Interaction-Continuous Predictors

Paul E. Johnson[1] [2]

[1]Department of Political Science

[2]Center for Research Methods and Data Analysis, University of Kansas

2014

Note: This demonstrates some features in rockchalk

Outline

Outline

Definition: Interaction

- Linear Model:

$$y_i = b_0 + b_1 x1_i + b_2 x2_i + e_i$$

- Social/Behavioral researchers often assert an additional “interaction effect”

$$y_i = b_0 + b_1 x1_i + b_2 x2_i + b_3 x1_i \cdot x2_i + e_i$$

Outline

Justification 1. The "Moderated Slope" Model

- My best explanation in English:

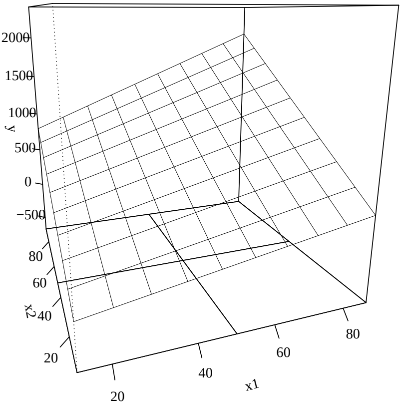
  The effect of one variable depends on another.

- Moderator: name for a variable that "moderates" (changes) the effect of a variable

Examples Of Interaction in Literature

- Corruption temptation caused by electoral uncertainty depends on intra-party competition.
  Nyblade, Benjamin and Steven Reed. 2008. Who Cheats?Who Loots? Political Competition and Corruption in Japan, 1947–1993. *American Journal of Political Science*, 52(4): 926–941.
- Effect of Parliamentarianism depeds on level of ethnic fragmentation
  Selway, Joel. 2011. The Myth of Consociationalism? Conflict Reduction in Divided Societies. *Comparative Political Studies*. 45: 1542-1571.

Visualize the Slope's Dependence: $x1_i$ influence depends



Suppose: $x2_i$ moderates $x1$.

Re-group the terms so that this:

$$y_i = b_o + b_1 x1_i + b_2 x2_i + b_3 x2_i \cdot x1_i + e_i \quad (1)$$

becomes this:

$$y_i = b_o + (b_1 + b_3 x2_i)\, x1_i + b_2 x2_i + e_i \quad (2)$$

- b_3 is an "interaction effect"
- b_1, b_2 often called the "main effects" of $x1$ and $x2$

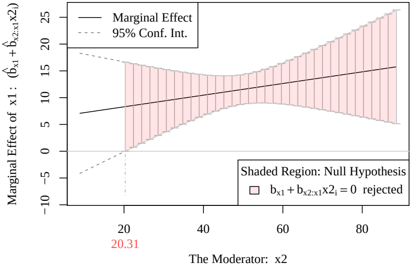
Concentrate: Interpret Coefficients!

$y_i = b_0 + (b_1 + b_3 x2_i)x1_i + b_2 x2_i + e_i$

Note substantive importance of $b_1 \gtreqless -b_3 x2_i$.

- If $>$, the marginal effect of $x1$ is positive
- if $<$, the marginal effect of $x1$ is negative
- If $=$, then $x1_i$ has no marginal effect

I'll explain how to make this later:



Warning: Always include $x1$ and $x2$ if you fit $x1 : x2$ as well

- If a model includes an interaction $x1 \cdot x2$, it should always include $x1$ and $x2$ (even if they appear to be "not statistically significant").
- $x1$ and $x2$ are said to be "marginal" to $x1 \cdot x2$
- Wm Venables, "Exegesis on Linear Models" `http://www.stats.ox.ac.uk/pub/MASS3/Exegeses.pdf`

Reason 2: Approximation of a Function

- Taylor's Theorem says that any function f at a point x, $f(x1, x2)$, can be approximated by a clever choice of coefficients.

$$
\begin{aligned}
y &= f(x1_0, x2_0) + \beta_1(x1 - x1_o) + \beta_2(x2 - x2_0) & (3)\\
&\quad +\beta_3(x1 - x1_0)(x2 - x2_0) + \frac{1}{2}\beta_4(x1 - x1_o)^2 + \ldots & (4)
\end{aligned}
$$

- $(x1_0, x2_0)$ is value where we "approximate from"
- If curvature of f is mild, then the Taylor approximation will stay close to true values.
- We throw away the higher order terms, asserting/hoping they are small.

An Identification Problem

- Identification: Ability to estimate parameters with data at hand.
- The theoretical model boils down to this:

$$y_i = b_o + b_1 x1_i + b_2 x2_i + b_3 x1_i \cdot x2_i + e_i \tag{5}$$

- Expression (5) is equivalent to both of these interpretations:
  - $x1_i$'s slope depends on $x2_i$

$$y_i = b_o + (b_1 + b_3 x2_i) \cdot x1_i + b_2 x2_i + e_i \tag{6}$$

  - $x2_i$'s slope depends on $x1_i$.

$$y_i = b_o + b_1 x1_i + (b_2 + b_3 x1_i) \cdot x2_i + e_i \tag{7}$$

- Data cannot differentiate those 2 models, hence we say there is an "identification problem".

Outline

Same Advice, many disciplines

Recent Chorus:

Must compare predicted values from various predictor combinations to understand their effects.
Each predictors "example values" must be set and understood while focusing on some particular values.

- Psychology
  - Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park: Sage.
  - Preacher, Kristopher J, Curran, Patrick J.,and Bauer, Daniel J. (2006). Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling, and Latent Curve Analysis. *Journal of Educational and Behavioral Statistics*. 31,4, 437-448.
- Political Science:

Same Advice, many disciplines ...

- Brambor, Thomas, Clark, William, & Golder, Matt. 2006. Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14, 63-82.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science* 44: 341–355.

Economics:

- Norton, E.C., Wang, H., and Ai, C. 2004 Computing interaction effects and standard errors in logit and probit models. *Stata Journal* 4(2): 154-167.

Choose Some "For Instance" Values of One Variable, Plot the other

- We want to make a 2D plot of y_i on $x1_i$
- User must supply some "substantively interesting values" of $x2_i$, say 1,2,3, so that this

$$y_i = b_0 + (b_1 + b_3 x2_i) \cdot x1_i + b_2 x2_i + e_i \qquad (8)$$

- Generates a family of lines,

$$x2_i = 1 \quad : \quad y_i = (b_0 + b_2 \cdot (1)) + (b_1 + b_3(1)) \cdot x1_i + e_i \qquad (9)$$
$$x2_i = 2 \quad : \quad y_i = (b_0 + b_2 \cdot (2)) + (b_1 + b_3(2)) \cdot x1_i + e_i \qquad (10)$$
$$x2_i = 3 \quad : \quad y_i = (b_0 + b_2 \cdot (3)) + (b_1 + b_3(3)) \cdot x1_i + e_i \qquad (11)$$

The Marginal Effect of x1 is $(b_1 + b_3 \times moderator_i)$

- The full model might be

$$y_i = b_0 + b_1 x1_i + b_2 x2_i + b_3 x1_i \cdot x2_i + e_i \quad (12)$$

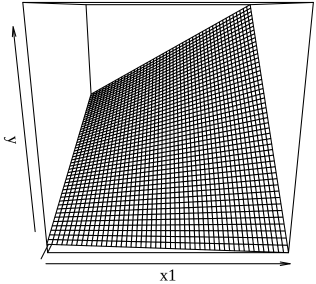
- You choose $x2_i = whatever\ value\ is\ interesting\ for\ the\ moderator$:

$$y_i = \underbrace{(b_0 + b_2 \cdot (whatever))}_{} + \underbrace{(b_1 + b_3 \cdot (whatever))}_{} \cdot x1_i + e_i \quad (13)$$

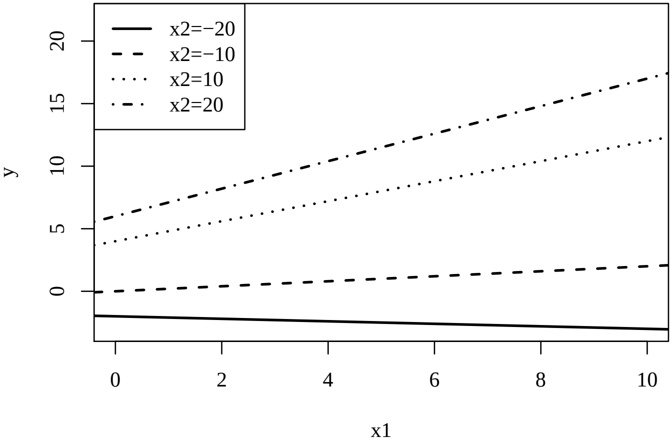
- After you set a particular $x2_i = whatever$, then the line for the "simple slope" has
  - The "New Intercept:" $(b_0 + b_2 whatever)$: AKA "shifted intercept"
  - The "New Slope:" $(b_1 + b_3 whatever)$: AKA the "marginal effect" of $x1_i$.

In 3 Dimensions

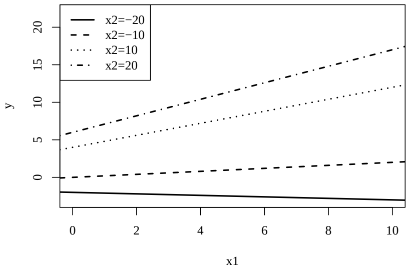
Suppose $y_i = 2 + 0.5 \cdot x1 + 0.2 \cdot x2 + 0.03 \cdot x1 \cdot x2$

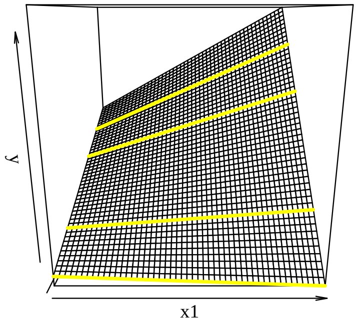


One Line Per Value of $x2_i$ while plotting Simple Slope



Translate Between 3-D and the Simple Slopes in 2-D





- These four lines are highlighted in yellow in the graph on the right.

Ways To Choose "whatever" Values

- Problem specific interesting cases that suit your project!
  - Fahrenheight temperatures? Pick {32, 100, 212}
  - Salary (dollars)? Pick {20,000, 100,000, 500,000, 1,000,000}
- The rockchalk package has routines to choose, based either on
  - quantiles (break a range into values that correspond with, for example, the lowest 25%, the median (50%), and the top 75%.
    - I originally developed the plotSlopes function with this in mind
  - standard deviation-based ranges.
    - psychologists suggest it is easier for them to conceptualize special values like the mean, the mean - 1 standard deviation, mean + 1 standard deviation, and so forth.

A Special Hypothesis Test

- The simple line $y_i = (2 + 0.2 * x2_i) + (0.5 + 0.03 * x2_i) * x1_i$
- Concentrate on the slope, the "marginal effect": For a given value of $x2_i$, of course, that is just a sum like

$$\text{for } x2_i \quad : \textit{whatever}, \quad \textit{the slope is}: \ 0.5 + 0.03 * \textit{whatever} \tag{14}$$

- Some people ask, "is that particular slope statistically significantly different from 0?"

And a Fancy T-Test Pops Out (Not Entirely Unexpected)

- Suppose *whatever* $= 10$. They are asking "is the estimate $(0.5 + 0.03 * 10)$ statistically significantly different from 0?"
- And if you put in estimates from a regression, that's a fancy t-test

$$H_0 : b_1 + b_3 x2_0 = 0 \tag{15}$$

$$\hat{t} = \frac{\hat{b}_1 + \hat{b}_3 x2}{\sqrt{Var(\hat{b}_1 + \hat{b}_3 x2_0)}}, \; x2_0 \textit{ is selected value} \tag{16}$$

$$\hat{t} = \frac{\hat{b}_1 + \hat{b}_3 x2}{\sqrt{Var(\hat{b}_1) + x2_0^2 Var(\hat{b}_3) + 2\, x2_0\, Cov(\hat{b}_1, \hat{b}_3)}} \tag{17}$$

J-N Interval: The *whatever* over and over problem

- Imagine letting your research director says, over and over again,
  - What if $x2_0 = 10$, Is it significant then?
  - What if $x2_0 = 13$, Is it significant then?
- That drives you crazy! Over and over, you calculate

$$\hat{t} = \frac{\hat{b}_1 + \hat{b}_3 x2}{\sqrt{Var(\hat{b}_1) + x2_0^2 Var(\hat{b}_3) + 2\, x2_0\, Cov(\hat{b}_1, \hat{b}_3)}} \qquad (18)$$

- Wish you could find a formula to say "$\hat{b}_1 + \hat{b}_3 x2$ is statistically significant if $x2$ is in "this range"?
- It is necessary to solve for $|\hat{t}| > 1.98$, to get the values of $x2_0$ that cause $\hat{t}$ to be statistically different from zero.
- That interval is known as the Johnson-Neyman interval.

rockchalk plotting approaches for both of these

- Described in vignette (run vignette("rockchalk") with rockchalk version 1.5.4 or later).
- Step 1: use regression to fit a model with multiplicative terms
- Plot Type 1: plotSlopes() will to draw the 2 dimensional plot with several lines, one for each value of a moderator
- Plot Type 2: The "J-N interval" plot.
  - The testSlopes() function finds an interval on which the marginal effect is not 0.
  - A plot method for testSlopes objects creates a "marginal effect" plot. I find these confusing, but some people love them!

Basic Idea Behind: plotSlopes, plotCurves, plotPlane

- Fit any regression with interactions and as many other variables you like

```
m1 <- lm(y ~ x1*x2 + x3 + x4, data=dat)
```

- You want to focus on the predictive effect of $x1$ and $x2$.
- draw a plot with $x2$ on the horizontal axis and a line for some focal values of *modx*.

```
m1ps <- plotSlopes(m1, plotx = "x2", modx = "x1")
```

- That creates a plot, but also an output object for further analysis.
- Give the output to testSlopes(), like so:

```
m1psts <- testSlopes(m1ps)
```

- There is a plot method for that type of object

```
plot(m1psts)
```

Basic Idea Behind: plotSlopes, plotCurves, plotPlane ...

- Make a reasonable 3D plot of the pair of variables:
  - the defaults are mostly good

```
m1pp <- plotPlane(m1, plotx1="x1", plotx2="x2")
```

  - there are many options that can customize

```
m1pp <- plotPlane(m1, plotx1="x1", plotx2="x2", plotPoints=
    F, drawArrows=F, ticktype="detailed", theta=-20, npp=7)
```

- plotSlopes has to generate a particular kind of output object for its eventual input into testSlopes(). If we ignore that problem, then we can have a more flexible line plotter. That is calle plotCurves().
- Unlike plotSlopes(), the plotCurves() and plotPlane() functions can handle many types of nonlinear functions. They do work (or should work) on formula like

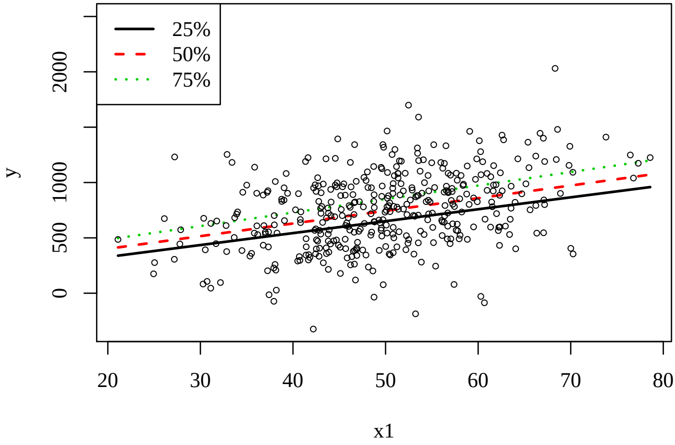
```
alpha <- 2
m2 <- lm(y ~ x1*log(x2 + alpha) + sin(x3) + x4, data=dat)
```

Basic Idea Behind: plotSlopes, plotCurves, plotPlane ...

- In rockchalk 1.6.3, a new function called addLines() was introduced. It can take the lines from a "plotSlopes()" or "plotCurves()" and superimpose them on the 3d output from plotPlane. That *should* make a plot of the sort displayed above "Translate Between 3-D and ..."

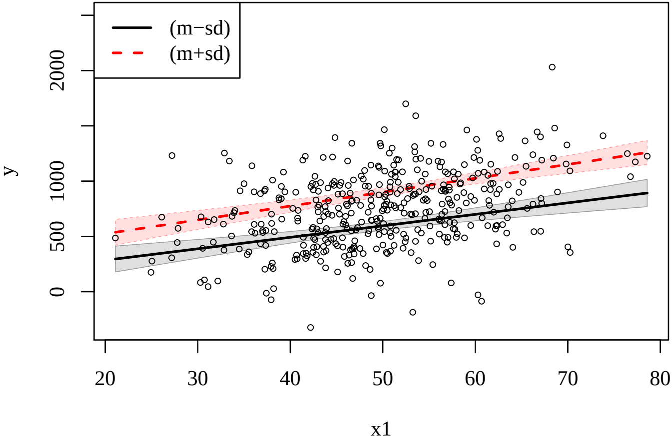
Default plotSlopes output

```
m1ps <- plotSlopes(m1, plotx="x1", modx="x2")
```



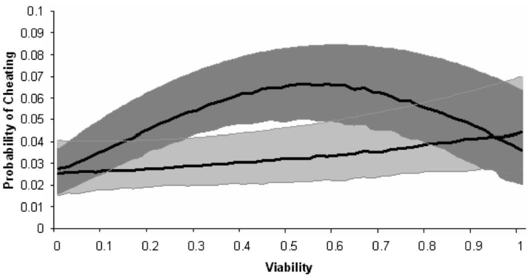
plotSlopes variations

```
m1ps <- plotSlopes(m1, plotx="x1", modx="x2", modxVals = "std.dev",
        n = 2, interval = "conf")
```



plotSlopes or plotCurves equivalents are in the literature

FIGURE 2 **The Impact of Viability and Intraparty Competition on the Likelihood of Cheating**

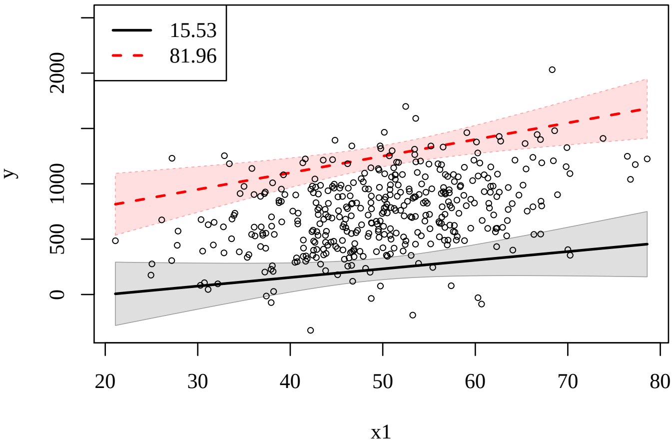


Notes: The figure reports simulations of the impact of the interaction of viability and intraparty competition on probability of cheating based on Table 2 Model 3, with other variables held at their median values. The lower line with a light gray 95% confidence interval represents the probability of cheating as viability varies with no intraparty competition. The upper line with dark gray confidence intervals represents the probability of cheating with high levels of intraparty competition (six other candidates in the same camp).

Admittedly, that's a nonlinear model, but it shows the same basic thing. predicted values for 2 moderator values.

Ask plotSlopes for Confidence Intervals if you want

```
m1ps2 <- plotSlopes(m1, plotx="x1", modx="x2", interval = "conf",
    modxVals = round(range(dat$x2, na.rm=TRUE), 2))
```



J-N confidence interval: testSlopes

```
m1psts <- testSlopes(m1ps)
```

```
Values of x2 INSIDE this interval:
       lo          hi
 20.30582 122.54847
cause the slope of (b1 + b2*x2)x1 to be statistically significant
```

Rather than calculating that interval, some would plot it

Selway and Templeman 1561

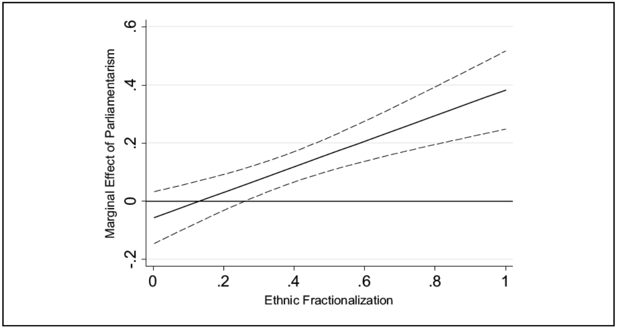
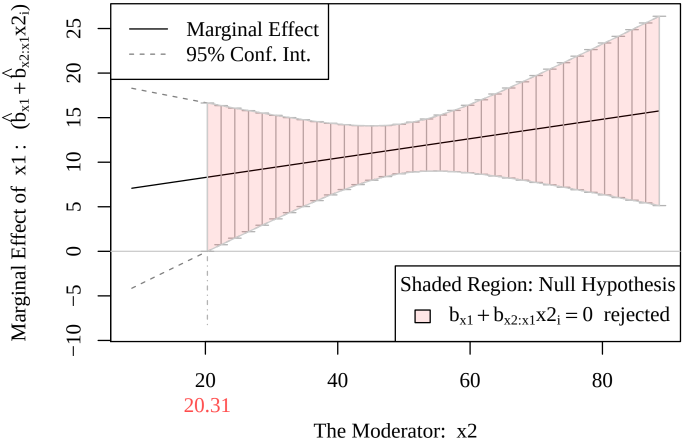


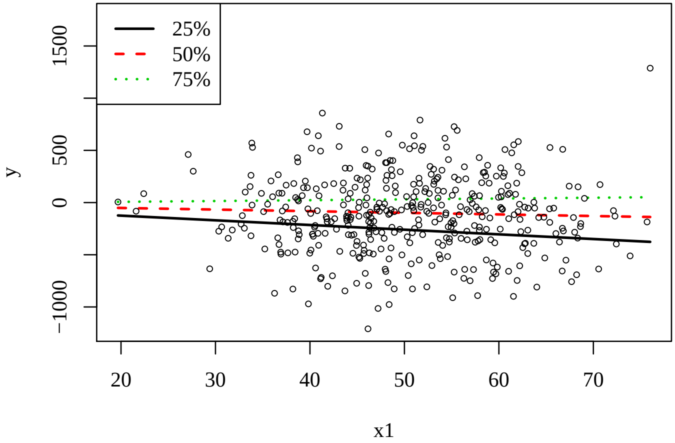
Figure 3. Marginal effect of Parliamentarism on *Political_Deaths* across levels of ethnic fractionalization

plot.testSlopes tries to make this type of plot more clear

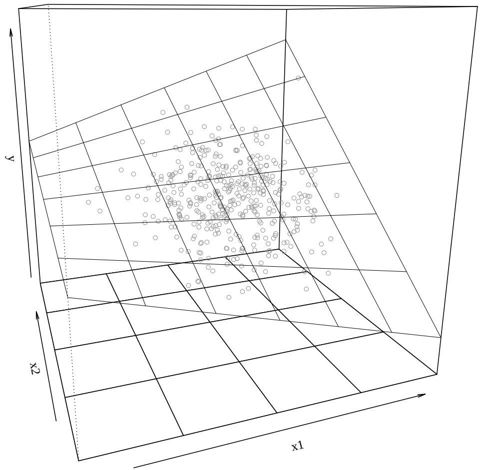
```
plot(m1psts)
```



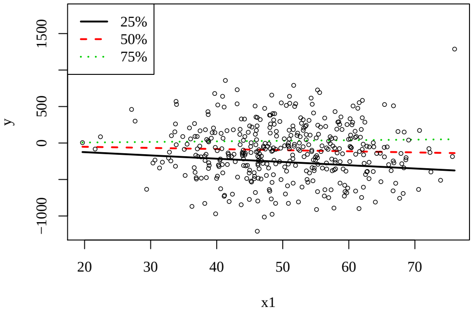
To me, this is a more understandable representation

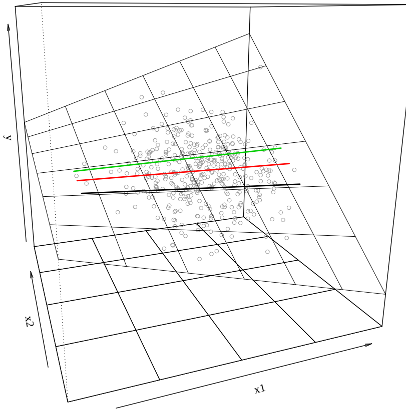


I like the 3d representation for a simple problem

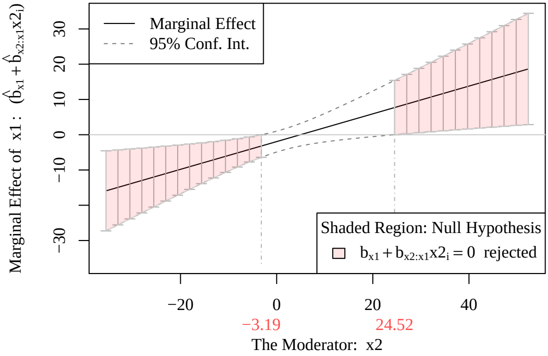


Can Superimpose one on the other

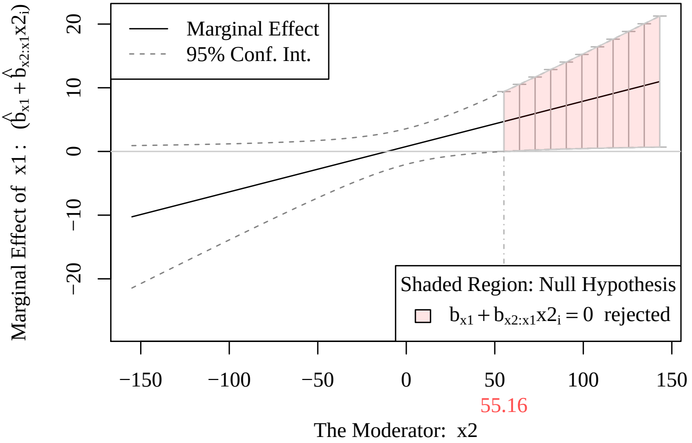




The Alternative J-N Plot is Preferred by Many



testSlopes can have various patterns of significant regions



Outline

Centering: My Theme

- rockchalk vignette has long-ish explanation of mean-centering
- Recall quadratic regression lecture, "centering does not really help" with multicollinearity between x and x^2.
- If you "really understand" regression, you should see that centering doesn't help here either
- It does not "fix" a multicollinearity problem, it was a mistake for its proponents to think so
- Centering does facilitate superficial interpretation in one situation:
  - centering of all x's has effect of making the intercept the predicted value for the "mean case".
  - Intercept is same as using non-centered model to calculate predicted value: $\hat{y}_i = \hat{b}_0 + \hat{b}_1 \overline{x1} + \hat{b}_2 \overline{x2} + \hat{b}_3 \overline{x3} + \ldots$

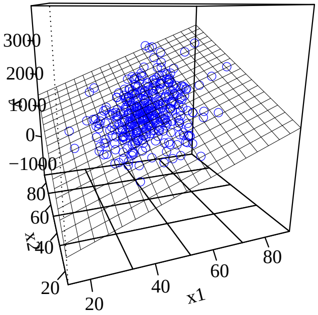
3 Variables I Found Lying About

```
dat <- genCorrelatedData(N=400, rho=.1, stde=250, beta=c(2,0.1,-0.1
     ,0.5))
m1 <- lm(y ~ x1 + x2, data=dat)
fit1 <- plotPlane(m1, plotx1="x1", plotx2="x2", ticktype="detailed"
     )
```

That manufactures data with the true coefficients

$$y = 2 + 0.1x1 - 0.1x2 + 0.3x1 \cdot x2_i + e_i,\ e_i \sim N(0, 300^2) \qquad (19)$$

The 3D Plot for that



The Fitted Linear Model

```
Call:
lm(formula = y ~ x1 + x2, data = dat)

Residuals:
    Min       1Q   Median       3Q      Max
-646.23  -179.72   -22.99   152.93   726.21

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1464.760     92.051  -15.91   <2e-16 ***
x1             26.539      1.388   19.12   <2e-16 ***
x2             27.837      1.339   20.79   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 259.4 on 397 degrees of freedom
Multiple R^2:  0.6972,  Adjusted R^2:  0.6957
F-statistic:   457 on 2 and 397 DF,  p-value: < 2.2e-16
```

That model seems persuasive!

	M1 Estimate (S.E.)
(Intercept)	-1464.760***
	(92.051)
x1	26.539***
	(1.388)
x2	27.837***
	(1.339)
N	400
RMSE	259.449
R^2	0.697
adj R^2	0.696

$*p \leq 0.05{*}{*}\ p \leq 0.01{*}{*}{*}p \leq 0.001$

- Plenty of "stars" indicating statistical significance!
- Easy to interpret parameter estimates

"Throw In" an Interaction to "See" if it is Needed

- Researcher wonders, "should I add $x1 \times x2$ as a predictor?"
- Code change
  - Change the model from

    ```
    m1 <- lm(y~x1+x2, data=dat)
    ```

  - To:

    ```
    m2 <- lm(y~x1*x2, data=dat)
    ```

- R will automatically return equivalent of

  ```
  m2 <- lm(y ~ x1 + x2 + x1:x2, data=dat)
  ```

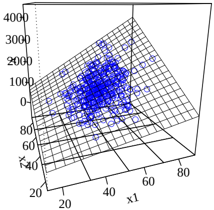
Oh My God! Your p's Exploded

```
m2 <- lm(y ~ x1*x2, data=dat)
summary(m2)
```

```
Call:
lm(formula = y ~ x1 * x2, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-634.81 -173.75  -17.76  160.99  715.48

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -424.2604   379.0010  -1.119  0.26364
x1             6.5226     7.2091   0.905  0.36613
x2             7.3553     7.3616   0.999  0.31833
x1:x2          0.3922     0.1387   2.829  0.00491 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 257.2 on 396 degrees of freedom
Multiple R^2:  0.7032, Adjusted R^2:  0.7009
F-statistic: 312.7 on 3 and 396 DF,  p-value: < 2.2e-16
```

The Interactive Model: Oh, Hell!

- Problem: "Nothing is significant anymore!"



	M1 Estimate (S.E.)
(Intercept)	-424.260
	(379.001)
x1	6.523
	(7.209)
x2	7.355
	(7.362)
x1:x2	0.392**
	(0.139)
N	400
RMSE	257.191
R^2	0.703
adj R^2	0.701

$*p \leq 0.05 ** p \leq 0.01 *** p \leq 0.001$

"Mean-Centering" to the Rescue

- Aiken & West (later, Cohen, Cohen, West and Aiken) claim the constructed variable $x1_i \cdot x2_i$ is multi-collinear with $x1_i$ and $x2_i$, thus causing the results to become "poor."
- As we recall, multicollinearity causes standard errors to inflate, and $\hat{t}$'s shrink.
- They recommend "mean-centering" as a way to ameliorate the "nonessential collinearity".

CCWA advice

- p. 266

  "We recommend that continuous predictors be centered before being entered into regression analyses containing interactions. ... Doing so yields two straightforward, meaningful interpretations of each first-order regression coefficient of predictors entered into the regression equation: (1) effects of the individual predictors at the mean of the sample, and (2) average effects of each individual predictors across the range of the other variables. Doing so also eliminates nonessential multicollinearity between first-order predictors and predictors that carry their interaction with other predictors."

- This advice has been followed VERY widely.
- My counter-argument will be that
  - benefits 1 and 2 are not wrong, but not beneficial either, and
  - the "nonessential multicollinearity" argument is just completely wrong.

Run Again, But Center the Data First

- We "manually" center predictors (either use R's scale() function or the more literal):

```
dat$x1c <- dat$x1 - mean(dat$x1, na.rm = TRUE)
dat$x2c <- dat$x2 - mean(dat$x2, na.rm = TRUE)
m3 <- lm(y ~ x1c * x2c, data=dat)
```

- meanCenter() in rockchalk will do this for us:

```
m2centered <- meanCenter(m2)
summary(m2centered)
fit1 <- plotPlane(m2centered, plotx1="x1c", plotx2="x2c",
     ticktype="detailed")
```

meanCenter Output

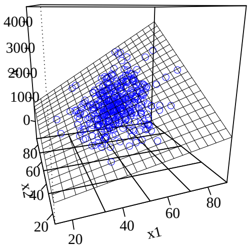
```
These variables were mean-centered before any transformations were
    made on the design matrix.
[1] "x1c" "x2c"
The centers and scale factors were
           x1c      x2c
mean  50.48552 50.69844
scale  1.00000  1.00000
The summary statistics of the variables in the design matrix (after
    centering).
             mean  std.dev.
y       1286.3620  470.3068
x1c        0.0000    9.4335
x2c        0.0000    9.7806
x1c:x2c   11.7519   94.5357

The following results were produced from:
meanCenter.default(model = m2)

Call:
lm(formula = y ~ x1c * x2c, data = stddat)

Residuals:
    Min       1Q   Median       3Q      Max
-634.81  -173.75   -17.76   160.99   715.48
```

meanCenter Output ...

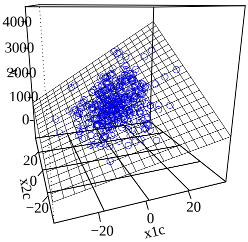
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1281.7531    12.9624  98.883  < 2e-16 ***
x1c           26.4058     1.3770  19.177  < 2e-16 ***
x2c           27.1550     1.3490  20.129  < 2e-16 ***
x1c:x2c        0.3922     0.1387   2.829  0.00491 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 257.2 on 396 degrees of freedom
Multiple R^2:  0.7032, Adjusted R^2:  0.7009
F-statistic: 312.7 on 3 and 396 DF,  p-value: < 2.2e-16
```

Plots Same, but moving Y-axis makes fitted models appear different!

The NON CENTERED FIT



The CENTERED FIT



- Recall beginning of course: rescaling by subtraction shifts intercept, does not change slope
- Find the "y axis" in each plot. Understand why centering seems to matter now?

Look at your p's. The Centered Fit Is Super Awesome!

```
outreg(m2centered)
```

	M1 Estimate (S.E.)
(Intercept)	1281.753***
	(12.962)
x1c	26.406***
	(1.377)
x2c	27.155***
	(1.349)
x1c:x2c	0.392**
	(0.139)
N	400
RMSE	257.191
R^2	0.703
adj R^2	0.701

$*p \leq 0.05 ** p \leq 0.01 *** p \leq 0.001$

Don't Get Carried Away: Its The SAME MODEL!

	Not Centered		Centered	
	Estimate	(S.E.)	Estimate	(S.E.)
(Intercept)	-424.260	(379.001)	1281.753***	(12.962)
x1	6.523	(7.209)	.	
x2	7.355	(7.362)	.	
x1:x2	0.392**	(0.139)	.	
x1c	.		26.406***	(1.377)
x2c	.		27.155***	(1.349)
x1c:x2c	.		0.392**	(0.139)
N	400		400	
RMSE	257.191		257.191	
R^2	0.703		0.703	
adj R^2	0.701		0.701	

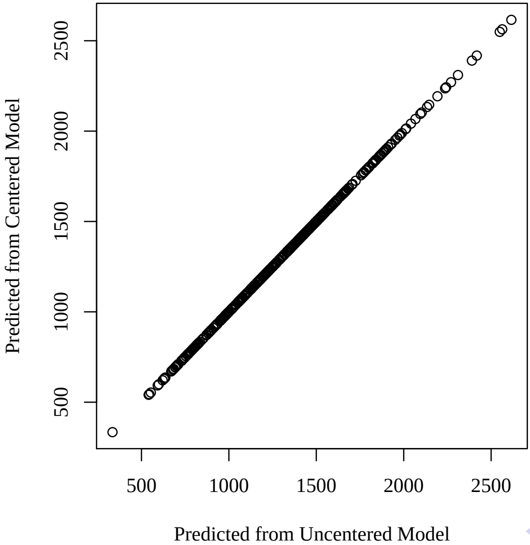
$*p \leq 0.05 ** p \leq 0.01 *** p \leq 0.001$

As CCWA observe, the estimates for the "highest order" slope coefficients are same in both models.

Don't Get Carried Away: Its The SAME MODEL! ...

So is R^2, $RMSE$, etc.

Predicted Values of The Two Models



See why they are the Same Model?

- Same predicted values from same values of input X values
- Same estimates of "slopes" at any given combination of X values
- Same uncertainty (variance) of slope estimates at any given combination of X values
- Why the interaction coefficient the same in the 2 models?
  - Answer: it is the only parameter that is a "constant" in the nonlinear model (cross partial derivative same at all points)

Can Reproduce "Mean Centered" Parameter Estimates From UnCentered Model Estimates

The Un Centered Model fit is

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x1_i + \hat{b}_2 x2_i + \hat{b}_3 x1_i \cdot x2_i$$

The predicted value AT THE MEANS $\overline{x1}$ and $\overline{x2}$ is

$$\hat{y}_{mean} = \hat{b}_0 + \hat{b}_1 \overline{x1} + \hat{b}_2 \overline{x2} + \hat{b}_3 \overline{x1} \cdot \overline{x2}$$

```
bs <- coef(m2)
yhatmean <- bs[1] + bs[2] * mean(dat$x1) + bs[3] * mean(dat$x2) +
     bs[4]*mean(dat$x1)*mean(dat$x2)
yhatmean
```

```
(Intercept)
   1281.753
```

Which is the estimated intercept of the "centered regression".

Can Reproduce ”Mean Centered” Parameter Estimates From UnCentered Model Estimates

Uncentered: $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x1_i + \hat{b}_2 x2_i + \hat{b}_3 x1_i \times x2_i$
The partial slope–the effect of a change in either IV–can be evaluated AT THE MEANS $\overline{x1}$ and $\overline{x2}$.
The effect of $x1_i$(for example) is:

$$\frac{\partial y_i}{\partial x_i} = \hat{b}_1 + \hat{b}_3 \overline{x2} \tag{20}$$

```
partialx1 <- bs[2] + bs[4] * mean(dat$x2)
partialx1
```

```
      x1
26.40583
```

Which is the estimated slope of x1 in the “centered regression”.

Can Reproduce "Mean Centered" Parameter Estimates From UnCentered Model Estimates

Previous showed the partial slope at the mean of x1,x2 is:

$$\frac{\partial y_i}{\partial x_i} = \hat{b}_1 + \hat{b}_3 \overline{x2}$$

Calculate the Variance of that estimated value:

$$Var[\hat{b}_1 + \hat{b}_2 \overline{x2}] = Var[\hat{b}_1] + \overline{x2}^2 Var[\hat{b}_3] + 2\overline{x2} Cov(\hat{b}_1, \hat{b}_3) \quad (21)$$

```
V <- vcov(m2)
varsum <- V[2,2]+mean(dat$x2)^2*V[4,4]+2*mean(dat$x2)*V[2,4]
varsum
```

```
[1] 1.895999
```

```
sqrt(varsum)
```

```
[1] 1.376953
```

Can Reproduce "Mean Centered" Parameter Estimates From UnCentered Model Estimates ...

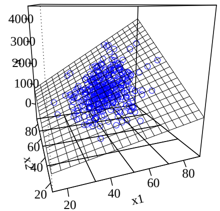
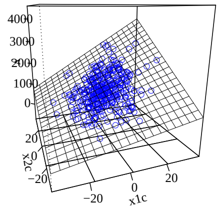
Notice that the square root of the estimated $Var[\hat{b}_1 + \hat{b}_2\overline{x2}]$ is EXACTLY the same standard error that is reported in the Centered Regression for the coefficient $x1c$.

How To Explain "Centered Mirage"?

Two components cause the illusion that the Centered Regression Line is somehow better

- Recall the uncertainty around a regression line is hour shaped. If we place the y axis into the center of the data, we are going to the smallest part of the hourglass, so the standard errors are at their smallest possible values.
- Centering (accidentally, really) may move from a "flat spot" on the curving surface to a place that has steeper slope. This will make the estimated coefficients "bigger" because we are at a steeper spot
  - intercept is y at $x1 = x2 = 0$
  - slope coefficients $\hat{b}_2$ and $\hat{b}_3$ are linear effects of $x1$ and $x2$ at $x1 = x2 = 0$.

Look Again

	
Not Centered Data	Centered Data

Where to Read More about this?

Echambadi, R., & Hess, J. D. (2007). Mean-Centering Does Not Alleviate Collinearity Problems in Moderated Multiple Regression Models. *Marketing Science*, 26(3), 438-445.

"Many empirical marketing researchers commonly mean-center their moderated regression data hoping that this will improve the precision of estimates from ill conditioned, collinear data, but unfortunately, this hope is futile. Therefore, researchers using moderated regression models should not mean-center in a specious attempt to mitigate collinearity between the linear and the interaction terms. Of course, researchers may wish to mean-center for interpretive purposes and other reasons."

Where to Read More about this? (cont)

"Specifically, we demonstrate that (1) in contrast to Aiken and West's (1991) suggestion, mean centering does not improve the accuracy of numerical computation of statistical parameters, (2) it does not change the sampling accuracy of main effects, simple effects, and/or interaction effects (point estimates and standard errors are identical with or without meancentering), and (3) it does not change overall measures of fit such as R2 and adjusted-R2. It does not hurt, but it does not help, not one iota."

See Also:

Kromrey, J. D., & Foster-Johnson, L. (1998). Mean Centering in Moderated Multiple Regression: Much Ado about Nothing. *Educational and Psychological Measurement*, 58(1), 42 -67.