



大數據資料處理實務

李水彬

2023-09-01

Chapter 06-02 統計量-百分位數

壽司店近48天銷售盤數

壽司店近48天銷售盤數排序後如下:

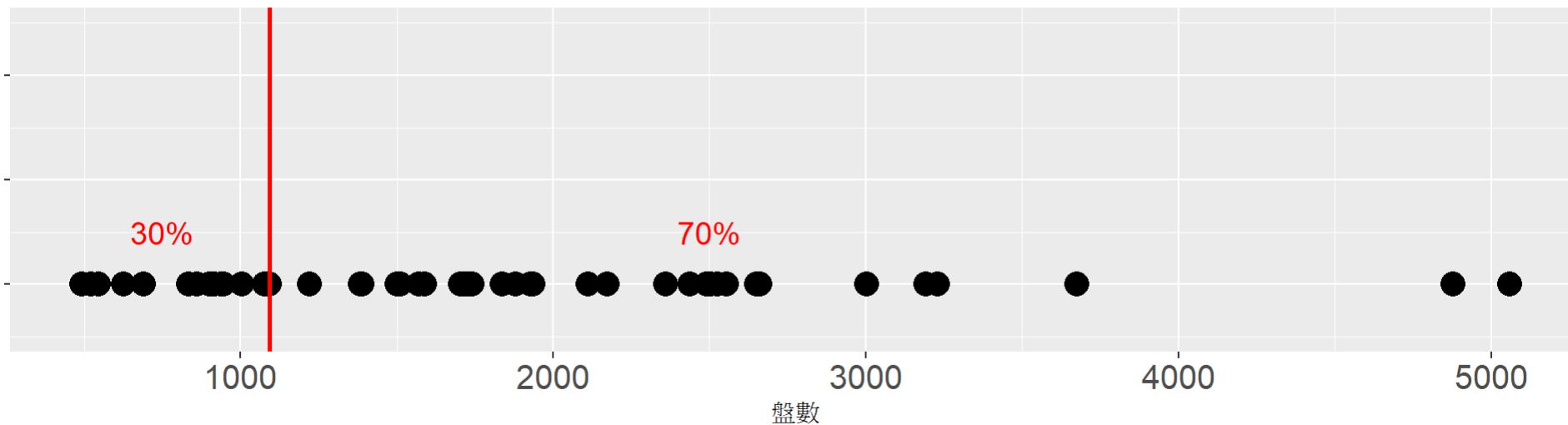
493	522	522	545	627	689	833	861
901	917	939	947	1004	1077	1094	1221
1383	1390	1500	1507	1511	1511	1571	1590
1703	1714	1729	1740	1837	1880	1927	1936
2111	2173	2359	2437	2491	2504	2526	2555
2650	2662	3003	3193	3229	3676	4878	5060

百分位數

將資料分割成100等份, 我們稱這些分割點為百分位數。

令 P_i 代表一組樣本的第 i 個百分位數, 在這組樣本中至少有 $(100 - i)\%$ 的數據大於或等於此數, 也至少有 $i\%$ 的數據小於或等於此數”。

盤數的第30百分位數(紅色線)



十分位數

十分位數是利用 9 個分割點將資料分割成 10 等分。這 9 個分割點以 D_1, D_2, \dots, D_9 表示之。根據定義, 可以用百分位數求十分位數, 即

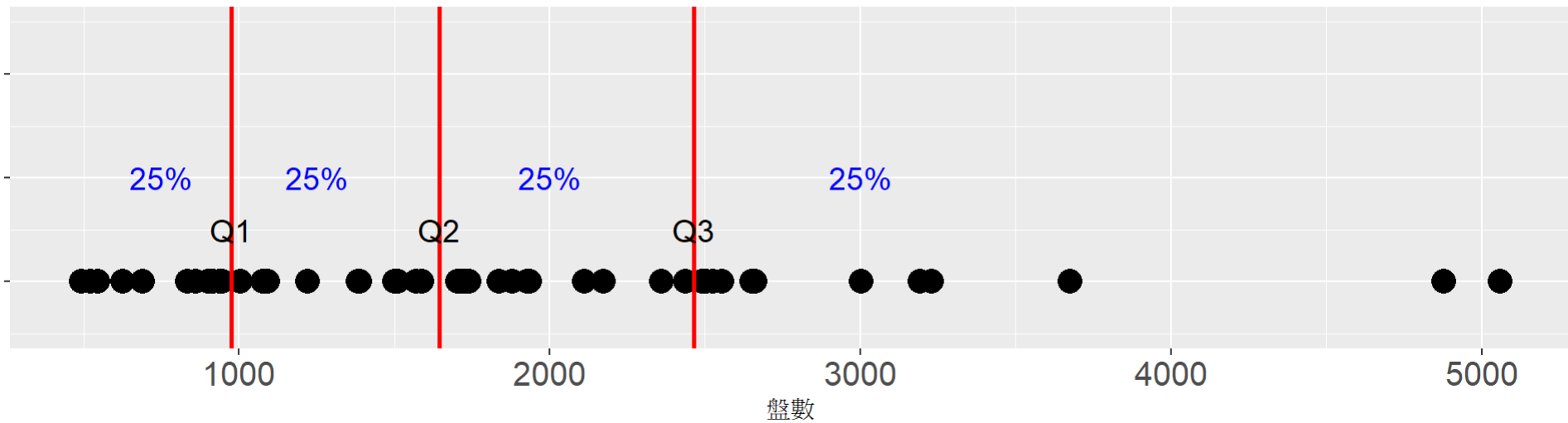
$$D_1 = P_{10}, D_2 = P_{20}, \dots, D_9 = P_{90}$$

四分位數

四分位數是利用 3 個分割點將資料分割成 4 等分。這 3 個分割點為 Q_1, Q_2, Q_3 。根據定義, 可以用百分位數求四分位數, 即

$$Q_1 = P_{25}, Q_2 = P_{50}, Q_3 = P_{75}$$

盤數的第25,50,75百分位數(紅色線)



幼兒的體重的百分位數

- `quantile(data,prob,type=2)`

```
x<-c(18, 17, 18, 16, 17, 15, 15, 16, 18, 15, 14, 16,  
      12, 14, 13, 16, 15, 19, 16, 16)
```

```
mean(x) # 求平均值
```

```
## [1] 15.8
```

```
median(x) # 求中位數
```

```
## [1] 16
```

```
quantile(x,prob=c(0.25,0.40),type=2) #求第25個和第40個百分位數
```

```
## 25% 40%
```

```
## 15.0 15.5
```

範例-百分位數

求晶圓厚度 3 個四分位數、 P_{30} 和平均值。

```
thickness<-c(1379, 1385, 1412, 1425, 1427, 1428, 1448, 1449, 1451,  
             1455, 1459, 1460, 1460, 1461, 1461, 1462, 1465, 1469,  
             1470, 1470, 1472, 1474, 1482, 1483, 1483, 1484, 1488,  
             1495, 1500, 1500, 1501, 1505, 1509, 1512, 1512, 1513,  
             1516, 1517, 1520, 1533, 1543, 1546, 1547, 1548, 1554,  
             1566, 1568, 1569, 1601, 1614)
```

```
quantile(thickness,prob=0.25,type=2) #Q1
```

```
## 25%
```

```
## 1460
```


範例-百分位數

```
quantile(thickness,prob=0.50,type=2) #Q2
```

```
##      50%  
## 1483.5
```

```
quantile(thickness,prob=0.75,type=2) #Q3
```

```
##      75%  
## 1517
```

```
quantile(thickness,prob=0.30,type=2) #P30
```

```
##      30%  
## 1461.5
```

範例-百分位數

- 可以同時計算多個百分位數

```
qq<-quantile(thickness, prob=c(0.25,0.5,0.75,0.30),type=2)#多個百分位數  
qq
```

```
##      25%      50%      75%      30%  
## 1460.0 1483.5 1517.0 1461.5
```

課堂練習 15

求新生兒體重的 3 個四分位數和 P_{45} 。

2.03, 2.21, 2.29, 2.43, 2.44, 2.51, 2.54, 2.54, 2.55, 2.61, 2.62, 2.63, 2.64, 2.64, 2.66,
2.69, 2.70, 2.71, 2.72, 2.80, 2.80, 2.82, 2.87, 2.88, 2.88, 2.90, 2.92, 2.93, 2.96, 3.02,
3.02, 3.03, 3.04, 3.04, 3.05, 3.08, 3.12, 3.15, 3.16, 3.16, 3.17, 3.21, 3.22, 3.23, 3.24,
3.27, 3.30, 3.33, 3.37, 3.38, 3.40, 3.45, 3.46, 3.46, 3.49, 3.53, 3.55, 3.56, 3.61, 3.68,
3.70, 3.71, 3.72, 3.88, 4.20

課堂練習 16

請讀取mobilephone.csv，求身高和體重的 P_5 和 P_{95} 。

- 註: 共有152筆資料，部分如下。

Height	Weight	CurrentMobile	DesiredMobile
173	62.0	Nokia	Apple
165	75.0	Samsung	ASUS
158	52.0	Apple	Apple
171	66.1	Samsung	Samsung
173	67.0	Samsung	Apple
