

Moderation, Interactions, and Polynomials

Statistics & Methodology Lecture 8

TILBURG
UNIVERSITY



Understanding
Society

Kyle M. Lang

Department of Methodology & Statistics
Tilburg University

Outline

1. Testing for moderation with MLR models
2. Polynomial regression



Moderation

So far we've been discussing *additive models*.

- Additive models allow us to examine the partial effects of several predictors on some outcome.
 - The effect of one predictor does not change based on the values of other predictors.

Now, we'll discuss *moderation*.

- Moderation allows us to ask *when* one variable, X , affects another variable, Y .
 - We're considering the conditional effects of X on Y given certain levels of a third variable Z .

Equations

In additive MLR, we might have the following equation:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$$

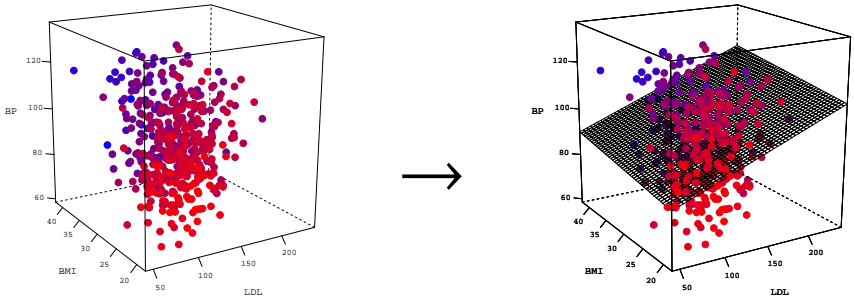
This additive equation assumes that X and Z are independent predictors of Y .

When X and Z are independent predictors, the following 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 .

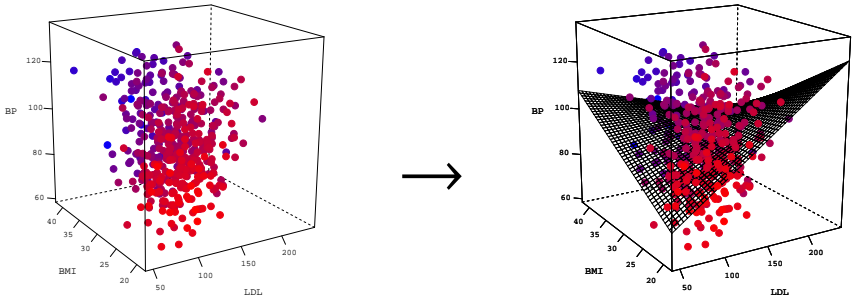
Additive Regression

The effect of X on Y is the same at **all levels** of Z .



Moderated Regression

The effect of X on Y varies as a function of Z .



Equations

The following derivation is adapted from Hayes (2017).

- When testing moderation, we hypothesize that the effect of X on Y varies as a function of Z .
- We can represent this concept with the following equation:

$$Y = \beta_0 + f(Z)X + \beta_2Z + \varepsilon \quad (1)$$



Equations

The following derivation is adapted from Hayes (2017).

- When testing moderation, we hypothesize that the effect of X on Y varies as a function of Z .
- We can represent this concept with the following equation:

$$Y = \beta_0 + f(Z)X + \beta_2Z + \varepsilon \quad (1)$$

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

$$f(Z) = \beta_1 + \beta_3Z \quad (2)$$

Equations

- Substituting Equation 2 into Equation 1 leads to:

$$Y = \beta_0 + (\beta_1 + \beta_3 Z)X + \beta_2 Z + \varepsilon$$



Equations

- Substituting Equation 2 into Equation 1 leads to:

$$Y = \beta_0 + (\beta_1 + \beta_3 Z)X + \beta_2 Z + \varepsilon$$

- Which, after distributing X and reordering terms, becomes:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$$

Testing Moderation

Now, we have an estimable regression model that quantifies the linear moderation we hypothesized.

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$$

- To test for significant moderation, we simply need to test the significance of the interaction term, XZ .
 - Check if $\hat{\beta}_3$ is significantly different from zero.



Interpretation

Given the following equation:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 Z + \hat{\beta}_3 XZ + \hat{\varepsilon}$$

- $\hat{\beta}_3$ quantifies the effect of Z on the focal effect (the $X \rightarrow Y$ effect).
 - For a unit change in Z , $\hat{\beta}_3$ is the expected change in the effect of X on Y .
- $\hat{\beta}_1$ and $\hat{\beta}_2$ are *conditional effects*.
 - Interpreted where the other predictor is zero.
 - For a unit change in X , $\hat{\beta}_1$ is the expected change in Y , when $Z = 0$.
 - For a unit change in Z , $\hat{\beta}_2$ is the expected change in Y , when $X = 0$.

Example

Still looking at the *diabetes* dataset.

- We suspect that patients' BMIs are predictive of their average blood pressure.
- We further suspect that this effect may be differentially expressed depending on the patients' LDL levels.



Example

```
## Focal Effect:
out0 <- lm(bp ~ bmi, data = dDat)
partSummary(out0, -c(1, 2))

## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   61.9973      3.6659   16.91  <2e-16
## bmi           1.2379      0.1371    9.03  <2e-16
##
## Residual standard error: 12.72 on 440 degrees of freedom
## Multiple R-squared:  0.1563, Adjusted R-squared:  0.1544
## F-statistic: 81.54 on 1 and 440 DF,  p-value: < 2.2e-16
```

Example

```
## Additive Model:
out1 <- lm(bp ~ bmi + ldl, data = dDat)
partSummary(out1, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  59.26577      3.91281  15.147 < 2e-16
## bmi          1.16567      0.14156   8.235 2.08e-15
## ldl          0.04016      0.02056   1.953  0.0515
##
## Residual standard error: 12.68 on 439 degrees of freedom
## Multiple R-squared:  0.1636, Adjusted R-squared:  0.1598
## F-statistic: 42.94 on 2 and 439 DF,  p-value: < 2.2e-16
```

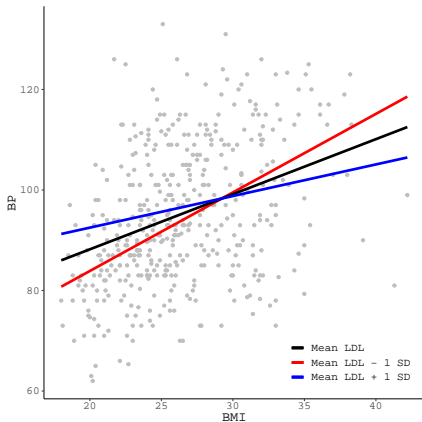
Example

```
## Moderated Model:
out2 <- lm(bp ~ bmi * ldl, data = dDat)
partSummary(out2, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.480616  14.291677   1.013 0.311514
## bmi         2.867825   0.541312   5.298 1.86e-07
## ldl         0.448771   0.127160   3.529 0.000461
## bmi:ldl     -0.015352   0.004716  -3.255 0.001221
##
## Residual standard error: 12.54 on 438 degrees of freedom
## Multiple R-squared:  0.1834, Adjusted R-squared:  0.1778
## F-statistic: 32.78 on 3 and 438 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 much about how they differ with 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 $X \rightarrow Y$ effect is significant and areas where it is not.

Probing the Interaction

The most popular method of probing interactions is to do a so-called *simple slopes* analysis.

- Pick-a-point approach
- Spotlight analysis

In simple slopes analysis, we test if the slopes of the conditional effects plotted above are significantly different from zero.

- To do so, we test the significance of *simple slopes*.



Simple Slopes

Recall the derivation of our moderated equation:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$$

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

$$Y = \beta_0 + (\beta_1 + \beta_3 Z)X + \beta_2 Z + \varepsilon$$

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 the *simple slope*.

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

Significance Testing of Simple Slopes

The values of Z used to define the simple slopes are arbitrary.

- The most common 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:

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

So, you can test the significance of a simple slope by constructing a Wald statistic or confidence interval using $\hat{f}(Z)$ and $SE_{f(Z)}$:

$$t = \frac{\hat{f}(Z)}{SE_{f(Z)}}, \quad CI = \hat{f}(Z) \pm t_{crit} \times SE_{f(Z)}$$

Interaction Probing

When probed the interaction with simple slopes analysis:

1. Choose interesting values of the moderator, Z .
2. Check the significance of the focal effect, $X \rightarrow Y$, at the Z values chosen in Step 1.
3. Use the results from Step 2 to get an idea of where in Z 's distribution the focal effect is or is not significant.



Interaction Probing

When probed the interaction with simple slopes analysis:

1. Choose interesting values of the moderator, Z .
2. Check the significance of the focal effect, $X \rightarrow Y$, at the Z values chosen in Step 1.
3. Use the results from Step 2 to get an idea of where in Z 's distribution the focal effect is or is not significant.

We saw manual calculations for the the quantities needed, but there is a simpler way:

- CENTERING

Centering

Centering shifts the scale of a variable up or down by subtracting a constant (e.g., the variable's mean) from each of its observations.

- The most familiar form of center is *mean centering*.
- We can center on any value.
 - When probing interactions, we can center Z on the interesting values we choose to define the simple slopes.
 - Due to the interpretation of conditional effects, running the model with Z centered on a specific value automatically provides a test of the simple slope for that value of Z .

Probing via Centering

Say we want to do a simple slopes analysis to test the conditional effect of X on Y at three levels of $Z = \{Z_1, Z_2, Z_3\}$.

- All we need to do is fit the following three models:

$$Y = \beta_0 + \beta_1 X + \beta_2(Z - Z_1) + \beta_3 X(Z - Z_1) + \varepsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_2(Z - Z_2) + \beta_3 X(Z - Z_2) + \varepsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_2(Z - Z_3) + \beta_3 X(Z - Z_3) + \varepsilon$$

Probing via Centering

Say we want to do a simple slopes analysis to test the conditional effect of X on Y at three levels of $Z = \{Z_1, Z_2, Z_3\}$.

- All we need to do is fit the following three models:

$$Y = \beta_0 + \beta_1 X + \beta_2(Z - Z_1) + \beta_3 X(Z - Z_1) + \varepsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_2(Z - Z_2) + \beta_3 X(Z - Z_2) + \varepsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_2(Z - Z_3) + \beta_3 X(Z - Z_3) + \varepsilon$$

- The default output for $\hat{\beta}_1$ provides tests of the simple slopes.

Example

Create transformed predictors by centering on critical values of the moderator, Z_{LDL} .

```
zMean      <- mean(dDat$ldl)
zSD        <- sd(dDat$ldl)
dDat$zCen  <- dDat$ldl - zMean
dDat$zHi   <- dDat$ldl - (zMean + zSD)
dDat$zLo   <- dDat$ldl - (zMean - zSD)
```

Example

Test the simple slope of $X_{BMI} \rightarrow Y_{BP}$ at 1 *SD* below the mean of Z_{LDL} .

```
out2.1 <- lm(bp ~ bmi*zLo, data = dDat)
partSummary(out2.1, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  52.637886   4.764883  11.047 < 2e-16
## bmi          1.562525   0.185667   8.416 5.59e-16
## zLo          0.448771   0.127160   3.529 0.000461
## bmi:zLo      -0.015352   0.004716  -3.255 0.001221
##
## Residual standard error: 12.54 on 438 degrees of freedom
## Multiple R-squared:  0.1834, Adjusted R-squared:  0.1778
## F-statistic: 32.78 on 3 and 438 DF,  p-value: < 2.2e-16
```

The estimated slope for *bmi*, $\hat{\beta}_1 = 1.563$, is the simple slope.

Example

Test the simple slope of $X_{BMI} \rightarrow Y_{BP}$ at the mean of Z_{LDL} .

```
out2.2 <- lm(bp ~ bmi*zCen, data = dDat)
partSummary(out2.2, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 66.286409   3.812480  17.387 < 2e-16
## bmi         1.095631   0.141679   7.733 7.27e-14
## zCen         0.448771   0.127160   3.529 0.000461
## bmi:zCen    -0.015352   0.004716  -3.255 0.001221
##
## Residual standard error: 12.54 on 438 degrees of freedom
## Multiple R-squared:  0.1834, Adjusted R-squared:  0.1778
## F-statistic: 32.78 on 3 and 438 DF,  p-value: < 2.2e-16
```

The estimated slope for bmi, $\hat{\beta}_1 = 1.096$, is the simple slope.

Example

Test the simple slope of $X_{BMI} \rightarrow Y_{BP}$ at 1 *SD* above the mean of Z_{LDL} .

```
out2.3 <- lm(bp ~ bmi*zHi, data = dDat)
partSummary(out2.3, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  79.934933   6.023132  13.271  < 2e-16
## bmi         0.628736    0.216369   2.906  0.003848
## zHi         0.448771    0.127160   3.529  0.000461
## bmi:zHi     -0.015352    0.004716  -3.255  0.001221
##
## Residual standard error: 12.54 on 438 degrees of freedom
## Multiple R-squared:  0.1834, Adjusted R-squared:  0.1778
## F-statistic: 32.78 on 3 and 438 DF,  p-value: < 2.2e-16
```

The estimated slope for *bmi*, $\hat{\beta}_1 = 0.629$, is the simple slope.

Compare Approaches

The manual and the centering approaches give identical answers, barring rounding errors:

	Z Low	Z Center	Z High
Manual	1.562525	1.095631	0.628736
Centering	1.562525	1.095631	0.628736

Simple Slopes

	Z Low	Z Center	Z High
Manual	0.185667	0.141679	0.216369
Centering	0.185667	0.141679	0.216369

Standard Errors

Alternative Probing Strategies

Simple slopes analysis is nice due to its simplicity and ease of interpretation, but the Z values we choose are totally arbitrary.

- We may be missing important nuances that occur in areas of Z 's distribution that we *did not* pick.



Alternative Probing Strategies

Simple slopes analysis is nice due to its simplicity and ease of interpretation, but the Z values we choose are totally arbitrary.

- We may be missing important nuances that occur in areas of Z 's distribution that we *did not* pick.

The *Johnson-Neyman* technique is an alternative approach that removes the arbitrary selection of conditional Z values.

- Johnson-Neyman finds the *region of significance* wherein the conditional effect of X on Y is statistically significant.
- Johnson-Neyman inverts the logic of simple slopes analysis to find what cut-points on the moderator correspond to a critical t value for the conditional effect.

Johnson-Neyman Technique

With simple slopes analysis, we:

1. Choose a conditional value of Z , say Z_1 .
2. Calculate the simple slope, $\hat{f}(Z_1)$, and standard error, $SE_{f(Z_1)}$, associated with Z_1 .
3. Test $\hat{f}(Z_1)$ for significance via a simple Wald-type test:

$$\hat{t} = \frac{\hat{f}(Z_1)}{SE_{f(Z_1)}} \quad (3)$$

Johnson-Neyman Technique

With Johnson-Neyman, we:

1. Choose an α level for our test.
 - Use $\{\alpha, df\}$ to define a critical t -value
 - E.g., $\alpha = 0.05$ and $df = 438 \Rightarrow t_{crit} = 1.97$
2. Replace \hat{t} in Equation 3 with t_{crit} and re-arrange into the following quadratic form:

$$t_{crit}^2 SE_{f(Z)}^2 - \hat{f}^2(Z) = 0 \quad (4)$$

3. Solve Equation 4 to find the two values of Z (i.e., the roots) that produce critical t statistics for the simple slope.

Deriving Equation 4

Recall the definitions of $\hat{f}(Z)$ and $SE_{f(Z)}$:

$$\hat{f}(Z) = \hat{\beta}_1 + \hat{\beta}_3 Z,$$

$$SE_{f(Z)} = \sqrt{SE_{\beta_1}^2 + 2Z \cdot \text{COV}(\beta_1, \beta_3) + Z^2 SE_{\beta_3}^2}.$$

So, our starting point is:

$$t_{crit} = \frac{\hat{\beta}_1 + \hat{\beta}_3 Z}{\sqrt{SE_{\beta_1}^2 + 2Z \cdot \text{COV}(\beta_1, \beta_3) + Z^2 SE_{\beta_3}^2}}. \quad (5)$$

Deriving Equation 4

Square both sides of Equation 5 to get:

$$t_{crit}^2 = \frac{\hat{\beta}_1^2 + 2\hat{\beta}_1\hat{\beta}_3Z + \hat{\beta}_3^2Z^2}{SE_{\beta_1}^2 + 2Z \cdot \text{COV}(\beta_1, \beta_3) + Z^2SE_{\beta_3}^2}. \quad (6)$$

Lastly, move all terms in Equation 6 to the left-hand side and simplify to get the final quadratic form:

$$\left(t_{crit}^2SE_{\beta_3}^2 - \hat{\beta}_3^2\right)Z^2 + \left(2t_{crit}^2\text{COV}(\beta_1, \beta_3) - 2\hat{\beta}_1\hat{\beta}_3\right)Z + \left(t_{crit}^2SE_{\beta_1}^2 - \hat{\beta}_1^2\right) = 0.$$

Solving Equation 4

Given our final quadratic equation:

$$\left(t_{crit}^2 SE_{\beta_3}^2 - \hat{\beta}_3^2\right) Z^2 + \left(2t_{crit}^2 \text{COV}(\beta_1, \beta_3) - 2\hat{\beta}_1\hat{\beta}_3\right) Z + \left(t_{crit}^2 SE_{\beta_1}^2 - \hat{\beta}_1^2\right) = 0,$$

we find our solutions by taking:

$$a = t_{crit}^2 SE_{\beta_3}^2 - \hat{\beta}_3^2,$$

$$b = 2t_{crit}^2 \text{COV}(\beta_1, \beta_3) - 2\hat{\beta}_1\hat{\beta}_3,$$

$$c = t_{crit}^2 SE_{\beta_1}^2 - \hat{\beta}_1^2,$$

and applying the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Example

```
## Estimate model:
out <- lm(bp ~ bmi*ldl, data = dDat)

## Get (squared) critical value of t:
t2 <- qt(0.975, df = out$df.residual)^2

## Extract pertinent elements from the asymptotic
## covariance matrix:
aCov <- vcov(out)
v1 <- diag(aCov) ["bmi"]
v3 <- diag(aCov) ["bmi:ldl"]
cv <- aCov["bmi", "bmi:ldl"]

## Extract pertinent slope coefficients:
b1 <- coef(out) ["bmi"]
b3 <- coef(out) ["bmi:ldl"]
```

Example

```
## Compute coefficients of the quadratic equation:
a <- as.numeric(t2 * v3 - b3^2)
b <- as.numeric(2 * t2 * cv - 2 * b1 * b3)
c <- as.numeric(t2 * v1 - b1^2)

## Compute roots:
myRoots <- c(
  (-b + sqrt(b^2 - 4 * a * c)) / (2 * a),
  (-b - sqrt(b^2 - 4 * a * c)) / (2 * a)
)

myRoots

## [1] 154.7983 305.9235
```


Interpreting Johnson-Neyman

The roots produced by the Johnson-Neyman technique delineate the *region of significance*.

- The conditional effect of X on Y is either significant everywhere inside or outside of the interval defined by these two points.
- If only one of the points falls within the observed range of Z , ignore the other point.
 - In this case, the region of significance is either everywhere above or everywhere below the legal root.
- If neither of the roots fall within the observed range of Z , then you are in one of two cases:
 1. The focal effect is significant across the entire range of Z .
 2. The focal effect is not significant anywhere within the range of Z .

Interpreting Johnson-Neyman

```
myRoots
```

```
## [1] 154.7983 305.9235
```

```
partSummary(out2, -c(1, 2))
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 14.480616  14.291677   1.013 0.311514
```

```
## bmi         2.867825   0.541312   5.298 1.86e-07
```

```
## ldl         0.448771   0.127160   3.529 0.000461
```

```
## bmi:ldl     -0.015352   0.004716  -3.255 0.001221
```

```
##
```

```
## Residual standard error: 12.54 on 438 degrees of freedom
```

```
## Multiple R-squared:  0.1834, Adjusted R-squared:  0.1778
```

```
## F-statistic: 32.78 on 3 and 438 DF,  p-value: < 2.2e-16
```

Perspectives on Simple Slopes

Recall the formula for a simple slope:

$$f(Z) = \beta_1 + \beta_3 Z$$

We can think about working with $f(Z)$ in, at least, two different ways:

1. Treat $f(Z)$ as a weight for X that we can use to evaluate the effect of X on Y at different (discrete) levels of Z .
2. Consider how $f(Z)$, itself, changes as Z (continuously) changes.

The first option gives rise to simple slopes analysis while the second embodies the spirit of Johnson-Neyman.

Confidence Bands

A natural quantity to consider is a confidence interval for $\hat{f}(Z)$:

$$CI = \hat{f}(Z) \pm t_{crit} \times SE_{\hat{f}(Z)}$$

- We can easily computed such intervals for the interesting values of Z that we chose for the simple slopes analysis.



Confidence Bands

A natural quantity to consider is a confidence interval for $\hat{f}(Z)$:

$$CI = \hat{f}(Z) \pm t_{crit} \times SE_{\hat{f}(Z)}$$

- We can easily computed such intervals for the interesting values of Z that we chose for the simple slopes analysis.

When doing Johnson-Neyman, we can compute the pointwise values of the CI for the entire range of Z .

- Interpolating between these pointwise CIs produces *confidence bands* for $\hat{f}(Z)$.
 - With these confidence bands, we can immediately check any value of Z for a significant simple slope.

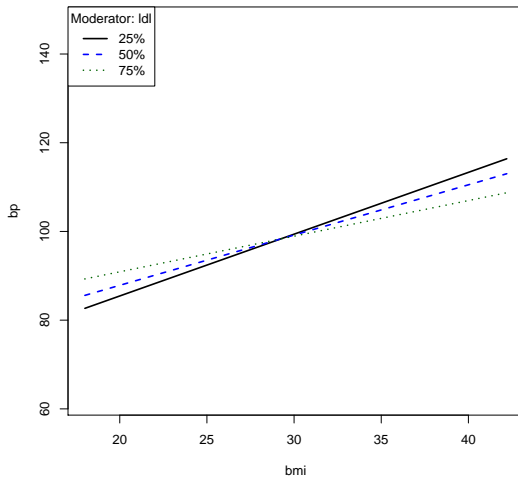
Example

Implementing the Johnson-Neyman technique by hand is a pain, but we can easily do so by using the **rockchalk** package in R.

```
library(rockchalk)

## First we need to create a 'plotSlopes' object:
plotOut <- plotSlopes(model      = out2,
                      plotx      = "bmi",
                      modx       = "ldl",
                      plotPoints = FALSE)
```

Example



Example

We conduct the Johnson-Neyman test by using the `testSlopes` function to modify the “`plotSlopes`” object:

```
## Implement the J-N test:
wrap(testOut <- testSlopes(plotOut))

## Values of ldl OUTSIDE this interval:
##      lo      hi
## 154.7983 305.9235
## cause the slope of (b1 + b2*ldl)bmi to be statistically
## significant

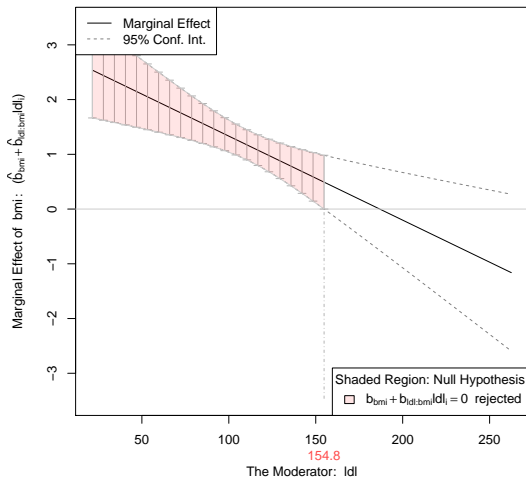
## Extract the significance boundaries:
testOut$jn$roots

##      lo      hi
## 154.7983 305.9235
```


Example

Finally, we can plot the result:

```
plot(testOut)
```



Compare Implementations

As expected, we get the same answers from our manual solution and from the **rockchalk** implementation.

	Low	High
Manual	154.80	305.92
Rockchalk	154.80	305.92

Johnson-Neyman Roots

Categorical Moderators

Categorical moderators encode *group-specific* effects.

- E.g., if we include *sex* as a moderator, we are modeling separate focal effects for males and females.

Given a set of codes representing our moderator, we specify the interactions as before:

$$Y_{total} = \beta_0 + \beta_1 X_{inten} + \beta_2 Z_{male} + \beta_3 X_{inten} Z_{male} + \varepsilon$$

$$Y_{total} = \beta_0 + \beta_1 X_{inten} + \beta_2 Z_{lo} + \beta_3 Z_{mid} + \beta_4 Z_{hi} \\ + \beta_5 X_{inten} Z_{lo} + \beta_6 X_{inten} Z_{mid} + \beta_7 X_{inten} Z_{hi} + \varepsilon$$

Example

```
## Load data:
socSup <- readRDS(paste0(dataDir, "social_support.rds"))

## Focal effect:
out3 <- lm(bdi ~ tanSat, data = socSup)
partSummary(out3, -c(1, 2))

## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   24.4089      5.3502   4.562 1.54e-05
## tanSat        -0.8100      0.3124  -2.593  0.0111
##
## Residual standard error: 9.278 on 93 degrees of freedom
## Multiple R-squared:  0.06742, Adjusted R-squared:  0.05739
## F-statistic: 6.723 on 1 and 93 DF,  p-value: 0.01105
```

Example

```
## Estimate the interaction:
out4 <- lm(bdi ~ tanSat * sex, data = socSup)
partSummary(out4, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    20.8478     6.2114   3.356  0.00115
## tanSat         -0.5772     0.3614  -1.597  0.11372
## sexmale        14.3667    12.2054   1.177  0.24223
## tanSat:sexmale -0.9482     0.7177  -1.321  0.18978
##
## Residual standard error: 9.267 on 91 degrees of freedom
## Multiple R-squared:  0.08955, Adjusted R-squared:  0.05954
## F-statistic: 2.984 on 3 and 91 DF,  p-value: 0.03537
```

Example

On the last slide, the estimated slope for `tanSat`, $\hat{\beta}_1 = -0.577$, is the simple slope for females.

- To estimate the simple slope for males, we simply change the reference group of the `sex` factor and re-estimate the model.

```
## Test the 'male' simple slope by changing reference group:  
socSup$sex2 <- relevel(socSup$sex, ref = "male")  
  
## Re-estimate the interaction:  
out5 <- lm(bdi ~ tanSat * sex2, data = socSup)
```

Example

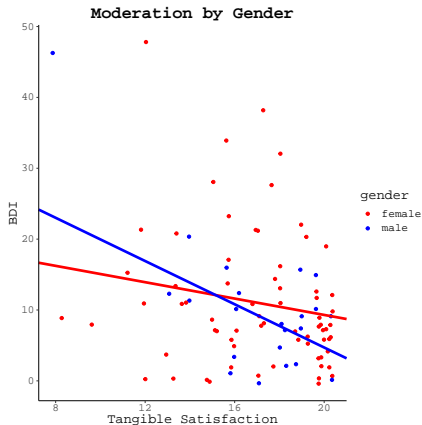
```
partSummary(out5, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    35.2146   10.5067   3.352  0.00117
## tanSat         -1.5254    0.6201  -2.460  0.01579
## sex2female     -14.3667   12.2054  -1.177  0.24223
## tanSat:sex2female  0.9482    0.7177   1.321  0.18978
##
## Residual standard error: 9.267 on 91 degrees of freedom
## Multiple R-squared:  0.08955, Adjusted R-squared:  0.05954
## F-statistic: 2.984 on 3 and 91 DF,  p-value: 0.03537
```

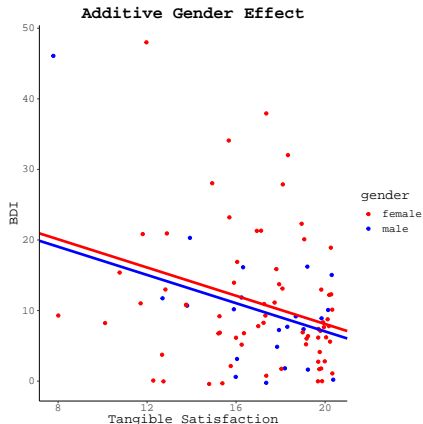
The estimated slope for tanSat, $\hat{\beta}_1 = -1.525$, is now the simple slope for males.

Visualizing Categorical Moderation

$$\hat{Y}_{BDI} = 20.85 - 0.58X_{tsat} + 14.37Z_{male} - 0.95X_{tsat}Z_{male}$$



$$\hat{Y}_{BDI} = 28.10 - 1.00X_{tsat} - 1.05Z_{male}$$



POLYNOMIAL REGRESSION



Polynomial Regression

Polynomial regression simply incorporates powered transformations of the predictors into the model.

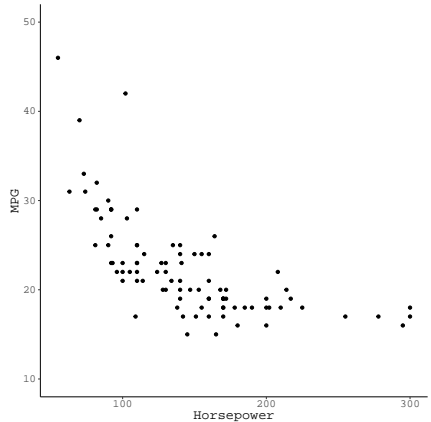
- Polynomial terms (i.e., power terms) model curvature in the relationships.

We can think about polynomial terms as interactions between a predictor and itself.

- Many of the rules that apply to interactions transfer directly to polynomials.

Polynomial Visualization

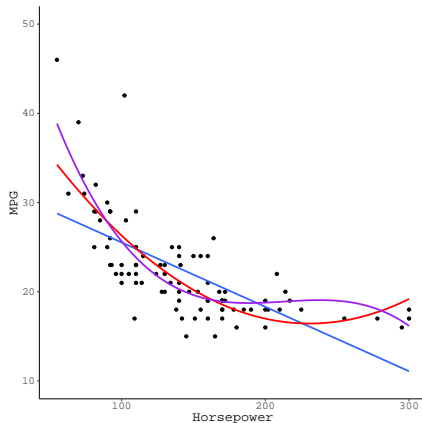
We may hypothesize a curvilinear relationship between X and Y .



Polynomial Visualization

Polynomials are one way to model curvilinear relationships.

- $\hat{Y}_{mpg} = \hat{\beta}_0 + \hat{\beta}_1 X_{hp}$
- $\hat{Y}_{mpg} = \hat{\beta}_0 + \hat{\beta}_1 X_{hp} + \hat{\beta}_2 X_{hp}^2$
- $\hat{Y}_{mpg} = \hat{\beta}_0 + \hat{\beta}_1 X_{hp} + \hat{\beta}_2 X_{hp}^2 + \hat{\beta}_3 X_{hp}^3$



Example

```
## Attach the data:
data(Cars93)

## Fit the linear model:
out6 <- lm(MPG.city ~ Horsepower, data = Cars93)

## Fit the quadratic model:
out7 <- lm(MPG.city ~ Horsepower + I(Horsepower^2),
           data = Cars93)
```

Example

```
partSummary(out6, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  32.746279   1.273229  25.719  < 2e-16
## Horsepower  -0.072174   0.008323  -8.671 1.54e-13
##
## Residual standard error: 4.181 on 91 degrees of freedom
## Multiple R-squared:  0.4524, Adjusted R-squared:  0.4464
## F-statistic: 75.19 on 1 and 91 DF,  p-value: 1.537e-13
```

- For each unit increase in horsepower, the expected change in fuel economy is $\hat{\beta}_1 = -0.0722$ units.

Example

```
partSummary(out7, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.714e+01  2.544e+00  18.528 < 2e-16
## Horsepower   -2.660e-01  3.186e-02  -8.350 7.71e-13
## I(Horsepower^2) 5.762e-04  9.239e-05   6.237 1.42e-08
##
## Residual standard error: 3.513 on 90 degrees of freedom
## Multiple R-squared:  0.6177, Adjusted R-squared:  0.6092
## F-statistic: 72.7 on 2 and 90 DF,  p-value: < 2.2e-16
```

- Extrapolating from powerless cars, each unit increase in horsepower, is expected to change fuel economy by $\hat{\beta}_1 = -0.266$ units.
- For a unit increase in horsepower, the effect of horsepower on fuel economy is expected to increase by $\hat{\beta}_2 = 5.76 \times 10^{-4}$ units.

Example

Does adding the quadratic term explain a significantly larger proportion of variability?

```
anova(out6, out7)

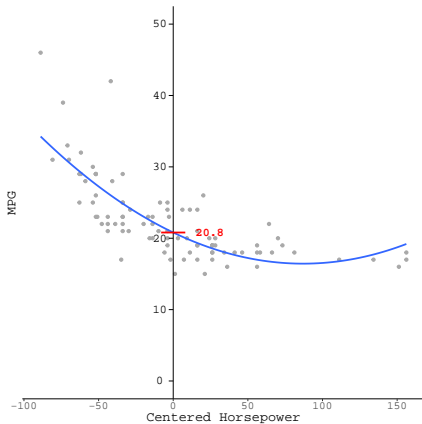
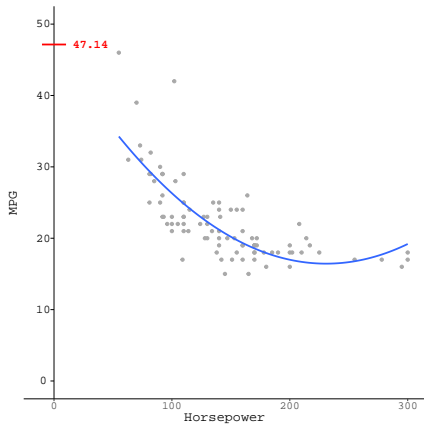
## Analysis of Variance Table
##
## Model 1: MPG.city ~ Horsepower
## Model 2: MPG.city ~ Horsepower + I(Horsepower^2)
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      91 1591.0
## 2      90 1110.9  1    480.11 38.897 1.424e-08 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Example

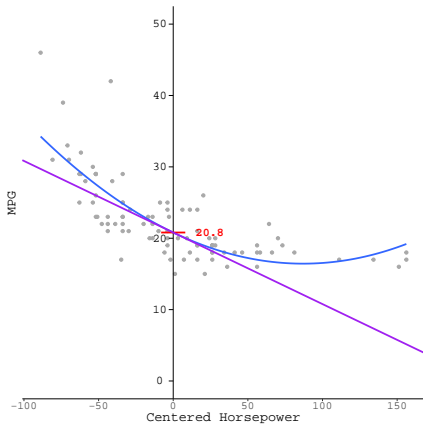
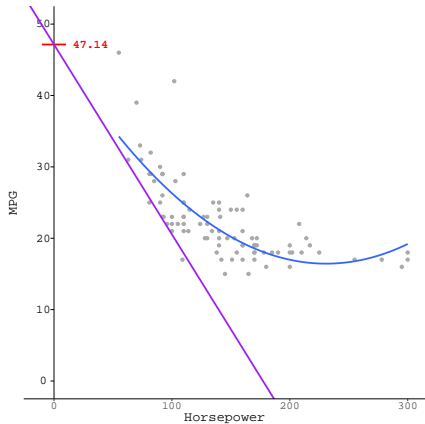
Does adding the quadratic term reduce prediction error?

```
cv.lm(data = Cars93,  
      models = c(formula(out6), formula(out7)),  
      K = 5,  
      names = c("line", "quad")  
      )  
  
##      line      quad  
## 17.93996 13.03084
```

Effects of Centering



Effects of Centering



Example

```
## Mean center horsepower:
Cars93$HorsepowerMC <-
  with(Cars93, Horsepower - mean(Horsepower))

## Fit the quadratic model:
out8 <- lm(MPG.city ~ HorsepowerMC + I(HorsepowerMC^2),
  data = Cars93)
```

Example

```
partSummary(out8, -c(1, 2))

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.080e+01  4.422e-01  47.038 < 2e-16
## HorsepowerMC   -1.003e-01  8.320e-03 -12.053 < 2e-16
## I(HorsepowerMC^2) 5.762e-04  9.239e-05   6.237 1.42e-08
##
## Residual standard error: 3.513 on 90 degrees of freedom
## Multiple R-squared:  0.6177, Adjusted R-squared:  0.6092
## F-statistic: 72.7 on 2 and 90 DF, p-value: < 2.2e-16
```

- Averaging over cars, each unit increase in horsepower, is expected to change fuel economy by $\hat{\beta}_1 = -0.1003$ units.
- For a unit increase in horsepower, the effect of horsepower on fuel economy is expected to increase by $\hat{\beta}_2 = 5.76 \times 10^{-4}$ units.

Choosing an “Optimal” Polynomial Order

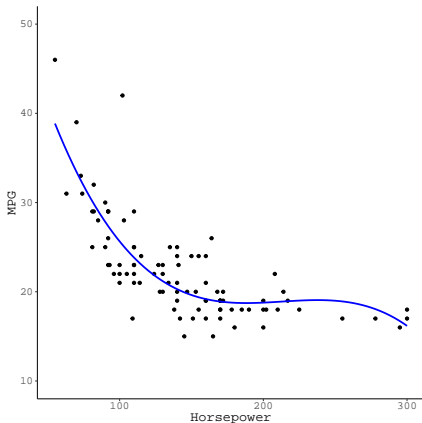
We can use K-fold cross-validation to estimate which polynomial order will give us the lowest prediction error.

```
cv.lm(data = Cars93,  
      models = polyList(y = "MPG.city",  
                        x = "Horsepower",  
                        n = 5),  
      K = 5,  
      names = c("line", "quad", "cube", "quart", "quint")  
      )  
  
##      line      quad      cube      quart      quint  
## 17.68113 12.76947 11.33363 11.85176 12.36846
```

Visualize the Best Model

The cubic model produced the lowest cross-validation error.

- $\hat{Y}_{mpg} = +\hat{\beta}_1 X_{hp} + \hat{\beta}_2 X_{hp}^2 + \hat{\beta}_3 X_{hp}^3$
- We should be cautious.
 - We have very few data.
 - The few cases in the tails are strongly influencing the curve.



Conclusion

- *Moderation* is a specific type of hypothesis.
 - The effect of one variable on the outcome (i.e., the focal effect) changes as a function of some third variable (i.e., the moderator).
- We test for significant moderation by modeling *interaction* terms.
 - If the slope associated with an interaction term is statistically significant, we infer significant moderation.
- A significant interaction term only tells us that the focal effect is dependent upon the moderator.
 - We must usually probe significant interactions to get any real insight into the patterns of moderation.

Conclusion

- We can probe an interaction by testing specific simple slopes.
 - *Simple slopes* analysis
- We can also probe an interaction by finding the region of the moderator's distribution wherein the focal effect is significant.
 - *Johnson-Neyman* approach
- Categorical moderators imply group-specific focal effects.
 - We use simple slopes analysis to probe categorical moderation.
 - Johnson-Neyman doesn't make any sense with discrete moderators.

Conclusion

- When we include interactions in our model, the first-order effects of variables involved in the interactions become *conditional effects*.
 - Conditional effects must be interpreted at the point where the other predictor involved in the interaction is equal to zero.
- We can use the interpretation of conditional effects to our advantage when testing simple slopes.
 - Center the moderator on a conditional value and re-run the model.
 - The estimated focal effect is a simple slope.

Conclusion

- Polynomial regression models curvilinear trends.
 - Each unit increase in the order of the polynomial adds another bend to the curve.
- We can think of polynomials as interactions between the focal predictor and itself.
 - Centering the focal predictors is especially important for interpretation in polynomial regression.
- We need to be very cautious and constantly watch for overfitting when using polynomial regression.
 - Even proper cross-validation may not protect us with small samples.

References

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

