Today, I am going to cover cross effects.

So far, we only consider main effects such as ReadingScore and WritingScore. We may want to build a model based on the following **math formula**:

$$MathScore = \beta_0 + \beta_1 * ReadingScore + \beta_2 * WritingScore + \beta 3 * ReadingScore * WritingScore$$

We already know how to include the main effects in the R formula. How can we include cross effect ReadingScore*WritingScore in the R formula?

We use the **:** operator to do it. It means include the **interaction between these variables**.

Let's run the model:

```
lm.result3 <- lm(MathScore ~ ReadingScore + WritingScore + ReadingScore:Writi
ngScore, data= StudentsPerformance)
coef(lm.result3)
```

Since both ReadingScore and WritingScore are **continuous variables**, so is the cross effect of reading score and writing score. The corresponding math formula is

$$MathScore = 3.7614864 + 0.6617578 * ReadingScore + 0.3076971 * WritingScore + -8.9160555 \times 10^{-4} * ReadingScore: WritingScore$$

Typically we use the **:** operator in R formula with something like

$$target \sim var1 + var2 + var3 + var1: var2 + var1: var3 + var2: var3$$

There is another operator * in R that can be used in the R formula to include the cross effects as well.

* in R formula means **include these variables and the interactions between them**. Note here * does not mean multiplication in the R formula.

We can use the following R formula

$$MathScore \sim ReadingScore * WritingScore$$

The R formula $ReadingScore * WritingScore$ above means to include the **variables/main effects** of ReadingScore, WritingScore and the **interactions between them**, which is ReadingScore : WritingScore (R formula) or ReadingScore * WritingScore (math formula).

Let's run the model:

```
lm.result4 <- lm(MathScore ~ ReadingScore * WritingScore, data= StudentsPerfo
rmance)
summary(lm.result4)
```

It produces the same results as before since $MathScore \sim ReadingScore * WritingScore$ is equivalent to $MathScore \sim ReadingScore + WritingScore + ReadingScore : WritingScore$.

If we have a R formula:

$$MathScore \sim ReadingScore * WritingScore * TestPreparationCourse$$

By the definition of *operator in the R formula, it includes the following items:

- Main effects: ReadingScore, WritingScore, and TestPreparationCourse
- interactions between them
  - Cross effects between two variables (There are $C_3^2 = 3$ cases.)
    - ReadingScore : WritingScore
    - ReadingScore : TestPreparationCourse
    - WritingScore : TestPreparationCourse
  - Cross effects between three variables (There is $C_3^3 = 1$ case.)
    - ReadingScore : WritingScore : TestPreparationCourse

Therefor the R formula

$$MathScore \sim ReadingScore * WritingScore * TestPreparationCourse$$

is equivalent to the following R formula:

$$\begin{aligned} MathScore \sim & ReadingScore + WritingScore + TestPreparationCourse \\ & + ReadingScore : WritingScore + ReadingScore : TestPreparationCourse + WritingScore : TestPreparationCourse \\ & + ReadingScore : WritingScore : TestPreparationCourse \end{aligned}$$

Let's run the R model based on the R formula above.

```
lm.result5 <- lm(MathScore ~ ReadingScore * WritingScore*TestPreparationCours
e, data= StudentsPerformance)
summary(lm.result5)
```

Let's write down the corresponding math formula. We notice that Reading Score and Writing score are continuous variables. But TestPreparationCourse is a factor with two cases: none and completed. We have to use a piece-wise function with two parts to denote it. When we look at the output of the summary function, we can only find the TestPreparationCoursenone which denotes the none case. Therefore, the completed cases is the base line hidden in the y-intercept.

$$MathScore = \begin{cases} 16.5047877 + 0.4940647 * ReadingScore + 0.2926499 * WritingScore + 0.0020214 * ReadingScore * WritingScore; & \text{TestPreparationCourse} = \text{completed} \\ 16.5047877 + 0.4940647 * ReadingScore + 0.0701232 * WritingScore \\ \quad + -16.0182093 + 0.0020214 * ReadingScore * WritingScore + 0.1407791 * ReadingScore \\ \quad + 0.3839524 * WritingScore + -0.0037856 * ReadingScore * WritingScore; & \text{TestPreparationCourse} = \text{none} \end{cases}$$

We combine the like terms and obtain the following formula:

$$MathScore = \begin{cases} 16.5047877 + 0.4940647 * ReadingScore + 0.2926499 * WritingScore + 0.0020214 * ReadingScore * WritingScore; & \text{TestPreparationCourse} = \text{completed} \\ 0.4865784 + 0.6348438 * ReadingScore + 0.4540757 * WritingScore \\ \quad + -0.0017642 * ReadingScore * WritingScore; & \text{TestPreparationCourse} = \text{none} \end{cases}$$