

# 集中趋势与离散趋势

## Central tendency and variability

Qingyao Zhang

2026-05-09

### 目录

<b>1 集中趋势</b>	<b>1</b>
1.1 众数	1
1.2 中位数	2
1.3 均值	2
<b>2 离散趋势</b>	<b>4</b>
2.1 全距	4
2.2 四分位距	5
2.3 方差	5
2.4 标准差	5
<b>3 多个变量的描述性统计</b>	<b>6</b>

```
1 data("well", package = "Keng")
```

## 1 集中趋势

集中趋势的统计量有三个：众数 (mode)，中位数 (median)，算数平均数 (mean)。

### 1.1 众数

众数指的是频数最大的数值。以题项 flourish101 为例，通过统计频数或者绘制条形图，我们容易发现在 flourish101 上得 6 分的人最多，即，flourish101 的众数是 6。

## 1.2 中位数

百分位

```
1 # 第 50 百分位数, 即中位数
2 quantile(well$flourish1, probs = 0.5)
3 ## 50%
4 ## 5.375
```

## 1.3 均值

$$M = \frac{\sum x_i}{biased}$$

```
1 # 幸福感的一般水平
2 mean(well$flourish1, na.rm = TRUE)
3 ## [1] 5.32125
```

数据框的行均值与列均值: 以第一个时间点的实现论幸福感为例

rowname	flourish101	flourish102	flourish103	flourish104	flourish105	flourish106	flourish107	flourish108	rowMeans
1	7	6	6	7	6	4	6	4	?
2	4	2	4	6	7	4	4	4	?
3	6	4	6	6	5	6	6	6	?
4	7	6	4	7	7	7	7	6	?
5	6	5	6	7	7	7	7	7	?
6	6	5	6	6	6	6	6	6	?
colMeans	?	?	?	?	?	?	?	?	?

```
1 # 计算维度或量表的得分
2 # 计算 flourish 在第 1 个时间点的得分
3 # 查看变量名
4 names(well)
```

```
1 # 根据列号计算行均值, 提前移除缺失值
2 well$flourish1_1 <- rowMeans(well[287:294], na.rm = TRUE)
3 well$flourish1_1
4 ## [1] 5.750 4.375 5.625 6.375 6.500 5.875 4.000 7.000 4.375 5.625 5.750 5.125
5 ## [13] 3.375 4.875 3.250 4.000 6.500 6.500 5.250 5.125 6.500 4.625 2.875 4.125
6 ## [25] 5.750 4.750 5.500 5.625 5.625 5.250 5.625 5.125 6.000 5.000 4.625 5.500
```

```

7  ## [37] 5.000 4.250 4.625 7.000 4.375 5.125 5.500 4.500 6.625 5.000 6.000 6.125
8  ## [49] 4.875 4.375 5.000 4.750 5.000 5.750 4.625 3.000 1.750 4.500 6.250 5.000
9  ## [61] 5.500 4.750 4.875 3.625 4.500 4.375 6.125 5.250 4.500 4.125 6.000 4.500
10 ## [73] 5.375 5.625 6.000 6.000 5.750 4.000 5.875 6.000 6.125 5.875 5.750 6.000
11 ## [85] 5.375 6.375 4.125 4.750 6.750 5.500 5.875 6.000 5.750 5.125 4.125 4.875
12 ## [97] 5.000 3.500 5.500 6.250 4.875 5.125 4.625 4.125 3.000 5.000 6.000 4.625
13 ## [109] 4.250 5.375 7.000 5.625 6.000 5.875 7.000 6.750 7.000 6.500 4.625 5.250
14 ## [121] 4.875 6.000 5.000 6.000 6.000 5.750 6.625 5.250 7.000 4.125 4.000 6.500
15 ## [133] 6.875 6.125 7.000 4.750 5.000 6.250 5.625 6.000 5.375 7.000 5.750 5.250
16 ## [145] 6.000 4.250 5.125 5.500 5.625 5.750 5.250 5.375 5.750 5.750 5.000 4.875
17 ## [157] 4.375 4.375 7.000 5.875 4.500 3.875 5.875 6.250 6.750 5.875 6.625 5.875
18 ## [169] 4.000 6.125 5.375 4.625 6.000 4.625 4.875 5.375 5.750 6.500 5.000 6.500
19 ## [181] 4.875 4.500 5.000 5.000 4.500 6.000 7.000 5.875 5.625 4.625 4.500 6.625
20 ## [193] 5.625 5.375 4.250 5.750 5.000 4.375 4.625 4.000
21
22 # 根据列名计算行均值，提前移除缺失值
23 flourish_items <- c("flourish101","flourish102","flourish103","flourish104",
24                    "flourish105","flourish106","flourish107","flourish108")
25
26 well$flourish1_2 <- rowMeans(well[flourish_items], na.rm = TRUE)
27 well$flourish1_2
28 ## [1] 5.750 4.375 5.625 6.375 6.500 5.875 4.000 7.000 4.375 5.625 5.750 5.125
29 ## [13] 3.375 4.875 3.250 4.000 6.500 6.500 5.250 5.125 6.500 4.625 2.875 4.125
30 ## [25] 5.750 4.750 5.500 5.625 5.625 5.250 5.625 5.125 6.000 5.000 4.625 5.500
31 ## [37] 5.000 4.250 4.625 7.000 4.375 5.125 5.500 4.500 6.625 5.000 6.000 6.125
32 ## [49] 4.875 4.375 5.000 4.750 5.000 5.750 4.625 3.000 1.750 4.500 6.250 5.000
33 ## [61] 5.500 4.750 4.875 3.625 4.500 4.375 6.125 5.250 4.500 4.125 6.000 4.500
34 ## [73] 5.375 5.625 6.000 6.000 5.750 4.000 5.875 6.000 6.125 5.875 5.750 6.000
35 ## [85] 5.375 6.375 4.125 4.750 6.750 5.500 5.875 6.000 5.750 5.125 4.125 4.875
36 ## [97] 5.000 3.500 5.500 6.250 4.875 5.125 4.625 4.125 3.000 5.000 6.000 4.625
37 ## [109] 4.250 5.375 7.000 5.625 6.000 5.875 7.000 6.750 7.000 6.500 4.625 5.250
38 ## [121] 4.875 6.000 5.000 6.000 6.000 5.750 6.625 5.250 7.000 4.125 4.000 6.500
39 ## [133] 6.875 6.125 7.000 4.750 5.000 6.250 5.625 6.000 5.375 7.000 5.750 5.250
40 ## [145] 6.000 4.250 5.125 5.500 5.625 5.750 5.250 5.375 5.750 5.750 5.000 4.875
41 ## [157] 4.375 4.375 7.000 5.875 4.500 3.875 5.875 6.250 6.750 5.875 6.625 5.875
42 ## [169] 4.000 6.125 5.375 4.625 6.000 4.625 4.875 5.375 5.750 6.500 5.000 6.500
43 ## [181] 4.875 4.500 5.000 5.000 4.500 6.000 7.000 5.875 5.625 4.625 4.500 6.625
44 ## [193] 5.625 5.375 4.250 5.750 5.000 4.375 4.625 4.000
45
46 # 根据列名计算行均值
47 # 通过 paste0() 生成题项变量名
48 paste0("flourish", 101:108)

```

```

49 ## [1] "flourish101" "flourish102" "flourish103" "flourish104" "flourish105"
50 ## [6] "flourish106" "flourish107" "flourish108"
51 well$flourish1_3 <- rowMeans(well[paste0("flourish", 101:108)], na.rm = TRUE)
52 well$flourish1_3
53 ## [1] 5.750 4.375 5.625 6.375 6.500 5.875 4.000 7.000 4.375 5.625 5.750 5.125
54 ## [13] 3.375 4.875 3.250 4.000 6.500 6.500 5.250 5.125 6.500 4.625 2.875 4.125
55 ## [25] 5.750 4.750 5.500 5.625 5.625 5.250 5.625 5.125 6.000 5.000 4.625 5.500
56 ## [37] 5.000 4.250 4.625 7.000 4.375 5.125 5.500 4.500 6.625 5.000 6.000 6.125
57 ## [49] 4.875 4.375 5.000 4.750 5.000 5.750 4.625 3.000 1.750 4.500 6.250 5.000
58 ## [61] 5.500 4.750 4.875 3.625 4.500 4.375 6.125 5.250 4.500 4.125 6.000 4.500
59 ## [73] 5.375 5.625 6.000 6.000 5.750 4.000 5.875 6.000 6.125 5.875 5.750 6.000
60 ## [85] 5.375 6.375 4.125 4.750 6.750 5.500 5.875 6.000 5.750 5.125 4.125 4.875
61 ## [97] 5.000 3.500 5.500 6.250 4.875 5.125 4.625 4.125 3.000 5.000 6.000 4.625
62 ## [109] 4.250 5.375 7.000 5.625 6.000 5.875 7.000 6.750 7.000 6.500 4.625 5.250
63 ## [121] 4.875 6.000 5.000 6.000 6.000 5.750 6.625 5.250 7.000 4.125 4.000 6.500
64 ## [133] 6.875 6.125 7.000 4.750 5.000 6.250 5.625 6.000 5.375 7.000 5.750 5.250
65 ## [145] 6.000 4.250 5.125 5.500 5.625 5.750 5.250 5.375 5.750 5.750 5.000 4.875
66 ## [157] 4.375 4.375 7.000 5.875 4.500 3.875 5.875 6.250 6.750 5.875 6.625 5.875
67 ## [169] 4.000 6.125 5.375 4.625 6.000 4.625 4.875 5.375 5.750 6.500 5.000 6.500
68 ## [181] 4.875 4.500 5.000 5.000 4.500 6.000 7.000 5.875 5.625 4.625 4.500 6.625
69 ## [193] 5.625 5.375 4.250 5.750 5.000 4.375 4.625 4.000
70
71 # 检查计算结果是否一致
72 all(well$flourish1_1 == well$flourish1_2)
73 ## [1] TRUE
74 all(well$flourish1_1 == well$flourish1_3)
75 ## [1] TRUE

```

## 2 离散趋势

### 2.1 全距

$$\max - \min$$

```

1 # flourish1 的全距
2 range(well$flourish1)
3 ## [1] 1.75 7.00

```

四分位

四分位数

## 2.2 四分位距

$$Q75 - Q25$$

```
1 # flourish1 的四分位距
2 Q25 <- quantile(well$flourish1, 0.25)
3 Q75 <- quantile(well$flourish1, 0.75)
4 as.numeric(Q75 - Q25)
5 ## [1] 1.375
```

## 2.3 方差

$$SS = \sum (x_i - M_x)^2 \quad (1)$$

$$s_{biased}^2 = \frac{SS}{biased} \quad (2)$$

$$s_{unbiased}^2 = \frac{SS}{unbiased} \quad (3)$$

在上面的公式中，SS 表示平方和。 $s_{biased}^2$  表示有偏方差， $s_{unbiased}^2$  表示无偏方差。

```
1 # flourish1 的方差
2 # 有偏方差，使用公式
3 sum((well$flourish1 - mean(well$flourish1))^2)/200
4 ## [1] 0.8582047
5 # 无偏方差，使用公式
6 sum((well$flourish1 - mean(well$flourish1))^2)/(200-1)
7 ## [1] 0.8625173
8 # 无偏方差，使用函数 var()
9 var(well$flourish1)
10 ## [1] 0.8625173
```

## 2.4 标准差

$$s_{biased} = \sqrt{s_{biased}^2} \quad (4)$$

$$s_{unbiased} = \sqrt{s_{unbiased}^2} \quad (5)$$

$s_{biased}$  表示有偏标准差， $s_{unbiased}$  表示无偏标准差。

```

1 # flourish1 的标准差
2 # 有偏标准差, 基于有偏方差的公式
3 sqrt(sum((well$flourish1 - mean(well$flourish1))^2)/200)
4 ## [1] 0.9263934
5 # 无偏标准差, 基于无偏方差的公式
6 sqrt(sum((well$flourish1 - mean(well$flourish1))^2)/(200-1))
7 ## [1] 0.9287181
8 # 无偏标准差, 使用函数 sd()
9 sd(well$flourish1)
10 ## [1] 0.9287181

```

### 3 多个变量的描述性统计

```

1 # 使用程序包 psych 中的函数 describe() 进行描述性统计
2 library(psych)
3 describe(well[347:349])
4 ##           vars    n mean   sd median trimmed  mad  min max range  skew kurtosis
5 ## flourish1    1  200 5.32 0.93   5.38   5.34 0.93 1.75   7  5.25 -0.40   0.36
6 ## flourish2    2  200 5.32 1.09   5.38   5.36 0.93 1.00   7  6.00 -0.81   1.54
7 ## flourish3    3  200 5.09 0.97   5.00   5.07 1.11 1.38   7  5.62 -0.15   0.23
8 ##           se
9 ## flourish1 0.07
10 ## flourish2 0.08
11 ## flourish3 0.07

```