Interaction Workshop
Presented by the CFDR ~ April 9, 2007

What is an interaction?

Basic multiple regression ("main effects") model:

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 Z + \ldots \beta_k X_k + e$$

- implies that *X* and *Z* have independent effects, which are added together to predict *Y*

- β_1 represents the change in the expected value of *Y* for a one-unit increase in *X*, controlling for the other regressors in the model.

Example

$$E(Y) = 2 + 4X + 6Z + e$$

- For every one unit increase in *X*, the expected value of *Y* increases by 4 units, when controlling for *Z* (Z is held at a fixed value).

- But what if the effect of *X* on E(*Y*) depends on the level of *Z*?
  - *X* and *Z* *interact* in their effects on E(Y)
  - The impact of *X* is *moderated* by *Z*

E(CONT) = 2 + 4*PTALK + 6*PCLOSE

CONT: frequency of adolescent contraceptive use
PTALK: parents talk about contraceptive use
PCLOSE: adolescent feels close to parents

- Does the effect of parents talking about contraception on contraceptive use depend on the level of closeness to parents? (e.g. Parents talking about contraception has a stronger effect on contraceptive use for adolescents who feel closer to their parents.)

How do I model an interaction?

- When *X* and *Z* interact in their effects on Y, this is captured by including a cross-product term (*X***Z*) as an additional regressor in the model.

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 Z + \gamma XZ$$

- Before running this model, you may want to center your continuous variables.

<u>Centering</u> (DeMaris, 2004, p. 106)

- To center a variable, subtract it's mean from it. The mean of your centered variable (X_c) is zero.

- Centering continuous variables in interaction models is advantageous for two main reasons:
  - Can make interaction models "more interpretable"
    - β_1 is the effect of X on Y when Z equals zero. If zero is not a legitimate value for Z, this interpretation is not as meaningful. However, if you center continuous variables, β_1 is then the effect of X on Y when Z is at its mean (zero).
  - Can reduce multicollinearity (problem that can arise because the cross-product term is highly correlated to its component variables)

<u>How do I interpret interaction…main effects, interaction term, significance?</u>

- In many cases, you will have first run a "main effects" model – one with no interactions. With this model, you will know whether X and Z are significantly related to Y.

- In the interaction model

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 Z + \gamma XZ$$

  - β_1 is the effect of X on Y when Z (or all other IVs in the model) equals zero
    - If β_1 is significant, this means that X is significantly related to Y when Z (or all other IVs in the model) equals zero.

  - β_2 is the effect of Z on Y when X (or all other IVs in the model) equals zero
    - If β_2 is significant, this means that Z is significantly related to Y when X (or all IVs in the model) equals zero.

    - Differences in values and interpretations for β's in the "main effects" and interaction models result from the fact that in the interaction model, they estimate conditional relationships…the case where all IVs in the model except for the one in question equal zero (Jaccard, Turrisi, & Wan, 1990, pp.26-27).

  - γ is the coefficient for the interaction term
    - Choose the variable whose effect you want to focus on. For example, if we think Z moderates the effect of X on Y, we want to focus on X

    - We want to find the impact of X as a function of Z, so we need factor the regression equation to isolate the common multipliers of X (Demaris, 2004, p. 104)

$$E(Y) = \beta_0 + \beta_2 Z + (\beta_1 + \gamma Z)X$$

$$(\beta_1 + \gamma Z)X$$

- The impact of *X* (it's partial slope) – is a function of the level of *Z*

- The partial slope of *X* in this model (a first-order interaction model) indicates the expected change in *Y* for a one-unit increase in *X*

- In order to better understand the nature of the interaction, it is suggested that you calculate the partial slope of *X* at a few different values of *Z*. Choice of values can be guided by theory, or choose "low" value (one standard deviation below the mean), "medium" value (the mean), and "high" value (one standard deviation above the mean).

- Significant γ means that the partial slope of *X* varies with *Z* – the effect of *X* on *Y* depends on the level of *Z*

- Significant γ does NOT mean that *X* is significant at a particular level of *Z* – *X* might not be significant at any level of *Z*, even when γ is significant.
    - To determine whether *X* is significant at a specific level of *Z*, must determine the significance of $b_1 + gZ$ at a given level of *Z*…targeted centering (Demaris, 2004, pp. 286-287)

    - Create new variables such that…
        - VARNEW1 = VAROLDc – (Mean + SD) …one sd above mean
        - VARNEW2 = VAROLDc – (Mean – SD) …one sd below mean
        - Interaction of VARNEW1 and/or VARNEW2 with focus variable

            - Since mean is already 0 from centering…
            - VARNEW1 = VAROLDc – SD
            - VARNEW2 = VAROLDc + SD

        - Use new variables and interactions with new variables in model

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 VARNEW1 + \gamma X{*}VARNEW1$$

<u>Example using OLS</u>

A student is interested in adolescent delinquency. In particular, he wants to know how friends' delinquency at Time 1 is related to respondents' delinquency at Time 2. He'll include controls for age and gender.

DELINQ	Respondent's delinquency scale	AGEc*FDELc
FDELc	Centered friends' delinquency scale	FEMALE*FDELc
AGEc	Centered age variable	MALE*FDELc
FEMALE	Female = 1, Male = 0	
MALE	Male = 1, Female = 0	

Table 1. Regression models for adolescent delinquency (n =1,151)				
Predictor	Model 1	Model 2	Model 3	Model 4
Intercept	14.44 ***	14.58 ***	14.42 ***	13.46 ***
Friends' delinquency	3.08 ***	3.25 ***	3.37 ***	2.70 ***
Age	0.34 ***	0.31 **	0.34 **	0.34 ***
Female	-0.96 **	-1.00 **	-0.96 **	
Age*Friends' delinquency		-0.26 *		
Female*Friends' delinquency			-0.67 †	
Male				0.96 **
Male*Friends' delinquency				0.67 †

The regression equation for the model with an age by friends' delinquency interaction is:

E(*DELINQ*) = 14.58 + 3.25 *FDELc* + 0.31 *AGEc* – 1.00 *FEMALE* – 0.26 *AGEc*FDELc*

How do we interpret main effects?

- For every one-unit increase in FDELc, the expected value of delinquency increases by 3.25 units, when all other variables equal zero (AGEc at its mean, gender = male)

- For every one unit increase in AGEc, the expected value of delinquency increases by 0.31 when all are variables equal zero (FDELc at its mean, gender = male)

- Females have an expected value of delinquency that is one-unit lower than males, when all other variables equal zero (FDELc at its mean, AGE at its mean)

How do we interpret the interactions?

Focus variable: friends' delinquency; moderator: age

- Significant γ: the effect of friends' delinquency on R's delinquency changes over levels of age

- What is the nature of this effect?
  - Partial slope of friends' delinquency…factor out FDELc
    - (3.25 – 0.26AGEc) FDELc
    - partial slope reflects expected change in Y for a unit increase in focus variable (for a given value of moderator variable)

  - Plug in some values…
    - At mean age (0), the partial slope of FDELc is 3.25
    - At one standard deviation below the mean age (-1.72), the partial slope of FDELc is 3.70
    - At one standard deviation above the mean age (1.72), the partial slope of FDELc is 2.80
      - The effect of friends' delinquency on R's delinquency is greater at lower ages

- Is the effect of friends' delinquency significant at specific values of age? (Use targeted centering to examine whether it is significant at one standard deviation above and below mean)
    - Create new variables such that
        - AGESDA = AGEc – 1.72
        - AGESDB = AGEc + 1.72
        - AAFDEL = AGESDA * FDELc
        - ABFDEL = AGESDB * FDELc

    - The main effect of friends' delinquency is significant at one standard deviation above and below the mean for age (Table 2).

Table 2. Regression models for adolescent delinquency (targeted centering)

Predictor	Model 4		Model 5	
Intercept	15.10	***	14.05	***
Friends' delinquency	2.80	***	3.69	***
Female	-1.00	**	-1.00	**
Age - SD above mean	0.31	**		
(Age - SD above mean)*Friends' delinquency	-0.26	*		
Age - SD below mean			0.31	**
(Age - SD below mean)*Friends' delinquency			-0.26	*

Let's try it again looking at a gender by friends' delinquency interaction…
The regression equation (Table 1, Model 3) for the model with gender by friends' delinquency interaction is:

$$E(DELINQ) = 14.42 + 3.37\ FDELc + 0.34\ AGEc - 0.96\ FEMALE - 0.67\ FEMALE*FDELc$$

Focus variable: friends' delinquency; moderator: gender

- Significant γ: The effect of friends' delinquency on R's delinquency changes over levels of gender (i.e. varies by gender)

- What is the nature of this effect?
    - Partial slope of friends' delinquency…factor out FDELc
        - $(3.37 - 0.67 FEMALE)\ FDELc$

    - Plug in some values…
        - When FEMALE = 1, the partial slope of FDELc is 2.70
        - When FEMALE = 0, the partial slope of FDELc is 3.37
        - The effect of friends' delinquency on R's delinquency is greater for males

    - Is the main effect of FDELc significant for males? Yes, the effect of friends' delinquency (main effect) is 3.37*** when all other variables in the model = 0 (age at its mean, female = 0)

- Is the main effect of FDELc significant for females? Must recode the dummy for gender to make male = 1 and female = 0 (Table 1, Model 4)
  - Yes, the effect of friends' delinquency (main effect) is 2.70*** when all other variables = 0 (age at its mean, male = 0)

Logistic regression

$$\ln O = \beta_0 + \beta_1 X + \beta_2 Z + \gamma XZ$$

- Interpretation is similar to linear regression, but the response is the log odds of event occurrence (DeMaris, 2004, p. 285).

- Factor out the focus variable (X):

$$\ln O = \beta_0 + \beta_2 Z + (\beta_1 + \gamma Z)X$$

$$\text{impact of X: } \exp(\beta_1 + \gamma Z)$$

(Demaris, 2004, p. 285)

Example using logistic regression

- We can dichotomize the delinquency variable from above (0 = no delinquency, 1 = any delinquency) and run logistic regression models.

Table 3. Logistic regression models for adolescent delinquency (n =1,151)

Predictor	Model 1		Model 2		Model 3	
	b	O.R.	b	O.R.	b	O.R.
Intercept	0.59 ***		0.74 ***		0.56 ***	
Friends' delinquency	0.99 ***	2.68	1.20 ***	3.33	0.82 ***	2.27
Age	0.29 ***	1.34	0.20 ***	1.22	0.29 ***	1.34
Female	0.02	1.02	0.00	1.00	0.11	1.12
Age*Friends' delinq			-0.34 ***	0.71		
Female*Friends' delinq					0.37	1.44

Model 2:

$$\ln O = 0.74 + 1.20\,FDELc + 0.20\,AGEc + 0.00\,FEMALE - 0.34\,AGE{*}FDELc$$

Interpretations

- Effect of friends' delinquency on odds of delinquency: $\exp(1.20 - 0.34 AGE) FDELc$
  - For those at the mean age, the effect of a one-unit increase in friends' delinquency is exp(1.20) = 3.33. A one-unit increase in friends' delinquency raises the odds of R's delinquency by 3.33 for those at the mean age.

  - For those at one standard deviation above the mean age, the effect of a one-unit increase in friends' delinquency is exp(1.20 – 0.34*1.72) = 1.85. A one-unit increase in friends' delinquency raises the odds of R's delinquency by 1.85 for those one-standard deviation above the mean age.

- The interaction term for female*friends' delinquency is not significant. Therefore, the effect of friends' delinquency on respondent delinquency does not vary by gender. For fun (or more practice), let's pretend it was significant.

Model 3:

$$\ln O = 0.56 + 0.82\, FDELc + 0.29\, AGEc + 0.11\, FEMALE + 0.37\, FEMALE*FDELc$$

- Effect of friends' delinquency on odds of delinquency: $\exp(0.82 + 0.37\, FEMALE)FDELc$
  - The effect of a one-unit increase in friend's delinquency for females: exp(1.19) = 3.29
    - A one-unit increase in friends' delinquency raises the odds of R's delinquency by 3.29 for females.
  - The effect of a one-unit increase in friend's delinquency for males: exp (0.82) = 2.27
    - A one-unit increase in friends' delinquency raises the odds of R's delinquency by 2.27 for males.

And so on…

<u>Model differences</u>

- Testing interactions allows you to see whether the impact of one variable depends on the level of another.

- What if you want to know if the whole model is different for different groups (e.g. males/female; blacks/whites; singles/marrieds, etc.)?

- If you run models separately and you may see differences in the magnitude, direction, and/or significance of your predictors. However, to determine whether the models are truly different for each group (and not just an artifact of sampling error), you can do a Chow test (DeMaris, 2004, p.111).
  - Run full model without splitting sample, including variable indicating your group of interest (i.e. female, black, married)

  - Run separate models for each group (taking out the variable indicating the group of interest)

$$F = \frac{[SSE_c - (SSE_1 + SSE_2]/p}{(SSE_1 + SSE_2)/(n - 2p)}$$

  - If F is significant, the regressors in the model work differently for group 1 than they do for group 2

  - Chow test analog for logistic regression:

$$X_2 = -2 \ln L_c - [(-2 \ln L_1) + (-2 \ln L_2)]$$

- df = (parameters model 1 + parameters model 2) – (parameters in combined model)

Coding for Interactions

SAS

- Some procedures (e.g. proc glm, proc logistic) will let you put the interaction term "var1*var2" right into the model:

```
proc glm;
model delinq = fdelc agec female agec*fdelc;
run;
```

- For proc reg, you must create the interaction term as a new variable (will not accept "var1*var2")

```
proc reg;
model delinq = fdelc agec female afdelc;
run;
```

Stata

- In most cases, you must create the interaction term as a new variable (will not accept "var1*var2")

```
regress delinq fdelc agec female afdel
```

- With the xi command, stata will create dummy variables and interaction terms for you.

```
xi: regress delinq agec i.female*fdelc
```

- In that example, it saved you from creating the interaction term variable, but female was already dummy coded.

- Xi can save you even more time when you have an ordinal variable that needs to be dummy coded.

- Example:
  R1: 1=white, 2=black, 3=Hispanic, 4=other

  - We want to interact R1 with friends' delinquency.

  ```
  xi: regress delinq agec i.r1*fdelc
  ```

  - Note: By default, the dummy variable corresponding to the lowest value is dropped (in this case, 1 = white). This can be changed.
  - Type help xi in Stata command window for more information.

- To interpret this regression equation, apply concepts from above…

$$E(DELINQ) = 14.43 + 3.43FDELc + 0.32AGEc - 1.84BLACK - 0.63HISP + 0.28OTHER - 1.03FDELc*BLACK - 0.08FDELc*HISP + 2.53FDELc*OTHER$$

  - What is the effect of friends' delinquency?

    $(3.43 - 1.03BLACK - 0.08HISP + 2.53OTHER)FDELc$
  - For blacks, the effect is (3.43 – 1.03) = 2.4
  - For Hispanics, the effect is (3.43 – 0.08) = 3.35
  - For other races, the effect is (3.43 + 2.53) = 5.96

*Remember…

- Include variables for main effects and interaction in model
  - e.g. A model examining the interaction between female and age should include variables female, age, and female*age

- If you are recoding a dummy variable (e.g. gender [1=female, 0=male] to [1=male, 0=female], you must use the recoded variable for both the main effects and interaction variables.

Resources:

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications, Inc.

DeMaris, A. (2004). *Regression with social data: Modeling continuous and limited response variables*. Hoboken, NJ: John Wiley & Sons, Inc.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage Publications, Inc.