

class06

Steven Gan

2022-10-14

Table of contents

Input the Student Grade Sheet	1
Q1	1
Q2	2
Q3	2
Q4	2
Method 1	2
Method 2	3
Comparing two methods	3

Input the Student Grade Sheet

```
student_grade <- read.csv(file = 'https://tinyurl.com/gradeinput', row.names = 1)
student_grade[is.na(student_grade)] <- 0
```

Q1

```
grade <- function (student) {
  student[is.na(student)] <- 0
  sort <- sort(student, na.last = FALSE)
  return(mean(sort[-1]))
}
```

```
}
```

```
overall_stu <- apply(student_grade, 1, grade)
overall_stu
```

```
student-1 student-2 student-3 student-4 student-5 student-6 student-7
      91.75      82.50      84.25      84.25      88.25      89.00      94.00
student-8 student-9 student-10 student-11 student-12 student-13 student-14
      93.75      87.75      79.00      86.00      91.75      92.25      87.75
student-15 student-16 student-17 student-18 student-19 student-20
      78.75      89.50      88.00      94.50      82.75      82.75
```

Q2

```
names(sort(overall_stu, decreasing = TRUE)[1])
```

```
[1] "student-18"
```

Q3

```
overall_hw <- apply(student_grade, 2, grade)
names(sort(overall_hw)[1])
```

```
[1] "hw2"
```

Q4

Method 1

y = x correlation

```
#Calculate the deviation in terms of the physical distances between student grades in each
diff <- (student_grade - overall_stu) ^ 2
dev <- sqrt(colSums(diff) / (ncol(diff) - 1))
```

```
names(sort(dev)[1])
```

```
[1] "hw1"
```

Method 2

$y = ax + b$ correlation

```
#Calculate the correlation coefficient by Pearson method (correlation based on linearity)
cor <- apply(student_grade, 2, cor, y = overall_stu)
names(sort(cor, decreasing = TRUE)[1])
```

```
[1] "hw5"
```

Comparing two methods

```
library(ggplot2)
library(patchwork)

com_grade <- cbind(student_grade, overall_stu)

#Student-15's score in homework 5 is removed (Score = 0)
scale1 = c(75, 100)
scale2 = c(60, 100)

p1 <- ggplot(com_grade) +
  aes(x = overall_stu, y = hw1) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(x = "Overall Student Grades", y = "Homework 1",
       title = "Method 1 Correlation") +
  geom_abline(intercept = 0, slope = 1,
              color = "red", linetype = "dashed", size = 0.5) +
  scale_x_continuous(limits = scale1) +
  scale_y_continuous(limits = scale1)

p2 <- ggplot(com_grade) +
  aes(x = overall_stu, y = hw5) +
  geom_point() +
```

```
geom_smooth(method = "lm") +
labs(x = "Overall Student Grades", y = "Homework 5",
     title = "Method 2 Correlation") +
geom_abline(intercept = 0, slope = 1,
            color = "red", linetype = "dashed", size = 0.5) +
scale_x_continuous(limits = scale2) +
scale_y_continuous(limits = scale2)
```

p1 | p2

```
`geom_smooth()` using formula 'y ~ x'
`geom_smooth()` using formula 'y ~ x'
```

Warning: Removed 1 rows containing non-finite values (stat_smooth).

Warning: Removed 1 rows containing missing values (geom_point).

