Interaction of Covariates in Statistics

a_statistician

March 24, 2017

Imagine this situation. After one year of hard work, it is time to increase the salary. Because there was no special situation last year, the boss wants to make the fair decision of increasing $2,000 for each employee. Staff A knows the average salary is $50,000 and his salary is $80,000. It seems other people get an increase of 4% and he gets an increase of only 2.5%. Staff A discusses this situation with his boss and his boss thinks staff A's idea is correct. The boss changes the decision to increase salary by 3% for each employee. Now the staff B think it is unfair because the average increase is $1,500 = $50,000 x 3% and her increase is $900 = $30,000 (her salary last year) x 3%. In this situation, the fairness is based on what measurement will be used, the absolute value or relative value.

In statistics, if the effect of covariate X_1 on response variable Y depends on the value of the another covariate X_2, we say that x_1 and X_2 have interaction. Then no interaction between covariates X_1 and X_2 on the response variable Y means that the effect of the covariate X_1 does not change along the values of covariate X_2. In the salary example, fairness depends on what measure will be use to describe the increase of salary. Similarly, the interaction depends on how the effect on response variable Y is measured.

Here we use the response variable Y following binomial distribution as an example to explain this situation. Suppose both covariates X_1 and X_2 are dichotomous. So there are four different combinations of X_1 and X_2, i.e., ($X_1 = 0$, X_2=0), ($X_1 = 0$, X_2=1), ($X_1 = 1$, X_2=0), and ($X_1 = 1$, $X_2 = 1$). Corresponding to four combinations of covariate values, there are 4 different binomial distributions. Let $\pi_{ij} = \Pr\{Y = 1|X_1 = i, X_2 = j\}, i, j = 0, 1$ be the parameters of the four binomial distributions and we call them rates.

For given $X_2 = j$, we have three common measurements to describe the effect of covariate $X_1 = 1$ vs $X_1 = 0$ on response variable Y.

1. Difference of rates: $D_j = \pi_{1j} - \pi_{0j}$

2. Rate ratio: $R_j = \dfrac{\pi_{1j}}{\pi_{0j}}$

3. Odds Ratio: $O_j = \dfrac{\pi_{1j}/(1-\pi_{1j})}{\pi_{0j}/(1-\pi_{0j)}}$

As we described before, no interaction between covariates X_1 and X_2 on response variable Y means the effect of X_1 on Y is constant across the different values of X_2. In the given binomial distribution, if $D_0 = D_1$, we say that X_1 and X_2 have no interaction when the effect is measured by the difference of rates; if $R_0 = R_1$, we say that X_1 and X_2 have no interaction when the effect is measured by the rate ratio; if $O_0 = O_1$, we say that X_1 and X_2 have no interaction when the effect is measured by the odds ratio. But we need to notice that for the given $\pi_{ij}, i, j = 0, 1$, if there is no interaction between X_1 and X_2 under one measurement (for example, odds ratio), there is an interaction between X_1 and X_2 under the other two measurement (for example, difference of rate and rate ratio).

Following table presents four sets of π_{ij} to illustrate this phenomenon.

π_{00}	π_{01}	π_{10}	π_{11}	D_0	D_1	R_0	R_1	O_0	O_1
0.1	0.2	0.3	0.4	0.2	0.2	3.0	2.0	3.857	2.667
0.1	0.2	0.3	0.6	0.2	0.4	3.0	3.0	3.857	6.000
0.1	0.2	0.3	0.491	0.2	0.291	3.0	2.454	3.857	3.857
0.1	0.2	0.3	0.7	0.2	0.5	3.0	3.5	3.857	9.333

For the first line, there is no interaction based on difference of rates, but there is interaction on the other measurements, The second line indicates that no interaction exists on rate ratio, but interactions exist when the other two measurements are used. The third line is used for the situation of no interaction on odds ratio, The last line describe the situation that interaction exists under any of three measurements, but with different values.

In summary, when we write/read about interaction between covariate X_1 and X_2 on response variable Y, we need to understand what measurement is used to measure the effect.