


```
#  
library(tidyverse) #  
library(magrittr) #  
library(stargazer)  
library(gtsummary)
```

$$X_i = X_1, X_2, \dots, X_n$$

$$\bar{X}$$

$$\bar{X} = (X_1 + X_2 + \dots + X_n) / n = \frac{1}{n} \sum_{i=1}^n X_i$$

```
data3_1 <- c(100,90,80,70,60,40,10) #<1>
xbar <- mean(data3_1)
xbar

#
xbar2 = sum(data3_1) / length(data3_1) #<2>
xbar2
```

① data3_1

② sum() () length() data3_1 7 7

[1] 64.28571

[1] 64.28571

```
median(data3_1)
```

```
[1] 70
```

```
x <- c(10, 10, 10, 10, 10, 10, 10, 20, 20, 30, 90, 100)
```

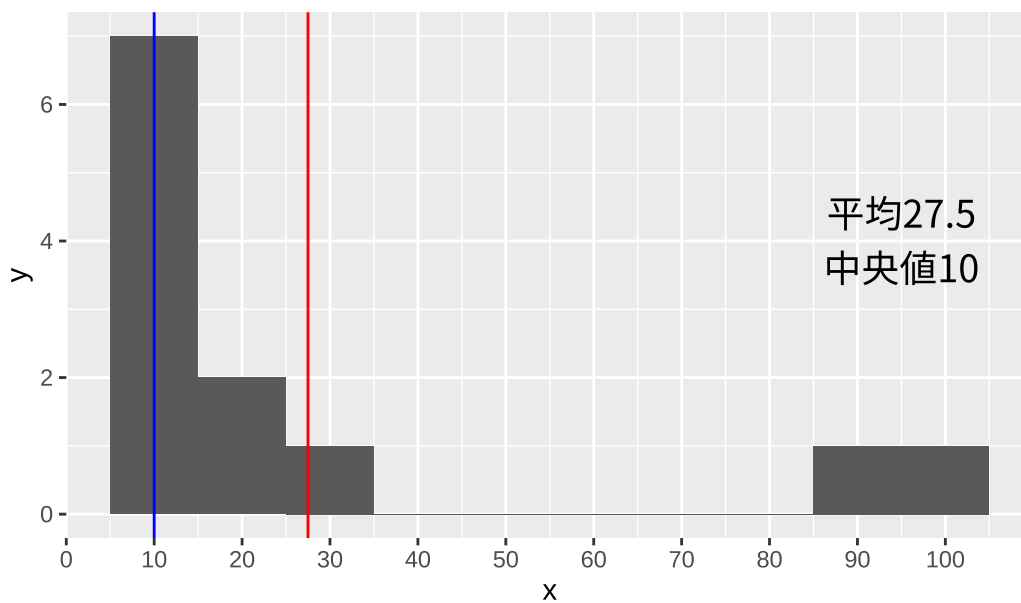
```
mean(x)
```

```
[1] 27.5
```

```
median(x)
```

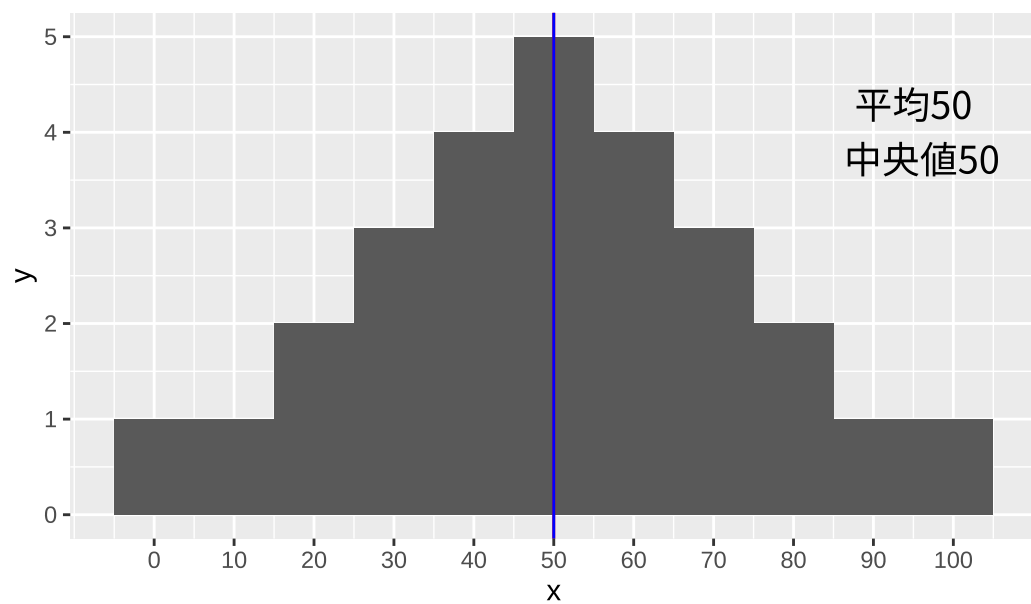
```
[1] 10
```

平均(赤)と中心値(青)が一致しない例



```
# A tibble: 27 x 1
  value
  <dbl>
1     0
2    10
3    20
4    20
5    30
6    30
7    30
8    40
9    40
10   40
# i 17 more rows
```

平均(赤)と中心値(青)が一致する例



```
max(data3_1)
```

[1] 100

```
min(data3_1)
```

[1] 10

A, B

60

A. 20,60,100,30,90

B. 60,60,60,60,60

$$\begin{aligned} \bar{X} - X_i & \quad 0 \quad 2 \quad (\bar{X} - X_i)^2 \\ & = \frac{1}{n} \sum_{i=1}^n (\bar{X} - X_i)^2 \end{aligned}$$

```
data3_2a <- c(20,60,100,30,90)
```

```
data3_2b <- c(60,60,60,60,60)
```

```
meana <- mean(data3_2a)
```

```
sa <- meana - data3_2a
```

```
sua <- sum(sa^2)
```

```
a <- sua / length(data3_2a) #<1>
```

```
b = sum((mean(data3_2b) - data3_2b)^2)/length(data3_2b)
```

```
a
```

```
b
```

①

[1] 1000

[1] 0

A 1000 B 0

$$100 \quad 2$$

$$= \sqrt{\quad}$$

```
A <- sqrt( a)
A
```

```
[1] 31.62278
```

```

data3 <- read_csv("data/data3.csv")
data3
table(data3$hometown)

data3 %>%
  select(.,hometown) %>% #<1>
  tbl_summary(., #<2>
    label = list(hometown ~ ""),
    sort = list(everything() ~ "frequency")
  ) %>%
  modify_header(label = "")

```

- ① select() (hometown)
- ② tbl_summary()

```

# A tibble: 30 x 5
      id class gender  test hometown
  <dbl> <dbl> <chr>   <dbl> <chr>
1     1     1     1     100
2     2     2     1      20
3     3     3     1      60
4     4     4     1      80
5     5     5     1      40
6     6     6     1      90
7     7     7     1      30
8     8     8     1      60
9     9     9     1      90

```



```
10    10    1          30
# i 20 more rows
```

```
7      6      1      10      3      1      2
```

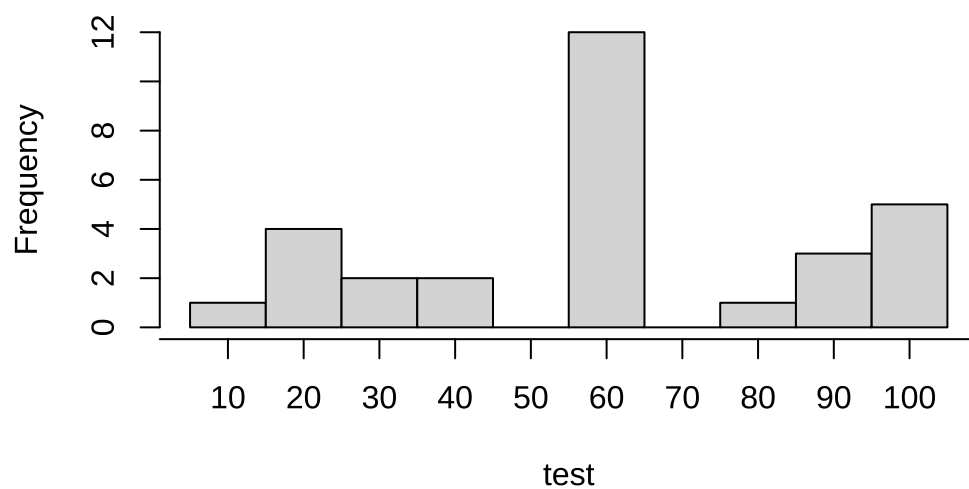
N = 30

```
10 (33%)
7 (23%)
6 (20%)
3 (10%)
2 (6.7%)
1 (3.3%)
1 (3.3%)
```

```
data3 %$%
  hist(test,
        breaks = seq(5,105,10), #<1>
        xaxt = "n" #<2>
        )
axis(1, at = 10*(0:100)) #<3>
```

- ① 10
- ②
- ③ 100 10

Histogram of test



ggplot2 (tidyverse)

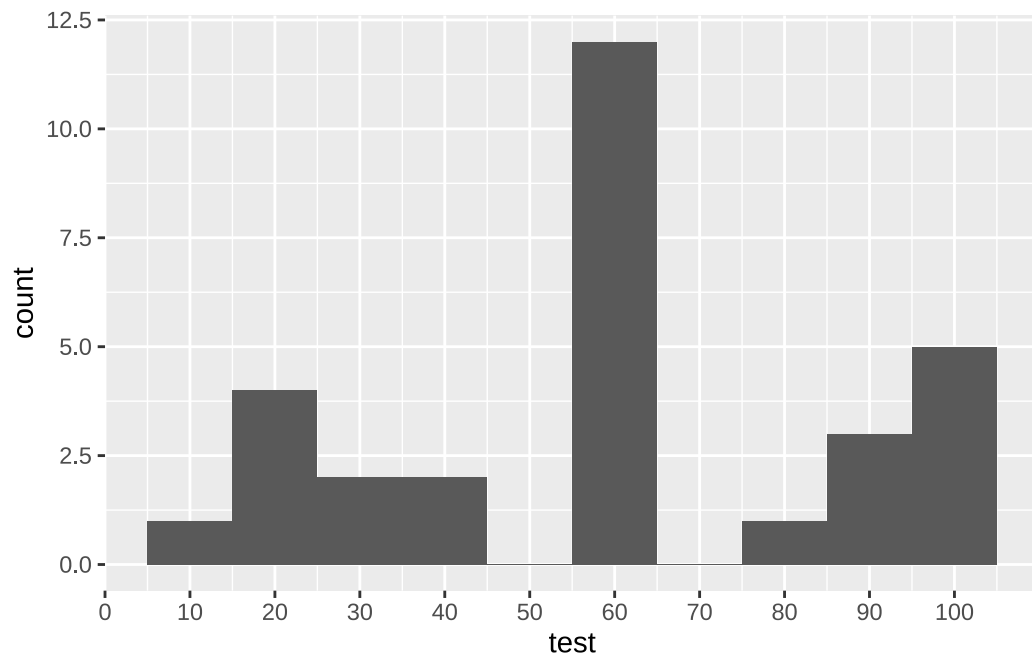
ggplot

```
data3 %$%  
  ggplot(., aes(test)) + #<1>  
  geom_histogram(breaks = seq(5,105,10)) + #<2>  
  scale_x_continuous(breaks=seq(0,100,10)) #<3>
```

① x

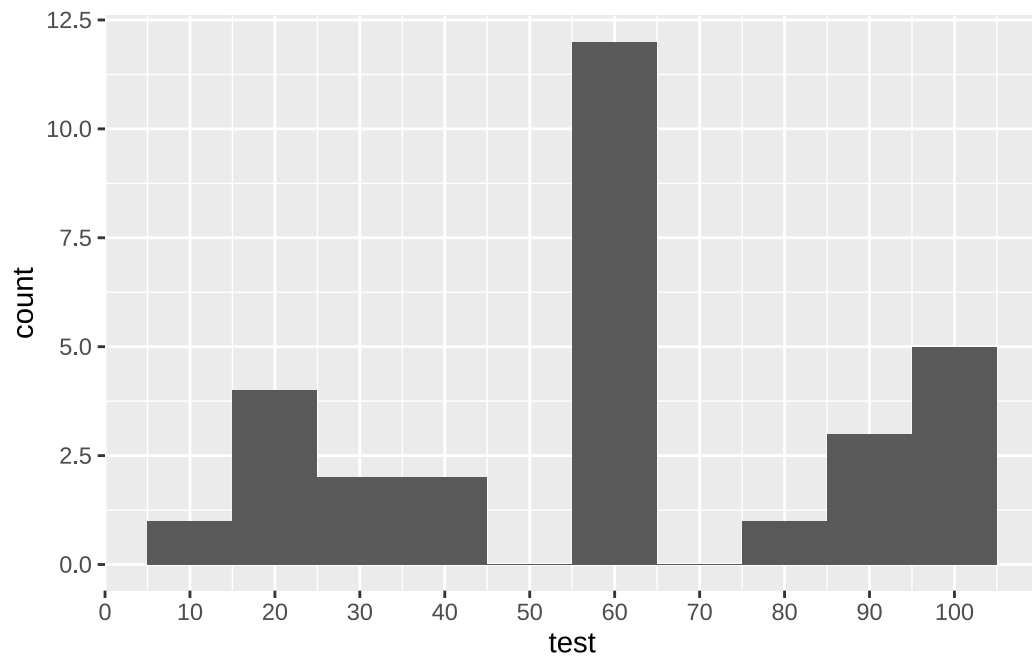
② 10

③ (10)



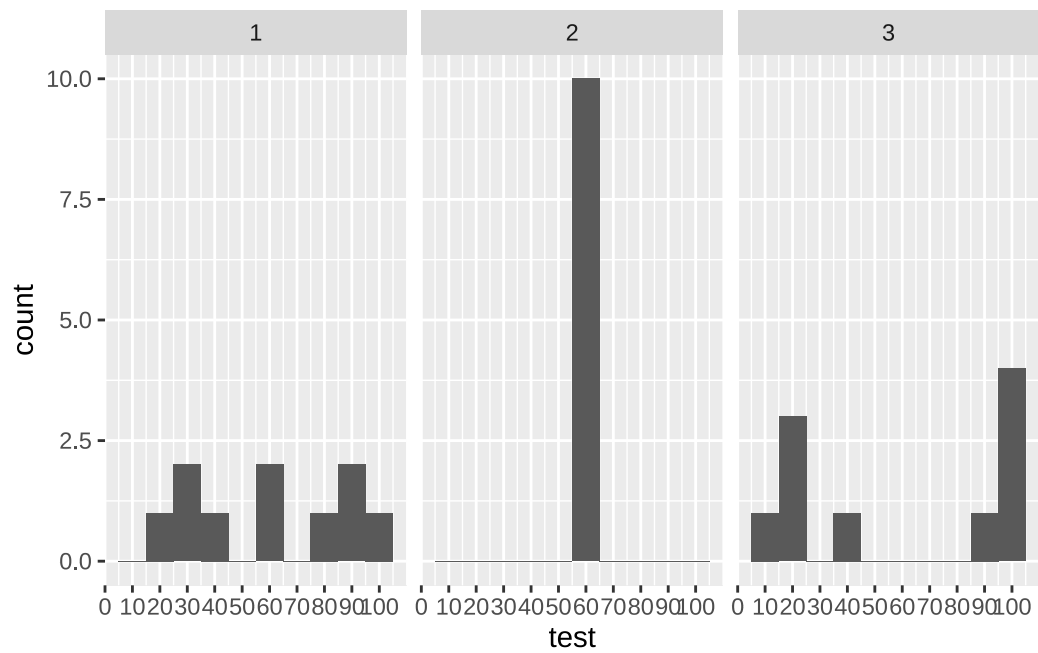
```
data3 %>%  
  ggplot(., aes(test, fill = class)) + #<1>  
  geom_histogram(breaks = seq(5,105,10)) +  
  scale_x_continuous(breaks=seq(0,100,10))
```

① fill



```
data3 %$%\n  ggplot(., aes(test)) +\n  geom_histogram(breaks = seq(5,105,10)) +\n  scale_x_continuous(breaks=seq(0,100,10)) +\n  facet_grid(~class) #<1>
```

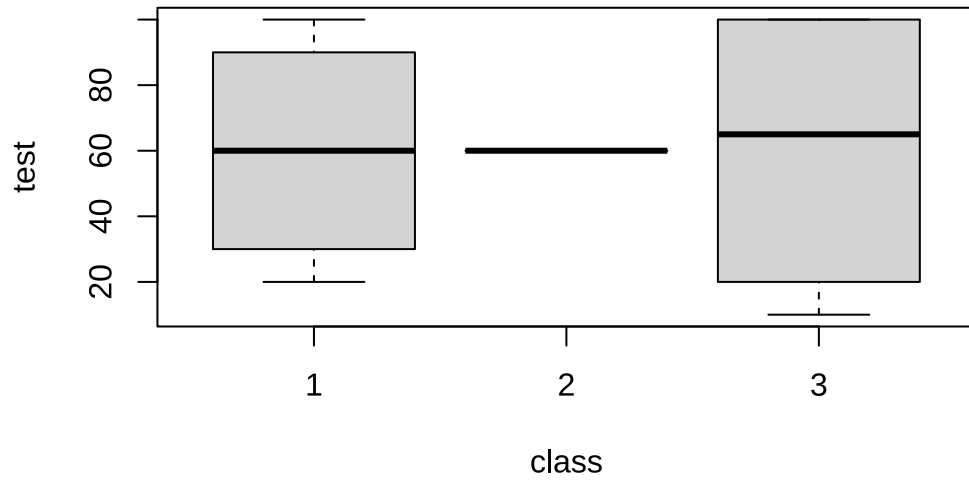
① class



60 3

```
data3 %$%
  boxplot(test~class) #<1>
```

① boxplot() () (~)



- 75 25 4 1
- 2 60

```
par(family= "jp")
data3 %$%
  boxplot(test~gender)
```

