zjin26_mini-assign3

Vincent Jin

2023-04-03

==========

Mini-assignment # 3

==========

Vincent Jin

(1) create a vector containing numbers 0 to 10, 100 to 200, and 900 to 1000 (without entering all numbers individually)

```
v <- c(0:10, 100:200, 900:1000)
```

(2) find the min, max, sum, average and quartiles of the last vector

```
min(v)
```

```
## [1] 0
```

```
max(v)
```

```
## [1] 1000
```

```
sum(v)
```

```
## [1] 111155
```

```
mean(v)
```

```
## [1] 521.8545
```

```
quantile(v)
```

```
##   0%  25%  50%  75% 100%
##    0  142  195  947 1000
```

(3) find the min, max, sum, average and quartiles of the following vector: c(rnorm(1000,500,10)) [note that rnorm command is generating random numbers every time you run it!]

```
v1 <- c(rnorm(1000,500,10))
min(v1)
```

```
## [1] 471.5826
```

```
max(v1)
```

```
## [1] 529.2073
```

```
sum(v1)
```

```
## [1] 500022
```

```
mean(v1)
```

```
## [1] 500.022
```

```
quantile(v1)
```

```
##       0%      25%      50%      75%     100%
## 471.5826 493.5263 499.8866 506.6264 529.2073
```

(4) create a data.frame with four columns (age, sex, cost, risk) with 1000 rows, as follow:

first column randomly picked from 18 to 65;

second column random 'F' or 'M' or NA (i.e., missing);

third column with random numbers between 1000 and 100,000 (round to 1 digit); and,

fourth column with random number between 0 and 1 (do NOT round)

```
df1 <- data.frame(age = sample(18:65, 1000, replace = T),
           sex = sample(c('M','F', NA), 1000, replace = T),
           cost = round(runif(1000, min = 1000, max = 1000000), 1),
           risk = runif(1000, min = 0, max = 1))
```

(5) assign the previous data.frame to a variable such as x

get the mean of all columns in the data.frame

count the number of F and M and NA (find online how to count NA in the table command)

set your seed number to 100

```
set.seed(100)
x <- data.frame(age = sample(18:65, 1000, replace = T),
           sex = sample(c('M','F', NA), 1000, replace = T),
           cost = round(runif(1000, min = 1000, max = 1000000), 1),
           risk = runif(1000, min = 0, max = 1))
col <- names(x)
for (i in col) {
  if (typeof(x[[i]]) != "character") {
    cat(paste("the mean of", i, "is:\n"))
    print(mean(x[[i]]))
  }
}
```

```
## the mean of age is:
## [1] 41.34
```

```
## the mean of cost is:
## [1] 491410.7
## the mean of risk is:
## [1] 0.5027846
```

```
table(x$sex, useNA = "always")
```

```
##
##    F    M <NA>
##  332  322  346
```

(6) get the quartiles of the cost and risk columns before and after log transformation

```
col <- c('cost', 'risk')
for (i in col) {
  cat(paste("the quartiles of", i, "is:\n"))
  print(quantile(x[[i]]))
  cat(paste("the quartiles of", i, "after log transformation is:\n"))
  print(quantile(log(x[[i]])))
}
```

```
## the quartiles of cost is:
##       0%      25%      50%      75%     100%
##   6076.3 233668.7 483777.8 728429.7 999426.0
## the quartiles of cost after log transformation is:
##        0%       25%       50%       75%      100%
##  8.712151 12.361660 13.089380 13.498646 13.814936
## the quartiles of risk is:
##           0%          25%          50%          75%         100%
## 0.0001990511 0.2536240069 0.4965564649 0.7595366460 0.9996303301
## the quartiles of risk after log transformation is:
##            0%           25%           50%           75%          100%
## -8.5219489410 -1.3719054916 -0.7000582680 -0.2750468509 -0.0003697382
```

(7) count how many of the costs are more than mean of cost

avoid entering a plain number in the comparison

```
sum(x$cost > mean(x$cost))
```

```
## [1] 494
```