dataframe

Nicholus Tint Zaw

2022-08-14

What we are going to cover in this exercise?

1. Function

```
set.seed(523535)

x <- rnorm(100)

# apply based functions
median(x)
```

```
## [1] 0.03011317
```

```
mean(x)
```

```
## [1] 0.01084897
```

```
min(x)
```

```
## [1] -1.991196
```

```
max(x)
```

```
## [1] 2.77532
```

```
quantile(x, 0.25)
```

```
##        25%
## -0.5333256
```

```
quantile(x, c(0.25, 0.75))
```

```
##        25%        75%
## -0.5333256  0.5431427
```

```
summary(x)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.99120 -0.53333  0.03011  0.01085  0.54314  2.77532
```

Now, it is time to create your own function!

$\bar{X} = \frac{\sum x_i}{N}$

What input parameter do we need to calculate mean? Of course, the numeric vector, it length (number of observation in the vector), and summation of all numbers from that vector. Right?

```
my_mean <- function(x){ # x = input parameter - numeric vector

  # n = number of observation in vector x
  n <- length(x)

  # summation of vector
  sumvect <- sum(x)

  # mean calculation
  cal_mean <- round(sumvect/n, 3)

  return(cal_mean)

}
```

Now, please working yourself on SD function!

$\sigma = \sqrt{\frac{\sum (x_i - \bar{X})^2}{N-1}}$

```
my_sd <- function(x){

  # n = number of observation in vector x
  n <- length(x)

  # summation of vector
  sumvect <- sum(x)

  # mean calculation
  cal_mean <- round(sumvect/n, 3)

  # variance calculation
  cal_var <- sum((x - cal_mean)^2)/(n-1)

  # sd calculation
  cal_sd = sqrt(cal_var)

  return(cal_sd)

}
```

Then, create a function which return the `dataframe` with contain `category variable` as index, `frequency` and `proportion` as column for each respective category. Please use `mtcars` data as sample dataset.

```r
df <- mtcars

table(df$cyl)
```

```
##
##  4  6  8
## 11  7 14
```

```r
prop.table(table(df$cyl))
```

```
##
##       4       6       8
## 0.34375 0.21875 0.43750
```