Class notes 6

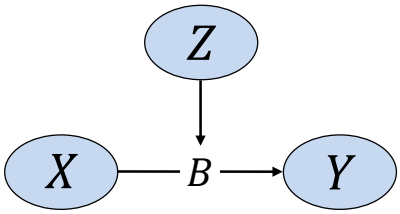
Moderation analysis

Continuous moderator

A popular and rather useful application of regression modeling is *moderation* analysis. First, consider the simple regression model given by

$$Y = B_0 + B_1 X + \varepsilon$$

Suppose we expand the model by introducing a so-called moderator variable Z. The aim is to specify a relationship (as measured by some parameter) between Y and X that *depends* on the variable Z. The figure below illustrates how



To accomplish this, we use the following specification

$$Y = B_0 + B_1 X + B_2 Z + B_3 Z \cdot X + \varepsilon$$

This specification is useful when implementing the model. However, it becomes somewhat to interpret the model if we rewrite it

$$Y = (B_0 + B_2 Z) + (B_1 + B_3 Z) X + \varepsilon$$

From this formulation, it is clear that the regression line depends on the value of Z (and of course the model parameters).

Example (taken from Zhang and Wang, 2017)

To make this clearer, we consider the following example. Suppose we want to investigate the relation between stress and depression

$$Depress = B_0 + B_1 Stress + \varepsilon$$

It has long been known that social support from family, friends etc. plays a critical role in well-being. We therefore extend the model to include social support as a moderating variable

$$Depress = B_0 + B_1 Stress + B_2 Support + B_3 Stress \cdot Support + \varepsilon$$

We estimate the model and obtain the following results

$$\widehat{Depress} = 29.26 + 2.00 \cdot Stress - 0.24 \cdot Support - 0.39 \cdot Stress \cdot Support$$

Re-writing the model for ease of interpretation gives

$$\widehat{Depress} = (29.26 - 0.24 \cdot Support) + (2.00 - 0.39 \cdot Support) \cdot Stress$$

Let us evaluate the level of depression for different values of social support. It is common to choose levels of the moderator one standard deviation below/above the mean of the moderator. Suppose that $Support$ has a sample mean of is 5.37 and a sample standard deviation of 2.81.

Case 1 (low support): $Support = 5.37 - 2.81 = 2.56$, which leads to

$$\begin{aligned}\widehat{Depress} &= (29.26 - 0.24 \cdot 2.56) + (2.00 - 0.39 \cdot 2.56) \cdot Stress \\ &= 28.65 + 1.00 \cdot Stress\end{aligned}$$

Case 2 (high support): $Support = 5.37 + 2.81 = 8.18$, which leads to

$$\begin{aligned}\widehat{Depress} &= (29.26 - 0.24 \cdot 8.18) + (2.00 - 0.39 \cdot 8.18) \cdot Stress \\ &= 27.30 - 1.19 \cdot Stress\end{aligned}$$

We have the following conclusion. For the low support condition, we find that increasing stress leads to higher levels of depression (since the slope coefficient is positive). For the high support condition, we find that increasing stress leads to lower levels of depression (since slope is negative).