Computer input and output to accompany *Statistical Methods for the Social and Behavioural Sciences: A Model-Based Approach* by David B. Flora

This document © 2017 by David B. Flora

R syntax and output for Chapter 4: Interactions in Multiple Regression

Much of the material in this document builds on the R syntax and output for Chapters 1 to 3. If necessary, refer back to those documents for reminders.

Begin by importing the ‘drink.txt’ data (for the heavy alcohol use example) into R:

#setwd("")  
ch4data <- read.table("drink.txt", header=T)

Preliminary descriptive statistics (Table 4.1):

Need to load the 'psych' package to use the 'describe' function.

library(psych)  
cor(ch4data)

## coa alcuse ext numalc  
## coa 1.0000000 0.2874121 0.2098107 0.9270248  
## alcuse 0.2874121 1.0000000 0.3362025 0.3129409  
## ext 0.2098107 0.3362025 1.0000000 0.1737163  
## numalc 0.9270248 0.3129409 0.1737163 1.0000000

describe(ch4data)

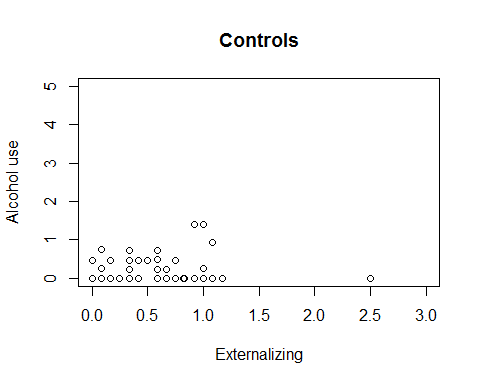
## vars n mean sd median trimmed mad min max range skew  
## coa 1 165 0.55 0.50 1.0 0.56 0.00 0 1.00 1.00 -0.18  
## alcuse 2 165 0.36 0.67 0.0 0.20 0.00 0 4.09 4.09 2.94  
## ext 3 165 0.56 0.44 0.5 0.51 0.37 0 2.50 2.50 1.14  
## numalc 4 165 0.60 0.59 1.0 0.56 1.48 0 2.00 2.00 0.40  
## kurtosis se  
## coa -1.98 0.04  
## alcuse 10.19 0.05  
## ext 1.64 0.03  
## numalc -0.72 0.05

The variable 'coa' is the dichotomous parental alcoholism variable, with 1 = child of alcoholic, 0 = control.  
'alcuse' is the heavy alcohol use variable.  
'ext' is externalizing behaviour.  
'numalc' indicates the number of alcoholic parents (0, 1, or 2) for each case.

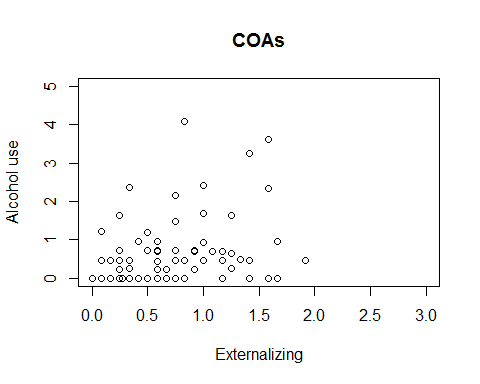
Scatterplots of alcohol use by externalizing, separately for COAs and controls:

It is important that the x- and y-axes have the same scale in both plots. The 'xlim' and 'ylim' options are used to scale the x- and y-axes in the plot.  
Specifically, 'xlim' is set so that the plot displays an x-axis with values from 0 to 3:

controls <- subset(ch4data, coa==0)  
plot(controls$alcuse ~ controls$ext, xlim=c(0,3), ylim=c(0,5),  
 ylab='Alcohol use',xlab='Externalizing',  
 main='Controls')



coas <- subset(ch4data, coa==1)  
plot(coas$alcuse ~ coas$ext, xlim=c(0,3), ylim=c(0,5),  
 ylab='Alcohol use',xlab='Externalizing',  
 main='COAs')



**Modeling an interaction with a dichotomous moderator**

Specify and estimate the regression of alcohol use on externalizing, parental alcoholism, and their interaction.  
In the 'lm' function, we can specify an interaction term as two variables multiplied together without having to literally create a new variable equal that equals the product of two variables:

ch4mod1 <- lm(alcuse~ext+coa+ext\*coa, data=ch4data)  
summary(ch4mod1)

##   
## Call:  
## lm(formula = alcuse ~ ext + coa + ext \* coa, data = ch4data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2206 -0.2731 -0.1382 0.1373 3.4305   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.104262 0.106595 0.978 0.3295   
## ext 0.101702 0.172869 0.588 0.5571   
## coa 0.001682 0.155166 0.011 0.9914   
## ext:coa 0.567081 0.224159 2.530 0.0124 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6114 on 161 degrees of freedom  
## Multiple R-squared: 0.1943, Adjusted R-squared: 0.1792   
## F-statistic: 12.94 on 3 and 161 DF, p-value: 1.291e-07

confint(ch4mod1)

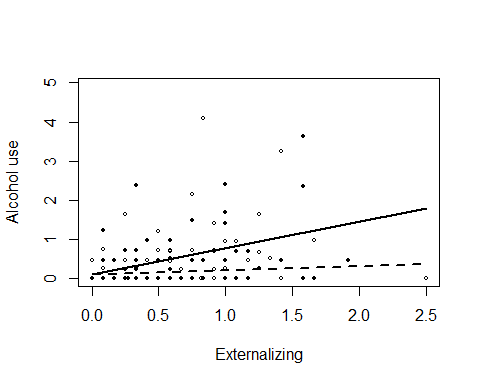
## 2.5 % 97.5 %  
## (Intercept) -0.1062430 0.3147663  
## ext -0.2396821 0.4430858  
## coa -0.3047413 0.3081058  
## ext:coa 0.1244105 1.0097520

The output above provides values matching those in Table 4.2.

Plot of alcohol use by externalizing showing implied simple regression lines for COAs and controls (Figure 4.2):

The 'plotSlopes' function from the 'rockchalk' package plots model-implied slopes of a focal predictor ('plotx' variables) across the levels of a moderator ('modx' variable).  
The 'pch' option is used to specify the symbols used for the data points (here, symbol 16 is open circles, symbol 1 is closed circles).

library(rockchalk)  
plotSlopes(ch4mod1, plotx=ext, modx=coa, plotLegend=F, pch=c(16,1),  
 col=c('black','black'), xlab='Externalizing', ylab='Alcohol use')



**Probing an interaction with simple-slope analysis**

Recode parental alcoholism for the purpose of simple-slope analysis:

ncoa <- abs(ch4data$coa-1)

Now, specify and estimate the interaction model using re-coded COA variable. Results should match those in Table 4.3:

ch4mod1r <- lm(alcuse~ext+ncoa+ext\*ncoa, data=ch4data)  
summary(ch4mod1r)

##   
## Call:  
## lm(formula = alcuse ~ ext + ncoa + ext \* ncoa, data = ch4data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2206 -0.2731 -0.1382 0.1373 3.4305   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.105944 0.112757 0.940 0.3488   
## ext 0.668783 0.142700 4.687 5.88e-06 \*\*\*  
## ncoa -0.001682 0.155166 -0.011 0.9914   
## ext:ncoa -0.567081 0.224159 -2.530 0.0124 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6114 on 161 degrees of freedom  
## Multiple R-squared: 0.1943, Adjusted R-squared: 0.1792   
## F-statistic: 12.94 on 3 and 161 DF, p-value: 1.291e-07

confint(ch4mod1r)

## 2.5 % 97.5 %  
## (Intercept) -0.1167289 0.3286167  
## ext 0.3869776 0.9505886  
## ncoa -0.3081058 0.3047413  
## ext:ncoa -1.0097520 -0.1244105

Get Bonferroni-corrected simultaneous confidence interval for simple slope of externalizing among COAs. The confidence level is set to .975, as explained in Chapter 4:

confint(ch4mod1r, level=.975)

## 1.25 % 98.75 %  
## (Intercept) -0.1491747 0.36106252  
## ext 0.3459155 0.99165065  
## ncoa -0.3527549 0.34939045  
## ext:ncoa -1.0742538 -0.05990871

Get Bonferroni-corrected simultaneous confidence interval for simple slope of externalizing among controls:

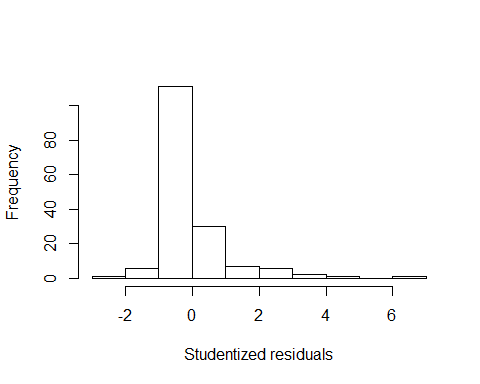
confint(ch4mod1, level=.975)

## 1.25 % 98.75 %  
## (Intercept) -0.13691571 0.3454391  
## ext -0.28942539 0.4928291  
## coa -0.34939045 0.3527549  
## ext:coa 0.05990871 1.0742538

**Regression diagnostics for the interaction model**

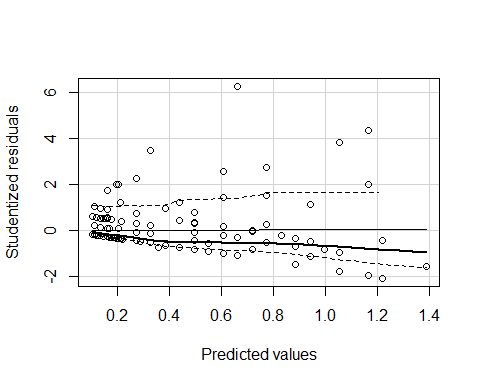
Histogram of Studentized residuals and scatterplot of Studentized residuals against predicted values (Figure 4.3):

hist(rstudent(ch4mod1), xlab='Studentized residuals', main='')



library(car)

scatterplot(rstudent(ch4mod1)~fitted(ch4mod1), boxplots=F, ylab='Studentized residuals',  
 xlab='Predicted values', col='black')



**Modeling an interaction with a three-category moderator**

Create dummy variables for number of alcoholic parents:

d1 <- (ch4data$numalc==1)\*1  
d2 <- (ch4data$numalc==2)\*1

To follow hierarchical regression strategy in Chapter 4, first specify and estimate the model without interactions:

ch4mod2 <- lm(alcuse~ext+d1+d2, data=ch4data)  
summary(ch4mod2)

##   
## Call:  
## lm(formula = alcuse ~ ext + d1 + d2, data = ch4data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.07109 -0.31884 -0.07404 0.11402 3.08686   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.05728 0.08771 -0.653 0.51463   
## ext 0.45135 0.11107 4.064 7.53e-05 \*\*\*  
## d1 0.26156 0.10112 2.587 0.01058 \*   
## d2 0.68804 0.21753 3.163 0.00187 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.616 on 161 degrees of freedom  
## Multiple R-squared: 0.1819, Adjusted R-squared: 0.1667   
## F-statistic: 11.93 on 3 and 161 DF, p-value: 4.265e-07

Then specify and estimate model including interactions with dummy variables. Use the 'anova' function to test whether R2 increases significantly when interaction terms are added to model:

ch4mod3 <- lm(alcuse~ext+d1+d2+(ext\*d1)+(ext\*d2), data=ch4data)  
summary(ch4mod3)

##   
## Call:  
## lm(formula = alcuse ~ ext + d1 + d2 + (ext \* d1) + (ext \* d2),   
## data = ch4data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2024 -0.2254 -0.1297 0.1373 2.9554   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.10426 0.10572 0.986 0.326   
## ext 0.10170 0.17145 0.593 0.554   
## d1 -0.04766 0.15840 -0.301 0.764   
## d2 0.27294 0.38020 0.718 0.474   
## ext:d1 0.57359 0.22572 2.541 0.012 \*  
## ext:d2 0.81166 0.58016 1.399 0.164   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6064 on 159 degrees of freedom  
## Multiple R-squared: 0.2172, Adjusted R-squared: 0.1926   
## F-statistic: 8.826 on 5 and 159 DF, p-value: 2.063e-07

confint(ch4mod3)

## 2.5 % 97.5 %  
## (Intercept) -0.1045396 0.3130630  
## ext -0.2369197 0.4403234  
## d1 -0.3604985 0.2651685  
## d2 -0.4779612 1.0238388  
## ext:d1 0.1277941 1.0193888  
## ext:d2 -0.3341506 1.9574673

anova(ch4mod2,ch4mod3)

## Analysis of Variance Table  
##   
## Model 1: alcuse ~ ext + d1 + d2  
## Model 2: alcuse ~ ext + d1 + d2 + (ext \* d1) + (ext \* d2)  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 161 61.098   
## 2 159 58.458 2 2.64 3.5902 0.02985 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Results above for R2 and significance test for R2 change match those in Chapter 4. Results for 'ch4mod3' match values in Table 4.5.

**Simple-slope analysis with three-category moderator**

Create new dummy variables from 'numalc' so that reference category is 1-alcoholic parent:

d12 <- (ch4data$numalc==0)\*1  
d22 <- (ch4data$numalc==2)\*1

Specify and estimate the interaction model with new dummy variables. Results should match Table 4.6:

ch4mod3b <- lm(alcuse~ext+d12+d22+(ext\*d12)+(ext\*d22), data=ch4data)  
summary(ch4mod3b)

##   
## Call:  
## lm(formula = alcuse ~ ext + d12 + d22 + (ext \* d12) + (ext \*   
## d22), data = ch4data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2024 -0.2254 -0.1297 0.1373 2.9554   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.05660 0.11795 0.480 0.632   
## ext 0.67529 0.14681 4.600 8.59e-06 \*\*\*  
## d12 0.04766 0.15840 0.301 0.764   
## d22 0.32060 0.38378 0.835 0.405   
## ext:d12 -0.57359 0.22572 -2.541 0.012 \*   
## ext:d22 0.23807 0.57336 0.415 0.679   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6064 on 159 degrees of freedom  
## Multiple R-squared: 0.2172, Adjusted R-squared: 0.1926   
## F-statistic: 8.826 on 5 and 159 DF, p-value: 2.063e-07

confint(ch4mod3b)

## 2.5 % 97.5 %  
## (Intercept) -0.1763557 0.2895491  
## ext 0.3853438 0.9652427  
## d12 -0.2651685 0.3604985  
## d22 -0.4373670 1.0785745  
## ext:d12 -1.0193888 -0.1277941  
## ext:d22 -0.8943130 1.3704469

Create new dummy variables from 'numalc' so that reference category is 2-alcoholic parents:

d13 <- (ch4data$numalc==0)\*1  
d23 <- (ch4data$numalc==1)\*1

Specify and estimate the interaction model with new dummy variables. Results should match Table 4.7:

ch4mod3c <- lm(alcuse~ext+d13+d23+(ext\*d13)+(ext\*d23), data=ch4data)  
summary(ch4mod3c)

##   
## Call:  
## lm(formula = alcuse ~ ext + d13 + d23 + (ext \* d13) + (ext \*   
## d23), data = ch4data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2024 -0.2254 -0.1297 0.1373 2.9554   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.3772 0.3652 1.033 0.303  
## ext 0.9134 0.5542 1.648 0.101  
## d13 -0.2729 0.3802 -0.718 0.474  
## d23 -0.3206 0.3838 -0.835 0.405  
## ext:d13 -0.8117 0.5802 -1.399 0.164  
## ext:d23 -0.2381 0.5734 -0.415 0.679  
##   
## Residual standard error: 0.6064 on 159 degrees of freedom  
## Multiple R-squared: 0.2172, Adjusted R-squared: 0.1926   
## F-statistic: 8.826 on 5 and 159 DF, p-value: 2.063e-07

confint(ch4mod3c)

## 2.5 % 97.5 %  
## (Intercept) -0.3440851 1.0984860  
## ext -0.1812692 2.0079897  
## d13 -1.0238388 0.4779612  
## d23 -1.0785745 0.4373670  
## ext:d13 -1.9574673 0.3341506  
## ext:d23 -1.3704469 0.8943130

**Modeling an interaction with a continuous moderator**

Import the ‘hasslesNA.txt’ data (for the psychological symptoms and daily hassles example) into R:

#setwd("")  
ch4data2 <- read.table("hasslesNA.txt", header=T)

Descriptive stats for hassles example:

cor(ch4data2)

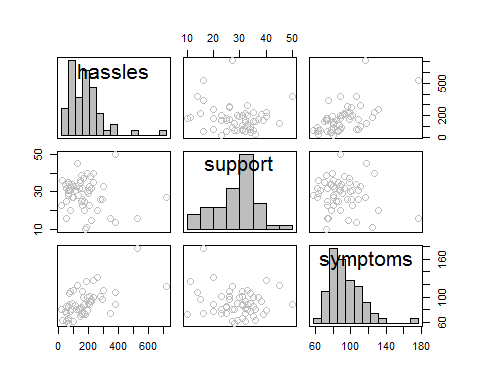
## hassles support symptoms  
## hassles 1.0000000 -0.1671141 0.5770951  
## support -0.1671141 1.0000000 -0.1336386  
## symptoms 0.5770951 -0.1336386 1.0000000

describe(ch4data2)

## vars n mean sd median trimmed mad min max range skew  
## hassles 1 56 170.20 124.34 154.5 152.37 96.37 16 717 701 1.96  
## support 2 56 28.96 8.16 31.0 29.17 6.67 10 50 40 -0.25  
## symptoms 3 56 90.43 21.01 87.5 88.63 19.27 58 177 119 1.35  
## kurtosis se  
## hassles 5.50 16.62  
## support 0.04 1.09  
## symptoms 3.30 2.81

The scatterplot matrix in Figure 4.4 can be made using the the 'scatterplotMatrix' function from the 'car' package. Most of its options are the same as those in the 'scatterplot' function. The 'diagonal' option is used to indicate the type of univariate plot for the diagonal:

scatterplotMatrix(ch4data2, smooth=F, reg.line=F, col='gray', pch=21, diagonal='histogram', cex=1.5)



Specify and estimate interaction model:

ch4mod4 <- lm(symptoms~hassles+support+(hassles\*support), data=ch4data2)  
summary(ch4mod4)

##   
## Call:  
## lm(formula = symptoms ~ hassles + support + (hassles \* support),   
## data = ch4data2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -38.734 -13.379 -0.444 8.689 35.523   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 45.750822 17.371529 2.634 0.011099 \*   
## hassles 0.232643 0.066214 3.514 0.000925 \*\*\*  
## support 1.008385 0.588870 1.712 0.092780 .   
## hassles:support -0.005065 0.002363 -2.144 0.036767 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 16.89 on 52 degrees of freedom  
## Multiple R-squared: 0.3885, Adjusted R-squared: 0.3532   
## F-statistic: 11.01 on 3 and 52 DF, p-value: 1.046e-05

confint(ch4mod4)

## 2.5 % 97.5 %  
## (Intercept) 10.892298901 80.6093446476  
## hassles 0.099774999 0.3655106419  
## support -0.173268312 2.1900386876  
## hassles:support -0.009806324 -0.0003234736

Results above match values in Table 4.9.

**Probe the interaction with a continuous moderator**

First, center 'support' at a low value (e.g., 25th percentile = 25 for 'support'):

suppL <- ch4data2$support - 25

Then estimate the regression model again:

ch4mod4b <- lm(symptoms~hassles+suppL+(hassles\*suppL), data=ch4data2)  
summary(ch4mod4b)

##   
## Call:  
## lm(formula = symptoms ~ hassles + suppL + (hassles \* suppL),   
## data = ch4data2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -38.734 -13.379 -0.444 8.689 35.523   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 70.960451 4.486888 15.815 < 2e-16 \*\*\*  
## hassles 0.106020 0.019114 5.547 9.91e-07 \*\*\*  
## suppL 1.008385 0.588870 1.712 0.0928 .   
## hassles:suppL -0.005065 0.002363 -2.144 0.0368 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 16.89 on 52 degrees of freedom  
## Multiple R-squared: 0.3885, Adjusted R-squared: 0.3532   
## F-statistic: 11.01 on 3 and 52 DF, p-value: 1.046e-05

confint(ch4mod4b)

## 2.5 % 97.5 %  
## (Intercept) 61.956852724 79.9640502051  
## hassles 0.067664760 0.1443759417  
## suppL -0.173268312 2.1900386876  
## hassles:suppL -0.009806324 -0.0003234736

Results above match values in Table 4.10.

Next, center 'support' at a medium value (e.g., 50th percentile = 31 for 'support'):

suppM <- ch4data2$support - 31

Then estimate the regression model again:

ch4mod4c <- lm(symptoms~hassles+suppM+(hassles\*suppM), data=ch4data2)  
summary(ch4mod4c)

##   
## Call:  
## lm(formula = symptoms ~ hassles + suppM + (hassles \* suppM),   
## data = ch4data2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -38.734 -13.379 -0.444 8.689 35.523   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 77.010763 4.127691 18.657 < 2e-16 \*\*\*  
## hassles 0.075631 0.020959 3.609 0.000692 \*\*\*  
## suppM 1.008385 0.588870 1.712 0.092780 .   
## hassles:suppM -0.005065 0.002363 -2.144 0.036767 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 16.89 on 52 degrees of freedom  
## Multiple R-squared: 0.3885, Adjusted R-squared: 0.3532   
## F-statistic: 11.01 on 3 and 52 DF, p-value: 1.046e-05

confint(ch4mod4c)

## 2.5 % 97.5 %  
## (Intercept) 68.727944182 85.2935809978  
## hassles 0.033573989 0.1176879275  
## suppM -0.173268312 2.1900386876  
## hassles:suppM -0.009806324 -0.0003234736

Results above match values in Table 4.11.

Finally, center 'support' at a high value (e.g., 75th percentile = 34 for 'support'):

suppH <- ch4data2$support - 34

Then estimate the regression model again:

ch4mod4d <- lm(symptoms~hassles+suppH+(hassles\*suppH), data=ch4data2)  
summary(ch4mod4d)

##   
## Call:  
## lm(formula = symptoms ~ hassles + suppH + (hassles \* suppH),   
## data = ch4data2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -38.734 -13.379 -0.444 8.689 35.523   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 80.035918 4.985318 16.054 <2e-16 \*\*\*  
## hassles 0.060436 0.025039 2.414 0.0194 \*   
## suppH 1.008385 0.588870 1.712 0.0928 .   
## hassles:suppH -0.005065 0.002363 -2.144 0.0368 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 16.89 on 52 degrees of freedom  
## Multiple R-squared: 0.3885, Adjusted R-squared: 0.3532   
## F-statistic: 11.01 on 3 and 52 DF, p-value: 1.046e-05

confint(ch4mod4d)

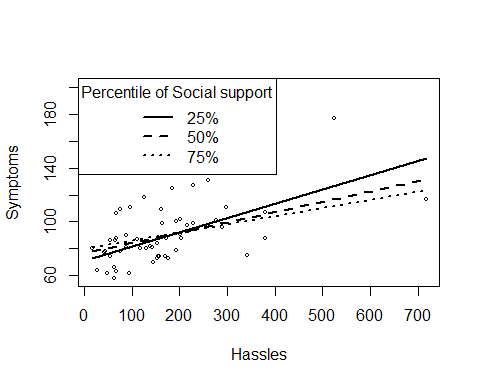
## 2.5 % 97.5 %  
## (Intercept) 70.032145821 90.0396904842  
## hassles 0.010190891 0.1106816334  
## suppH -0.173268312 2.1900386876  
## hassles:suppH -0.009806324 -0.0003234736

Results above match values in Table 4.12.

Plot simple regression lines implied at the low, medium, and high levels of social support (Figure 4.5).

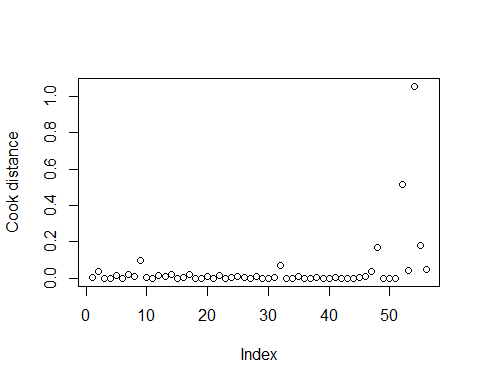
The 'plotSlopes' function from the 'rockchalk' package can do so nicely. By default, this function takes low, medium, and high values of the moderator as the 25th, 50th, and 75th percentiles (it is possible to change this default using the 'modxVals' option):

plotSlopes(ch4mod4, plotx=hassles, modx=support, legendTitle='Percentile of Social support',  
 col=c('black','black','black'), xlab='Hassles', ylab='Symptoms')



But, as reported in Chapter 4, there is a case with a large value for Cook's distance:

plot(cooks.distance(ch4mod4), ylab='Cook distance')



So we need to figure out which case that is. Here is one approach:

The command below creates a new data frame 'ch4data2b' which merges the individual Cook's distance values with the original 'ch4data2' data frame:

ch4data2b <- cbind(cooks.distance(ch4mod4), ch4data2)

The next command sorts this new data frame by the Cook's distance values, then the 'tail' command prints out the last 6 cases (i.e., those with the highest values of Cook's distance):

ch4data2b <- ch4data2b[order(cooks.distance(ch4mod4)),]   
tail(ch4data2b)

## cooks.distance(ch4mod4) hassles support symptoms  
## 32 0.0733583 229 40 127  
## 9 0.1006684 126 45 118  
## 48 0.1696378 185 11 125  
## 55 0.1821190 342 16 75  
## 52 0.5175062 717 27 117  
## 54 1.0534657 525 16 177

In the output above, we can see that the last case (case 54) is the one with Cook's distance = 1.05. This case has an extremely high value for the 'symptoms' outcome.

The syntax below creates a new data frame from which the outlying case has been removed (i.e., only cases with a value of 'symptoms' less than 177 are retained):

ch4data2r <- subset(ch4data2, symptoms < 177)

Now re-estimate the model using this new data frame:

ch4mod4 <- lm(symptoms~hassles+support+(hassles\*support), data=ch4data2r)  
summary(ch4mod4)

##   
## Call:  
## lm(formula = symptoms ~ hassles + support + (hassles \* support),   
## data = ch4data2r)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -28.375 -12.587 -1.797 5.939 36.634   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 65.264594 17.704000 3.686 0.000552 \*\*\*  
## hassles 0.118839 0.074024 1.605 0.114579   
## support 0.407644 0.592114 0.688 0.494286   
## hassles:support -0.001655 0.002524 -0.656 0.515015   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 15.86 on 51 degrees of freedom  
## Multiple R-squared: 0.2291, Adjusted R-squared: 0.1838   
## F-statistic: 5.053 on 3 and 51 DF, p-value: 0.00385

In the output above, the interaction is no longer significant, consistent with the results presented in Chapter 4 for the model with the influential case removed.