Python R

gg_hatano

2020-08-23

Contents

Chapter 1

1.1

Python pandas R
GitHub

1.2

R

Chapter 2

chapter

2.1 (1-1)

1-1

2.1.1

IPSS

```
library(readr)
library(dplyr)
url = 'http://www.ipss.go.jp/p-toukei/JMD/00/STATS/Births.txt'
dat = read.table(url, skip=2, header = TRUE)
dat %>% head
```

```
##   Year  Female    Male   Total
## 1 1947 1301806 1376986 2678792
## 2 1948 1303060 1378564 2681624
## 3 1949 1316630 1380008 2696638
## 4 1950 1134396 1203111 2337507
## 5 1951 1043048 1094641 2137689
## 6 1952  977101 1028061 2005162
```

Female Male

```
library(tidyr)
library(magrittr)
dat %>%
        pivot_longer(cols = c("Male", "Female"), names_to = "Sex", values_to = "Life") %>%
        mutate(Sex = if_else(Sex == "Female", "F", "M")) -> dat
dat
```

```
## # A tibble: 140 x 4
##     Year   Total Sex      Life
##    <int>   <int> <chr>   <int>
##  1  1947 2678792 M     1376986
##  2  1947 2678792 F     1301806
##  3  1948 2681624 M     1378564
##  4  1948 2681624 F     1303060
##  5  1949 2696638 M     1380008
##  6  1949 2696638 F     1316630
##  7  1950 2337507 M     1203111
##  8  1950 2337507 F     1134396
##  9  1951 2137689 M     1094641
## 10  1951 2137689 F     1043048
## # ... with 130 more rows
```

```
dat %<>%
        arrange(Sex, Year)
dat %<>%
        mutate(ratio = Life / sum(Life)) %>%
        mutate(cum_sum = cumsum(ratio))

dat %>% head
```

```
## # A tibble: 6 x 6
##    Year   Total Sex      Life   ratio cum_sum
##   <int>   <int> <chr>   <int>   <dbl>   <dbl>
## 1  1947 2678792 F     1301806 0.0121   0.0121
## 2  1948 2681624 F     1303060 0.0122   0.0243
## 3  1949 2696638 F     1316630 0.0123   0.0366
## 4  1950 2337507 F     1134396 0.0106   0.0472
## 5  1951 2137689 F     1043048 0.00973  0.0569
## 6  1952 2005162 F      977101 0.00912  0.0660
```

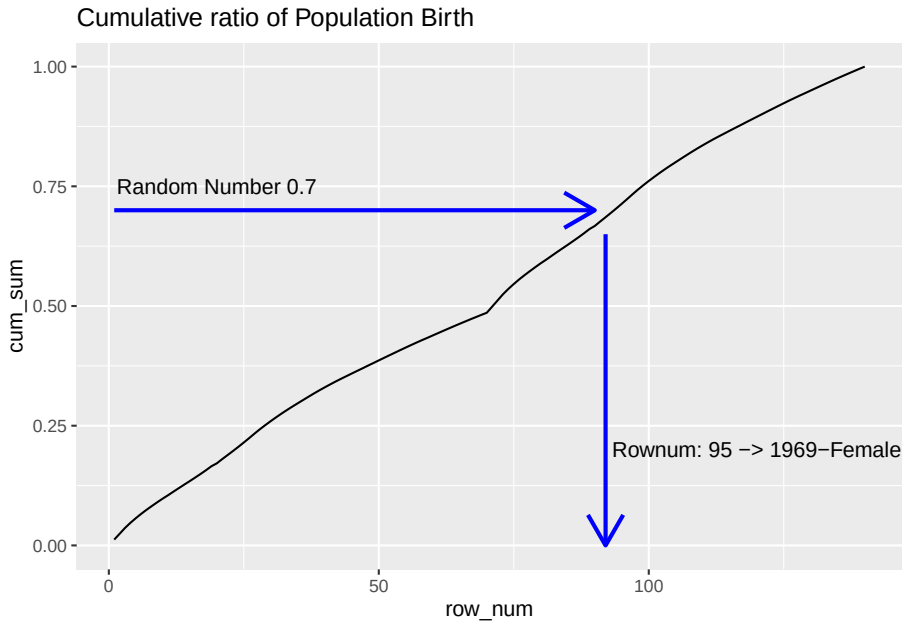
```
dat %>%
        select(-Total) %>%
        write.csv("./data/ipss_birth.csv", row.names=FALSE, quote = FALSE)
```

(y :0-1) (,)

```
library(ggplot2)
dat = read_csv("./data/ipss_birth.csv")
```

```
## Parsed with column specification:
## cols(
##   Year = col_double(),
##   Sex = col_character(),
##   Life = col_double(),
##   ratio = col_double(),
##   cum_sum = col_double()
## )
```

```
dat %>%
        mutate(x_axis = paste(Year, Sex, sep="")) %>%
        mutate(row_num = 1:nrow(.)) %>%
        ggplot() +
        geom_line(aes(x = row_num, y = cum_sum), stat = "identity") +
        annotate("segment",x=1,xend=90,y=0.7,yend=0.7,colour="blue",
             size=1, arrow=arrow()) +
        annotate("segment",x=92,xend=92,y=0.65,yend=0.0,colour="blue",
                    size=1,arrow=arrow()) +
        annotate("text", x=20, y=0.75, parse=TRUE, label="'Random Number 0.7'") +
        annotate("text", x=120, y=0.2, parse=TRUE, label="'Rownum: 95 -> 1969-Female'") +
        ggtitle("Cumulative ratio of Population Birth")
```



2.1.2

```
url_death_rate = "http://www.ipss.go.jp/p-toukei/JMD/00/STATS/Mx_1x1.txt"
dat = read.table(url_death_rate, skip=2, header = TRUE)
dat %>% str
```

```
## 'data.frame': 7770 obs. of  5 variables:
##  $ Year  : int  1947 1947 1947 1947 1947 1947 1947 1947 1947 1947 ...
##  $ Age   : chr  "0" "1" "2" "3" ...
##  $ Female: chr  "0.087401" "0.033723" "0.016994" "0.011412" ...
##  $ Male  : chr  "0.099181" "0.034697" "0.016804" "0.011461" ...
##  $ Total : chr  "0.093432" "0.034220" "0.016897" "0.011437" ...
```

Year X 1 ( )

2016

```
dat %<>%
        filter(Year == 2016) %>%
        select(-Year) %>%
        select(-Total)
```

Age char

```
dat$Age %>% table
```

```
## .
##    0    1   10  100  101  102  103  104  105  106  107  108  109   11 110+   12
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
##   13   14   15   16   17   18   19    2   20   21   22   23   24   25   26   27
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
##   28   29    3   30   31   32   33   34   35   36   37   38   39    4   40   41
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
##   42   43   44   45   46   47   48   49    5   50   51   52   53   54   55   56
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
##   57   58   59    6   60   61   62   63   64   65   66   67   68   69    7   70
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
##   71   72   73   74   75   76   77   78   79    8   80   81   82   83   84   85
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
##   86   87   88   89    9   90   91   92   93   94   95   96   97   98   99
##    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
```

110+ 111

```
dat %<>%
        mutate(Age = if_else(Age == "110+", "111", Age)) %>%
        mutate(Age = as.integer(Age))

dat %<>%
        mutate(Anb = Age) %>%
        select(-Age)

dat %>% head
```

```
##     Female     Male Anb
## 1 0.002028 0.001995   0
## 2 0.000313 0.000340   1
## 3 0.000174 0.000178   2
## 4 0.000098 0.000133   3
## 5 0.000087 0.000095   4
## 6 0.000084 0.000101   5
```

Anb Alb Anb Alb

x Anb q_x Alb $\frac{q_x+q_{x+1}}{2}$

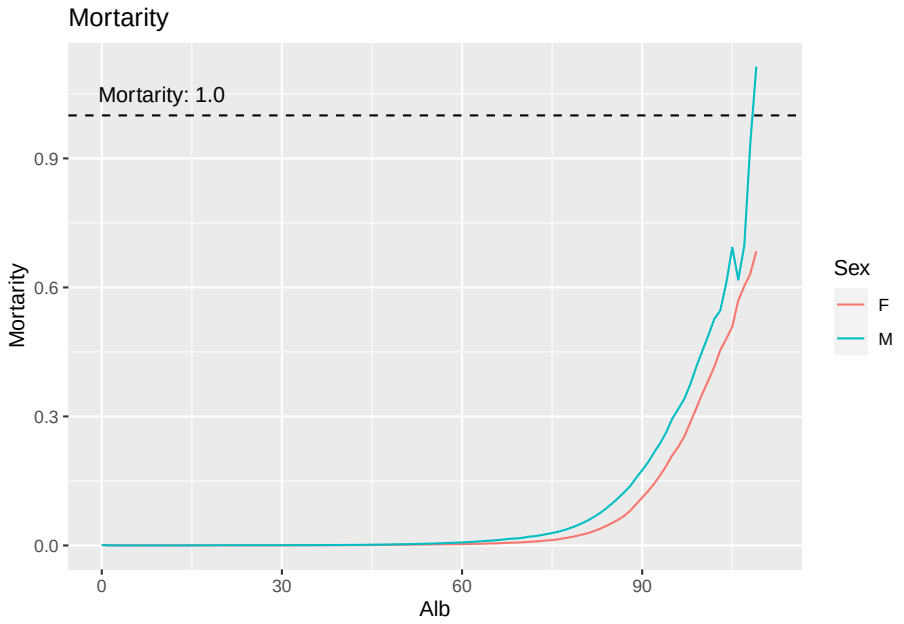
```
dat %<>%
        mutate(Female = as.numeric(Female)) %>%
        mutate(Male = as.numeric(Male)) %>%
        mutate(lead_Female = lead(Female)) %>%
        mutate(lead_Male = lead(Male)) %>%
        mutate(F = (Female + lead_Female)/2) %>%
        mutate(M = (Male + lead_Male)/2) %>%
        mutate(Alb = Anb) %>%
        select(Alb,F,M)

dat %>% head
```

```
##   Alb         F         M
## 1   0 0.0011705 0.0011675
## 2   1 0.0002435 0.0002590
## 3   2 0.0001360 0.0001555
## 4   3 0.0000925 0.0001140
## 5   4 0.0000855 0.0000980
## 6   5 0.0000815 0.0001060
```

1

```
dat %>%
        pivot_longer(cols=c("F","M"), names_to = "Sex", values_to = "Mortarity") %>%
        ggplot(aes(x = Alb, y = Mortarity, group = Sex, color = Sex)) +
        geom_line() +
        geom_hline(yintercept = 1.0, linetype = "dashed") +
        annotate("text", x = 10, y = 1.05, label = 'Mortarity: 1.0') +
        ggtitle("Mortarity")
```

```
## Warning: Removed 2 row(s) containing missing values (geom_path).
```



100

y x

$y = 1 - (1 - x)^{12}$ $x = 1 - (1 - y)^{0.08\ldots}$

```
dat %<>%
        filter(Alb < 100)

dat %<>%
        mutate(F = 1 - (1-F)**(1/12)) %>%
        mutate(M = 1 - (1-M)**(1/12))

dat %>% write.csv("./data/ipss_mortality.csv", quote=F, row.names = F)
dat %>% head
```

```
##   Alb            F            M
## 1   0 9.759403e-05 9.734377e-05
## 2   1 2.029393e-05 2.158590e-05
## 3   2 1.133404e-05 1.295926e-05
## 4   3 7.708660e-06 9.500496e-06
## 5   4 7.125279e-06 8.167034e-06
## 6   5 6.791920e-06 8.833763e-06
```

2.1.3

-
-
-

2.1.4

… - -

2.2 (1-2)

2.2.1

2.2.2

2.2.3