

Interactions/Moderation in Regression : Example II

Joel S Steele

Data

We will use the same example data from our last example. Again, these data are completely made-up. The variables in these data are *spend*, *sex*, *pol*, and *inc* which correspond to the amount of money spent on campaign contributions (*spend*), the gender of the participant (*sex*), the political party affiliation (*pol*), and the income of the participant (*inc*).

```
# you should be able to cut-and-paste this into your console.
# example data
exdat = structure(list(spend = c(20.09, 18.8, 19.98, 19.16, 20.65, 18.63,
20.25, 18.92, 20.38, 19.64, 22.42, 19.37, 23.01, 20.63, 22.36,
20.01, 22.02, 19.65, 21.81, 20.23), sex = structure(c(1L, 1L,
1L, 1L, 1L, 1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
2L, 2L), .Label = c("male", "female"), class = "factor"), pol = structure(c(1L,
2L, 1L, 2L, 1L, 2L, 1L, 2L, 1L, 2L, 1L, 2L, 1L, 2L, 1L,
2L, 1L, 2L), .Label = c("dem", "other"), class = "factor"), inc = c(180000,
220000, 130000, 120000, 230000, 250000, 50000, 190000, 1e+05,
80000, 250000, 60000, 140000, 80000, 210000, 90000, 110000, 50000,
190000, 140000)), .Names = c("spend", "sex", "pol", "inc"), row.names = c(NA,
-20L), class = "data.frame")

# I'm going to add a scaled and centered version of income.
exdat$sinc = exdat$inc/10000
exdat$sinc = exdat$sinc - mean(exdat$sinc)
```

Again we may need the mean *spending* based on *sex* and *political party*.

```
# aggregate based on sex and politics, get means per combination
spmeans = with(exdat, aggregate( spend ~ sex + pol, FUN=mean ))
#let's have a look
kable(spmeans,digits=3)
```

sex	pol	spend
male	dem	20.270
female	dem	22.324
male	other	19.030
female	other	19.978

We use these values again for calculations below, so we create some variables for each value.

```
# we will need these for later
md = spmeans[1,3] # male democrat
fd = spmeans[2,3] # female democrat
mo = spmeans[3,3] # male other
fo = spmeans[4,3] # female other
am = mean(c(md,mo)) # all males (dem and other)
af = mean(c(fd,fo)) # all females
ad = mean(c(md,fd)) # all democrats (male and female)
ao = mean(c(mo,fo)) # all others
```

A key might help to keep these straight in your head.

- f = *female*
- m = *male*
- d = *democrat*
- o = *other*
- a = *all/average*

For example, **ao** and **md**, which correspond to **all other** and **male democrat**. These values will come into our computations below.

Interactions with Categorical Predictors

We saw last time that, with the inclusion of the interaction term, all of our effects go from being *main* effects to *simple* effects. Meaning, for example, that the coefficient for *sexfemale* is no longer all females vs all males, but instead is conditional (*depends*) on the value of the other variables in the equation and when that value is zero. Let's see with the following model.

$$Spend \sim \beta_0 + \beta_1 Sex + \beta_2 Pol + \beta_3 (Sex \times Pol)$$

```
# let's add in the interaction term.
kable(coef(summary(lm(spend ~ sex * pol, exdat)->lm0)),digits=3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	20.270	0.183	110.500	0.000
sexfemale	2.054	0.259	7.918	0.000
polother	-1.240	0.259	-4.780	0.000
sexfemale:polother	-1.106	0.367	-3.015	0.008

```
# calculate from the cell means
kable(cbind('model'=coef(lm0),
'direct'=c(
'(Intercept)' = md,
'SexFemale' = fd - md,
'PolOther' = mo - md,
'SexFemale:PolOther' = (fo-fd)-(mo-md)
)),digits=3)
```

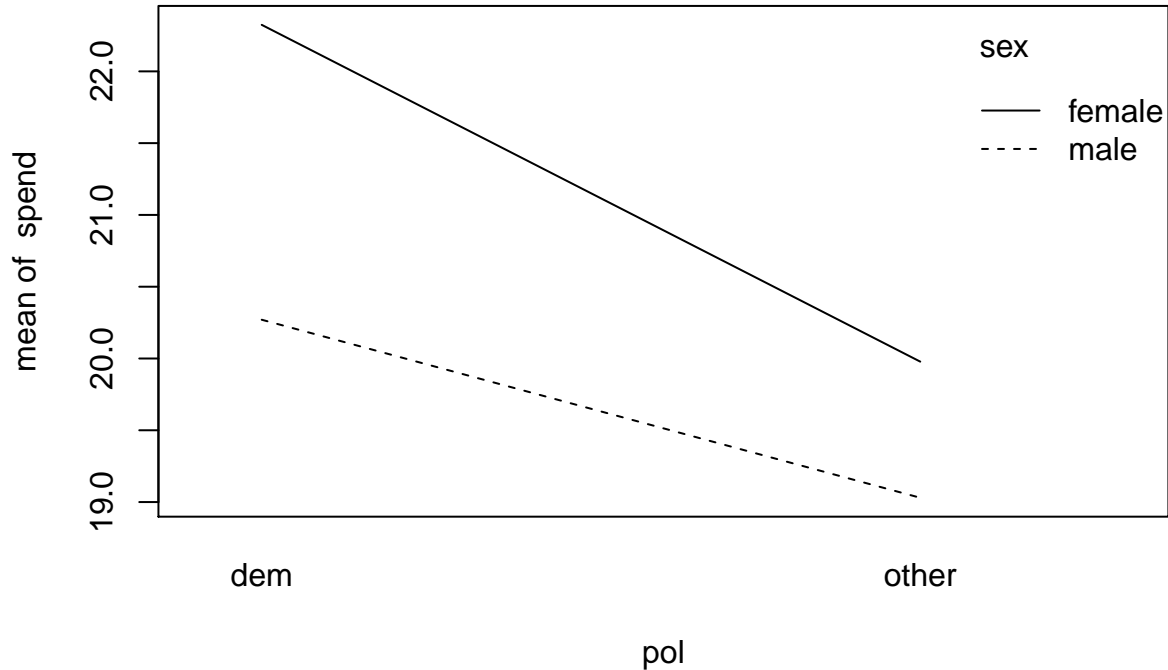
	model	direct
(Intercept)	20.270	20.270
sexfemale	2.054	2.054
polother	-1.240	-1.240
sexfemale:polother	-1.106	-1.106

We can see that the (*Intercept*) is still the mean for **male democrats** ($\beta_0 = 20.27$). The *sexfemale* estimate the difference between *male* and *female democrats* ($\beta_1 = 2.054$). *polother* is only for males as well. It's the difference between *other* and *democrats* for **males** only ($\beta_2 = -1.24$).

The part we want to focus on is the interaction term the *SexFemale:PolOther* part ($\beta_3 = -1.106$). This represents a slope, of sorts, in that it estimates the difference in change from being **democrat** to being **other** for **females** relative to **males**. That is, we know the change in spending for **males** between **democrat** and **other**, it's the *PolOther* estimate ($\beta_2 = -1.24$). The fact that this value is negative, means that there is a

greater drop in spending between *female democrats* and *others*, than the drop for the *males*. Both estimates are negative, meaning that for both genders the *other* political group spends less. But the interaction term tells us that the slope is steeper for *females*. It may help to see a picture.

```
with(exdat, interaction.plot(pol,sex,spend))
```



This is helpful, but we will understand even more if we can reproduce it ourselves. To do this we will need to pull out each estimate and construct an equation for each combination in the data. From this point on we will be including the continuous predictor of *income* as well.

To begin we scale the income variable based on the actual data. From low to high in 50K steps. Then divide by 10,000.

```
# create a vector of income values based on the actual data.
inc = seq(min(exdat$inc),max(exdat$inc),50000)
inc = inc/10000
inc
```

```
[1] 5 10 15 20 25
```

Now we fit the model,

$$Spend \sim \beta_0 + \beta_1 Incom + \beta_2 Sex + \beta_3 Pol + \beta_4 (Sex \times Pol)$$

```
# categorical by categorical interactions
kable(coef(summary(lm(spend ~ sinc + sex*pol, exdat)->lm2)),digits=3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	20.266	0.188	107.569	0.000
sinc	-0.008	0.018	-0.458	0.654
sexfemale	2.088	0.276	7.559	0.000
polother	-1.212	0.273	-4.445	0.000
sexfemale:polother	-1.211	0.441	-2.747	0.015

Now we capture the coefficients to plot our own lines.

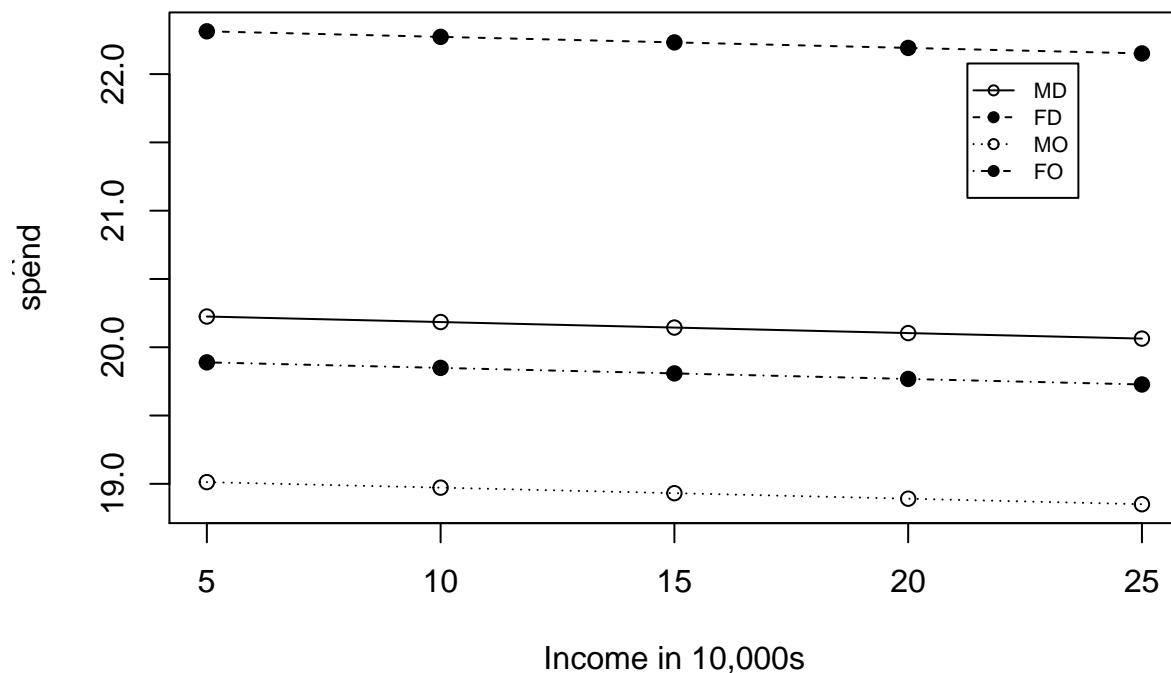
```

b0 = coef(lm2)[[1]]
binc = coef(lm2)[[2]]
bsex = coef(lm2)[[3]]
bpol = coef(lm2)[[4]]
bsp = coef(lm2)[[5]]

# simulate data using 0 and 1 for binary outcomes,
# with the simulated income data we created earlier
fit.md = b0 + bsex*0 + bpol*0 + binc*inc + bsp*0*0
fit.fd = b0 + bsex*1 + bpol*0 + binc*inc + bsp*1*0
fit.mo = b0 + bsex*0 + bpol*1 + binc*inc + bsp*0*1
fit.fo = b0 + bsex*1 + bpol*1 + binc*inc + bsp*1*1

# Plot!
plot(inc, fit.md,
     ylim=range(fit.md,fit.fd,fit.mo,fit.fo),
     type='o',xlab='Income in 10,000s', ylab=expression(hat(spend)))
# one line each
lines(inc, fit.fd, type='o', lty=2,pch=19)
lines(inc, fit.mo, type='o', lty=3)
lines(inc, fit.fo, type='o', lty=4,pch=19)
legend('topright',c('MD','FD','MO','FO'),
      lty=1:4,
      pch=c(1,19,1,19),inset=.1, cex=.7)

```



Here we interact *sex* with both *income* and *pol*. So we can see if the influence of *income* is the same for **males** and **females**.

```

# categorical by continuous interaction as well
kable(coef(summary(lm(spend ~ sex*(sinc+pol), data=exdat)->lm3)),digits=3)

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	20.259	0.188	107.780	0.000
sexfemale	1.994	0.290	6.876	0.000
sinc	-0.020	0.021	-0.959	0.354
polother	-1.171	0.275	-4.260	0.001
sexfemale:sinc	0.040	0.038	1.042	0.315
sexfemale:polother	-0.987	0.489	-2.016	0.063

```
anova(lm2,lm3)
```

Analysis of Variance Table

Model 1: spend ~ sinc + sex * pol

Model 2: spend ~ sex * (sinc + pol)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	15	2.6549				
2	14	2.4637	1	0.19119	1.0865	0.3149

```
# collect coefficients
```

```
int3=coef(lm3)[[1]]
```

```
bfem=coef(lm3)[[2]]
```

```
binc=coef(lm3)[[3]]
```

```
bpol=coef(lm3)[[4]]
```

```
bsxi=coef(lm3)[[5]]
```

```
bsxp=coef(lm3)[[6]]
```

```
# use the equation for illustrative predictions
```

```
# not as explicit here but you understand the point
```

```
# females
```

```
pred.fo = int3 + bfem + binc*inc + bpol + bsxi*inc + bsxp
```

```
pred.fd = int3 + bfem + binc*inc + bsxi*inc
```

```
# males
```

```
pred.mo = int3 + binc*inc + bpol
```

```
pred.md = int3 + binc*inc
```

```
# plot
```

```
plot(NA,
```

```
  xlim=range(inc),ylim=c(18,25),
```

```
  xlab='Income',ylab='Spending')
```

```
lines(inc,pred.md, pch=19, lty=1, type='o')
```

```
lines(inc,pred.fd, pch=5, lty=2, type='o')
```

```
lines(inc, pred.mo, pch=19, lty=3, type='o')
```

```
lines(inc, pred.fo, pch=5, lty=4, type='o')
```

```
legend('topleft',
```

```
  c('MaleDem', 'FemDem', 'MaleOthr', 'FemOthr'),
```

```
  pch=c(19,5,19,5),
```

```
  lty=1:4,
```

```
  inset=.02, cex=.7)
```

