Introducing Interaction Terms: example of STATA

PA 8181

Guest Lecture by Juhyun Bae

What are Interaction Terms?

- An interaction term is the multiplicative product of two variables that we think "condition" each other's effects ("$x_1 * x_2$" or just "$x_1 x_2$")

- With quadratics, when we think a variable's effect depends on its own value, and we multiply it by itself.

- With interactions, we do the same thing, except in this case the multiplication is with another variable.

Interpretation with Interaction Terms

- Implementing interaction terms is really easy – the pitfalls come with **interpretation**.

- The biggest pitfall: inclusion of an interaction term *changes the interpretation for the constituent terms.*

STATA example

```
. reg trumpft male#white
```

Source	SS	df	MS
Model	539231.343	3	179743.781
Residual	12366541.5	7,924	1560.6438
Total	12905772.9	7,927	1628.07782

Number of obs	= 7,928
F(3, 7924)	= 115.17
Prob > F	= 0.0000
R-squared	= 0.0418
Adj R-squared	= 0.0414
Root MSE	= 39.505

trumpft	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
male#white						
0#1. White, non-Hispanic	19.31665	1.349197	14.32	0.000	16.67187	21.96143
1. Male#0	9.675035	1.725856	5.61	0.000	6.291902	13.05817
1. Male#1. White, non-Hispanic	23.90399	1.374551	17.39	0.000	21.20951	26.59847
_cons	23.61921	1.146639	20.60	0.000	21.37149	25.86692

- DV: thermometer toward Trump

- IV: two dummies – gender and race

What is the reference group?

```
. reg trumpft male#white

      Source │       SS           df       MS              Number of obs   =      7,928
─────────────┼──────────────────────────────────           F(3, 7924)      =     115.17
       Model │   539231.343          3   179743.781         Prob > F        =     0.0000
    Residual │   12366541.5      7,924    1560.6438         R-squared       =     0.0418
─────────────┼──────────────────────────────────           Adj R-squared   =     0.0414
       Total │   12905772.9      7,927   1628.07782         Root MSE        =     39.505

─────────────────────────┬──────────────────────────────────────────────────────────────────
                  trumpft │      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────────────────┼──────────────────────────────────────────────────────────────────
               male#white │
 0#1. White, non-Hispanic │   19.31665   1.349197    14.32   0.000     16.67187    21.96143
                1. Male#0 │   9.675035   1.725856     5.61   0.000     6.291902    13.05817
1. Male#1. White, non-Hispanic │ 23.90399   1.374551    17.39   0.000     21.20951    26.59847
                    _cons │   23.61921   1.146639    20.60   0.000     21.37149    25.86692
─────────────────────────┴──────────────────────────────────────────────────────────────────
```

	Female (male = 0)	Male (male =1)
Non-white (white = 0)	Reference group: Baseline: 23.62	+9.68
White (white = 1)	+19.32	+23.9

STATA example

```
. reg trumpft male#white
```

Source	SS	df	MS
Model	539231.343	3	179743.781
Residual	12366541.5	7,924	1560.6438
Total	12905772.9	7,927	1628.07782

Number of obs	= 7,928
F(3, 7924)	= 115.17
Prob > F	= 0.0000
R-squared	= 0.0418
Adj R-squared	= 0.0414
Root MSE	= 39.505

trumpft	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
male#white						
0#1. White, non-Hispanic	19.31665	1.349197	14.32	0.000	16.67187	21.96143
1. Male#0	9.675035	1.725856	5.61	0.000	6.291902	13.05817
1. Male#1. White, non-Hispanic	23.90399	1.374551	17.39	0.000	21.20951	26.59847
_cons	23.61921	1.146639	20.60	0.000	21.37149	25.86692

Dataset: 2020 ANES

$$Trumpft = 23.62 + 19.32 white female + 9.68 nonwhite male + 23.9 white male + u$$