Applied Regression Analysis using R

L01-Basic

use ’_’ in name, to avoid confusion with the member function - help

```
rm(list = ls()) # clean environment
?is.logical()
library(help = "stats") # for function list in a package
is.integer(7L) # integer, with 'L' after number
```

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

```
is.integer(7)
```

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

```
x <- 1 # assignment
1 -> x1
x = 1

# vector c()
month <- c('Jan','Feb','Dec')
class(month)
```

```
## [1] "character"
```

```
num <- 1:50 # return a vector
num_seq <- seq(1, 50, by = 2) # create sequence: seq(from, to, by)
seq(stats::rnorm(20))
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

```
num_seq[which(num_seq>9)]
```

```
##  [1] 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
```

```
# categorical data
month.fac <- factor(month, order=TRUE, levels = c("Jan", "Feb", "Dec"))
month.fac.in <- factor(month) # default: mathematical order

# matrix, special vectors in R
A = matrix(month, nrow = 6, byrow = TRUE, ncol = 9)
B = matrix(month, nrow = 6, byrow = FALSE, ncol = 9)

# list, combine different data type
listAB = list(num = num_seq, Amatrix = A, Bmatrix = B)
class(listAB)
```

```
## [1] "list"
```

```
# data.frame
df.A <- data.frame(A)
df.A$X1
```

```
## [1] "Jan" "Jan" "Jan" "Jan" "Jan" "Jan"
```

```
class(df.A)
```

```
## [1] "data.frame"
```

```
class(df.A$X3)
```

```
## [1] "character"
```

```
df.A$X1 <- factor(df.A$X1)

# function
myfuc = function(x) { # key word function
    s = 0
    for (eps in x) {
      if (eps <= 0){
        next
      }
      s = s + eps
      if (s > 20) {
        break
      }
    }
    s
  }
```

e ## implicit coercion to mixed types logical -> integer -> numeric -> complex -> character

as transform from logical to integer is easier

`c(11, month)` will convert integer to a character

L02-slr

simple linear regression basic

- regression analysis: statistical model that involve one dependent variable and one or more independent variables
- regression: find a rule of picking distribution for Y from a space of infinitely many distribution that agrees with the data

Primary Assumption: (for now) a sequence of random variables is independent and identically distributed (i.i.d.)

Binomial Example

```
*L02-16/55 pages, US presidents *

- if we become US presidents, we are more likely to have sons -- T/F depend on the p-value
- US presidents are more likely to have sons -- independent of the p-value, because the current data ha
```

~~~
```r
theta <- 1/2
n <- 158
x <- 0:n
fx <- dbinom(x, size = n, prob = theta)
~~~

```r
myp_func = function(p){
  n = 190
  p125 = pbinom(125, size = n, prob = p)

}
```