

wang minjie

关于如何使用 *easyuse* 宏包

To my daughter,
without whom I should have finished this book two years earlier

目录

表格	v
插图	vii
前言	ix
关于作者	xi
1 如何安装	1
2 混合线性模型	3
3 给列变量一个阈值	7
4 行方向的操作	11
4.1 行方向求和	11
4.2 行方向求均值	12
4.3 行方向赋权重再求和	13
4.4 行方向赋权重再求均值	14
4.5 行方向探测最先匹配	15
附录	17
A 更多	17
索引	19



表格

2.1	The test data.	4
2.2	The summary data.	5



插图

2.1 Hello World!	5
----------------------------	---



前言

事实上，你根本不需要这个宏包。

为什么有这个宏包

A Package to make social statistics more easy and more tidyverse. It is very important to me.

它有那些函数

```
library(easyuse)
lsf.str("package:easyuse")

## add_above_avg_num : function (.data, ..., .name = "above_avg_num",
##     na.rm = TRUE)
## add_row_means : function (.data, ..., .name = "row_mean", na.rm = TRUE)
## add_row_sums : function (.data, ..., .name = "row_sum", na.rm = FALSE)
## add_weighted_mean : function (.data, ..., .weights, .name = "weighted_mean",
##     na.rm = TRUE)
## add_weighted_sum : function (.data, ..., .weights, .name = "weighted_sum",
##     na.rm = TRUE)
## cutoffs_modify_at : function (df, .vars, cutoffs)
```

```
## get_ran_vals : function (.data, .var_school, .var_class, .var_score_pre,  
##      .var_score_post, effects = "class")  
## row_first_match : function (.f, ...)
```



Acknowledgments

A lot of people helped me when I was writing the book.

王敏杰
川师图书馆某角落

关于作者

王敏杰，四川师范大学研究生公选课《数据科学中的 R 语言》授课老师，西南交通大学量子物理学博士，爱好数据科学，喜欢用 R 和 Raku 编程，联系方式 38552109@qq.com¹

¹<mailto:38552109@qq.com>



1

如何安装

没打算放在 CRAN，所以只能通过 github 上安装

```
install.packages("devtools")  
devtools::install_github("perlatex/easyuse")
```



2

混合线性模型

在开始介绍混合线性模型之前，我们先看看一个份学生考试成绩[2.1](#)

```
library(tidyverse)
library(easyuse)
data(chengdu)

d <- chengdu
```

```
d %>%
  head(8) %>%
  knitr::kable(
    caption = 'The test data.',
    booktabs = TRUE
  )
```

- score_pre 是前一次考试成绩
- score_post 是后一次考试成绩

我们想看看两次考试成绩是否有关联，或者想用第一次考试成绩，预测下一次考试成绩，常用的模型是混合线性模型

```
lme4::lmer(score_post ~ 1 + score_pre + (1 | effect))
```

为了方便计算不同分组（学校或者班级）下的随机效应（增值分），可以使

表 2.1: The test data.

school	id	class	score_pre	score_post
成都市大弯中学校	1301	13011714	570	630
四川师范大学附属中学	0401	04011712	616	670
四川师范大学附属中学	0401	04011708	622	600
四川师范大学附属中学	0401	04011712	597	630
四川师范大学附属中学	0401	04011708	630	620
四川师范大学附属中学	0401	04011711	599	640
四川师范大学附属中学	0401	04011711	659	630
四川师范大学附属中学	0401	04011711	619	630

用 `easyuse` 包的 `get_ran_vals()` 函数。该函数返回分组下随机效应和统计描述

```
d %>%
  get_ran_vals(
    .var_school = school,
    .var_class = class,
    .var_score_pre = score_pre,
    .var_score_post = score_post,
    effects = "school"
  ) %>%
  head(8) %>%
  knitr::kable(
    caption = 'The summary data.',
    booktabs = TRUE
  )
```

We have a nice figure in Figure 2.1

表 2.2: The summary data.

level	estimate	num_of_students	mean_score_pre	mean_score_post
安仁中学	-32.96	408	460.8	441.5
北大成都附属实验学校	0.17	31	543.2	523.5
北大附中成都实验学校	-16.33	224	384.9	413.2
北大附中新津实验学校	-76.05	138	380.0	349.4
成都大学附属中学	-4.41	106	560.5	529.1
成都第 11 中学	-22.85	133	519.9	486.3
成都第 12 中学	47.68	110	633.6	625.6
成都第 17 中学	-2.59	101	574.9	539.5

```
par(mar = c(4, 4, 1, .1))
plot(cars, pch = 19)
```

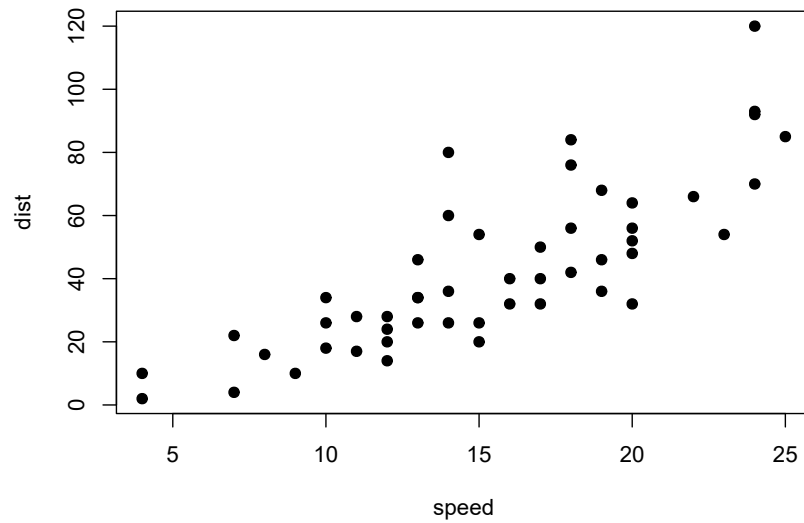


图 2.1: Hello World!



3

给列变量一个阈值

有时候希望获得一个剥夺矩阵，具体来说，就是给列变量一个阈值，超过阈值赋值为 0，低于阈值赋值 1

```
library(tidyverse)
library(easyuse)
```

```
df <- tribble(
  ~id, ~x, ~y, ~z, ~g,
  #--/--/--/--/--
  "a", 13.1, 14, 4, 1,
  "b", 11.2, 7, 5, 0,
  "c", 12.5, 10, 1, 0,
  "d", 20, 11, 3, 1
)

cutoffs <- list(x = 13, y = 12, z = 3)
```

一般来说，可以使用 `mutate() + if_else()`

```
df %>%
  mutate(x = if_else(x < cutoffs$x, 1, 0)) %>%
  mutate(y = if_else(y < cutoffs$y, 1, 0)) %>%
  mutate(z = if_else(z < cutoffs$z, 1, 0))
```

```
## # A tibble: 4 x 5
##   id      x      y      z      g
##   <chr> <dbl> <dbl> <dbl> <dbl>
## 1 a      0      0      0      1
## 2 b      1      1      0      0
## 3 c      1      1      1      0
## 4 d      0      1      0      1
```

或者使用 `pivot_longer()` , 然后 `pivot_wider()`

```
df %>%
  pivot_longer(cols = x:z,
               names_to = "var",
               values_to = "val") %>%
  mutate(below_cutoff = as.integer(val < cutoffs[var])) %>%
  select(id, g, var, below_cutoff) %>%
  pivot_wider(names_from = var, values_from = below_cutoff)
```

```
## # A tibble: 4 x 5
##   id      g      x      y      z
##   <chr> <dbl> <int> <int> <int>
## 1 a      1      0      0      0
## 2 b      0      1      1      0
## 3 c      0      1      1      1
## 4 d      1      0      1      0
```

当然, 我们可以用 `cutoffs_modify_at()` 函数写的更直观一些

```
df %>%
  cutoffs_modify_at(.vars = c(x, y), cutoffs = c(x = 13, y = 12))
```

```
## # A tibble: 4 x 5
##   id      x      y      z      g
##   <chr> <dbl> <dbl> <dbl> <dbl>
```

3.0

9

## 1 a	0	0	4	1
## 2 b	1	1	5	0
## 3 c	1	1	1	0
## 4 d	0	1	3	1



4

行方向的操作

tidyverse 行方向的操作，喜欢用 `purrr::pmap()`，但遇到特殊的需求，需要写较多的语句，比如，给指定的若干列，赋予权重，然后计算行方向的均值或者求和等等。为了简化语句增强可读性，我定义了下面一些函数

4.1 行方向求和

```
iris %>%  
  add_row_sums(starts_with("Sepal"), .name = "Sepal.sum") %>%  
  head()
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width  
## 1          5.1          3.5          1.4          0.2  
## 2          4.9          3.0          1.4          0.2  
## 3          4.7          3.2          1.3          0.2  
## 4          4.6          3.1          1.5          0.2  
## 5          5.0          3.6          1.4          0.2  
## 6          5.4          3.9          1.7          0.4  
##   Species Sepal.sum  
## 1  setosa      8.6  
## 2  setosa      7.9  
## 3  setosa      7.9  
## 4  setosa      7.7
```

```
## 5  setosa      8.6
## 6  setosa      9.3
```

4.2 行方向求均值

```
iris %>%
  add_row_means(starts_with("Sepal"), .name = "Sepal.Mean") %>%
  head()
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1           5.1           3.5           1.4           0.2
## 2           4.9           3.0           1.4           0.2
## 3           4.7           3.2           1.3           0.2
## 4           4.6           3.1           1.5           0.2
## 5           5.0           3.6           1.4           0.2
## 6           5.4           3.9           1.7           0.4
##   Species Sepal.Mean
## 1  setosa      4.30
## 2  setosa      3.95
## 3  setosa      3.95
## 4  setosa      3.85
## 5  setosa      4.30
## 6  setosa      4.65
```

4.3 行方向赋权重再求和

```
library(tidyverse)
library(easyuse)
```

```
df <- tribble(
  ~id, ~x, ~y, ~z, ~g,
  #--/--/--/--/--
  "a", 13.1, 14, 4, 1,
  "b", 11.2, 7, 5, 0,
  "c", 12.5, 10, 1, 0,
  "d", 20, 11, 3, 1
)
df
```

```
## # A tibble: 4 x 5
##   id      x      y      z      g
##   <chr> <dbl> <dbl> <dbl> <dbl>
## 1 a      13.1    14      4      1
## 2 b      11.2     7      5      0
## 3 c      12.5    10      1      0
## 4 d       20     11      3      1
```

```
weights <- c(
  x = 0.25,
  y = 0.25,
  z = 0.25,
  g = 0.25
)
```

```
df %>%
  add_weighted_sum(x:g, .name = "wt_sum", .weights = weights)
```

A tibble: 4 x 6

	id	x	y	z	g	wt_sum
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	a	13.1	14	4	1	8.02
## 2	b	11.2	7	5	0	5.8
## 3	c	12.5	10	1	0	5.88
## 4	d	20	11	3	1	8.75

4.4 行方向赋权重再求均值

```
df %>%
  add_weighted_mean(x:g, .name = "wt_mean", .weights = weights)
```

A tibble: 4 x 6

	id	x	y	z	g	wt_mean
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	a	13.1	14	4	1	2.01
## 2	b	11.2	7	5	0	1.45
## 3	c	12.5	10	1	0	1.47
## 4	d	20	11	3	1	2.19

4.5 行方向探测最先匹配

```
df <- tibble(  
  a = c("b", "d", "l", "m"),  
  x = c(1, 1, 1, 2),  
  y = c(5, 1, 2, 3),  
  z = c(1, 1, 0, 1)  
)  
df
```

```
## # A tibble: 4 x 4  
##   a         x     y     z  
##   <chr> <dbl> <dbl> <dbl>  
## 1 b         1     5     1  
## 2 d         1     1     1  
## 3 l         1     2     0  
## 4 m         2     3     1
```

x, y, z 三列，行方向选出最先大于 1 的值，构成新的一列

```
df %>% mutate(new_col = row_first_match(~ . > 1, x, y, z))
```

```
## # A tibble: 4 x 5  
##   a         x     y     z new_col  
##   <chr> <dbl> <dbl> <dbl>   <dbl>  
## 1 b         1     5     1       5  
## 2 d         1     1     1      NA  
## 3 l         1     2     0       2  
## 4 m         2     3     1       2
```



A

更多

Yeah! I have finished my book, but I have more to say about some topics.
Let me explain them in this appendix.

To know more about **bookdown**, see <https://github.com/yihui/bookdown-crc>.



索引

cutoffs, 7

lme4, 3

lmm, 3

剥夺矩阵, 7