TEXAS TECH UNIVERSITY™

STATS CAMP

statscamp.org

Lecture 7: Basic Moderation

Kyle M. Lang

Institute for Measurement, Methodology, Analysis & Policy
Texas Tech University
Lubbock, TX

2016 Stats Camp

Outline

- Review the moderation hypothesis
- Doing basic moderation analysis
- Visualizing the moderation (a little)
- Probing the moderation (also a little)

Intuition

So far we've been discussing *mediation*

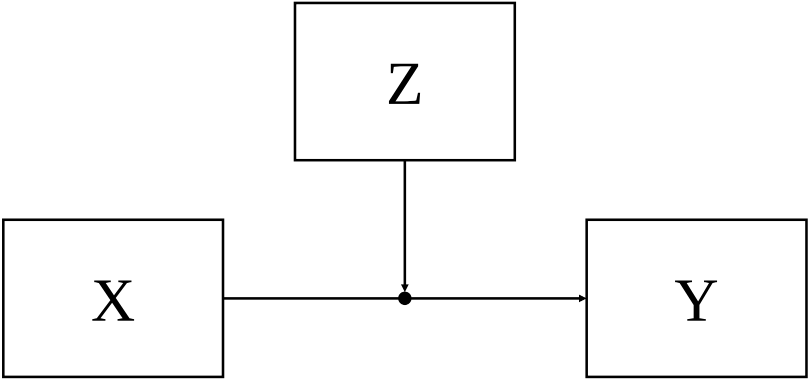
- Mediation allows us to ask *how* one variable (X) affects another variable (Y).
  - Namely, through the intermediary influence of a third variable (M).

Now, we're stepping into the realm of *moderation*

- Moderation allows us to ask *when* one variable (X) affects another variable (Y).
  - Here, we're considering the effect of X on Y conditional on certain levels of a third variable Z.

Conceptual Diagram

We can diagrammatically represent the above intuition with:



Equations

In simple additive MLR, we might have the following equation:

$$Y = \alpha + \beta_1 X + \beta_2 Z + e_i \tag{1}$$

This additive equation assumes that X and Z are independent predictors of Y.

When X and Z are independent predictors, the following points are true:

- X and Z *can* be correlated
- β_1 and β_2 are *partial* regression coefficients
- The effect of X on Y is the same at **all levels** of Z, and the effect of Z on Y is the same at **all levels** of X

Equations

When testing moderation, we hypothesize that the effect of X on Y in Equation 1 varies as a function of Z.

We can represent this concept with the following equation:

$$Y = \alpha + f(Z)X + \beta_2 Z + e_i \tag{2}$$

Equations

When testing moderation, we hypothesize that the effect of X on Y in Equation 1 varies as a function of Z.

We can represent this concept with the following equation:

$$Y = \alpha + f(Z)X + \beta_2 Z + e_i \tag{2}$$

If we assume that Z linearly affects the relationship between X and Y, then we can take:

$$f(Z) = \beta_1 + \beta_3 Z \tag{3}$$

Equations

When testing moderation, we hypothesize that the effect of X on Y in Equation 1 varies as a function of Z.

We can represent this concept with the following equation:

$$Y = \alpha + f(Z)X + \beta_2 Z + e_i \tag{2}$$

If we assume that Z linearly affects the relationship between X and Y, then we can take:

$$f(Z) = \beta_1 + \beta_3 Z \tag{3}$$

Which, after substitution, leads to:

$$Y = \alpha + (\beta_1 + \beta_3 Z)X + \beta_2 Z + e_i \tag{4}$$

Equations

When testing moderation, we hypothesize that the effect of X on Y in Equation 1 varies as a function of Z.

We can represent this concept with the following equation:

$$Y = \alpha + f(Z)X + \beta_2 Z + e_i \tag{2}$$

If we assume that Z linearly affects the relationship between X and Y, then we can take:

$$f(Z) = \beta_1 + \beta_3 Z \tag{3}$$

Which, after substitution, leads to:

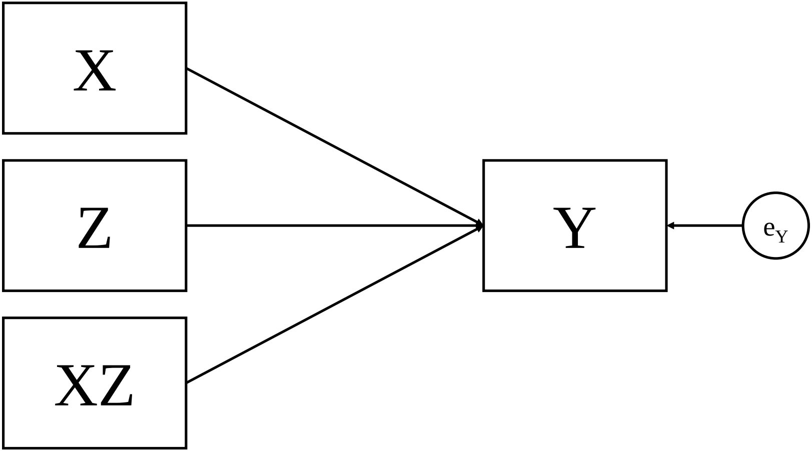
$$Y = \alpha + (\beta_1 + \beta_3 Z)X + \beta_2 Z + e_i \tag{4}$$

Which, after distributing X and reordering terms, becomes:

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ + e_i \tag{5}$$

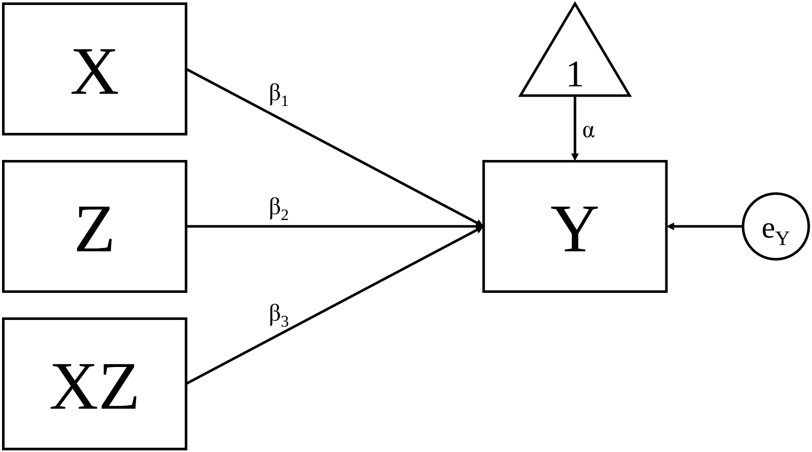
Analytical Model

We can diagrammatically represent the analytical model we'll actually be fitting with:



Analytical Model

By adding the appropriate path labels, we get:



Testing Moderation

This is the equation we'll be working with:

$$\boxed{Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ + e_i}$$

Or, after fitting the above to some data:

$$\boxed{\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X + \hat{\beta}_2 Z + \hat{\beta}_3 XZ}$$

To test for significant moderation, we simply need to see if $\hat{\beta}_3$ is significantly different from zero.

We do so using simple linear regression modeling.

Example

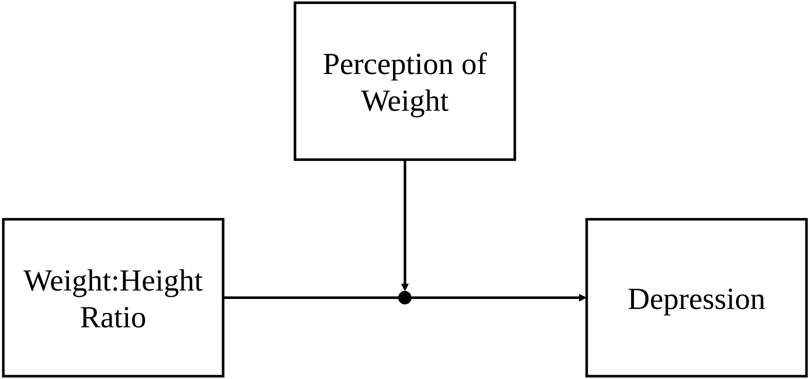
Data from the *National Longitudinal Survey of Youth*

We suspect that participants' weight to height ratio is predictive of their levels of depression.

We further suspect that this effect may be differentially expressed depending on how the participants perceive their own weight.

Example

This is the conceptual diagram for the model we'll fit:



Example

```r
## Focal Effect:
out1 <- lm(depress1 ~ ratio1, data = dat1)
summary(out1)
```

```
Call:
lm(formula = depress1 ~ ratio1, data = dat1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1229 -0.2712  0.1148  0.3452  0.9866

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.94773    0.02555 115.360  < 2e-16 ***
ratio1       0.05095    0.01081   4.715 2.45e-06 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05
       .    0.1          1

Residual standard error: 0.5116 on 8982 degrees of freedom
Multiple R^2:  0.002469, Adjusted R^2:  0.002358
F-statistic: 22.23 on 1 and 8982 DF,  p-value: 2.45e-06
```

Example

```
## Additive Model:
out2 <- lm(depress1 ~ ratio1 + perception1, data = dat1)
summary(out2)
```

```
Call:
lm(formula = depress1 ~ ratio1 + perception1, data = dat1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.09842 -0.28788  0.09376  0.34951  1.08428

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.047115   0.028211 108.011  < 2e-16 ***
ratio1       0.102258   0.012460   8.207 2.59e-16 ***
perception1 -0.068113   0.008328  -8.178 3.27e-16 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05
        .   0.1       1

Residual standard error: 0.5097 on 8981 degrees of freedom
Multiple R^2:  0.009843, Adjusted R^2:  0.009623
```

Example

```
F-statistic: 44.64 on 2 and 8981 DF,  p-value: < 2.2e-16
```

Example

```
## Moderated Model:
out3 <- lm(depress1 ~ ratio1*perception1, data = dat1)
summary(out3)
```

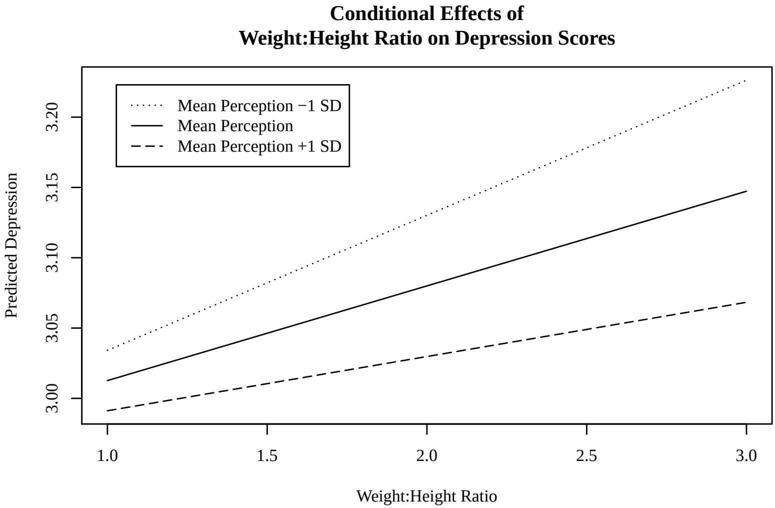
```
Call:
lm(formula = depress1 ~ ratio1 * perception1, data = dat1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1218 -0.2830  0.0881  0.3515  1.0835

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.41374    0.10375  23.265  < 2e-16 ***
ratio1              0.38126    0.04571   8.341  < 2e-16 ***
perception1         0.11083    0.02941   3.768 0.000165 ***
ratio1:perception1 -0.07702    0.01214  -6.343 2.37e-10 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05
        .    0.1          1

Residual standard error: 0.5086 on 8980 degrees of freedom
```

Example

```
Multiple R^2:  0.01426,  Adjusted R^2:  0.01393
F-statistic:  43.3 on 3 and 8980 DF,  p-value: < 2.2e-16
```

Visualizing the Interaction

We can get a better idea of the patterns of moderation by plotting the focal effect at conditional values of the moderator:



Probing the Interaction

A significant estimate of β_3 tells us that the effect of X on Y depends on the level of Z, but nothing more.

The plot on the previous slide gives a descriptive illustration of the pattern, but does not support statistical inference.

- The three conditional effects we plotted look different, but we cannot say that they differ in any meaningful way by only the plot and $\hat{\beta}_3$.

This is the purpose of *probing* the interaction.

- Try to isolate areas of Z's distribution in which $\hat{\beta}_3$ is significant and areas where it is not.

Probing the Interaction

The most popular approach to probing the interaction is the *pick-a-point* approach AKA *simple slopes analysis* or *spotlight analysis*.

The pick-a-point approach tests if the slopes of the conditional effects plotted above are significantly different from zero.

To do so, pick-a-point tests the significance of *simple slopes*.

Simple Slopes

Recall the derivation of our moderated equation:

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ + e_i$$

We can reverse the process by factoring out X and reordering terms to get back to:

$$Y = \alpha + (\beta_1 + \beta_3 Z)X + \beta_2 Z + e_i$$

Where $f(Z) = \beta_1 + \beta_3 Z$ is the linear function that shows how the relationship between X and Y changes as a function of Z.

$f(Z)$ is actually our *simple slope*.

- By plugging different values of Z into $f(Z)$, we get the slope of the conditional effect of X on Y at the chosen value of Z.

Significance Testing of Simple Slopes

The conditional values of Z used to define the simple slopes in the pick-a-point approach are totally arbitrary

- The most popular choice is: $\{(\bar{Z} - SD_Z), \bar{Z}, (\bar{Z} + SD_Z)\}$
- You could also use interesting percentiles of Z's distribution

The standard error of a simple slope is given by:

$$SE_{SS} = \sqrt{SE_{\beta_1}^2 + 2Z \cdot \text{COV}(\beta_1, \beta_3) + Z^2 SE_{\beta_3}^2} \tag{6}$$

So, you can test the significance of a simple slope by constructing a Wald statistic or confidence interval using SE_{SS}:

$$Wald_{SS} = \frac{\hat{f}(Z)}{SE_{SS}}$$

$$95\% CI_{SS} = \hat{f}(Z) \pm 1.96 \cdot SE_{SS}$$

Example

```
## Specify function to compute simple slopes:
getSS <- function(z, lmOut) {
    tmp <- coef(lmOut)
    tmp[2] + tmp[4]*z
}
##
## Specify function to compute SE for simple slopes:
getSE <- function(z, lmOut) {
    tmp <- vcov(lmOut)
    varB1 <- tmp[2, 2]
    varB3 <- tmp[4, 4]
    covB13 <- tmp[4, 2]

    sqrt(varB1 + 2 * z * covB13 + z^2 * varB3)
}
```

Example

```r
## Compute vector of simple slopes:
ssVec <- sapply(c(meanZ - sdZ,
                  meanZ,
                  meanZ + sdZ),
                FUN = getSS,
                lmOut = out3)
##
## Compute vector of SEs for simple slopes:
seVec <- sapply(c(meanZ - sdZ,
                  meanZ,
                  meanZ + sdZ),
                FUN = getSE,
                lmOut = out3)
```

Example

```r
## Compute Wald Statistics:
waldVec <- ssVec / seVec
names(waldVec) <- c("Mean - SD", "Mean", "Mean + SD")
waldVec
```

```
Mean - SD      Mean Mean + SD
10.189798 10.021285  5.918862
```

```r
##
## Compute CIs:
ciMat <- cbind(ssVec - 1.96 * seVec,
               ssVec + 1.96 * seVec)
rownames(ciMat) <- c("Mean - SD", "Mean", "Mean + SD")
colnames(ciMat) <- c("LB", "UB")
ciMat
```

```
                  LB        UB
Mean - SD 0.15533968 0.2293307
Mean      0.10841462 0.1611339
Mean + SD 0.05164454 0.1027821
```