Graphing Multiple Regression Interactions in R

A. Alexander Beaujean[1]*

Abstract
This document shows how to plot regression interactions using the **R** (R Development Core Team, 2013) statistical programming language.

Keywords
Moderation; Interaction; **R**

[1] *Educational Psychology Department, Baylor University, Waco, TX*
***Corresponding author**: Alex_Beaujean@baylor.edu

Contents

1. Data

X is the predictor, Z is the moderator, and Y is the outcome.

2. Moderator is Continuous

For these cases you have to make your moderating variable categorical. Often this is done by selecting the mean, and ±1, 2 or 3 SDs from the mean. I will use the mean and $\pm$ 2 SDs for all of these examples.

2.1 Outcome is Continuous

2.1.1 Predictor is Continuous

A graph is given in Figure 1.

Table 1. Regression Results for Continuous Predictor, Continuous Outcome and Continuous Moderator.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	48.5249	0.1771	274.05	0.0000
X.cont	0.3505	0.0123	28.52	0.0000
Z.cont	1.0414	0.0570	18.26	0.0000
X.cont:Z.cont	0.1155	0.0036	31.88	0.0000

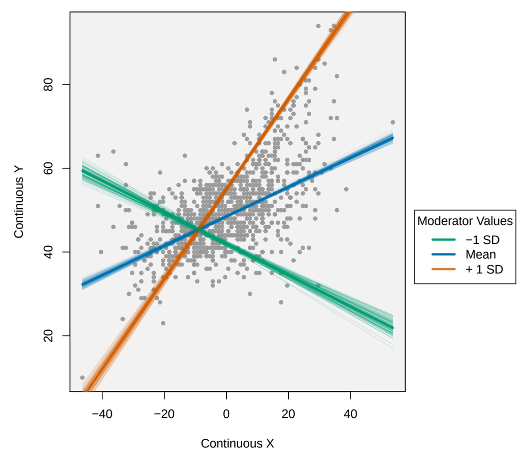

Figure 1. Continuous predictor, continuous outcome, continuous moderator.

2.1.2 Predictor is Categorical

This is an ANCOVA with interactions (i.e., no homogeneity of slopes). See Section 2.1.1 to create the graph.

2.2 Outcome is Categorical

If the outcome has two categories, then a logistic/probit regression is often a good model to use. I will use a logistic regression.

2.2.1 Predictor is Continuous

A graph is given in Figure 2.

2.2.2 Predictor is Categorical

See Section 2.2.1 to create the graph.

3. Moderator is Categorical

3.1 Outcome in Continuous

3.1.1 Predictor is Continuous

See Section 2.1.1 to create the graph.

Table 2. Regression Results for Continuous Predictor, Categorical Outcome and Continuous Moderator.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.2906	0.0836	-3.47	0.0005
X.cont	0.1088	0.0088	12.41	0.0000
Z.cont	0.2915	0.0345	8.45	0.0000
X.cont:Z.cont	0.0335	0.0031	10.83	0.0000

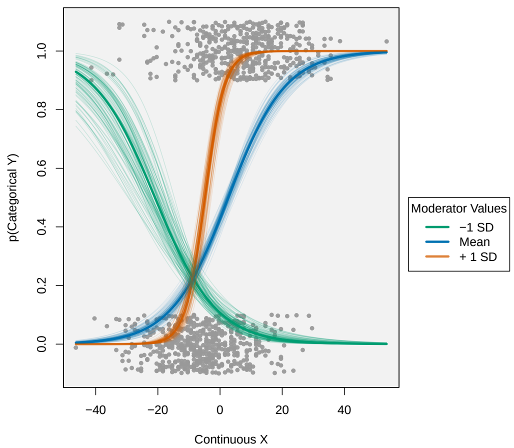

Figure 2. Continuous predictor, categorical outcome, continuous moderator.

3.1.2 Predictor is Categorical

This is a 2-factor ANOVA with interaction. A graph is given in Figure 3.

Table 3. Regression Results for Categorical Predictor, Continuous Outcome and Categorical Moderator.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	45.9814	0.4820	95.39	0.0000
X.cat	1.0610	0.7817	1.36	0.1750
Z.cat	-1.0133	0.6938	-1.46	0.1445
X.cat:Z.cat	12.9582	1.0289	12.59	0.0000

3.2 Outcome is Categorical

3.2.1 Predictor is Continuous

See Section 2.2.1 to create the graph.

3.2.2 Predictor is Categorical

This can be thought of as a logistic regression with two categorical predictors. A graph is given in Figure 4.

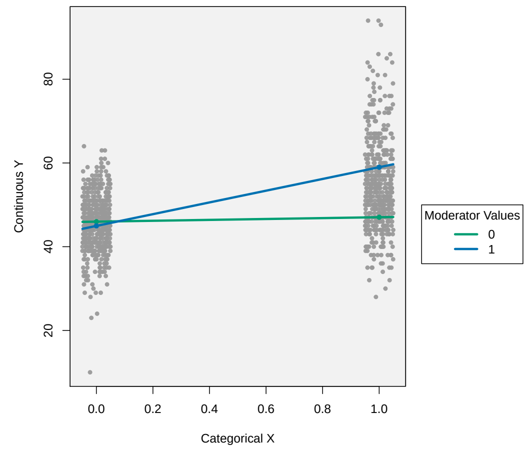

Figure 3. Categorical predictor, continuous outcome, categorical moderator.

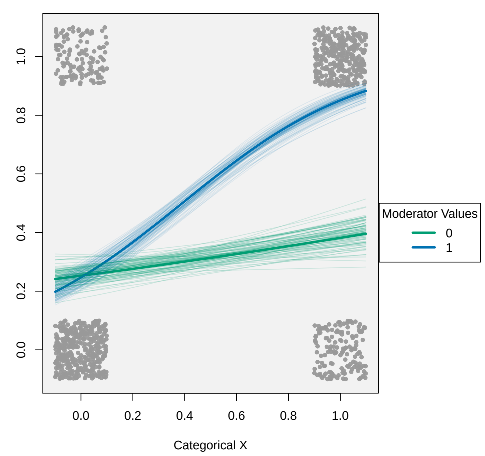

Figure 4. Categorical predictor, categorical outcome, categorical moderator.

Table 4. Regression Results for Categorical Predictor, Categorical Outcome and Categorical Moderator.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.0838	0.1403	-7.73	0.0000
X.cat	0.6020	0.2130	2.83	0.0047
Z.cat	-0.0308	0.2027	-0.15	0.8792
X.cat:Z.cat	2.2535	0.3030	7.44	0.0000

References & Further Reading

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

Beaujean, A. A. (2008). Mediation, moderation, and the study of individual differences. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (p. 422-442). Thousand Oaks, CA: Sage.

R Development Core Team. (2013). ***R**: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437-448. doi: 10.3102/10769986031004437

1. R Syntax

```
# Simulate Data
set.seed(56445)
X.cont = rnorm(1000)
Z.cont <- X.cont * 0.3 + rnorm(1000)
Y.cont <- 0.3 * X.cont + 0.2 * Z.cont + 0.3 * X.cont * Z.cont + rnorm(1000, 0, 0.3)

# Put X, Z and Y on IQ, T, and standard score scales; make them integers
X.cont <- round(X.cont * 15 + 100, 0)
Z.cont <- round(Z.cont * 3 + 10, 0)
Y.cont <- round(scale(Y.cont) * 10 + 50, 0)

# Mean center continuous predictors
X.cont <- scale(X.cont, scale = FALSE)
Z.cont <- scale(Z.cont, scale = FALSE)


# Make categorial vairables from continuous
X.cat <- ifelse(X.cont < mean(X.cont), 0, 1)
Y.cat <- ifelse(Y.cont < mean(Y.cont), 0, 1)
Z.cat <- ifelse(Z.cont < mean(Z.cont), 0, 1)
```

```
# Save the values for mean and +/- 2 SDs for the moderator

# -- note that apply(,2,sd) is how to calculate the SD in R
Z.mean <- mean(Z.cont)
Z.pos2SD <- Z.mean + 2 * apply(Z.cont, 2, sd)
Z.neg2SD <- Z.mean - 2 * apply(Z.cont, 2, sd)
```

```
# Regression when outcome, predictor and moderator are continuous
allCont.fit <- lm(Y.cont ~ X.cont * Z.cont)
```

```
# Create a object of colors for the graphs (taken after ggplot's color blind freindly colors)
plot.colors<-c(
# Gray
rgb(242,242,242,alpha=255,maxColorValue = 255),
#greens
rgb(0,158,115,alpha=255,maxColorValue = 255),
rgb(0,158,115,alpha=40,maxColorValue = 255),
# Blues
rgb(0,114,178,alpha=255,maxColorValue = 255),
rgb(0,114,178,alpha=25,maxColorValue = 255),
# Oranges
rgb(213,94,0,alpha=200,maxColorValue = 255),
rgb(213,94,0,alpha=30,maxColorValue = 255),
# light gray
rgb(153,153,153,alpha=240,maxColorValue = 255)
)
```

```
# Graph for continuous outcome, predictor, and moderator

par(mar=par()$mar+c(0,0,0,6))
plot(X.cont, Y.cont, ylab="Continuous Y", xlab="Continuous X", pch=20,
col=plot.colors[8],panel.first=rect(par("usr")[1], par("usr")[3], par("usr")[2], par("usr")[4],
col = plot.colors[1]))

# Simple  Regressions
```

```
curve (cbind (1,x, Z.neg2SD, x*Z.neg2SD) %*% coef(allCont.fit), add=TRUE, col=plot.colors[2], lwd=3)
curve (cbind (1,x, Z.mean, x*Z.mean) %*% coef(allCont.fit), add=TRUE, col=plot.colors[4], lwd=3)
curve (cbind (1,x, Z.pos2SD, x*Z.pos2SD) %*% coef(allCont.fit), add=TRUE, col=plot.colors[6], lwd=3)

# "Confidence bands" (Using Gelman & Hill's Method)
library(arm)
allCont.sim <- sim(allCont.fit)

for (i in 1:100){
curve (cbind (1,x, Z.neg2SD, x*Z.neg2SD) %*% coef(allCont.sim)[i,], add=TRUE, col=plot.colors[3])
curve (cbind (1,x, Z.mean, x*Z.mean) %*% coef(allCont.sim)[i,], add=TRUE, col=plot.colors[5])
curve (cbind (1,x, Z.pos2SD, x*Z.pos2SD) %*% coef(allCont.sim)[i,], add=TRUE, col=plot.colors[7])
}

legend(max(X.cont)+7,50,c("-1 SD", "Mean", "+ 1 SD"), title="Moderator Values", lwd = 3,
col=plot.colors[c(2,4,6)],xpd=TRUE)
```

```
# Logistic Regression
logContCont.fit <- glm(Y.cat ~ X.cont * Z.cont, family = binomial(link = "logit"))
```

```
# Plot when Predictor and  Moderator are continuous, with a Categorical outcome

# Jitter the outcome for a "prettier" plot
Y.jitter <- jitter(Y.cat,amount=.1)

par(mar=par()$mar+c(0,0,0,6))
plot(X.cont, Y.jitter, ylab="p(Categorical Y)", xlab="Continuous X", pch=20, col=plot.colors[8],
panel.first=rect(par("usr")[1], par("usr")[3], par("usr")[2], par("usr")[4], col = plot.colors[1]))

# Simple  Regressions
curve ( invlogit ( cbind (1,x, Z.neg2SD, x*Z.neg2SD) %*% coef(logContCont.fit)), add=TRUE,
col=plot.colors[2], lwd=3)
curve (invlogit ( cbind (1,x, Z.mean, x*Z.mean) %*% coef(logContCont.fit)), add=TRUE,
col=plot.colors[4], lwd=3)
curve (invlogit ( cbind (1,x, Z.pos2SD, x*Z.pos2SD) %*% coef(logContCont.fit)), add=TRUE,
col=plot.colors[6], lwd=3)

# Confidence bands (Using Gelman & Hill's Method)
logContCont.sim <- sim(logContCont.fit)

for (i in 1:100){
curve (invlogit ( cbind (1,x, Z.neg2SD, x*Z.neg2SD) %*% coef(logContCont.sim)[i,]), add=TRUE,
col=plot.colors[3])
curve (invlogit ( cbind (1,x, Z.mean, x*Z.mean) %*% coef(logContCont.sim)[i,]), add=TRUE,
col=plot.colors[5])
curve (invlogit ( cbind (1,x, Z.pos2SD, x*Z.pos2SD) %*% coef(logContCont.sim)[i,]), add=TRUE,
col=plot.colors[7])
}

legend(max(X.cont)+7,.5,c("-1 SD", "Mean", "+ 1 SD"), title="Moderator Values", lwd = 3,
col=plot.colors[c(2,4,6)],xpd=TRUE)
```

```
# Regression
ANOVA.fit <- lm(Y.cont ~ X.cat * Z.cat)
```

```
# Plot Categorical Predictor and Moderator, with a Continuous Outcome

# Jitter X for a "prettier" output
```

```
X.jitter <- jitter(X.cat, amount=.05)


par(mar=par()$mar+c(0,0,0,6))

plot(X.jitter, Y.cont, ylab="Continuous Y", xlab="Categorical X", pch=20, col=plot.colors[8],
panel.first=rect(par("usr")[1], par("usr")[3], par("usr")[2], par("usr")[4], col = plot.colors[1]))

# Simple  mean plots
curve (  cbind (1, x, 0, x*0) %*% coef(ANOVA.fit), add=TRUE, col=plot.colors[2], lwd=3)
points(0,mean(Y.cont[X.cat==0 & Z.cat==0]), col=plot.colors[2],lwd=2, pch=20)
points(1,mean(Y.cont[X.cat==1 & Z.cat==0]), col=plot.colors[2], lwd=2, pch=20)

curve (  cbind (1, x, 1, x*1) %*% coef(ANOVA.fit), add=TRUE, col=plot.colors[4], lwd=3)
points(0,mean(Y.cont[X.cat==0 & Z.cat==1]), col=plot.colors[4],lwd=2, pch=20)
points(1,mean(Y.cont[X.cat==1 & Z.cat==1]), col=plot.colors[4], lwd=2, pch=20)


# Confidence bands (Using Gelman & Hill's Method)
ANOVA.sim <- sim(ANOVA.fit)

for (i in 1:100){
curve (cbind (1,x, 0, x*0) %*% coef(ANOVA.sim)[i,], from=0, to=0, add=TRUE, col=plot.colors[3], lwd=3,
 type="l")
curve (cbind (1,x, 0, x*0) %*% coef(ANOVA.sim)[i,], from=1, to=1, add=TRUE, col=plot.colors[3], lwd=3,
 type="l")
curve (cbind (1,x, 1, x*1) %*% coef(ANOVA.sim)[i,], from=0, to=0, add=TRUE, col=plot.colors[5], lwd=3,
 type="l")
curve (cbind (1,x, 1, x*1) %*% coef(ANOVA.sim)[i,], from=1, to=1, add=TRUE, col=plot.colors[5], lwd=3,
 type="l")
}

legend(max(X.cat)+.15,50,c("0", "1"), title="Moderator Values", lwd = 3, col=plot.colors[c(2,4)], xpd=TRUE)
```

```
# Logistic Regression with two categorical predictors
logCatCat.fit <- glm(Y.cat ~ X.cat * Z.cat, family = binomial(link = "logit"))
```

```
# Plot for Categorical Predictor, outcome, and moderator

# Jitter the predictor and the outcome for a "prettier" plot
X.jitter <- jitter(X.cat, amount=.1)
Y.jitter <- jitter(Y.cat,amount=.1)

par(mar=par()$mar+c(0,0,0,6))

plot(X.jitter, Y.jitter, ylab="p(Categorical Y)", xlab="Categorical X", pch=20, col=plot.colors[8],
panel.first=rect(par("usr")[1], par("usr")[3], par("usr")[2], par("usr")[4], col = plot.colors[1]))


# Simple  Regressions

curve (invlogit ( cbind (1,x, 0, x*0) %*% coef(logCatCat.fit)), add=TRUE, col=plot.colors[2], lwd=3)
curve (invlogit ( cbind (1,x, 1, x*1) %*% coef(logCatCat.fit)), add=TRUE, col=plot.colors[4], lwd=3)


# Confidence bands (Using Gelman & Hill's Method)
logCatCat.sim <- sim(logCatCat.fit)

for (i in 1:100){
curve (invlogit ( cbind (1,x, 0, x*0) %*% coef(logCatCat.sim)[i,]), add=TRUE, col=plot.colors[3])
```

```
curve (invlogit ( cbind (1,x, 1, x*1) %*% coef(logCatCat.sim)[i,]), add=TRUE, col=plot.colors[5])
}

legend(max(X.cat)+.15,.5,c("0", "1"), title="Moderator Values", lwd = 3,
col=plot.colors[c(2,4)],xpd=TRUE)
```