

## I. INTERACTIONS: BASIC IDEA

- In multiple regression models, we interpret a particular regression coefficient as the “effect” of that  $X$  variable (e.g. of  $X_1$ ) on  $Y$ , holding the other  $X$ s in the model constant. When we first introduced the idea of “holding constant,” we said that this was restrictive in the sense that the effect of  $X_1$  on  $Y$  **was constrained to be constant across the range of other  $X$  values.** (e.g., the effect of POLICE on CRIMES was constrained to be constant across the different values of CRIMINAL; yet in some of the specifications, we saw that the effect of police on crimes was in fact different across different levels of criminals; our model just didn’t account for that difference).
- Using **interactions**, we have a way of allowing the effect of one variable,  $X_1$ , on  $Y$  to vary, depending on the value of another variable,  $X_2$ .
- We can have interactions between (1) two indicator variables; (2) an indicator and an interval/ratio variable; or (3) two interval ratio variables. The idea is the same: The interaction term allows the effect of one variable to vary, depending on the value of the other variable.
- It is possible to interact three or more variables; however, it can become difficult to interpret these coefficients. They are not used much in practice, and we won’t cover them here.

## II. INTERACTIONS WITH TWO INDICATOR VARIABLES

- First, let’s start out with a simple model of *bwght* on *white* and *male*:

```
. reg bwght white male
```

Source	SS	df	MS			
Model	12499.855	2	6249.92752	Number of obs =	1388	
Residual	562111.865	1385	405.856942	F( 2, 1385) =	15.40	
Total	574611.72	1387	414.283864	Prob > F =	0.0000	
				R-squared =	0.0218	
				Adj R-squared =	0.0203	
				Root MSE =	20.146	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
bwght						
white	6.365519	1.315637	4.84	0.000	3.784662	8.946376
male	3.057256	1.082694	2.82	0.005	.9333584	5.181153
_cons	112.1128	1.304413	85.95	0.000	109.554	114.6716

- This model constrains the “effect” of being MALE to be the same for WHITES as well as for NONWHITES. (“Holding race constant....”)
- And, this model constraints the “effect” of being WHITE to be the same for MALES as well as FEMALES. (“Holding gender constant....”)
- But, what if the effect of gender varies by the race of the baby (is the effect of gender different for whites and nonwhites)? Alternatively, does the effect of race vary by gender (is the effect of race different for males and females)? Another way we say this: is there a *gender\*race* effect on *birthweight*?
- We can accomplish this in two ways:

*First option:* Include main effects for *white* and *male*, and one interaction term for *whmale*:

To compute an interaction term, simply multiply the two main effect variables:

```
. gen whmale = white*male
. reg bwght white male whmale
```

Source	SS	df	MS			
Model	12525.2107	3	4175.07023	Number of obs =	1388	
Residual	562086.509	1384	406.131871	F( 3, 1384) =	10.28	
Total	574611.72	1387	414.283864	Prob > F =	0.0000	
				R-squared =	0.0218	
				Adj R-squared =	0.0197	
				Root MSE =	20.153	

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
white	6.71902	1.932267	3.48	0.001	2.928531	10.50951
male	3.575291	2.339114	1.53	0.127	-1.013301	8.163882
whmale	<b>-.6594038</b>	2.63905	-0.25	0.803	-5.836374	4.517567
_cons	111.8321	1.721763	64.95	0.000	108.4546	115.2097

What is the baseline category?

E.g.: predicted birthweight from this model for:

- female babies who are not white = **111.8 ounces**
- male babies who are not white =  $111.8 + (3.6*1) = \mathbf{115.4 \text{ ounces}}$
- female babies who are white =  $111.8 + (6.7*1) = \mathbf{118.5 \text{ oz}}$
- male babies who are white =  $111.8 + (6.7*1) + (3.6*1) - (0.66*1) = \mathbf{121.4 \text{ oz}}$

*Second option:* We can use the two different indicator variables (e.g., male/nonmale and white/nonwhite) to create a mutually exclusive set of indicators:

```
. reg bwght whfemale nonwhmale whmale
```

Source	SS	df	MS			
Model	12525.2107	3	4175.07023	Number of obs =	1388	
Residual	562086.509	1384	406.131871	F( 3, 1384) =	10.28	
Total	574611.72	1387	414.283864	Prob > F =	0.0000	
				R-squared =	0.0218	
				Adj R-squared =	0.0197	
				Root MSE =	20.153	

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
whfemale	6.71902	1.932267	3.48	0.001	2.928531	10.50951
nonwhmale	3.575291	2.339114	1.53	0.127	-1.013301	8.163882
whmale	9.634906	1.920523	5.02	0.000	5.867456	13.40236
_cons	111.8321	1.721763	64.95	0.000	108.4546	115.2097

where: WHFEMALE = 1 if white female, =0 otherwise  
 NONWHMALE = 1 if nonwhite male, =0 otherwise  
 WHMALE = 1 if white male, =0 otherwise

- What is the reference category (reflected in the constant) for this conceptual set of indicators?
- Are the coefficients statistically significant?
- Interpret the coefficient on WHFEMALE:
- Interpret the coefficient on NONWHMALE:
- Interpret the coefficient on WHMALE:
- What is the predicted birthweight for
  - (a) nonwhite, female babies?
  - (b) white, female babies?
  - (c) nonwhite, male babies?
  - (d) white, male babies?

These predictions are exactly equal to those obtained under method 1 and both are equivalent to the group means:

```
. * Statistics by gender and race  
. bys male white: summarize bwght
```

```
-----  
-> male = 0, white = 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bwght	137	111.8321	20.12665	30	158

```
-----  
-> male = 0, white = 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bwght	528	118.5511	20.16924	35	271

```
-----  
-> male = 1, white = 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bwght	162	115.4074	22.18095	23	176

```
-----  
-> male = 1, white = 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bwght	561	121.467	19.52113	50	192

- Note that the means above are exactly the same predicted birthweights that we got when we explicitly included the interaction term *whmale* in the model above under option 1.

### III. INTERACTIONS WITH ONE INDICATOR AND ONE INTERVAL/RATIO VAR

- What if the “effect” of the indicator variable (e.g., race) on Y (birthweight) is not constant throughout the range of the ratio/interval variable (e.g., mother’s age), or equivalently, the “effect” of the interval/ratio variable is not the same for both levels of the indicator variable?

Using the data set **BWGHT2.DTA**, how can we allow for the effect of mother’s age (*mage*) on baby’s birthweight to vary by race (where the mother’s race is indicated by *mwhite*). By including an interaction (*mwhmage*) between these two variables, we allow the effect of *mage* on *bwght* to be different for white and nonwhite mothers:

- To allow for this, we need to create an **interaction term** between the *mwhite* and *mage* variables and include this new interaction term in the equation:

```
. use bwght2.dta
.
. gen mwhmage = mwhite*mage
.
. reg bwght mage mwhite mwhmage
```

Source	SS	df	MS	Number of obs = 1832		
Model	3646382.99	3	1215461	F( 3, 1828)	=	3.67
Residual	604984392	1828	330954.262	Prob > F	=	0.0118
Total	608630775	1831	332403.481	R-squared	=	0.0060
				Adj R-squared	=	0.0044
				Root MSE	=	575.29

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mage	18.57864	7.92029	2.35	0.019	3.044867	34.1124
mwhite	586.3524	251.3175	2.33	0.020	93.45283	1079.252
mwhmage	-16.82123	8.475335	-1.98	0.047	-33.44359	-.198877
_cons	2773.682	234.4771	11.83	0.000	2313.811	3233.553

- What group is reflected in the intercept?
- predicted birthweight for babies born to: [note: do these predictions make sense to you?]
  - (a) nonwhite mothers who are zero years old = **2773.68 grams**
  - (b) white mothers who are zero years old =  $2773.68 + 586.35 =$  **3360.03 grams**
  - (c) nonwhite mothers who are 30 years old =  $2773.68 + (18.58*30) =$  **3331.08 grams**
  - (d) white mothers who are 30 years old =  $2773.68 + (18.58*30) + 586.35 + (-16.82*30) =$  **3412.83 grams**

- The interaction term allows the effect of mother's age to vary depending on race of the mother. The coefficient estimate on the interaction  $\hat{\beta}_{MWHMAGE} = -16.82$ , and it is statistically significant at conventional levels ( $p=0.0473$ ).

Because this interaction is statistically significant, we can conclude that the effect of mother's age *does* vary by race.

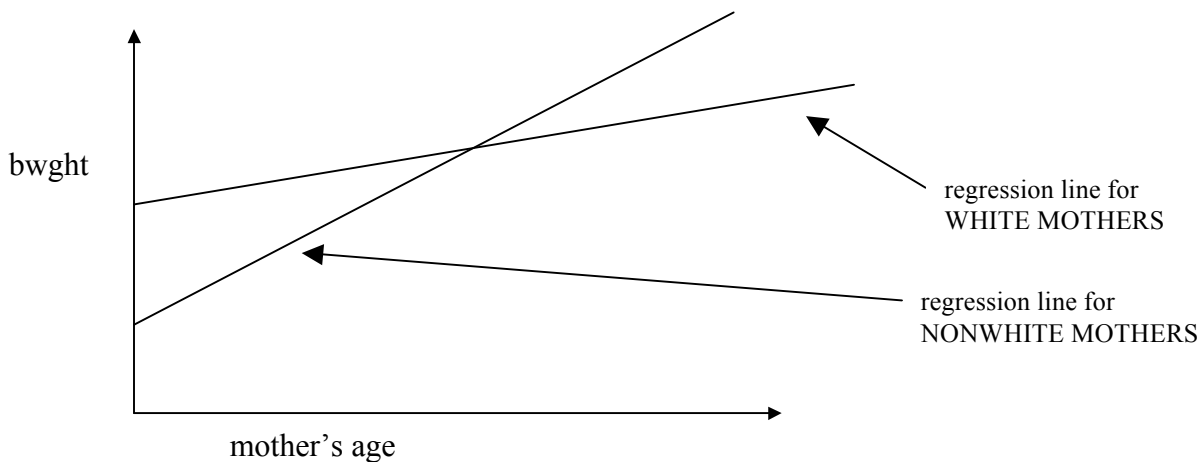
Alternatively, it could be interpreted as meaning that the effect of race on birthweight differs depending on the mother's age.

This is like the "return to education" example in Wooldridge (p. 241) where he is interested in whether an extra year of education has the same effect for men as for women.

In the birthweight example, we are interested in whether an additional year of age for the mother has the same effect on birthweight for whites compared to nonwhites (note that we are not controlling for anything else in this model: if we were, we would interpret the estimates as "holding constant the other variables in the model," as usual).

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mage	18.57864	7.92029	2.35	0.019	3.044867	34.1124
mwhite	586.3524	251.3175	2.33	0.020	93.45283	1079.252
mwhmage	-16.82123	8.475335	-1.98	0.047	-33.44359	-.198877
_cons	2773.682	234.4771	11.83	0.000	2313.811	3233.553

A picture....



- What is the effect of age on birthweight?

Answer—it *depends* on the *race* of the mother:

FOR NONWHITE MOTHERS: 18.57864 (i.e., just the coeff for *mage*)

FOR WHITE MOTHERS: 18.57864 + (- 16.2123) = 2.36

Interpretation: We predict that birthweight will increase by about 18.6 grams for each additional year of mother's age *for nonwhite mothers*, but only by about 2.4 grams *for white mothers*.

- What is the effect of race on birthweight?

Answer—it *depends* on the *age* of the mother (which ranges from 16 to 44 in the sample)

Direct interpretation of *mwhite* coefficient:

A baby born to a white mother who is zero years old (not possible, but hang tight!) has a predicted birthweight of 586 grams *more* than a baby born to a black mother who is zero years old. This is not very useful information, because  $mage=0$  is clearly outside the range of data.

It would be more meaningful to examine the effect of race for realistic ages of motherhood. For example, a baby born to a white mother who is the average age in the sample (29.5578603 years old) is predicted to have a birthweight that is  $(586.35235) + (-16.82123 * 29.5578603) = \mathbf{89.153}$  *grams more, compared to nonwhite babies whose mothers are this same age.*

You could also obtain this answer directly by re-defining terms in the model, and re-estimating the model. In particular, define the interaction term to equal zero when evaluated at the value of interest (in this case, the mean value of *mage*). [see full Stata program at end of notes]

```
*the following will set ave_mage=29.557 for all observations
egen ave_mage = mean(mage)

gen mwhmagenew = (mage-ave_mage)*mwhite

reg bwght mage mwhite mwhmagenew
```

- By doing this, you can look directly at the coefficient on *mwhite* to measure the difference between whites and nonwhites at the value of interest (e.g., average age of 29.557).
- This difference should be exactly equal to the prediction that we just calculated above, and we can look at the t-statistics on the *mwhite* coefficient to test its significance.
- Redefining the interaction term in this way just gives a convenient way to evaluate the effects of one term in the interaction when the other term in the interaction is equal to a particular value.
- If we wanted to evaluate the difference between whites compared to nonwhites at some other value of *mage*, we would need to either (1) redefine a new interaction term by subtracting off this new value of interest from *mage* in the interaction term and rerunning the model or (2) calculate that effect by hand using the original model.

The advantage of the first strategy (subtracting off the particular value of interest, e.g. the mean) is that it automatically calculates the relevant standard error and t-statistic (on the coefficient for *mwhite* in this case) and thus easily allows us to conduct hypothesis tests and calculate confidence intervals.

THE STATA CODE AND OUTPUT AT END OF THIS SECTION (III) OF THE NOTES SHOWS HOW TO DO THIS. This method is also shown in example 6.3 on pp. 198-199 of Wooldridge, and is more generally introduced in Section 4.4 (pp. 140-143).



- At what point is the predicted birthweight for white and nonwhite mothers the same (i.e., at what Y do the regression lines cross?)

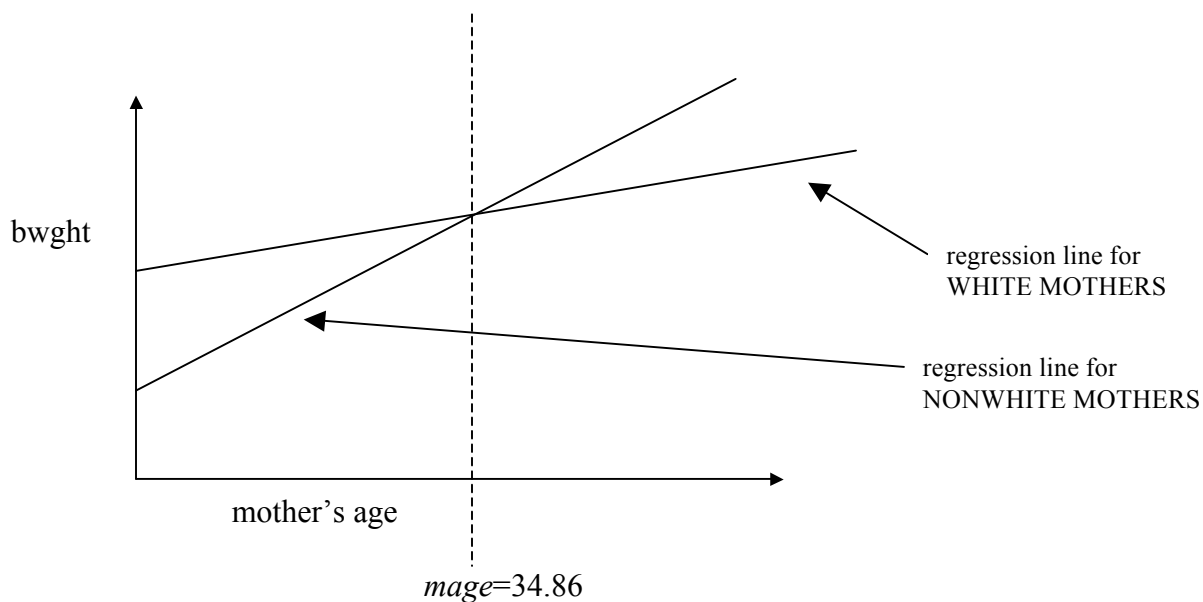
$$\hat{Y}_{white} = \hat{Y}_{nonwhite}$$

$$2773.68 + 18.58MAGE + 586.35 - 16.82MAGE = 2773.68 + 18.58MAGE =$$

$$586.35 - 16.82MAGE = 0$$

$$MAGE = 34.86$$

-- This says that when moms are about 35 years old, we predict that birthweight of babies born to white and nonwhite moms will be the same (plug into the above equation, and that predicted birthweight is 3421.38 grams)



#### IV. TESTING FOR DIFFERENCES ACROSS GROUPS IN REGRESSION FUNCTIONS (AKA, “Chow Test”): A FULLY INTERACTED MODEL

- In the examples above, we interacted only two variables; the slope of one variable was allowed to vary according to different values of another variable. What if we want **all** the coefficients in a model to be different according to different values of another variable?
- First, we’ll talk about this in terms of two separate regression models; then we’ll make the connection to a model that uses interactions, and show that these two different methods produce exactly the same results.
- For example, what if we think that males and females have different coefficient estimates on **all** the other variables in the equation?

If the effects of male and female babies essentially have different “processes” that predict *birthweight* from these variables, then we could think about estimating two separate regressions:

$$\text{females: } BWGHT = \beta_0 + \beta_1CIGS + \beta_2FAMINC + \beta_3MOTHEDEC + \beta_4FATHEDEC + \eta$$

$$\text{males: } BWGHT = \alpha_0 + \alpha_1CIGS + \alpha_2FAMINC + \alpha_3MOTHEDEC + \alpha_4FATHEDEC + \omega$$

- We’ll test whether estimating these regressions separately provides greater explanatory power than if we were to pool both females and males and estimate only one regression:

$$\text{pooled: } BWGHT = \pi_0 + \pi_1CIGS + \pi_2FAMINC + \pi_3MOTHEDEC + \pi_4FATHEDEC + \nu$$

- This is known as a **Chow Test**, and the test statistic is:

$$F = \left( \frac{[SSR_{pooled} - (SSR_1 + SSR_2)]}{SSR_1 + SSR_2} \right) * \left( \frac{n - 2(k + 1)}{k + 1} \right) \text{ with numerator d.f.} = k + 1$$

and denominator d.f. =  $n - 2(k + 1)$

[note: think about why these are the d.f.!]

. bys male: reg bwght cigs faminc motheduc fatheduc

-> male = 0

Source	SS	df	MS	Number of obs =	573
Model	8757.85824	4	2189.46456	F( 4, 568) =	5.27
Residual	235977.779	568	415.453836	Prob > F =	0.0004
Total	244735.637	572	427.859505	R-squared =	0.0358
				Adj R-squared =	0.0290
				Root MSE =	20.383

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cigs	-.5858689	.1574105	-3.72	0.000	-.8950466 - .2766912
faminc	.0727672	.0551728	1.32	0.188	-.0356005 .1811349
motheduc	-.9634324	.4832006	-1.99	0.047	-1.91251 -.0143543
fatheduc	.66321	.4149273	1.60	0.111	-.1517691 1.478189
_cons	120.2174	4.901743	24.53	0.000	110.5897 129.8452

-> male = 1

Source	SS	df	MS	Number of obs =	618
Model	7828.3217	4	1957.08042	F( 4, 613) =	5.31
Residual	225981.997	613	368.649261	Prob > F =	0.0003
Total	233810.319	617	378.947032	R-squared =	0.0335
				Adj R-squared =	0.0272
				Root MSE =	19.2

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cigs	-.5485773	.1553238	-3.53	0.000	-.8536086 - .2435461
faminc	.0515484	.0487796	1.06	0.291	-.0442471 .1473438
motheduc	.0277863	.4241947	0.07	0.948	-.805265 .8608375
fatheduc	.3554441	.3876648	0.92	0.360	-.405868 1.116756
_cons	115.544	5.001263	23.10	0.000	105.7223 125.3656

\* pooled regression with males & females

. reg bwght cigs faminc motheduc fatheduc

Source	SS	df	MS	Number of obs =	1191
Model	15827.6593	4	3956.91482	F( 4, 1186) =	10.05
Residual	466919.033	1186	393.69227	Prob > F =	0.0000
Total	482746.692	1190	405.669489	R-squared =	0.0328
				Adj R-squared =	0.0295
				Root MSE =	19.842

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cigs	-.5894954	.1106172	-5.33	0.000	-.8065225 - .3724682
faminc	.0538254	.0366502	1.47	0.142	-.0180811 .1257319
motheduc	-.4379234	.3197377	-1.37	0.171	-1.065238 .1893912
fatheduc	.4936695	.2832896	1.74	0.082	-.0621351 1.049474
_cons	118.0741	3.500291	33.73	0.000	111.2066 124.9415

- Calculate the F-statistic:

$$F = \left( \frac{[SSR_{pooled} - (SSR_1 + SSR_2)]}{SSR_1 + SSR_2} \right) * \left( \frac{n - 2(k + 1)}{k + 1} \right)$$

$$F = \left( \frac{[466,919 - (235,978 + 225,982)]}{235,978 + 225,982} \right) * \left( \frac{1,191 - 2(4 + 1)}{4 + 1} \right)$$

$$F = \left( \frac{4,959}{461,960} \right) * \left( \frac{1,181}{5} \right) = 2.5355$$

F-critical = (5, 1181) at alpha=0.05: 2.21; at alpha=0.01: 3.02. Therefore, reject the null at the 0.05 significance level; fail to reject at the 0.01 level.

- This test doesn't allow us to know WHICH coefficient(s) are different across the two models (this is no different from what we could learn from earlier F-tests).

*ALTERNATIVE, USING INTERACTIONS:*

- The same kind of test can be achieved using interaction terms. When **all** variables in a model are interacted with one of the indicator variables, it is referred to as a **fully-interacted model**, and it is equivalent to estimating two separate models.

$$\text{BWGHT} = \delta_0 + \delta_1\text{CIGS} + \delta_2\text{FAMINC} + \delta_3\text{MOTHEDEC} + \delta_4\text{FATHEDUC} + \\ + \delta_5\text{MALE} + \delta_6\text{MCIGS} + \delta_7\text{MFAMINC} + \delta_8\text{MMOTHEDEC} + \delta_9\text{MFATHED} + \varphi$$

- Does gender matter in explaining birthweight? The null to be tested using the fully-interacted model is:

$$H_0: \delta_5=0 \text{ and } \delta_6=0 \text{ and } \delta_7=0 \text{ and } \delta_8=0 \text{ and } \delta_9=0$$

$$H_1: \text{at least one of these coefficients is different from zero}$$

- To test the hypothesis above, we can just use an F-test that we already know. Because we are testing all the coefficients that are interacted with *male*, this is also a **Chow test**:

```
. reg bwght cigs faminc motheduc fatheduc male mcigs mfaminc mmothed mfathed
```

Source	SS	df	MS	Number of obs =	1191
Model	20786.916	9	2309.65734	F( 9, 1181) =	5.90
Residual	461959.776	1181	391.159844	Prob > F =	0.0000
				R-squared =	0.0431
				Adj R-squared =	0.0358
Total	482746.692	1190	405.669489	Root MSE =	19.778

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cigs	-.5858689	.1527388	-3.84	0.000	-.8855386 -.2861992
faminc	.0727672	.0535354	1.36	0.174	-.0322679 .1778023
motheduc	-.9634324	.46886	-2.05	0.040	-1.883324 -.0435409
fatheduc	.66321	.4026129	1.65	0.100	-.1267064 1.453126
male	-4.673477	7.011565	-0.67	0.505	-18.42999 9.083036
mcigs	.0372915	.2211962	0.17	0.866	-.3966899 .471273
mfaminc	-.0212188	.073422	-0.29	0.773	-.1652709 .1228332
mmothed	.9912187	.6409045	1.55	0.122	-.2662197 2.248657
mfathed	-.3077659	.5670607	-0.54	0.587	-1.420325 .8047928
_cons	120.2174	4.756268	25.28	0.000	110.8858 129.5491

```
. * test male and all interactions equal to zero
. test male = mcigs = mfaminc = mmothed = mfathed = 0
```

- ( 1) male - mcigs = 0
- ( 2) male - mfaminc = 0
- ( 3) male - mmothed = 0
- ( 4) male - mfathed = 0
- ( 5) male = 0

```
F( 5, 1181) = 2.54
Prob > F = 0.0271
```

- What variables are in the RESTRICTED model? In the UNRESTRICTED model?

- We reject the null that all the coefficients are equal to zero: we conclude that it's likely that the regression functions **do** differ for males and females.

Happily, this is the same F-stat (and conclusion) that Stata calculated when we ran the Chow test using separate regressions and a pooled regression.

- Note that the null hypothesis we test doesn't have to be all-or-nothing. We can test subsets of the interacted terms, using the F-test we already know.

- For example, another null of interest is: Do the effects of CIGS, FAMINC, MOTHEDEC, and FATHEDUC on *birthweight* vary by gender? (i.e., still allow for a difference in intercept).

H0:  $\delta_6=0$  and  $\delta_7=0$  and  $\delta_8=0$  and  $\delta_9=0$

H1: at least one of these coefficients is different from zero

```
. * test male interactions equal to zero (not including male itself)
. test mcigs = mfaminc = mmother = mfatheduc = 0

( 1) mcigs - mfaminc = 0
( 2) mcigs - mmother = 0
( 3) mcigs - mfatheduc = 0
( 4) mcigs = 0

F( 4, 1181) = 0.64
Prob > F = 0.6314
```

### Example 2: (back to regression of birthweight on mother's age and race):

In a sense the following is a “fully interacted” model because there are only 2 independent variables and there is an interaction (*mwhmage*) between those variables included in the model.

```
. reg bwght mage mwhte mwhmage
```

Source	SS	df	MS	Number of obs = 1832		
Model	3646382.99	3	1215461	F( 3, 1828) =	3.67	
Residual	604984392	1828	330954.262	Prob > F =	0.0118	
-----				R-squared =	0.0060	
-----				Adj R-squared =	0.0044	
Total	608630775	1831	332403.481	Root MSE =	575.29	

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mage	18.57864	7.92029	2.35	0.019	3.044867	34.1124
mwhte	586.3524	251.3175	2.33	0.020	93.45283	1079.252
mwhmage	-16.82123	8.475335	-1.98	0.047	-33.44359	-.198877
_cons	2773.682	234.4771	11.83	0.000	2313.811	3233.553

- Using the model above, how do you test the following question:

Controlling for the mother's age, is average birthweight the same regardless of race? (i.e., holding *mage* constant, is there a race effect?; think of this as: does race matter at all?)

We have to use an F-test to test this hypothesis:

H<sub>0</sub>:  $\beta_{mwhte} = 0$  and  $\beta_{mwhmage} = 0$ ;

H<sub>1</sub>:  $\beta_{mwhte} \neq 0$  and/or  $\beta_{mwhmage} \neq 0$ ;

```
. test mwhte = mwhmage = 0

( 1) mwhte - mwhmage = 0
```

```
( 2)  mwhite = 0

      F( 2, 1828) = 4.47
      Prob > F = 0.0116
```

- Note: Similarly, if we wanted to test whether age had no effect on birthweight *for mothers of the same race (i.e., holding race constant)*, we would need to test the joint null:

$$H_0: \beta_{mage} = 0 \text{ and } \beta_{mwhmage} = 0$$

What variables are in the RESTRICTED model?    In the UNRESTRICTED model?

We reject the null at the 0.10 level, but not at the 0.05 level (F= 2.92, and p=0.0541).

## V. INTERACTIONS WITH TWO INTERVAL-RATIO VARIABLES

- Using the same kind of logic, we can interact two interval-ratio variables, e.g., mother's education (*meduc*) and father's education (*feduc*). The interaction term (*mfe*) will tell us whether the effect of mother's education varies by level of father's education; or alternatively, whether the effect of father's education varies by mother's education.

```
. * create interval-ratio interaction
. gen mfe = meduc * feduc
(51 missing values generated)

. reg bwght mage meduc feduc mfe
```

Source	SS	df	MS	Number of obs = 1781		
Model	2807424.06	4	701856.015	F( 4, 1776)	=	2.15
Residual	579096570	1776	326067.889	Prob > F	=	0.0721
Total	581903994	1780	326912.356	R-squared	=	0.0048
				Adj R-squared	=	0.0026
				Root MSE	=	571.02

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mage	1.694888	3.018027	0.56	0.574	-4.22437	7.614145
meduc	-57.9106	30.63663	-1.89	0.059	-117.9982	2.177041
feduc	-50.98418	29.53053	-1.73	0.084	-108.9024	6.934059
mfe	4.525061	2.182147	2.07	0.038	.2452137	8.804907
_cons	3979.003	401.2162	9.92	0.000	3192.098	4765.909

- The main effects are each statistically significant at the .10 level (p for *meduc*=.059), and the interaction is statistically significant (p=.038). The fact that the interaction is significant tells us that effects of mother's education varies by

level of father's education; or alternatively, the effect of father's education varies by level of mother's education:

*Computation of Effects at Selected Levels of Other Interacted Variable*

	<b>Effect of <i>feduc</i></b>	<b>Computation of <i>feduc</i> effect (at set level of <i>meduc</i>)</b>
<i>meduc</i> =8	-14.7837	= - 50.9842 + (4.52506*8)
<i>meduc</i> =12	3.31654	= - 50.9842 + (4.52506*12)
<i>meduc</i> =16	21.41678	= - 50.9842 + (4.52506*16)
	<b>Effect of <i>meduc</i></b>	<b>Computation of <i>meduc</i> effect (at set level of <i>feduc</i>)</b>
<i>feduc</i> =8	-21.7101	= - 57.9106 + (4.52506*8)
<i>feduc</i> =12	-3.60988	= - 57.9106 + (4.52506*12)
<i>feduc</i> =16	14.49036	= - 57.9106 + (4.52506*16)



```

/*****
PPOL 509
Course Notes 11: Interactions
*****/
cd "C:\...\PPOL509\Stata datasets"

capture: log close
log using "..\do files\notes14.txt", text replace
set more off

/*****
Race and Education: bwght.dta data set
*****/
clear
use bwght.dta

* Create Interaction Variables
gen female      = 1-male
gen whmale      = white*male
gen whfemale    = white*female
gen nonwh       = 1-white
gen nonwhmale   = nonwh*male
gen mcigs       = male*cigs
gen mfaminc     = male*faminc
gen mmothed     = male*motheduc
gen mfatheduc   = male*fatheduc

*** Interactions with two indicator variables
reg bwght white male
reg bwght white male whmale
reg bwght whfemale nonwhmale whmale

* Statistics by gender and race
bys male white: summarize bwght

*** Fully Interacted Model: Chow Test
bys male: reg bwght cigs faminc motheduc fatheduc
* pooled regression with males & females
reg bwght cigs faminc motheduc fatheduc

reg bwght cigs faminc motheduc fatheduc male mcigs mfaminc mmothed mfatheduc
* test male and all interactions equal to zero
test male = mcigs = mfaminc = mmothed = mfatheduc = 0
* test male interactions equal to zero (not including male itself)
test mcigs = mfaminc = mmothed = mfatheduc = 0

/*****
Mother's Age: bwght2.dta data set
*****/
clear
use bwght2.dta

*** Interactions with one indicator and one interval-ratio variable
gen mwhmage = mwhte*mage

reg bwght mage mwhte mwhmage

```

```

test mwhite = mwhmage = 0
test mage = mwhmage = 0

* Effect of whmale at average value of Mage
egen ave_mage = mean(mage)    // create new variable with mean age
sum ave_mage                  // check value of mean
gen mwhmagenew = (mage-ave_mage)*mwhite
list bwght mage mwhite mwhmage ave_mage mwhmagenew in 1/10
reg bwght mage mwhite mwhmagenew

*** Interactions with two interval-ratio variables
* create interval-ratio interaction
gen mfe = meduc * feduc
reg bwght mage meduc feduc mfe

log close

```

### STATA CODE TO PRODUCE EFFECT OF *MWHITE* AT AVERAGE *MAGE*:

```

* Effect of whmale at average value of Mage
egen ave_mage = mean(mage)    // create new variable with mean age
sum ave_mage                  // check value of mean
gen mwhmagenew = (mage-ave_mage)*mwhite
list bwght mage mwhite mwhmage mmage ave_mage mwhmagenew in 1/10
reg bwght mage mwhite mwhmagenew

```

### STATA OUTPUT:

```
. sum ave_mage
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ave_mage	1832	29.55786	0	29.55786	29.55786

```
. list bwght mage mwhite mwhmage ave_mage mwhmagenew in 1/10
```

	bwght	mage	mwhite	mwhmage	ave_mage	mwhmage~w
1.	3060	26	0	0	29.55786	0
2.	3730	29	1	29	29.55786	-.5578594
3.	2530	33	1	33	29.55786	3.442141
4.	3289	28	1	28	29.55786	-1.557859
5.	3590	23	1	23	29.55786	-6.557859
6.	3420	28	1	28	29.55786	-1.557859
7.	3355	27	1	27	29.55786	-2.557859
8.	3459	41	0	0	29.55786	0
9.	3590	32	1	32	29.55786	2.442141
10.	4410	16	1	16	29.55786	-13.55786

```
. reg bwght mage mwhte mwhmagenew
```

Source	SS	df	MS			
Model	3646382.99	3	1215461	Number of obs =	1832	
Residual	604984392	1828	330954.262	F( 3, 1828) =	3.67	
				Prob > F =	0.0118	
				R-squared =	0.0060	
				Adj R-squared =	0.0044	
				Root MSE =	575.29	
Total	608630775	1831	332403.481			

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
bwght						
mage	18.57864	7.92029	2.35	0.019	3.044867	34.1124
mwhte	<b>89.15266</b>	42.47618	2.10	0.036	5.845718	172.4596
mwhmagenew	-16.82123	8.475335	-1.98	0.047	-33.44359	-.198877
_cons	2773.682	234.4771	11.83	0.000	2313.811	3233.553

NOTE: Notice that all the coefficients, t-statistics, F-statistic, R-squared, etc. values in this model are exactly the same as those in the original model *except for* the coefficient and t-state on *mwhte*.. For the reasons stated above, the coefficient on *mwhte* now gives you the effect of the mother being white (compared to being nonwhite) on the baby's birthweight *when **mage** is equal to its mean value in the sample*. The coefficient estimate of  $\hat{\beta}_{mwhte} = 89.153$ , exactly what we calculated by hand above from the regression estimated without subtracting 29.557 from *mage* when creating the interaction term *mwhmagenew*.