# Setting options for displaying number, to minimize

# use of scientific notation and to only show a

# limited number of decimals

options(scipen=100)

options(digits=3)

# SeriousStatsAllfunctions script - from Thom

# Baguley. We will use cm.ci.mixed function

# to produce Cousineau-Morey confidence intervals

# for a mixed design

source("http://www2.ntupsychology.net/seriousstats/SeriousStatsAllfunctions.txt")

# afex package "aov\_car" function to run our

# mixed-design ANOVA

# effects package "effect" function to calculate

# predictions from a regression

# ggplot2 package "ggplot" function to make graphs

# GGally package "ggpairs" function to make

# exploratory data analysis matrix

# jtools package "sim\_slopes" function to probe

# interaction in regression

# pastecs package "stat.desc" function to get summary

# statistics

# phia package "testInteractions" function to do

# contrast analyses in our ANOVA

# reshape2 "melt" to go from a person-level

# dataframe to a person-period

# dataframe--necessary for mixed-

# design ANOVA

# stargazer "stargazer" to make nice regression

# tables

install.packages("afex", dependencies=TRUE)

install.packages("effects", dependencies=TRUE)

install.packages("ggplot2", dependencies=TRUE)

install.packages("GGally", dependencies=TRUE)

install.packages("jtools", dependencies=TRUE)

install.packages("pastecs", dependencies=TRUE)

install.packages("phia", dependencies=TRUE)

install.packages("reshape2", dependencies=TRUE)

install.packages("stargazer", dependencies=TRUE)

library("afex")

library("effects")

library("ggplot2")

library("GGally")

library("jtools")

library("pastecs")

library("phia")

library("reshape2")

library("stargazer")

# Reading in Data

# tachistoscope experiment - subjects see either an "I"

# or a "T" and they have to correctly identify what

# the letter is.

# outcome is response time in milliseconds

# between-group condition: age (young vs old)

# 1=young, 2=old

# (I prefer dummy coding (0,1) but the cm.ci.mixed

# function will only work when categories are coded

# starting at 1)

# within-group condition: angle at which the letter

# appears off-center.

# seq function - we will ask R to make ID variable

# from 1 to 40

# rep function - we will ask R to assign the first

# 20 cases to the "young" condition

# (old=1) and the next 20 cases to

# the "old" condition (old=2)

tach<-data.frame(id=c(seq(from=1, to=40)),

old=c(rep(1,20), rep(2,20)),

angle0=c(450,390,570,450,510,360,

510,510,510,510,450,510,

480,470,460,490,570,450,

520,390,420,600,450,630,

420,600,630,480,690,510,

640,740,350,650,680,410,

570,570,410,590),

angle4=c(510,480,630,660,660,450,

600,660,660,540,550,650,

530,570,640,510,630,780,

440,550,570,720,540,660,

570,780,690,570,750,690,

640,750,620,670,700,560,

720,480,780,640),

angle8=c(630,540,660,720,630,450,

720,780,660,660,700,560,

670,670,580,490,440,760,

690,640,690,810,690,780,

780,870,870,720,900,810,

840,850,810,690,600,640,

920,830,820,730))

# creating factor variable for "old"

tach$oldf<-factor(tach$old, levels=c(1,2),

labels=c("young", "old"))

# Looking at Data - note it is in "wide" format--

# one line per participant

|  |
| --- |
| **head(tach)**  **id old angle0 angle4 angle8 oldf**  **1 1 1 450 510 630 young**  **2 2 1 390 480 540 young**  **3 3 1 570 630 660 young**  **4 4 1 450 660 720 young**  **5 5 1 510 660 630 young**  **6 6 1 360 450 450 young**  **tail(tach)**  **id old angle0 angle4 angle8 oldf**  **35 35 2 680 700 600 old**  **36 36 2 410 560 640 old**  **37 37 2 570 720 920 old**  **38 38 2 570 480 830 old**  **39 39 2 410 780 820 old**  **40 40 2 590 640 730 old** |

# Getting summary statistics by the old grouping

# variable

by(tach[c("angle0", "angle4", "angle8")],

tach$old, stat.desc)

|  |
| --- |
| **tach$old: 1**  **angle0 angle4 angle8**  **nbr.val 20.000 20.000 20.000**  **nbr.null 0.000 0.000 0.000**  **nbr.na 0.000 0.000 0.000**  **min 360.000 440.000 440.000**  **max 570.000 780.000 780.000**  **range 210.000 340.000 340.000**  **sum 9560.000 11700.000 12650.000**  **median 485.000 585.000 660.000**  **mean 478.000 585.000 632.500**  **SE.mean 12.346 19.283 21.372**  **CI.mean.0.95 25.840 40.360 44.733**  **var 3048.421 7436.842 9135.526**  **std.dev 55.213 86.237 95.580**  **coef.var 0.116 0.147 0.151**  **-------------------------------------------**  **tach$old: 2**  **angle0 angle4 angle8**  **nbr.val 20.000 20.00 20.000**  **nbr.null 0.000 0.00 0.000**  **nbr.na 0.000 0.00 0.000**  **min 350.000 480.00 600.000**  **max 740.000 780.00 920.000**  **range 390.000 300.00 320.000**  **sum 11040.000 13100.00 15650.000**  **median 580.000 665.00 810.000**  **mean 552.000 655.00 782.500**  **SE.mean 25.148 19.10 19.720**  **CI.mean.0.95 52.635 39.99 41.275**  **var 12648.421 7300.00 7777.632**  **std.dev 112.465 85.44 88.191**  **coef.var 0.204 0.13 0.113** |

# Getting Cousineau-Morey confidence intervals

# for within-subject designs

# Cousineau-Morey CIs essentially subtract the

# subject mean from the raw scores and then

# add the between-subject group mean, and

# applies Morey adjustment

# see Baguley 2012 for details

# http://dx.doi.org/10.3758/s13428-011-0123-7

# last variable listed has to be grouping

# variable

cm.ci.mixed(tach[c("angle0", "angle4", "angle8",

"old")], difference=FALSE)

**[[1]]**

**lower upper**

**angle0 448 508**

**angle4 556 614**

**angle8 597 668**

**[[2]]**

**lower upper**

**angle0 512 592**

**angle4 626 684**

**angle8 746 819**

# Creating a new data frame containing sample

# means and CIs

gtach<-data.frame(old=c(rep("Young",3),

rep("Old", 3)),

angle=c(0,4,8,0,4,8),

meanvar=c(478,585,632,552,655,782),

lower=c(448,556,597,512,626,746),

upper=c(508,614,668,592,684,819))

gtach

**old angle meanvar lower upper**

**1 Young 0 478 448 508**

**2 Young 4 585 556 614**

**3 Young 8 632 597 668**

**4 Old 0 552 512 592**

**5 Old 4 655 626 684**

**6 Old 8 782 746 819**

# Graphing Means

# ggplot function is based on "grammar of graphics"

# the following is a combination of paraphrases and

# verbatim quotes from Hadley Wickham's book

# "ggplot2: Elegant Graphics for Data Analysis: Second

# Edition", pp. 4-5

# plots are composed of:

# DATA and AESthetic mappings linking data to

# aesthetic attributes

# Layers made up of GEOMetric elements--

# represent what we actually see on the plot

# SCALEs mapping values in data space to values

# in an aesthetic space

# FACETs describe how to break up the data into

# subsets

# themes controlling finer points of display,

# like LABels and TITLEs

# Note that I "dodged" (aka "jittered") the error

# bars a bit--just so they are distinguishable in

# cases where the lines are too close to each

# other.

g<-ggplot(data=gtach, aes(x=angle, y=meanvar,

linetype=old, color=old))+

geom\_line(size=1)+geom\_point()+

geom\_errorbar(aes(ymin=lower, ymax=upper), width=.25,

position=position\_dodge(width=.2))+

labs(linetype=" ", color=" ")+

xlab("Angle")+

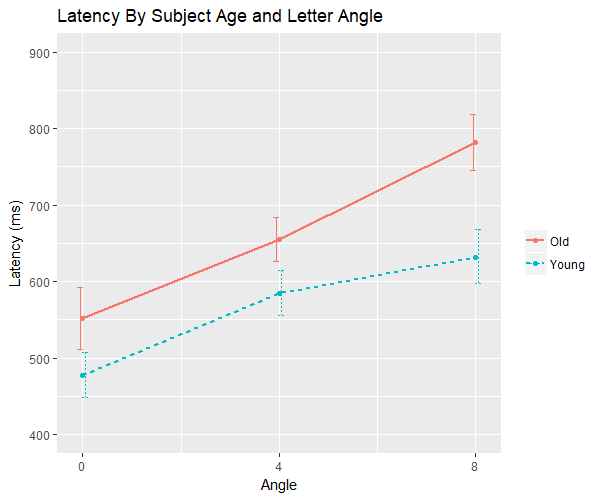
ylab("Latency (ms)")+

ggtitle("Latency By Subject Age and Letter Angle")+

scale\_x\_continuous(breaks=c(0,4,8))+

ylim(c(400,900))

g



# Reshaping tach dataframe into long format--one line

# per observation

tach\_long<-melt(tach, id=c("id","old","oldf"),

Measured=c("angle0","angle4","angle8"))

# Looking at melted data

**head(tach\_long)**

**id old oldf angle latency**

**1 1 1 young angle0 450**

**2 2 1 young angle0 390**

**3 3 1 young angle0 570**

**4 4 1 young angle0 450**

**5 5 1 young angle0 510**

**6 6 1 young angle0 360**

**tail(tach\_long)**

**id old oldf angle latency**

**115 35 2 old angle8 600**

**116 36 2 old angle8 640**

**117 37 2 old angle8 920**

**118 38 2 old angle8 830**

**119 39 2 old angle8 820**

**120 40 2 old angle8 730**

# renaming "variable" to "angle" and "value" to

# "latency"

names(tach\_long)[names(tach\_long)=="variable"]<-"angle"

names(tach\_long)[names(tach\_long)=="value"]<-"latency"

# doing omnibus test

m1<-aov\_car(latency~oldf+Error(id/angle),

data=tach\_long)

m1

**Anova Table (Type 3 tests)**

**Response: latency**

**Effect df MSE F ges p.value**

**1 oldf 1, 38 13368.33 21.55 \*\*\* .24 <.0001**

**2 angle 1.88, 71.34 5489.46 72.12 \*\*\* .45 <.0001**

**3 oldf:angle 1.88, 71.34 5489.46 3.94 \* .04 .03**

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘+’ 0.1 ‘ ’ 1**

**Sphericity correction method: GG**

# checking levels for angle and old factors to make

# sure we have correct order

levels(tach\_long$angle)

**[1] "angle0" "angle4" "angle8"**

levels(tach\_long$oldf)

**[1] "young" "old"**

# main effects contrasts, old vs young

# multiple comparisons not an issue

testInteractions(m1$lm,

custom=list(oldf=c(-1,1)),

idata=m1$data[["idata"]],

adjustments="none")

**Multivariate Test: Pillai test statistic**

**P-value adjustment method: holm**

**Value Df test stat approx F num Df den Df Pr(>F)**

**oldf1 98 1 0.362 21.6 1 38 0.00004 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

# main effects contrasts, angle 0 vs angle 4

# note: sometimes phia will use alphabetical order

# for your factors, I think this occurs for within-

# subject factors for some reason.

# not an issue here because angle already is in

# alphanumerical order

testInteractions(m1$lm,

custom=list(angle=c(-1,1,0)),

idata=m1$data[["idata"]],

adjustments="none")

**Multivariate Test: Pillai test statistic**

**P-value adjustment method: holm**

**Value Df test stat approx F num Df den Df Pr(>F)**

**angle1 105 1 0.567 49.7 1 38 0.000000021 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

# Tukey-adjusted F crit value:

((qtukey(1-(.05), 3, 38))^2)/(2)

|  |
| --- |
| **[1] 5.95** |

# simple effects contrast:

# old vs young at angle 0

testInteractions(m1$lm,

custom=list(oldf=c(-1,1),

angle=c(1,0,0)),

idata=m1$data[["idata"]],

adjustments="none")

**Multivariate Test: Pillai test statistic**

**P-value adjustment method: holm**

**Value Df test stat approx F num Df den Df Pr(>F)**

**oldf1 : angle1 74 1 0.155 6.98 1 38 0.012 \***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

# Tukey-adjusted F crit value:

((qtukey(1-(.05/3), 2, 38))^2)/(2)

**[1] 6.27**

# simple effects contrast:

# angle4 vs angle 8 for young people

testInteractions(m1$lm,

custom=list(oldf=c(1,0),

angle=c(0,-1,1)),

idata=m1$data[["idata"]],

adjustments="none")

**Multivariate Test: Pillai test statistic**

**P-value adjustment method: holm**

**Value Df test stat approx F num Df den Df Pr(>F)**

**oldf1 : angle1 47.5 1 0.115 4.95 1 38 0.032 \***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

# Tukey-adjusted F crit value:

((qtukey(1-(.05/2), 3, 38))^2)/(2)

**[1] 7.48**

# interaction contrast:

# angle 4 vs angle 8 for young vs old people

testInteractions(m1$lm,

custom=list(oldf=c(-1,1),

angle=c(0,-1,1)),

idata=m1$data[["idata"]],

adjustments="none")

**Multivariate Test: Pillai test statistic**

**P-value adjustment method: holm**

**Value Df test stat approx F num Df den Df Pr(>F)**

**oldf1 : angle1 80 1 0.156 7.03 1 38 0.012 \***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

# Scheffe-adjusted crit value:

2\*qf(.95,2,38)

**[1] 6.49**

# regression example

# this is data on scientists

# timeout - years since PhD

# numpubs - # of publications

# numcites - citation ccount

# female - 1=female, 0=male

# reading in dataframe

sci<-data.frame(timeout=c(3,6,3,8,9,6,16,

10,2,5,5,6,7,11,

18,6,9,7,7,3,7,

5,7,13,5,8,8,7,

2,13,5,3,1,3,9,

3,9,3,4,10,1,11,

5,1,21,7,5,16,5,

4,5,11,16,3,4,4,

5,6,4,8,3,4),

numpubs=c(18,3,2,17,11,6,38,

48,9,22,30,21,10,27,

37,8,13,6,12,29,29,

7,6,69,11,9,20,41,

3,27,14,23,1,7,19,

11,31,9,12,32,26,12,

9,6,39,16,12,50,18,

16,5,20,50,6,19,11,

13,3,8,11,25,4),

female=c(1,1,1,0,1,0,0,0,0,0,

1,0,1,0,0,0,1,0,1,1,

1,0,0,0,0,1,1,1,1,0,

0,0,0,0,0,0,0,0,1,0,

0,0,0,0,0,1,1,0,0,1,

0,0,1,1,1,1,0,1,1,1,

1,1),

numcite=c(50,26,50,34,41,37,48,

56,19,29,28,31,25,40,

61,32,36,69,47,29,35,

35,18,90,60,30,27,35,

14,56,50,25,35,1,69,

69,27,50,32,33,45,54,

47,29,69,47,43,55,33,

28,42,24,31,27,83,49,

14,36,34,70,27,28),

salary=c(51876,54511,53425,61863,

52926,47034,66432,61100,

41934,47454,49832,47047,

39115,59677,61458,54528,

60327,56600,52542,50455,

51647,62895,53740,75822,

56596,55682,62091,42162,

52646,74199,50729,70011,

37939,39652,68987,55579,

54671,57704,44045,51122,

47082,60009,58632,38340,

71219,53712,54782,83503,

47212,52840,53650,50931,

66784,49751,74343,57710,

52676,41195,45662,47606,

44301,58582))

# creating factor for female

sci$femalef<-factor(sci$female, levels=c(0,1),

labels=c("Male", "Female"))

# ggpairs - making exploratory data analysis

# note I have found it is best to specify

# factor variables LAST in list of

# variables.

# matrix

# lower argument - specifying we want loess

# scatterplots below diagonal

# showstrips - will add "male" and "female"

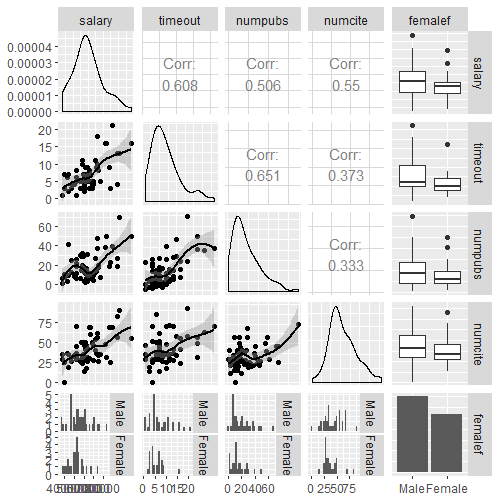
# labels to bivariate histograms

ggpairs(data=sci[c("salary", "timeout", "numpubs",

"numcite", "femalef")],

lower = list(continuous = "smooth\_loess"),

showStrips=TRUE)



# regressions - main effects only

m2<-lm(salary~timeout+numpubs+numcite+femalef,

data=sci)

# now let's add three-way interaction

# note "\*" operator adds interaction

# and all component parts (all component

# two-way interactions and main effects)

# ":" operator just adds interaction

m3<-lm(salary~timeout\*femalef\*numcite+numpubs+

numpubs:femalef, data=sci)

# get nice table for these two models

stargazer(m2, m3, type="text")

**==========================================================================**

**Dependent variable:**

**--------------------------------------------**

**salary**

**(1) (2)**

**--------------------------------------------------------------------------**

**timeout 857.000\*\*\* 1,218.000**

**(288.000) (830.000)**

**numpubs 92.700 151.000**

**(85.900) (117.000)**

**timeout:femalefFemale 2,658.000**

**(1,903.000)**

**timeout:numcite -7.220**

**(15.400)**

**femalefFemale:numcite 406.000**

**(281.000)**

**femalefFemale:numpubs -186.000**

**(180.000)**

**timeout:femalefFemale:numcite -84.700\***

**(49.000)**

**numcite 202.000\*\*\* 238.000\***

**(57.500) (129.000)**

**femalefFemale -918.000 -10,155.000**

**(1,860.000) (10,609.000)**

**Constant 39,587.000\*\*\* 36,784.000\*\*\***

**(2,717.000) (5,939.000)**

**--------------------------------------------------------------------------**

**Observations 62 62**

**R2 0.503 0.548**

**Adjusted R2 0.468 0.470**

**Residual Std. Error 7,077.000 (df = 57) 7,069.000 (df = 52)**

**F Statistic 14.400\*\*\* (df = 4; 57) 7.000\*\*\* (df = 9; 52)**

**==========================================================================**

**Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01**

# that's nice, but let's change the order so the

# interaction terms are at the bottom but above

# the intercept

stargazer(m2, m3, type="text",

order=c(8,9,1,2,5,3,4,7,6,10))

**==========================================================================**

**Dependent variable:**

**--------------------------------------------**

**salary**

**(1) (2)**

**--------------------------------------------------------------------------**

**numcite 202.000\*\*\* 238.000\***

**(57.500) (129.000)**

**femalefFemale -918.000 -10,155.000**

**(1,860.000) (10,609.000)**

**timeout 857.000\*\*\* 1,218.000**

**(288.000) (830.000)**

**numpubs 92.700 151.000**

**(85.900) (117.000)**

**femalefFemale:numcite 406.000**

**(281.000)**

**timeout:femalefFemale 2,658.000**

**(1,903.000)**

**timeout:numcite -7.220**

**(15.400)**

**timeout:femalefFemale:numcite -84.700\***

**(49.000)**

**femalefFemale:numpubs -186.000**

**(180.000)**

**Constant 39,587.000\*\*\* 36,784.000\*\*\***

**(2,717.000) (5,939.000)**

**--------------------------------------------------------------------------**

**Observations 62 62**

**R2 0.503 0.548**

**Adjusted R2 0.468 0.470**

**Residual Std. Error 7,077.000 (df = 57) 7,069.000 (df = 52)**

**F Statistic 14.400\*\*\* (df = 4; 57) 7.000\*\*\* (df = 9; 52)**

**==========================================================================**

**Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01**

# let us get summary stats for numcite and timeout

# pay attention to first and third quartiles for

# those variables

summary(sci$numcite)

**Min. 1st Qu. Median Mean 3rd Qu. Max.**

**1.0 28.2 35.0 40.2 50.0 90.0**

summary(sci$timeout)

**Min. 1st Qu. Median Mean 3rd Qu. Max.**

**1.00 4.00 6.00 6.79 8.75 21.00**

# using effect function in effects package to

# calculate predictions showing three-way

# interaction

# using Q1 and Q3 for timeout and numcite

# to get

m3eff<-effect(term="timeout\*femalef\*numcite",

mod=m3,

xlevels=list(female=c(0,1),

numcite=c(28.2,50),

timeout=c(4,8.75)))

m3eff

**timeout\*femalef\*numcite effect**

**, , numcite = 28.2**

**femalef**

**timeout Male Female**

**4 50296 49291**

**8.75 55115 55389**

**, , numcite = 50**

**femalef**

**timeout Male Female**

**4 54856 55319**

**8.75 58927 51900**

#convert eff object into dataframe so we can

# graph the effects

m3effdf<-data.frame(m3eff)

#creating factor variable of the timeout variable for

#graphing purposes

m3effdf$timeoutf<-factor(m3effdf$timeout,

levels=c(4,8.75),

labels=c("4 (Q1)",

"8.8 (Q3)"))

m3effdf

|  |
| --- |
| **timeout femalef numcite fit se lower upper timeoutf**  **1 4.00 Male 28.2 50296 1813 46659 53934 4 (Q1)**  **2 8.75 Male 28.2 55115 2070 50961 59269 8.8 (Q3)**  **3 4.00 Female 28.2 49291 1977 45324 53258 4 (Q1)**  **4 8.75 Female 28.2 55389 2538 50296 60481 8.8 (Q3)**  **5 4.00 Male 50.0 54856 1932 50978 58733 4 (Q1)**  **6 8.75 Male 50.0 58927 1515 55887 61966 8.8 (Q3)**  **7 4.00 Female 50.0 55319 2301 50703 59936 4 (Q1)**  **8 8.75 Female 50.0 51900 3489 44899 58901 8.8 (Q3)** |

g3<-ggplot(data=m3effdf,

aes(x=numcite, y=fit, color=timeoutf,

linetype=timeoutf))+

facet\_wrap(~femalef)+

geom\_line(size=1)+

ggtitle("Salary By Gender, Citations, and

Years Since PhD")+

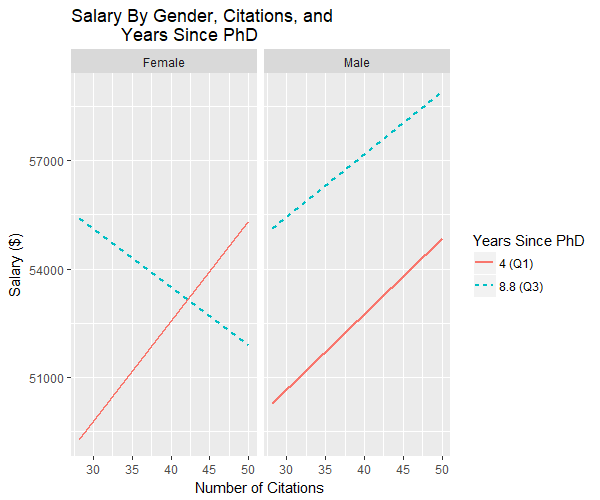
labs(color="Years Since PhD",

linetype="Years Since PhD")+

xlab("Number of Citations")+

ylab("Salary ($)")

g3



# let's probe the interaction

# examining effect of numcite contingent on

# timeout and sex

sim\_slopes(model=m3, pred=numcite, modx=timeout,

mod2=femalef, modxvals=c(4,8.75),

mod2vals=c(0,1))

**#######################################################**

**While femalef (2nd moderator) = 0**

**#######################################################**

**JOHNSON-NEYMAN INTERVAL**

**The slope of numcite is p < .05 when timeout is INSIDE this interval:**

**[0.926, 9.63]**

**Note: The range of observed values of timeout is [1, 21]**

**SIMPLE SLOPES ANALYSIS**

**Slope of numcite when timeout = 4:**

**Est. S.E. p**

**209.138 86.753 0.019**

**Slope of numcite when timeout = 8.75:**

**Est. S.E. p**

**174.863 80.183 0.034**

**#######################################################**

**While femalef (2nd moderator) = 1**

**#######################################################**

**JOHNSON-NEYMAN INTERVAL**

**The slope of numcite is p < .05 when timeout is OUTSIDE this interval:**

**[4.97, 254]**

**Note: The range of observed values of timeout is [1, 21]**

**SIMPLE SLOPES ANALYSIS**

**Slope of numcite when timeout = 4:**

**Est. S.E. p**

**276.545 105.365 0.011**

**Slope of numcite when timeout = 8.75:**

**Est. S.E. p**

**-160.02 201.14 0.43**