Interaction/Mediation Regression

1 Create a new Project

Create a new project in Github called "reg_interactions" and pull it to your computer. Place the file "lectureData6060.csv" in this directory. Create all the scripts for today as part of this project.

2 Caveat

This document provides you with a brief overview. For a full treatment of these topics consult the two books below:

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences.* Routledge.

Aguinis, H. (2004). *Regression analysis for categorical moderators.* Guilford Press.

Note that recent evidence (Westfall et al., 2016) suggests that regression should not be used for these analyses but rather SEM. Nonetheless, we present the standard (likely flawed) approach below. You should understand this approach because it is ubiquitous in the literature. The SEM approach will be covered in PSYC*6380 Psychological Applications of Multivariate Analysis.

Westfall, J., & Yarkoni, T. (2016). Statistically Controlling for Confounding Constructs Is Harder than You Think. PloS One, 11(3), e0152719. [http://doi.org/10.1371/journal.pone.0152719]

3 Interactions among continuous variables

An interaction occurs when the effect of one predictor on the criterion, depends on the level of another predictor. Consider, for example, students preparing for an exam. If the relation between amount of prepartion (hours spent studying) and exam grade depends on the level of anxiety each student experiences then there is an interaction. We test interactions using product terms (details below).

A common problem when attempting to find an interaction is low statistical power. When this occurs, the interaction exists, but the researcher fails to find it because he/she did not have enough participants in the study. Therefore, before conducting a study, in which you hypothesize an interaction, be sure to conduct a power analysis as per a previous weeks assignment. You would use the incremental prediction for sr^2 and should probably assume $sr^2 = .02$ as per Frazier et al. (2004). With continuous variable interactions sr^2 values in this range are common. If you cannot meet the required sample size, conduct an analysis to determine the power you will have for the available sample size. You can then make an informed decision as to whether it even makes sense to conduct the study.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology*, 51(1), 115.

Load the data:

```
library(tidyverse)
my_data <- read.table(file="lectureData6060.csv",header=TRUE,sep=",",na.strings=c("NA"))

glimpse(my_data)
```

```
## Observations: 100
## Variables: 3
## $ Anxiety     <dbl> 22.29, 21.14, 19.93, 17.23, 22.12, 21.51, 23.35, 1...
## $ Preparation <dbl> 7.94, 8.15, 8.17, 8.21, 8.43, 8.52, 8.96, 9.18, 9....
## $ Exam        <dbl> 43.958, 60.813, 37.874, 56.054, 29.500, 15.549, 22...
```

3.1 Extract just the key columns and keep complete cases

In the steps below we will do some "processing" of the data. To ensure we do that correctly, we need to make sure that we only "process" cases that will be used in the final analysis. If we don't do that we introduce error. In the lines below we select just the columns we need, then drop any cases with missing values. Since we will be creating a product term we can't use pair-wise deletion.

We will make extensive use of the *mutate* and *select* commands from the tidyverse package.

```
analytic.data <- my_data %>% select(Exam,Anxiety,Preparation)

#keep complete cases only, list-wise deletion of cases
analytic.data <- na.omit(analytic.data)
```

3.2 Create mean centered versions of the variables

The data set has is composed of Exam scores (i.e., Y) which we will predict using Anxiety (i.e., X) and Preparation (i.e., Z).

To facilitate interpretation of the graph relative to the regression equation (and the main effects in presence of the interaction) we will mean center the predictors. To do this we calculate the mean for X and the mean for Z. We then subtract those means from every score in the respective column to create x.centered and z.centered. The command *scale* creates the z-score version of the scores. The *as.numeric* commands strips off extra information bits added by the scale command that we don't need.

```
# add column with mean center anxiety
analytic.data <- analytic.data %>%
                 mutate(x.centered=as.numeric(scale(Anxiety,center=T,scale=F)) )

# add column with mean center preparation
analytic.data <- analytic.data %>%
                 mutate(z.centered=as.numeric( scale(Preparation,center=T,scale=F)) )
```

3.3 Run the regression, notice how we create the product term:

In this analysis we use x.centered*z.centered when specifying the regression. This creates three predictors for the regression: x.centered, z.centered, and x.centered*z.centered.

```
interaction.regression <- lm(Exam ~ x.centered + z.centered + I(x.centered*z.centered),
                             data=analytic.data, na.action=na.exclude)

summary(interaction.regression)
```

```
##
## Call:
## lm(formula = Exam ~ x.centered + z.centered + I(x.centered *
##     z.centered), data = analytic.data, na.action = na.exclude)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34.71 -12.55   1.01  12.55  29.24
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                55.5794     1.5739  35.313  < 2e-16 ***
## x.centered                 -2.1503     0.7949  -2.705  0.00808 **
## z.centered                  4.5648     0.7952   5.740  1.1e-07 ***
## I(x.centered * z.centered)  0.9077     0.4165   2.180  0.03174 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.74 on 96 degrees of freedom
## Multiple R-squared:  0.3396, Adjusted R-squared:  0.319
## F-statistic: 16.46 on 3 and 96 DF,  p-value: 1.047e-08
```

Note that the product of x.centered and z.centered is **not** equal to the interaction. The interaction and other information is contained in the product term. It is only when the product term is placed in a regression, and the effects of x.centered and z.centered are controlled for, that the b-weight (or semi-partial correlation) for the product term represents the interaction.

Note the interesting contradiction here. Although the interaction is significant, the confidence interval includes zero. This is not an error. The significance test was for sr (or the b-weight) but the confidence interval was created around sr^2. In this type of of scenario the 90% confidence interval for sr^2 (rather than 95%) more closely approximates the results of significance testing for sr using an alpha of .05).

Better than using the summary command is creating a table using apa.reg.table. This table gives your the squared semi-partial correlations (i.e., delta R-squared for the interaction terms, and all other terms).

```
library(apaTables)
apa.reg.table(interaction.regression)
```

Regression results using Exam as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	55.58**	[52.46, 58.70]						
x.centered	-2.15**	[-3.73, -0.57]	-0.23	[-0.39, -0.06]	.05	[-.02, .12]	-.24*	
z.centered	4.56**	[2.99, 6.14]	0.48	[0.31, 0.64]	.23	[.09, .36]	.50**	
I(x.centered * z.centered)	0.91*	[0.08, 1.73]	0.18	[0.02, 0.35]	.03	[-.02, .09]		
								R^2 = .340** 95% CI[.18,.45]

Note. * indicates $p < .05$; ** indicates $p < .01$. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights; *beta* indicates the standardized regression weights; sr^2 represents the semi-partial correlation squared; *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

If the slope for the product term is significant, there is a significant interaction between these two continuous variables. (Interactions among categorical variables will be covered in the ANOVA chapter).

Once you have established there is a significant interaction, you need to determine the pattern of results (i.e., explore the nature of the interaction). We will do in this in the next section.

3.4 Blocks approach

Note that some people are more comfortable with the equivalent approach below.

```
block1 <- lm(Exam ~ x.centered + z.centered,
                             data=analytic.data, na.action=na.exclude)

block2 <- lm(Exam ~ x.centered + z.centered + I(x.centered*z.centered),
                             data=analytic.data, na.action=na.exclude)
```

They conduct two regression (one with and without the product term). They then compare the regressions. Notice the delta-RSQ is the same as the sr^2 value from the above approach.

```
apa.reg.table(block1, block2)
```

Regression results using Exam as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit	Difference
(Intercept)	55.58**	[52.39, 58.76]							
x.centered	-2.33**	[-3.93, -0.73]	-0.24	[-0.41, -0.08]	.06	[-.02, .14]	-.24*		
z.centered	4.74**	[3.14, 6.34]	0.50	[0.33, 0.67]	.25	[.10, .39]	.50**		
								R^2 = .307** 95% CI[.15,.43]	
(Intercept)	55.58**	[52.46, 58.70]							
x.centered	-2.15**	[-3.73, -0.57]	-0.23	[-0.39, -0.06]	.05	[-.02, .12]	-.24*		
z.centered	4.56**	[2.99, 6.14]	0.48	[0.31, 0.64]	.23	[.09, .36]	.50**		
I(x.centered * z.centered)	0.91*	[0.08, 1.73]	0.18	[0.02, 0.35]	.03	[-.02, .09]			
								R^2 = .340** 95% CI[.18,.45]	ΔR^2 = .03* 95% CI[-.02, .09]

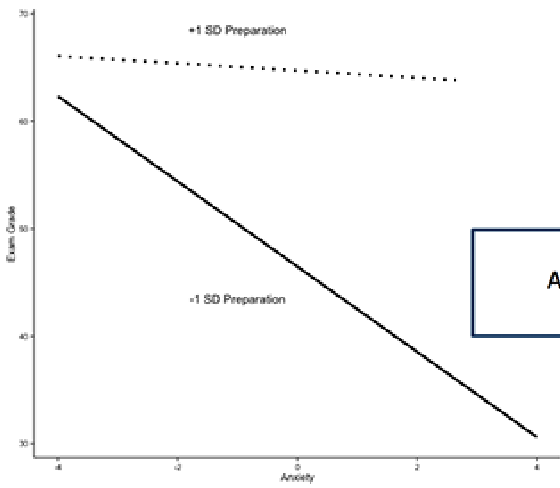
Note. * indicates $p < .05$; ** indicates $p < .01$. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights; *beta* indicates the standardized regression weights; sr^2 represents the semi-partial correlation squared; *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

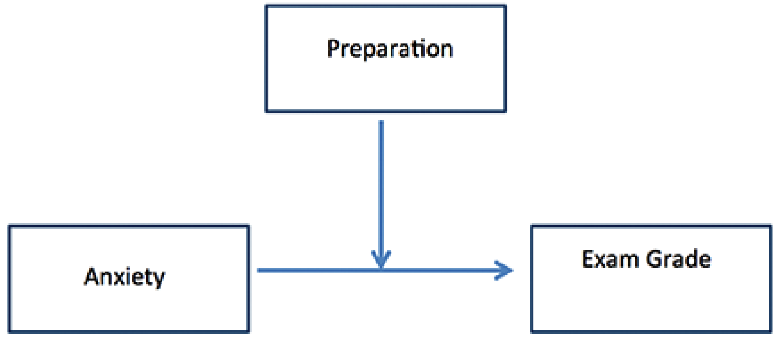
4 Interaction among continuous variables: Exploring the interaction

An interaction suggests that the relation between Anxiety (X) and Exam performance (Y) depends on the level of Preparation (Z). Therefore, we try to get an indication of the relation between Anxiety and Exam performance at low Preparation (-1 SD Preparation). As well, we try to get an indication of the relation between Anxiety and Exam performance at high Preparation (+1 SD Preparation). The significant interaction suggests the slopes are different at these points. We want to see what each slope is and whether it is significant. That is, we want to know if there is a relation between relation between Anxiety and Exam performance at each of these points.

Note that not entirely accurate to think in terms of just high/low preparation. Rather it is better to think in more general terms: the slope for Anxiety predicting Exam scores changes with the level of Preparation.

The graph illustrates the interaction – notice how the slopes are different. We will make this graph a bit later. Right now we want to determine the relation (i.e., slope) between Anxiety and Exam performance for each level of Preparation. In the context we say, the relation between Anxiety (X) and Exam performance (Y) is moderated by Preparation (Z); as indicated to the right of the graph.





Because we are using Anxiety (X) as a predictor and Preparation (Z) as the moderator we put Anxiety on the x-axis (because it is the predictor). Note the the distinction between which variable is the predictor or moderator is entirely arbitrary. As will be seen later, this is equivalent to simply picking one side or another to view the same 3-dimensional graph.

While using Preparation (Z) as the moderator, we want to know the relation between Anxiety (X) and Exam performance (Y) is at high and low levels of Preparation. More specifically, we want to know the relation between Anxiety (X) and Exam performance (Y) at 1 SD below the mean of Preparation (Z) and at 1 SD above the mean of Preparation (Z). Even more specifically, we want to know if the slope for Anxiety is significant at high and/or low Preparation.

4.1 Relation between Anxiety and Exam scores at high Preparation: Simple slope plus 1 SD Preparation

We want to get a regression line between for Anxiety predicting Exam scores (at +1SD Preparation) and determine if the slope for Anxiety is significant.

The first step is to transform Preparation scores so that zero corresponds to a spot 1 SD above the mean. When this is done, assigning a value of zero to Preparation in the regression equation gives us the relation between Anxiety and Exam performance at 1 SD above the mean. That is, because zero is used, Preparation (i.e., moderator terms / Z terms) drop out of the equation. Practically, this just means that to get the simple slope, you can just ignore any part of the output with Preparation in it after the transformation.

```
sd.z <- sd(analytic.data$z.centered, na.rm=TRUE)

analytic.data <- analytic.data %>%
     mutate(z.centered.at.plus.1SD = z.centered - sd.z)

#This may seem counter intuitive, but we lower the scores to increase the zero point to +1 SD

simple.slope.plus.1SD <- lm(Exam ~ x.centered + z.centered.at.plus.1SD
                               + I(x.centered*z.centered.at.plus.1SD),
                            data=analytic.data, na.action=na.exclude)

summary(simple.slope.plus.1SD)
```

```
##
## Call:
## lm(formula = Exam ~ x.centered + z.centered.at.plus.1SD + I(x.centered *
##     z.centered.at.plus.1SD), data = analytic.data, na.action = na.exclude)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34.71 -12.55   1.01  12.55  29.24
##
## Coefficients:
##                                      Estimate Std. Error t value
## (Intercept)                           64.7082     2.2374  28.922
## x.centered                            -0.3352     1.2081  -0.277
## z.centered.at.plus.1SD                 4.5648     0.7952   5.740
## I(x.centered * z.centered.at.plus.1SD)   0.9077     0.4165   2.180
##                                      Pr(>|t|)
## (Intercept)                           < 2e-16 ***
## x.centered                             0.7820
## z.centered.at.plus.1SD                1.1e-07 ***
## I(x.centered * z.centered.at.plus.1SD)   0.0317 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.74 on 96 degrees of freedom
## Multiple R-squared:  0.3396, Adjusted R-squared:  0.319
## F-statistic: 16.46 on 3 and 96 DF,  p-value: 1.047e-08
```

The summary output tells us the relation between Anxiety and Exam performance for individuals that prepared extensively (i.e., those at 1 SD above the mean on preparation). Because of the centering we can ignore lines with z.centered.at.plus.1SD in this output (i.e., ignore slopes that involve Preparation).

In this case at High Preparation (+1 SD): $\widehat{Y} = -.34(X) + 64.71$

In this case at High Preparation (+1 SD): $\widehat{Exam} = -.34(Anxiety) + 64.71$

Go back to the graph and notice how this near zero slope is depicted. The non-significant weight for Anxiety indicates that there is no relation between Anxiety and Exam performance for high preparation participants.

4.2 Relation between Anxiety and Exam scores at low Preparation: Simple slope minus 1 SD Preparation

We want to get a regression line between for Anxiety predicting Exam scores (at -1SD Preparation) and determine if the slope for Anxiety is significant.

We use the same process as above, but we now transform preparation scores so that zero corresponds to a spot 1 SD below the mean. This may seem counter intuitive, but we raise the scores to decrease the zero point to -1 SD.

```
analytic.data <- analytic.data %>%
                 mutate(z.centered.at.minus.1SD=z.centered + sd.z)

simple.slope.minus.1SD <- lm(Exam ~ x.centered + z.centered.at.minus.1SD
                             + I(x.centered*z.centered.at.minus.1SD),
                        data=analytic.data,na.action=na.exclude)

summary(simple.slope.minus.1SD)
```

```
##
## Call:
## lm(formula = Exam ~ x.centered + z.centered.at.minus.1SD + I(x.centered *
##     z.centered.at.minus.1SD), data = analytic.data, na.action = na.exclude)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34.71 -12.55   1.01  12.55  29.24
##
## Coefficients:
##                                         Estimate Std. Error t value
## (Intercept)                              46.4506     2.2375  20.760
## x.centered                               -3.9655     1.0916  -3.633
## z.centered.at.minus.1SD                   4.5648     0.7952   5.740
## I(x.centered * z.centered.at.minus.1SD)   0.9077     0.4165   2.180
##                                         Pr(>|t|)
## (Intercept)                              < 2e-16 ***
## x.centered                              0.000453 ***
## z.centered.at.minus.1SD                  1.1e-07 ***
## I(x.centered * z.centered.at.minus.1SD) 0.031739 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.74 on 96 degrees of freedom
## Multiple R-squared:  0.3396, Adjusted R-squared:  0.319
## F-statistic: 16.46 on 3 and 96 DF,  p-value: 1.047e-08
```

As before, the summary output tells us the relation between Anxiety and Exam performance for individuals that did not prepare extensively (i.e., those at 1 SD below the mean on preparation). Because of the centering we can ignore slopes with z.centered.at.minus.1SD in this output.

In this case at Low Preparation (-1 SD): $\widehat{Y} = -3.97(X) + 46.45$

In this case at Low Preparation (-1 SD): $\widehat{Exam} = -3.97(Anxiety) + 46.45$

Go back to the graph and notice how this near large negative slope is depicted.

The significant weight for anxiety indicates that there is a relation between Anxiety and Exam performance for low Preparation participants. Specifically, for low participants low in Preparation, as Anxiety scores increase by 1 Exam performance decreases by 3.97 percent.

5 Making a 2D graph (Part 1)

5.1 Indicate range of scores on the X-axis for the two lines

We pick Anxiety to be X.

```
sd.x <- sd(analytic.data$x.centered, na.rm=TRUE)

#we want the x-axis of the graph to range from -2 SD to +2 SD
x.axis.range <-seq(-2*sd.x,2*sd.x,by=.25*sd.x)
```

5.2 Indicate the values of Z that indicate high/low (preparation) for the two lines.

We pick Preparation to be Z.

We want two lines. One line representing high preparation (+1 SD preparation) and one line representing low preparation (-1 SD preparation).

```
sd.z<-sd(analytic.data$z.centered, na.rm=TRUE)
z.line.hi=  1*sd.z
z.line.lo= -1*sd.z
```

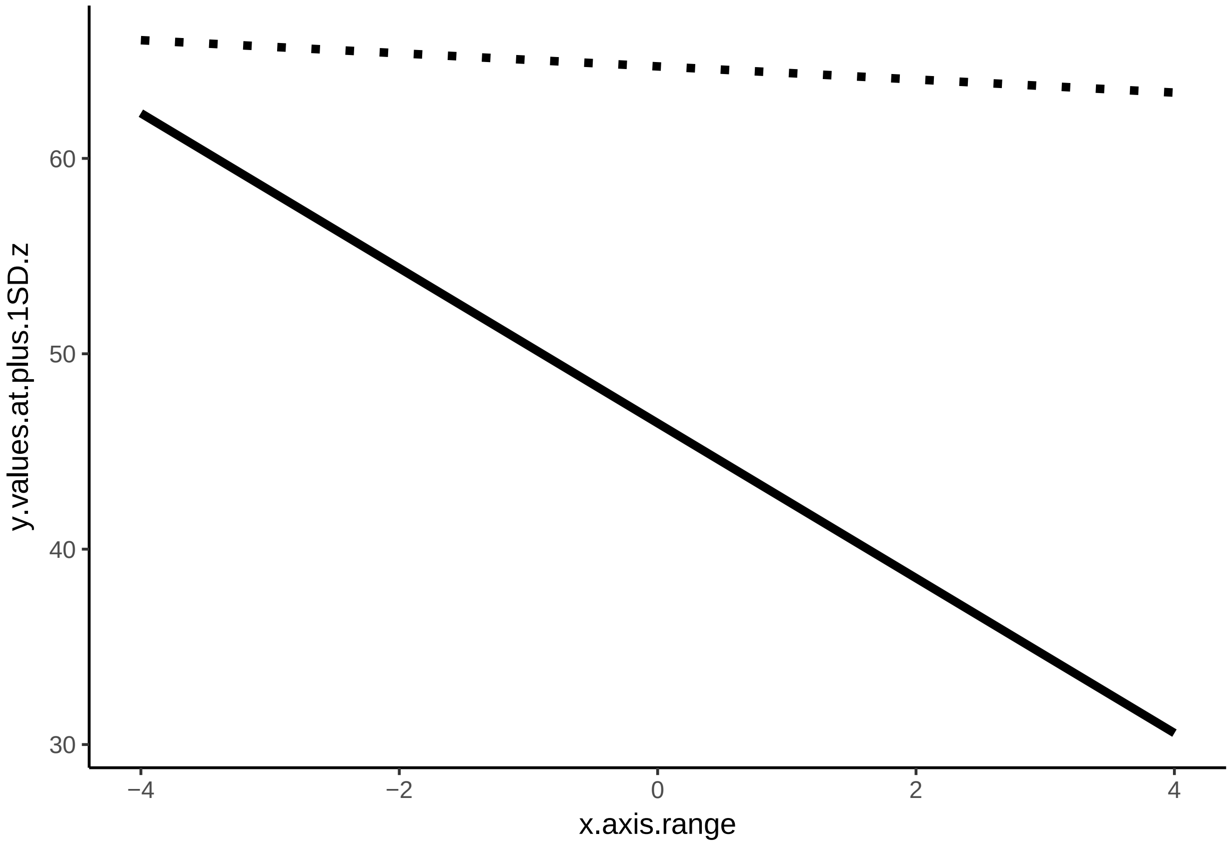
5.3 Created predicted values for each line

```
#+1SD Line
predictor.x.range.line.hi <- expand.grid(x.centered=x.axis.range, z.centered=z.line.hi)
y.values.at.plus.1SD.z <- predict(interaction.regression,newdata=predictor.x.range.line.hi)

#-1SD Line
predictor.x.range.line.lo <- expand.grid(x.centered=x.axis.range, z.centered=z.line.lo)
y.values.at.minus.1SD.z <- predict(interaction.regression,newdata=predictor.x.range.line.lo)

#Put the information describing the lines into a data frame
line.data <- data.frame(x.axis.range, y.values.at.plus.1SD.z, y.values.at.minus.1SD.z)
```

6 Make the 2D Graph (Part 2)

```
library(ggplot2)
#set default (x,y) variables
my.plot <- ggplot(line.data, aes(x=x.axis.range, y=y.values.at.plus.1SD.z))

#make +1 SD Z line (because it is the one in the aes statement above)
my.plot <- my.plot + geom_line(color="black",linetype="dotted",size=1.5)

#make the -1 SD Z line
my.plot <- my.plot + geom_line(aes(x=x.axis.range,y=y.values.at.minus.1SD.z),
                           color="black",linetype="solid",size=1.5)

#set APA part of graph below
```

```r
my.plot <- my.plot + theme_classic()

print(my.plot)
```



6.1 Label the lines.

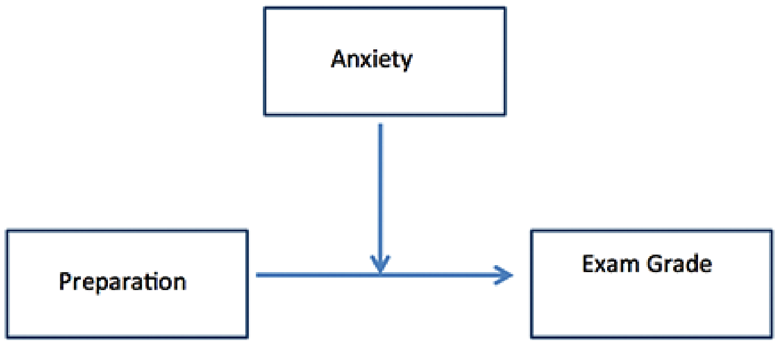
```r
my.plot <- my.plot+annotate("text", x = -1, y = 68.5, label = "+1 SD Preparation")
my.plot <- my.plot+annotate("text", x = -1, y =43.5, label = "-1 SD Preparation")
```

If you add the annotations more than once you will need to run the plot part of the script again (i.e., everything after library(ggplot2)). Try using the *angle* argument to put these parallel to each line. For example, just add angle=-45 in the brackets. Perhaps include the regression equation in each annotation.

7 Try it again: Make Preparation the Predictor and Anxiety the Moderator

As noted above, the distinction between predictor and moderator is arbitrary. Therefore, run the above analyses again but make Preparation the predictor and Anxiety the Moderator. What does the graph look like? How do you reconcile the two graphs?

Go beyond the last graph. Create lines for -2SD, -1SD, +1SD, and +2SD Anxiety.



8 Other Non-linear terms

8.1 Initial regressions:

Sometimes you want to look for other product terms. You would run the regression with those terms.

```
interaction.regression <- lm(Exam ~ x.centered + z.centered
                             + I(x.centered*z.centered)
                             + I(x.centered*x.centered) + I(z.centered*z.centered),
                             data=analytic.data,na.action=na.exclude)
summary(interaction.regression)
```

```
##
## Call:
## lm(formula = Exam ~ x.centered + z.centered + I(x.centered *
##     z.centered) + I(x.centered * x.centered) + I(z.centered *
##     z.centered), data = analytic.data, na.action = na.exclude)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.303 -12.557   1.222  12.710  30.322
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                56.31362    2.47519  22.751  < 2e-16 ***
## x.centered                 -2.18813    0.85227  -2.567   0.0118 *
## z.centered                  4.55344    0.80571   5.651 1.69e-07 ***
## I(x.centered * z.centered)  0.89832    0.42358   2.121   0.0366 *
## I(x.centered * x.centered) -0.06564    0.31502  -0.208   0.8354
## I(z.centered * z.centered) -0.11977    0.35758  -0.335   0.7384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.89 on 94 degrees of freedom
## Multiple R-squared:  0.3407, Adjusted R-squared:  0.3057
## F-statistic: 9.716 on 5 and 94 DF,  p-value: 1.669e-07
```

8.2 Simple slopes:

Activity: Figure out the two simple slope equations for this case. Hint: there will be more than one slope in each equation.

8.3 Graph data:

Exactly the same as before:

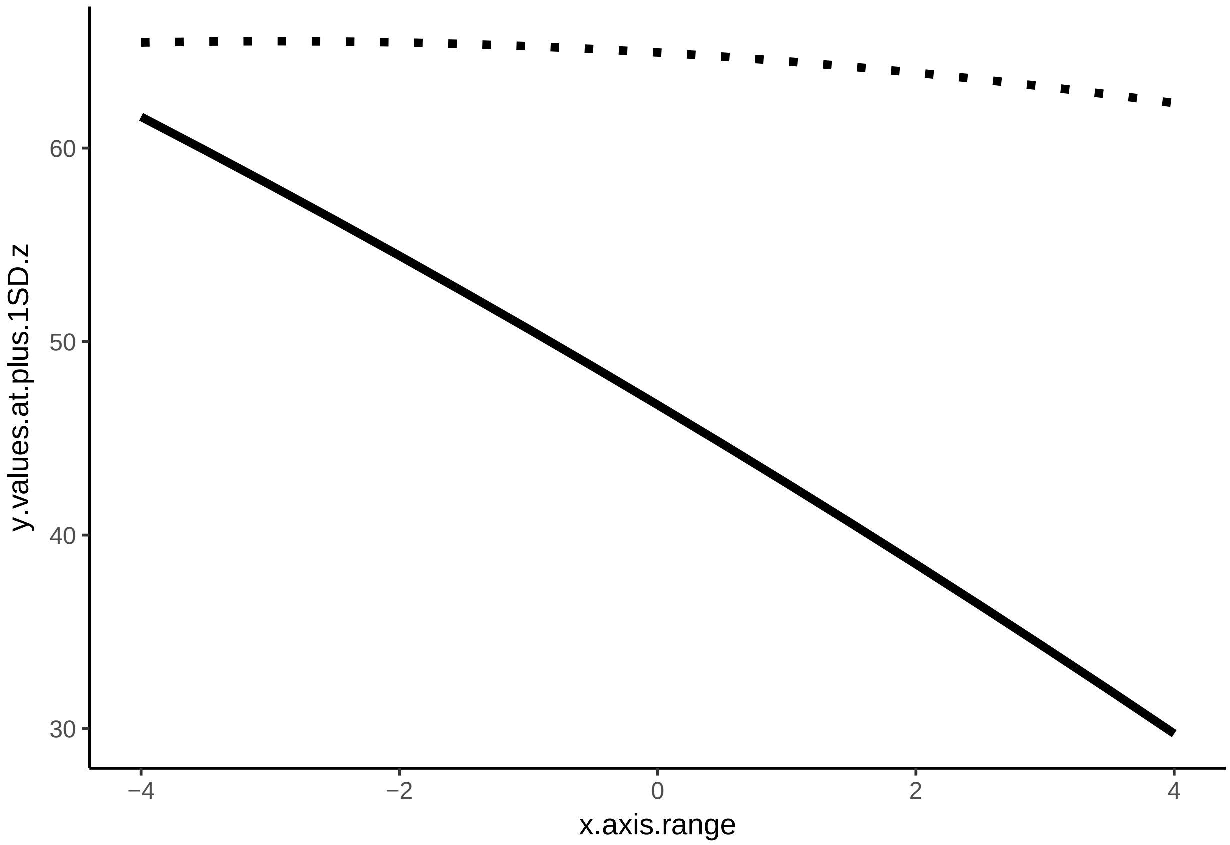
```
sd.x <- sd(analytic.data$x.centered, na.rm=TRUE)
#we want the x-axis of the graph to range from -2 SD to +2 SD
x.axis.range <-seq(-2*sd.x,2*sd.x,by=.25*sd.x)

sd.z<-sd(analytic.data$z.centered, na.rm=TRUE)
z.line.hi=  1*sd.z
z.line.lo= -1*sd.z
```

```r
#+1SD Line
predictor.x.range.line.hi <- expand.grid(x.centered=x.axis.range, z.centered=z.line.hi)
y.values.at.plus.1SD.z <- predict(interaction.regression,newdata=predictor.x.range.line.hi)

#-1SD Line
predictor.x.range.line.lo <- expand.grid(x.centered=x.axis.range, z.centered=z.line.lo)
y.values.at.minus.1SD.z <- predict(interaction.regression,newdata=predictor.x.range.line.lo)

#Put the information describing the lines into a data frame
line.data <- data.frame(x.axis.range, y.values.at.plus.1SD.z, y.values.at.minus.1SD.z)
```

8.4 Making the graph:

Exactly the same as before:

```r
library(ggplot2)
#set default (x,y) variables
my.plot <- ggplot(line.data, aes(x=x.axis.range, y=y.values.at.plus.1SD.z))

#make +1 SD Z line (because it is the one in the aes statement above)
my.plot <- my.plot + geom_line(color="black",linetype="dotted",size=1.5)

#make the -1 SD Z line
my.plot <- my.plot + geom_line(aes(x=x.axis.range,y=y.values.at.minus.1SD.z),
                            color="black",linetype="solid",size=1.5)

#set APA part of graph below
my.plot <- my.plot + theme_classic()

print(my.plot)
```



Notice how the lines are now curved to reflect the non-linear terms.

9 Making a 3D Graph

9.1 Look at the regression weights from the interaction regression.

Let's return to just looking at the X-Z product term.

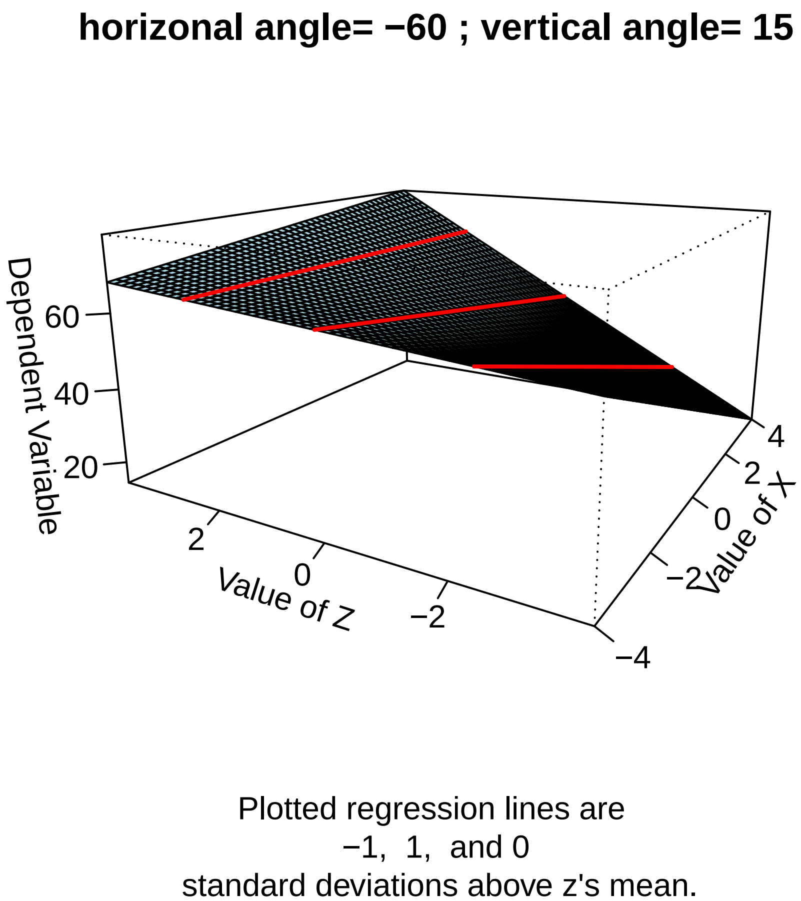
```
interaction.regression <- lm(Exam ~ x.centered + z.centered
                             + I(x.centered*z.centered),
                             data=analytic.data,na.action=na.exclude)
```

Notice the intercept and b-weights. Enter them into the intr.plot function on the next page.

```
summary(interaction.regression)
```

```
##
## Call:
## lm(formula = Exam ~ x.centered + z.centered + I(x.centered *
##     z.centered), data = analytic.data, na.action = na.exclude)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34.71 -12.55   1.01  12.55  29.24
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 55.5794     1.5739  35.313  < 2e-16 ***
## x.centered                  -2.1503     0.7949  -2.705  0.00808 **
## z.centered                   4.5648     0.7952   5.740  1.1e-07 ***
## I(x.centered * z.centered)   0.9077     0.4165   2.180  0.03174 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.74 on 96 degrees of freedom
## Multiple R-squared:  0.3396, Adjusted R-squared:  0.319
## F-statistic: 16.46 on 3 and 96 DF,  p-value: 1.047e-08
```

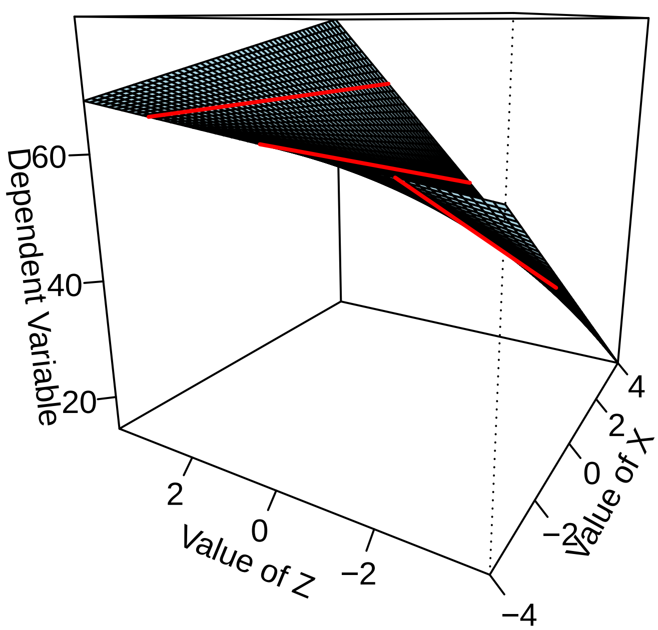
```
library(MBESS)
intr.plot(b.0=55.5794,b.x=-2.1503,b.z=4.5648,b.xz=0.9077,
          x.min=-2*sd.x,x.max=2*sd.x,z.min=-2*sd.z,z.max=2*sd.z)
```



9.2 Make the plot square by adding expand=1 (this increases the y-axis size to be proportional to the other).

```
intr.plot(b.0=55.5794,b.x=-2.1503,b.z=4.5648,b.xz=0.9077,
          x.min=-2*sd.x,x.max=2*sd.x,z.min=-2*sd.z,z.max=2*sd.z, expand=1)
```

horizonal angle= −60 ; vertical angle= 15

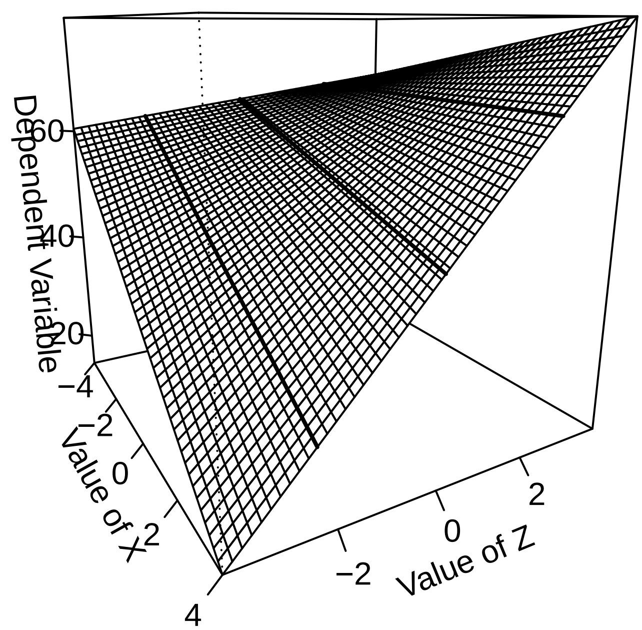


Plotted regression lines are
−1, 1, and 0
standard deviations above z's mean.

9.3 Set the angle and make the plot grayscale.

```
intr.plot(b.0=55.5794,b.x=-2.1503,b.z=4.5648,b.xz=0.9077,
          x.min=-2*sd.x,x.max=2*sd.x,z.min=-2*sd.z,z.max=2*sd.z,
          expand=1, ,hor.angle=60,gray.scale=TRUE)
```

horizonal angle= 60 ; vertical angle= 15

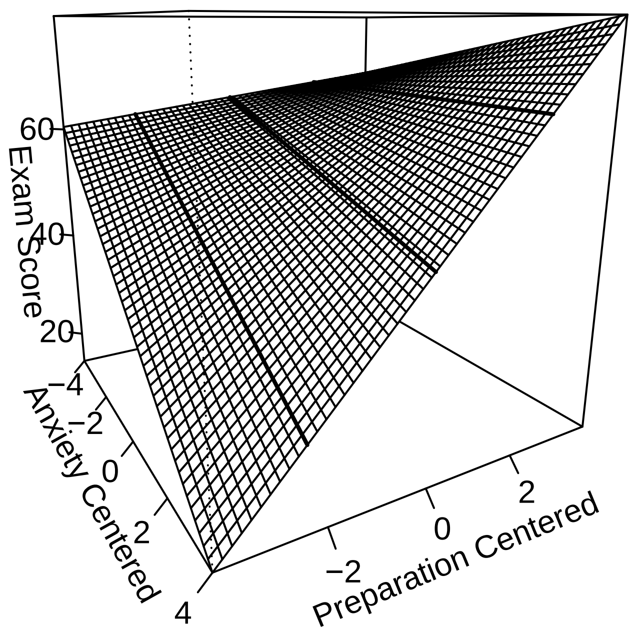


Plotted regression lines are
−1, 1, and 0
standard deviations above z's mean.

9.4 Now add axis labels.

```
intr.plot(b.0=55.5794,b.x=-2.1503,b.z=4.5648,b.xz=0.9077,
          x.min=-2*sd.x,x.max=2*sd.x,z.min=-2*sd.z,z.max=2*sd.z,
          xlab="Anxiety Centered",zlab="Preparation Centered",ylab="Exam Score",
          expand=1,hor.angle=60,gray.scale=TRUE)
```

horizonal angle= 60 ; vertical angle= 15



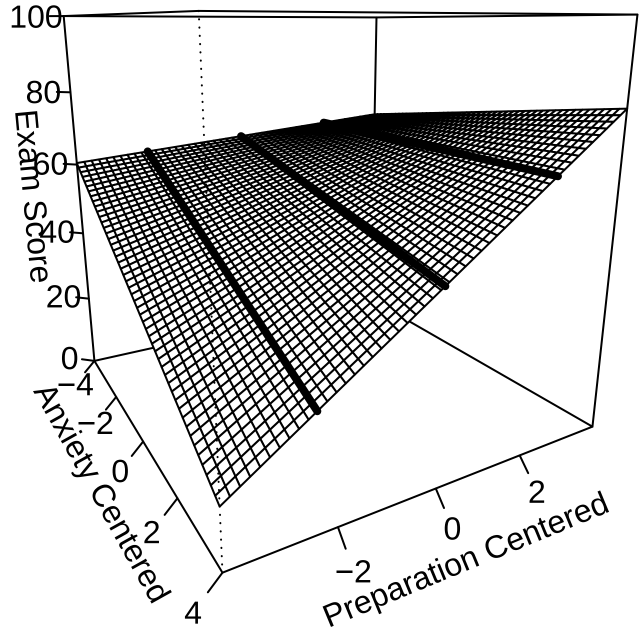
Plotted regression lines are
−1, 1, and 0
standard deviations above z's mean.

9.5 Here is the tricky part.

You want to adjust the y-axis (i.e., vertical axis) for Exam to the range of grades (0 to 100). Unfortunately, due to a naming convention issue with different parts of R, you need to refer to the y-axis (i.e., vertical axis) as the z-axis (even though this conflicts with the rest of the information we are using). In the line below zlim=c(0,100) is actually setting the y-axis from 0 to 100. Also note we increased the line width (line.wd).

```
intr.plot(b.0=55.5794,b.x=-2.1503,b.z=4.5648,b.xz=0.9077,
          x.min=-2*sd.x,x.max=2*sd.x,z.min=-2*sd.z,z.max=2*sd.z,
          xlab="Anxiety Centered",zlab="Preparation Centered",ylab="Exam Score",
          expand=1,hor.angle=60,gray.scale=TRUE,line.wd=4,zlim=c(0,100))
```

horizonal angle= 60 ; vertical angle= 15



Plotted regression lines are
−1, 1, and 0
standard deviations above z's mean.

Note: ggsave won't work with the graph. But if you are using R Studio – the saving options there will work.

10 If you have a MAC, there is an easier way.

Use FastInteraction see http://www.fastinteraction.com. It is a free download from the Mac App Store.

Warning. If you have missing data you need to save it in a particular way before using Fast Interaction. Save your data from with R using the command below:

```
write_csv(x=analytic.data,path="myDataOut.csv", na="")
```

The na="" is essential - it ensure missing values are written the file correctly. Load "myDataOut.csv" using Fast Interaction and quickly make the plot.

11 Mediation

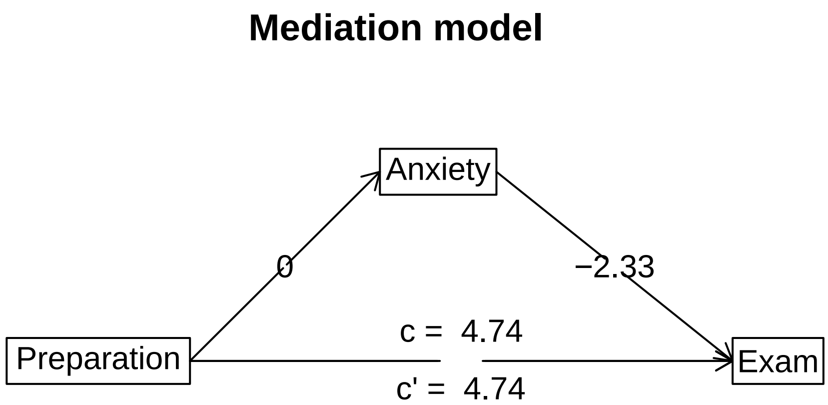
11.1 Mediation if you insist on regression

We didn't a chance to talk about mediation. As your other readings indicate, it's generally not a good idea to do mediation analyses with regression unless you have an extraordinarily large sample size (thousands of people).

Furthermore, methods/statistics researchers are starting to reject all techniques of this nature given the problems associated with them. The current trend is away from establishing mediation via statistics to establishing mediation via experimental interventions.

Imagine we reconceptualize our data as representing mediation. That is, Exam grade is still the DV, but we think that preparation reduces anxiety, which in turn increases Exam grade.

```
mediation.data <- analytic.data %>% select(Exam, Preparation, Anxiety)

psych::mediate(y=1,x=2, m=3, data=mediation.data)
```




```
## Call: psych::mediate(y = 1, x = 2, m = 3, data = mediation.data)
##
## The DV (Y) was  Exam . The IV (X) was  Preparation . The mediating variable(s) =  Anxiety .
##
## Total Direct effect(c) of  Preparation  on  Exam  =  4.74   S.E. =  0.84  t direct =  5.68   with pro
## Direct effect (c') of  Preparation  on  Exam  removing  Anxiety  =  4.74   S.E. =  0.81  t direct =
## Indirect effect (ab) of  Preparation  on  Exam  through  Anxiety   =  0
## Mean bootstrapped indirect effect =  0  with standard error =  0.24  Lower CI =  -0.49    Upper CI =
## R2 of model =  0.31
##  To see the longer output, specify short = FALSE in the print statement
##
##  Full output
##
##  Total effect estimates (c)
##             Exam   se    t    Prob
## Preparation 4.74 0.84 5.68 1.4e-07
##
## Direct effect estimates     (c')
##              Exam   se     t     Prob
## Preparation  4.74 0.81  5.88 5.73e-08
## Anxiety     -2.33 0.81 -2.89 4.82e-03
##
##  'a'  effect estimates
##         Preparation  se     t  Prob
## Anxiety           0 0.1 -0.01 0.994
##
##  'b'  effect estimates
##          Exam   se     t    Prob
## Anxiety -2.33 0.81 -2.89 0.00482
##
##  'ab'  effect estimates
```

```
##             Exam boot   sd lower upper
## Preparation    0    0 0.24 -0.49   0.5
```

```
#column numbers are 1 2 and 3 for Exam, Prep, and Anx respectively
```

11.2 Mediation using structural equation modelling (illustration only, topic covered in PSYC*6380 Psychological Applications of Multivariate Analysis)

A better approach to testing mediation models is to use structural equation modeling (SEM). See the paper below.

James, L. R., Mulaik, S. A., & Brett, J. M. (2006). A tale of two methods. Organizational research methods, 9(2), 233-244. Available here: [http://www.aipass.org/files/12.pdf]

Mediation using SEM is easily done in R using the lavaan package.

Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL [http://www.jstatsoft.org/v48/i02/]