Statistics for CSAI II
8 – Interactions

Travis J. Wiltshire, Ph.D. |

Tilburg University

Modules

1. Introduction and Probability
2. Sampling Theory
3. Revisiting Hypothesis Testing & Intro to Correlation
4. Correlation
5. Intro to Regression
6. More Regression Centering and Checking Assumptions
7. Multiple Regression and Assumptions
8. *Interactions*
9. Multiple Regression with Categories
10. Multiple Regression with Polynomials
11. Mixed Models
12. Growth Curve Analysis

Outline

1. PE updates
2. Simple vs multiple regression
3. Multiple regression with interactions
4. Creating interaction terms
5. Understanding interactions
   - Simple slopes analysis

Simple vs. Multiple Regression

$$Y_i = b_0 + b_1X_i + \varepsilon_i$$

- One dependent variable Y predicted from one independent variable X
- One regression coefficient
- **R^2**: proportion of variation in dependent variable Y predictable from X

$$Y_i = b_0 + b_1X1_i + b_2X2_i + \ldots + b_kXk_i + \varepsilon_i$$

- One dependent variable Y predicted from **a set of** independent variables (X1, X2 ….Xk)
- One regression coefficient for each independent variable
- **R^2**: proportion of variation in dependent variable Y predictable by **set of** independent variables (X's)

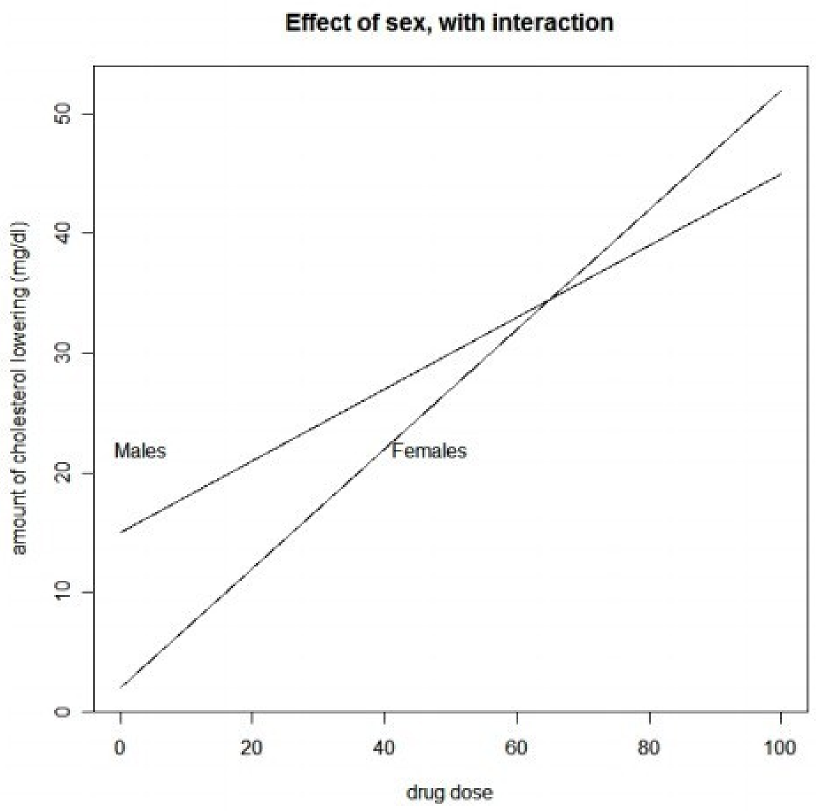
A more in depth comparison here

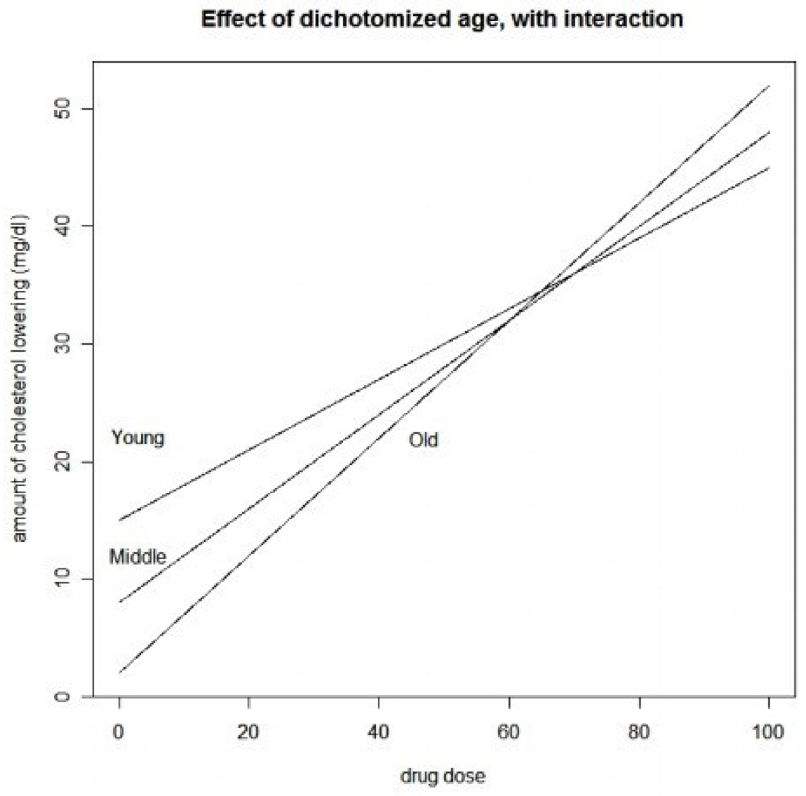
Multiple Regression with an Interaction

$$Y_i = b_0 + b_1X_i + b_2Z_i + \mathbf{b_3X_i{*}Z_i} + \varepsilon_i$$

- **Interaction:** When the effect of one predictor differs based on the level or magnitude of another predictor variable (~ *moderation effect*)
- **Interpretation of coefficients:**
  - b_0 : the intercept, or the predicted outcome when $X = 0$ and $Z = 0$
  - b_1 : simple effect (or slope) of X; predicted change in Y for a one unit change in X at $Z=0$
  - b_2 : simple effect (or slope) of Z; predicted change in Y for a one unit change in Z at $X=0$
  - b_3 : interaction of X and Z; predicted change in the slope of X for a one unit change in Z (or vice versa)

Some interaction examples





What kind of research questions can we answer with interactions?

What is the *predicted* Y given a particular X and Z? (predicted value)

What is *relationship* of X on Y at particular values of Z? (simple slopes/effects)

Is there a *difference* in the relationship of X on Y for different values of Z? (comparing simple slopes)

Creating interaction terms

- We could simply create a new variable by multiplying our two predictors

```
Newvar <- var1 * var2
```

  - Then we can include this in our linear model lm()
  - This could be useful for checking out issues with multicollinearity (but it's not how we would typically model interactions)
- Standard way to incorporate interaction terms in linear regression:

```
lm(y ~ var1 * var2, data = dat)
```

  - Or, equivalently:

```
lm(y ~ var1 + var2 + var1:var2, data = dat)
```

Avoiding Multicollinearity?

- Multicollinearity may be an issue because we are creating a new variable from two independent variables
- We can use **centering** of the individual independent variables to avoid issues with multicollinearity arising from interaction terms

```
var1Centered <- var1 – mean(var1)

var2Centered <- var2 – mean(var2)

Newvar <- var1Centered * var2Centered
```

- But… Not always necessary
  - Iacobucci, D., Schneider, M. J., Popovich, D. L., & Bakamitsos, G. A. (2016). Mean centering helps alleviate “micro” but not “macro” multicollinearity. *Behavior research methods*, *48*(4), 1308-1317.
  - Olvera Astivia, O. L., & Kroc, E. (2019). Centering in multiple regression does not always reduce multicollinearity: How to tell when your estimates will not benefit from centering. *Educational and Psychological Measurement*, *79*(5), 813-826.

How do we further our understanding of interactions?

decompose: to break down the interaction into its lower order components (i.e., predicted means or simple slopes)

probe: to use hypothesis testing to assess the statistical significance of simple slopes and simple slope differences (i.e., interactions)

plot: to visually display the interaction in the form of simple slopes such that values of the dependent variable are on the y-axis, values of the predictor on the x-axis, and the moderator separates the lines or bar graph

Ideas from: https://stats.idre.ucla.edu/r/seminars/interactions-r/

Calculating “Simple Slopes”

Aiken and West (1991)

$$Y_i = b_0 + b_1X + b_2Z + \mathbf{b_3X*Z} + \varepsilon_i$$

- For a significant interaction, we can plot predicted values of Y as a function of X when keeping Z constant at certain values
- The objective of a Simple Slopes analysis is to quantify the slopes of these *effect displays*.

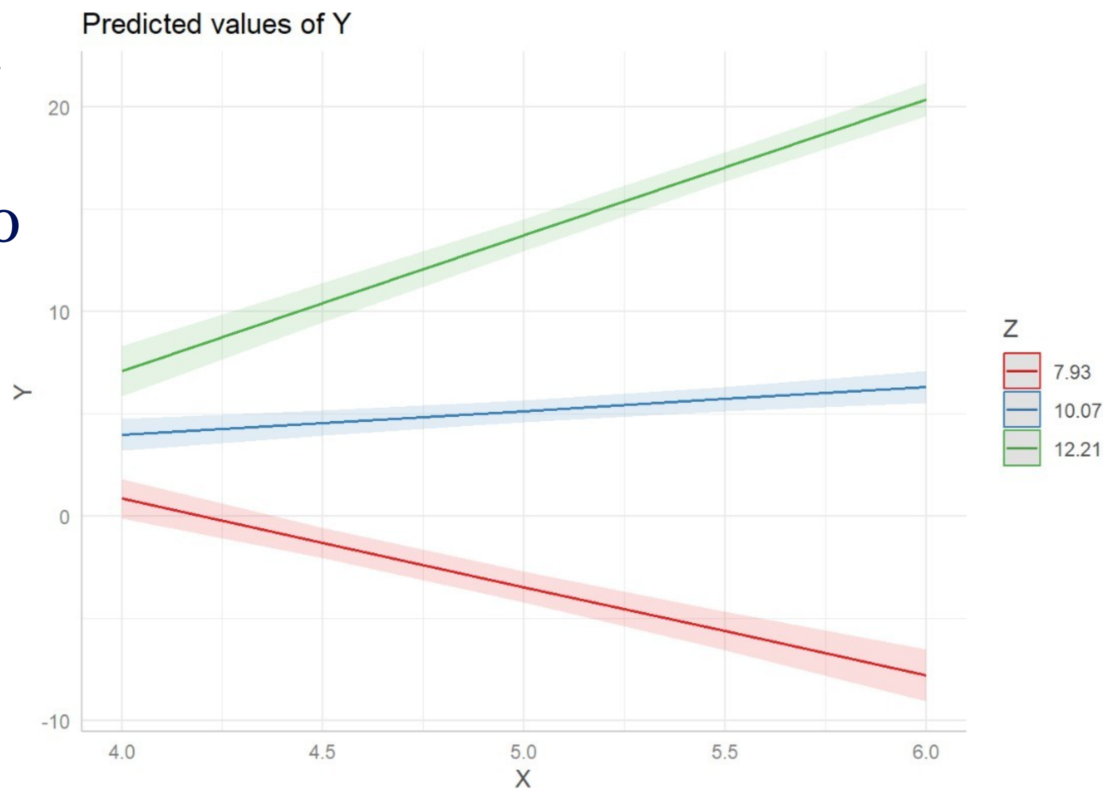
Simple slope

Simple intercept

$$\hat{Y}_i = (b_1 + b_3Z)\,X + (b_0 + b_2Z)$$

$$\hat{Y} = 90.15 - 24.68(X) - 9.33(Z) + 2.58(XZ)$$

$$\hat{Y} = (b_1 + b_3Z)X + (b_0 + b_2(Z))$$
$$= (-24.68 + 2.58(Z))X + (90.15 - 9.33(Z))$$



Understanding interactions using Simple Slopes

- Choose certain values of one of our predictors to compute simple slopes
  - Commonly use +/- 1 SD from the mean
  - Can also choose your own (or may be limited if there are categories)
  - Or use quantiles

MEAN of Z = 10.0
STDEV of Z = 2.2
Z_{low} = 7.8
Z_{mid} = 10.0
Z_{high} = 12.2

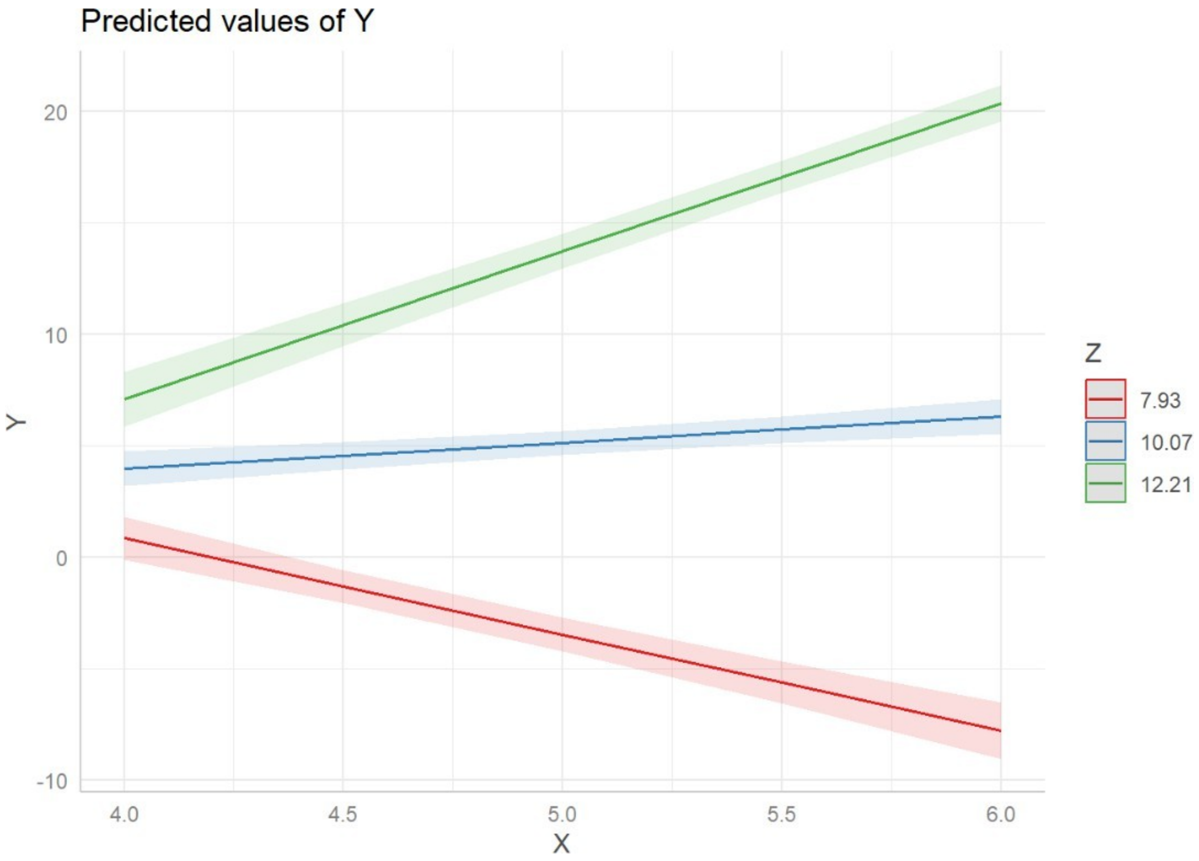
So...

Z_{low} line:
= (-24.68 + 2.58(7.8))X + (90.15 - 9.33(7.8))
= -4.556(X) + 17.376

Z_{mid} line:
= (-24.68 + 2.58(10.0))X + (90.15 - 9.33(10.0))
= 1.12(X) - 3.15

Z_{high} line:
= (-24.68 + 2.58(12.2))X + (90.15 - 9.33(12.2))
= 6.796(X) - 23.676

Understanding interactions using Simple Slopes



MEAN of Z = 10.0
STDEV of Z = 2.2
Z_{low} = 7.8
Z_{mid} = 10.0
Z_{high} = 12.2

So...

Z_{low} line:
= (-24.68 + 2.58(7.8))X + (90.15 - 9.33(7.8))
= -4.556(X) + 17.376

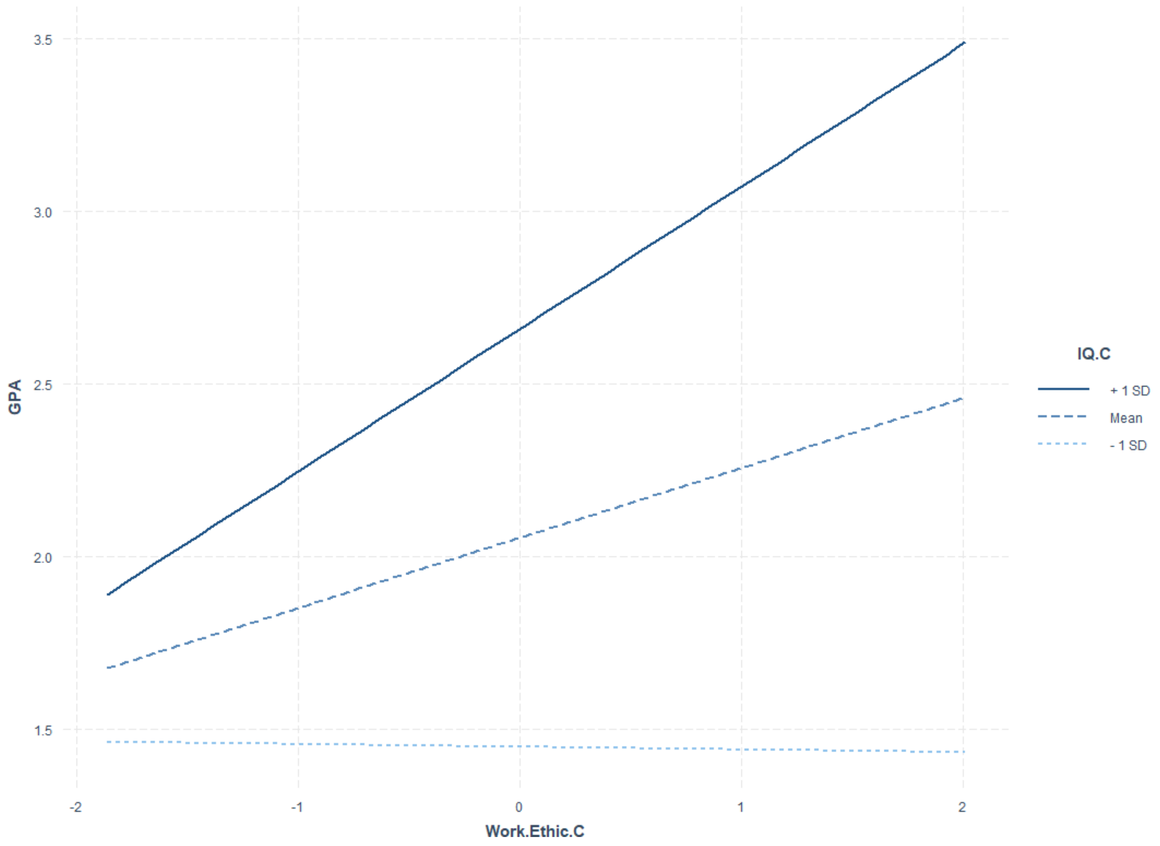
Z_{mid} line:
= (-24.68 + 2.58(10.0))X + (90.15 - 9.33(10.0))
= 1.12(X) - 3.15

Z_{high} line:
= (-24.68 + 2.58(12.2))X + (90.15 - 9.33(12.2))
= 6.796(X) - 23.676

Plot from: https://library.virginia.edu/data/articles/getting-started-with-simple-slopes-analysis

Easy plotting of simple slopes in R

- Using “rockchalk” package:
  - plotSlopes(model.object, plotx=X1, modx=X2, modxVals='std.dev', legendTitle=‘X2’)
- Using “interactions” package:
  - interact_plot(model.object, pred = X1, modx = X2)



emmeans package: Decompose into simple slopes

- Pick three representative values of Z variable (here: ‘effort’)

```
effa <- mean(dat$effort) + sd(dat$effort)
eff <- mean(dat$effort)
effb <- mean(dat$effort) - sd(dat$effort)
```

- Rounded variables (for presentation)

```
> (effar <- round(effa,1))
[1] 34.8
> (effr <- round(eff,1))
[1] 29.7
> (effbr <- round(effb,1))
[1] 24.5
```

- Spotlight analysis of simple slopes based on the three defined values

```
> mylist <- list(effort=c(effbr,effr,effar))
> emtrends(contcont, ~effort, var="hours",at=mylist)
 effort hours.trend    SE  df lower.CL upper.CL
   24.5       0.261 1.352 896   -2.392     2.91
   29.7       2.307 0.915 896    0.511     4.10
   34.8       4.313 1.308 896    1.745     6.88

Confidence level used: 0.95
```

3 separate simple slopes for our X variable (‘hours’)

From:

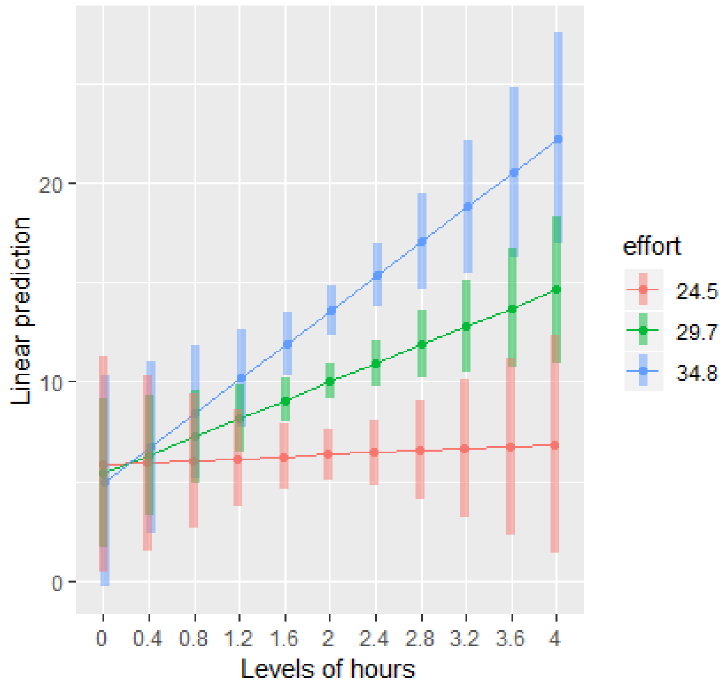
https://stats.idre.ucla.edu/r/seminars/interactions-r/

emmeans package: Plotting & testing simple slopes

- Fixing our variables for plotting and plot graph

```
(mylist <- list(hours=seq(0,4,by=0.4),effort=c(effbr,effr,effar)))
```

```
emmip(contcont,effort~hours,at=mylist, CIs=TRUE)
```



- Testing pairwise differences of simple slopes

```
emtrends(contcont, pairwise ~effort, var="hours",at=mylist, adjust="none")
```

```
$contrasts
 contrast    estimate    SE  df t.ratio p.value
 24.5 - 29.7    -2.05 0.975 896 -2.098  0.0362
 24.5 - 34.8    -4.05 1.931 896 -2.098  0.0362
 29.7 - 34.8    -2.01 0.956 896 -2.098  0.0362

Results are averaged over the levels of: hours
```

Same p values for all comparisons
(matches p value of interaction term!)

From: https://stats.idre.ucla.edu/r/seminars/interactions-r/

Preparing for Module 9: Multiple Regression with Categories

Attend Practical Session: Work on Regression with Interactions

Read:
Cohen, Cohen, West, & Aiken – CH 8

Thanks!