34	0.6598232

第 27 章练习

```
library(tidyverse)
library(lubridate)
d.patients<-read_csv("patients.csv", locale = locale(encoding = "GBK"),
                     col_types = cols(
                       `出生日期`=col_date(format = "%Y/%m/%d"),
                       `发病日期`=col_date(format = "%Y/%m/%d"),
                       `诊断时间`=col_datetime(format="%Y/%m/%d%.%R")
                     ))
d.patients[[" 发病年龄"]]<-as.numeric(
  difftime(d.patients[[" 发病日期"]],
    d.patients[[" 出生日期"]], units='days')/365.25)# 发病年龄
head(year(d.patients$发病日期))# 发病年
```

```
## [1] 2007 2007 2007 2007 2007 2007
```

```
head(month(d.patients$发病日期))# 发病月
```

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

```
head(as.character(d.patients$发病日期,format="%Y%m"))# 发病年月
```

```
## [1] "200701" "200701" "200701" "200701" "200701" "200701"
```

```
# 地址编码
```

```
d.patients[["地址编码"]]=substr(as.character(d.patients$现住地址国标),1,6)
head(d.patients$地址编码,5)
```

```
## [1] "440103" "440105" "440113" "440113" "440113"
```

```
# 分区分年月统计
```

```
d.patients[["发病年月"]]=as.character(d.patients$发病日期,format="%Y%m")
d.pas1<-d.patients %>%
  count(`地址编码`,`发病年月`)
names(d.pas1)[3]="发病人数"
head(d.pas1)
```

```
## Source: local data frame [6 x 3]
```

```
## Groups: 地址编码 [6]
```

```
##
```

```
## # A tibble: 6 x 3
```

```
## 地址编码 发病年月 发病人数
```

```
## <chr> <chr> <int>
```

```
## 1 360721 200708 1
```

```
## 2 360722 200707 1
```

```
## 3 360730 200710 1
```

```
## 4 361128 200702 1
```

```
## 5 369999 200708 1
```

```
## 6 421083 200712 1
```

```
write.csv(d.pas1," 分区分年月统计.csv",row.names=FALSE)
# 分区分月统计
d.patients[[" 发病月"]]=month(d.patients$发病日期)
d.pas2<-as.data.frame.matrix(table(unlist(d.patients[, ' 地址编码']),
                                   unlist(d.patients[, ' 发病月']))))
head(d.pas2)
```

```
##           1 2 3 4 5 6 7 8 9 10 11 12
## 360721 0 0 0 0 0 0 0 1 0 0 0 0
## 360722 0 0 0 0 0 0 1 0 0 0 0 0
## 360730 0 0 0 0 0 0 0 0 0 1 0 0
## 361128 0 1 0 0 0 0 0 0 0 0 0 0
## 369999 0 0 0 0 0 0 0 1 0 0 0 0
## 421083 0 0 0 0 0 0 0 0 0 0 0 1
```

```
write.csv(d.pas2," 分区分月统计.csv")
# 分年月分性别统计
d.pas3<-as.data.frame.matrix(table(unlist(d.patients[, ' 发病年月']),
                                   unlist(d.patients[, ' 性别']))))
d.pas3[[" 总计"]]=d.pas3$男+d.pas3$女
d.pas3
```

```
##           男 女 总计
## 200701 107 26 133
## 200702 71 16 87
## 200703 109 14 123
## 200704 109 21 130
## 200705 115 15 130
## 200706 115 19 134
## 200707 145 28 173
## 200708 169 37 206
## 200709 220 51 271
## 200710 180 44 224
## 200711 206 53 259
```

```
## 200712 199 42 241
```

```
write.csv(d.pas3," 分年月分性别统计.csv")
# 职业构成
d.pas4<-count(d.patients,d.patients$' 职业')
names(d.pas4)<-c(" 职业"," 发病人数")
d.pas4[[" 百分比"]]<-round(100*d.pas4$发病人数/sum(d.pas4$发病人数),digits = 1)
d.pas4
```

```
## # A tibble: 16 x 3
##   职业      发病人数 百分比
##   <chr>      <int>  <dbl>
## 1 不详          241  11.4
## 2 餐饮食品业      24   1.10
## 3 干部职工      153   7.20
## 4 工人          256  12.1
## 5 家务及待业     309  14.6
## 6 教师           19   0.900
## 7 离退人员      242  11.5
## 8 民工           34   1.60
## 9 农民          469  22.2
## 10 其它:        199   9.40
## 11 散居儿童        3   0.100
## 12 商业服务      114   5.40
## 13 学生           24   1.10
## 14 医务人员        13   0.600
## 15 幼托儿童         2   0.100
## 16 渔(船)民         9   0.400
```

```
write.csv(d.pas4," 职业构成.csv")
# 年龄性别分布
age_classes<-d.patients$发病年龄 %/% 10
age_classes<-ifelse(age_classes<=7,age_classes,7)
d.patients[[" 年龄段"]]<-as.factor(paste(age_classes*10, "-",
```

```

        ifelse(age_classes<7,10*(age_classes+1)," 以上"))
d.pas5<-as.data.frame.matrix(table(unlist(d.patients[, ' 年龄段']),
                                   unlist(d.patients[, ' 性别']))))
names(d.pas5)<-c(" 男性发病人数"," 女性发病人数")
d.pas5[[' 发病率 (男) ']]<-round(
  100*d.pas5$男性发病人数/sum(d.pas5$男性发病人数),digits=1)
d.pas5[[' 发病率 (女) ']]<-round(
  100*d.pas5$女性发病人数/sum(d.pas5$女性发病人数),digits=1)
d.pas5

```

##	男性发病人数	女性发病人数	发病率 (男)	发病率 (女)
## 0 - 10	4	2	0.2	0.5
## 10 - 20	17	9	1.0	2.5
## 20 - 30	97	43	5.6	11.7
## 30 - 40	360	80	20.6	21.9
## 40 - 50	401	89	23.0	24.3
## 50 - 60	357	58	20.5	15.8
## 60 - 70	248	35	14.2	9.6
## 70 - 以上	261	50	15.0	13.7

```

write.csv(d.pas5," 年龄性别分布.csv")

```