

# Homework Interactions

December 3, 2013

## Question 1

You have a theory that feeling thermometer scores for Bush depend on ideology, age, education, income, sex, and whether or not the respondent is African-American. You theorize that race and sex interact.

- Using the 2004 NES:
- Make an (an, one single, not several) appropriate dummy for race. Make the appropriate interaction term. Show me the code or otherwise indicate how you made it.
- Run the regression, nice table, interpretation, etc.
- For respondents with moderate ideology, who are pure independents, who are 35, who have some college but not a BA, and whose households earned \$37000, what are the expected feeling thermometer scores for:
  - Nonblack men
  - Nonblack women
  - Black men
  - Black women

## 1 – Answer

To make the appropriate variables, I made single dummies `male` and `black`, and then interacted them by multiplying them together. Alternately, if you kept your data or code, you could use whatever variable for African-Americans that you used before. Note that this does mean that for the racial variable, the excluded category is all people who are not African-American.

The regression results are reported in Table 1

The regression performs quite well, with good goodness of fit measures. Among the control variables, ideology and education are statistically significant predictors and in the expected direction, but age and income do not approach any reasonable standard of significance. The effects of race and sex are, given the interaction, complex. The coefficient on `Male` indicates that there is no gender effect for nonblacks. The coefficient on `Black` indicates that, among women, African-American respondents feel 27 points less warmly towards Bush. The interaction term shows that there is a strong difference in gender effects between African-Americans and others. The combined effect shows that among black

Table 1: EFFECTS OF SEX AND RACE ON BUSH FT SCORES

Variable	Coef	SE
Constant	11.707	5.062*
Ideology (conservatism)	12.697	0.681**
Age	-0.042	0.060
Education	-1.635	0.681*
Household income	0.109	0.191
Men	-0.951	2.103
Black	-26.829	4.314*
Men × Black	15.677	6.083**
N	804	
F( 7, 796)	62.32	
$P > F$	0.000	
$R^2$	0.354	
*	$p < 0.05$	
**	$p < 0.01$	

respondents, men feel about 15 points more warmly towards Bush than do women. Alternatively, the effect of race is reduced for men.

To do the second part, all you need to do is plug the appropriate values into the regression equation. The initial baseline will be

$$bushft = 11.707 + 12.697(4) - 0.042(35) - 1.635(5) + 0.109(14) = 54.376 \quad (1)$$

This baseline value is also the predicted FT score for nonblack women (that is, zero on both black and men). For black women, we add the coefficient on black, taking us to 27.547. For nonblack men, we add the coefficient for men, taking us to 53.425, or hardly different from the predicted value for nonblack women. For African-American men, we add the coefficients for black, men, and menXblack, since black men satisfy all three conditions. This takes us to 42.273.

tl;dr:

Nonblack women	54.376
Black women	27.547
Nonblack men	53.425
Black men	42.273

## Question 2

You have a theory that says that richer people should like Bush better, and that this effect should be particularly strong among conservatives. Test this theory in a multiple regression using the NES, and dummies to distinguish between liberals, moderates, and conservatives instead of using the raw ideology measure. You do not need to present a fancy table here; just show me the raw output. What is the predicted effect of income for liberals? For moderates? For conservatives?

## 2 – Answer

First, I need to set up data. I created variables “liberal” and “conserv” to denote liberal and conservatives from the ideology variable. I then created “libinc” and “consinc” by

multiplying liberal times income and conservative times income.

I then ran the following regression:

```
. reg bushft liberal conserv house libX conX
```

Source	SS	df	MS	Number of obs = 804		
Model	299800.919	5	59960.1839	F( 5, 798)	=	74.07
Residual	645961.643	798	809.475743	Prob > F	=	0.0000
-----				R-squared	=	0.3170
-----				Adj R-squared	=	0.3127
Total	945762.562	803	1177.7865	Root MSE	=	28.451

  

bushft	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
liberal	-13.53192	6.86665	-1.97	0.049	-27.01075	-.0530914
conserv	13.72105	7.260774	1.89	0.059	-.5314254	27.97352
householdin~e	.1532629	.3039309	0.50	0.614	-.4433356	.7498613
libXinc	-.6360084	.4273368	-1.49	0.137	-1.474845	.2028286
conXinc	.5755787	.4327375	1.33	0.184	-.2738595	1.425017
_cons	49.38658	4.810635	10.27	0.000	39.94359	58.82958

In R, it would look like this, including an embarrassing error:

```
> nes2004 <- read.csv("C:/courses/old/508 fall 07/data/nes2004/nes2004.csv")
> View(nes2004)
> attach(nes2004)
> liberal<-1*(lib_con<4)
Error: object 'lib_con' not found
> liberal<-1*(lib_cons<4)
> conserv<-1*(lib_cons>4)
> model<-lm(bushft~liberal*householdincome+ conserv*householdincome)
> summary(model)
```

Call:

```
lm(formula = bushft ~ liberal * householdincome + conserv * householdincome)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-78.413 -21.686   5.129  20.129  75.248
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    49.3866   4.8106  10.266 <2e-16
liberal       -13.5319   6.8666  -1.971  0.0491
householdincome  0.1533   0.3039   0.504  0.6142
conserv        13.7210   7.2608   1.890  0.0592
liberal:householdincome -0.6360  0.4273  -1.488  0.1371
householdincome:conserv  0.5756   0.4327   1.330  0.1839
```

```
(Intercept)    ***
liberal        *
```

```

householdincome
conserv .
liberal:householdincome
householdincome:conserv
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28.45 on 798 degrees of freedom
(372 observations deleted due to missingness)
Multiple R-squared:  0.317, Adjusted R-squared:  0.3127
F-statistic: 74.07 on 5 and 798 DF,  p-value: < 2.2e-16

```

The effect of income for the reference category – moderates in my case – is the coefficient reported for income, which is 0.153, and is non statistically significant. To get the effect of income for liberals, subtract 0.636 from the coefficient for income, getting -0.484. To get the effect for conservatives, add 0.576 to the coefficient for income, arriving at 0.729.

ADDED BONUS!! What if we want measures of uncertainty around our estimates of the effects for income for conservatives and liberals? And we do! Remember that the coefficient reported for household income (a) is the effect of income for the reference category, and (b) displays a p-value against a null of no difference. So all we have to do to get p-values for our estimates of the effects of income for liberals or conservatives is *re-run the equation with liberals or conservatives as the reference category!*. If you do this, you'll find that the p-value for liberals is 0.108 and for conservatives is 0.018. So our original theory, that the effect of income is strongest for conservatives, is right! It's almost like I made up that "theory" after the fact, making me worse than Hitler and Pol Pot put together.