

# 作业一

## R代码

```
# 一. 简答题（共1题）
# 1. (简答题)
# (1) 把data0106 length和data0107 yield合并为一个文件，保存到data0108 length yield（包括三个变量：group、length、yield）。
# (2) 计算变量length的平方根存入变量sqrtlg，计算yield的对数（LN）存入变量lnyd，变量sqrtlg和lnyd包括2位小数，保存到文件data0108 length yield。
# (3) 把length和yield变量按如下方式重新编码（length: 35-39 -> 1, 40-44 -> 2, 45-49 -> 3, 50-54 -> 4, 55-59 -> 5; yield: 50-54 -> 1, 55-59 -> 2, 60-64 -> 3, 65-69 -> 4）
# (4) 计算length和yield的描述性统计量，包括N、Range、Minimum、Maximum、Mean、SE of mean、SD、Variance、Skewness、Kurtosis。
# 注意：
# 如用R，需提交R命令和结果都需提交，保存pdf格式，命名"homework1_姓名"，作为附件提交。
```

```
library(haven)
```

```
# 1)
d_length <- read_sav("./R/DATA/data0106+length.sav"); head(d_length)
d_yield <- read_sav("./R/DATA/data0107+yield.sav"); head(d_yield)

length_yield <- merge(d_length, d_yield, by="group"); head(length_yield)
# write_sav(length_yield, "./R/DATA/length_yield.sav")
```

```
# 2)
length_yield$sqrtlg <- round(length_yield$length ** 0.5, 2)
# log() 求的自然对数，底数是 e (≈ 2.71828)
length_yield$lnyd <- round(log(length_yield$yield), 2)
head(length_yield)
```

```
# 3)
library(haven) # 读写 .sav
library(dplyr) # 管道与 mutate
```

```
length_yield <- length_yield %>% # 你的数据框已存在
mutate(
  ## ----- length category -----
  lgcat = case_when(
    between(length, 35, 39) ~ 1L,
    between(length, 40, 44) ~ 2L,
    between(length, 45, 49) ~ 3L,
    between(length, 50, 54) ~ 4L,
    between(length, 55, 59) ~ 5L
  ),
  ## ----- yield category -----
  ydcat = case_when(
    between(yield, 50, 54) ~ 1L,
    between(yield, 55, 59) ~ 2L,
    between(yield, 60, 64) ~ 3L,
    between(yield, 65, 69) ~ 4L,
    between(yield, 70, 74) ~ 5L,
    between(yield, 75, 79) ~ 6L,
    between(yield, 80, 84) ~ 7L,
    between(yield, 85, 89) ~ 8L,
    between(yield, 90, 94) ~ 9L,
    between(yield, 95, 99) ~ 10L
  )
) %>%
## 转换为有序因子，并加标签
mutate(
  lgcat = ordered(lgcat) %>%
    labelled(
      labels = c("35-39" = 1, "40-44" = 2, "45-49" = 3,
                 "50-54" = 4, "55-59" = 5),
      label = "length category"
    ),
  ydcat = ordered(ydcat) %>%
    labelled(
      labels = c("50-54" = 1, "55-59" = 2, "60-64" = 3, "65-69" = 4,
                 "70-74" = 5, "75-79" = 6, "80-84" = 7, "85-89" = 8,
                 "90-94" = 9, "95-99" = 10),
      label = "yield category"
    )
)
```

```

head(length_yield)
write_sav(length_yield, "./R/DATA/data0108_length_yield.sav")

# 4)
library(dplyr)
library(e1071)

desc_stats <- function(x){
  n <- length(x)
  m <- mean(x, na.rm = TRUE)
  s <- sd(x, na.rm = TRUE)
  c(N          = n,
     Range      = max(x, na.rm = TRUE) - min(x, na.rm = TRUE),
     Minimum    = min(x, na.rm = TRUE),
     Maximum    = max(x, na.rm = TRUE),
     Mean       = m,
     `SE of mean` = s / sqrt(n),
     SD         = s,
     Variance    = var(x, na.rm = TRUE),
     Skewness    = skewness(x, na.rm = TRUE),
     Kurtosis    = kurtosis(x, na.rm = TRUE))
}

result_length <- desc_stats(length_yield$length); print(result_length)
result_yield <- desc_stats(length_yield$yield); print(result_yield)

```

结果

```
D:\APP_cs\YD_learn\github\youth4science>Rscript ./R/HOMEWORK/work1.R
Warning message:
package 'haven' was built under R version 4.4.3
# A tibble: 6 × 2
  group length
  <dbl> <dbl>
1     1     36
2     2     37
3     1     38
4     2     38
5     1     39
6     2     39
# A tibble: 6 × 2
  group yield
  <dbl> <dbl>
1     1     51
2     2     51
3     1     51
4     2     53
5     1     53
6     2     53
  group length yield
1     1     36     51
2     1     36     61
3     1     36     51
4     1     36     57
5     1     36     53
6     1     36     57
  group length yield sqrtlg lnyd
1     1     36     51     6 3.93
2     1     36     61     6 4.11
3     1     36     51     6 3.93
4     1     36     57     6 4.04
5     1     36     53     6 3.97
6     1     36     57     6 4.04

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

  filter, lag

The following objects are masked from 'package:base':

  intersect, setdiff, setequal, union

Warning message:
package 'dplyr' was built under R version 4.4.3
  group length yield sqrtlg lnyd lgcat ydcat
1     1     36     51     6 3.93     1     1
2     1     36     61     6 4.11     1     3
3     1     36     51     6 3.93     1     1
4     1     36     57     6 4.04     1     2
5     1     36     53     6 3.97     1     1
6     1     36     57     6 4.04     1     2

Warning message:
package 'e1071' was built under R version 4.4.3
      N      Range    Minimum    Maximum      Mean
5000.00000000 19.00000000 36.00000000 55.00000000 45.50000000
SE of mean      SD    Variance    Skewness    Kurtosis
 0.05515985  3.90039006 15.21304261  0.00000000 -0.33895881
      N      Range    Minimum    Maximum      Mean  SE of mean
5.000000e+03 4.700000e+01 5.100000e+01 9.800000e+01 6.661000e+01 1.577991e-01
      SD    Variance    Skewness    Kurtosis
1.115808e+01 1.245028e+02 9.263295e-01 5.231171e-02
```