Interaction

February 7, 2017
Biostatistics 208

Announcement about homework assignments:

- Please include relevant Stata code and output for every question
- File name prefixes for submitted assignments: “LastName HW#" (e.g. "Shiboski HW2")

Interaction in Linear Regression Models

- Defining interaction
- Regression models including interaction terms
- Estimating effects in subgroups using `lincom` in Stata
- Interpretation of coefficients

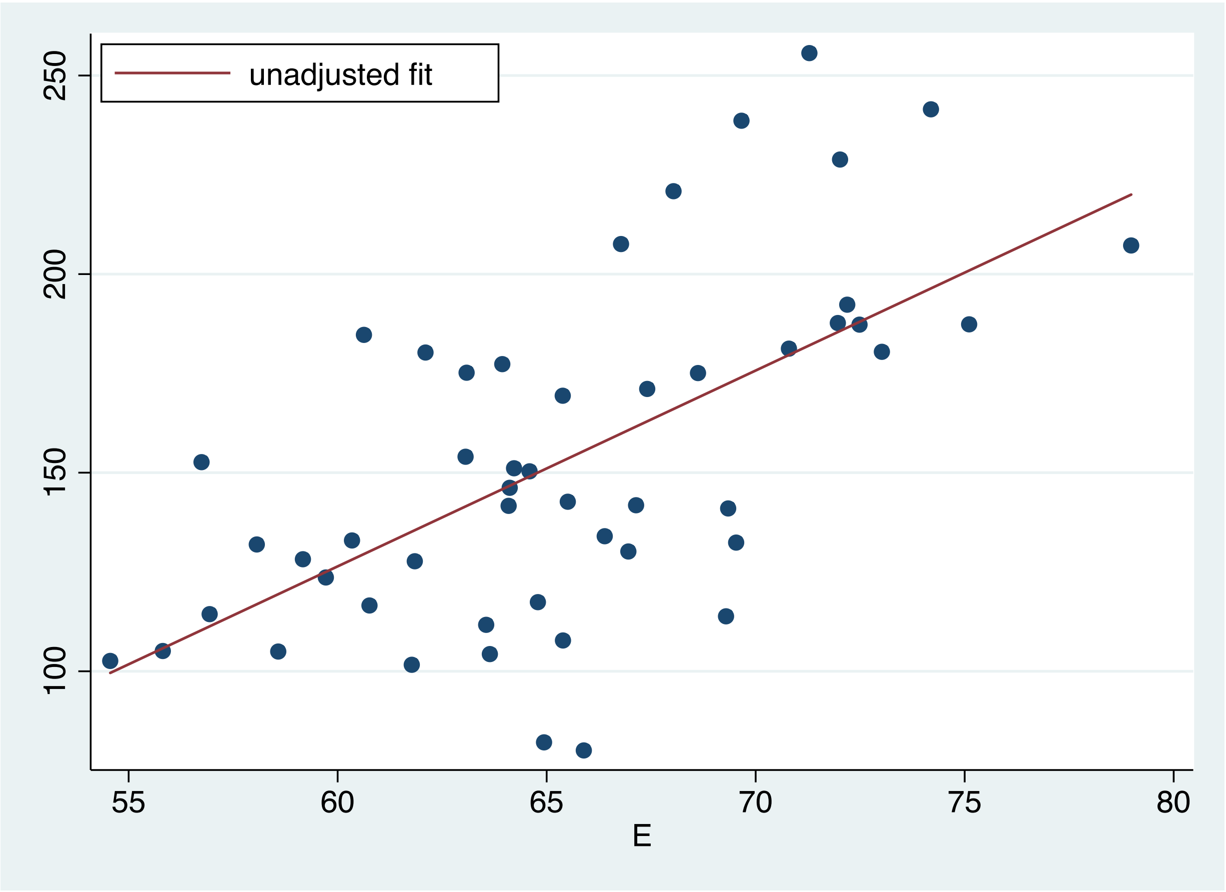
Interaction

- **Confounding and mediation**: association of primary predictor with outcome *differs after adjustment*
  - adjustment variable *explains* or *masks* unadjusted effect of predictor
- **Interaction**: association of primary predictor with the outcome *differs across levels of the interaction variable*
  - interaction variable *modifies/moderates* estimated effect of primary predictor

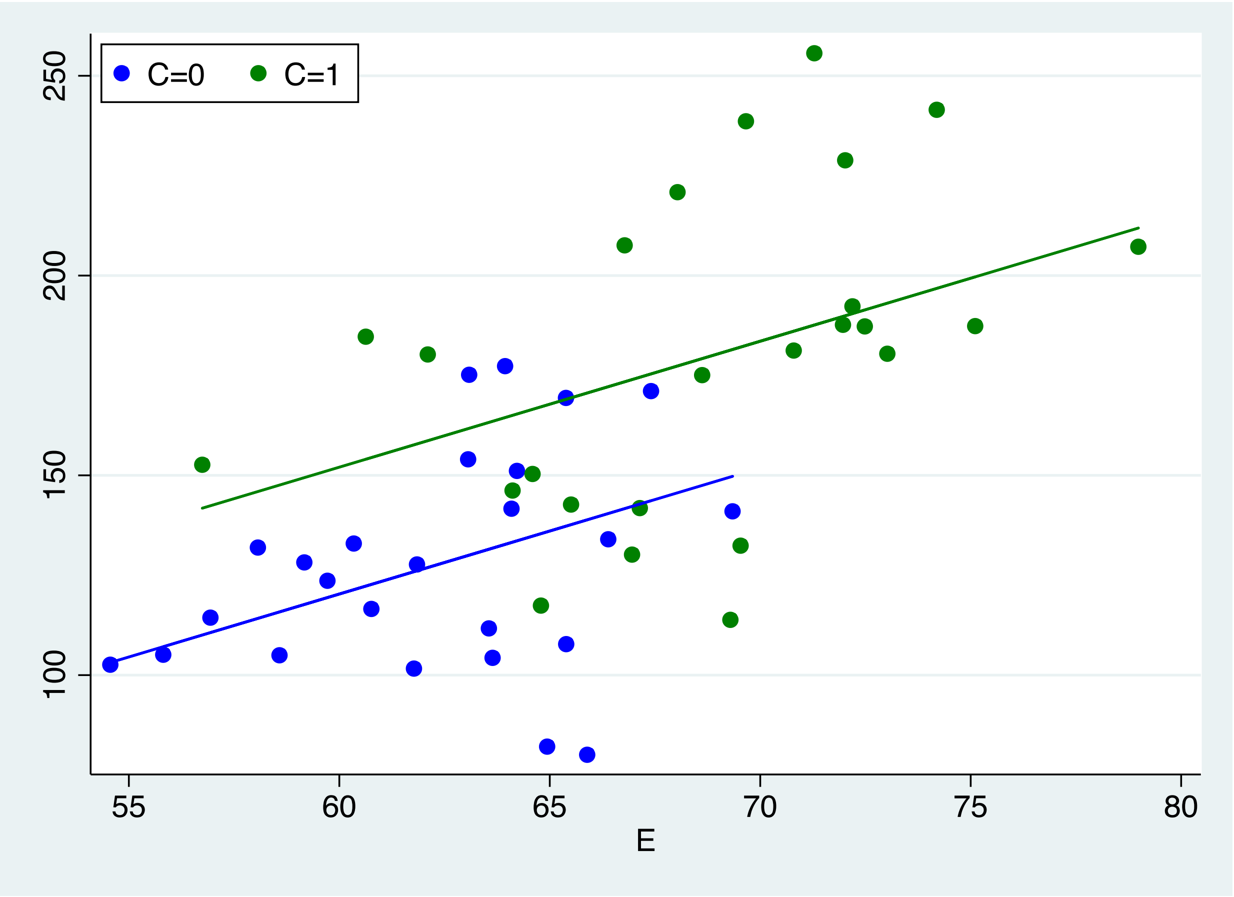
Confounding example based on simulated data (slide #9 - last lecture)

- continuous exposure E
- binary confounder C - no interaction with E
- continuous outcome $y = \beta_0 + \beta_1 E + \beta_2 C + \varepsilon$

Unadjusted



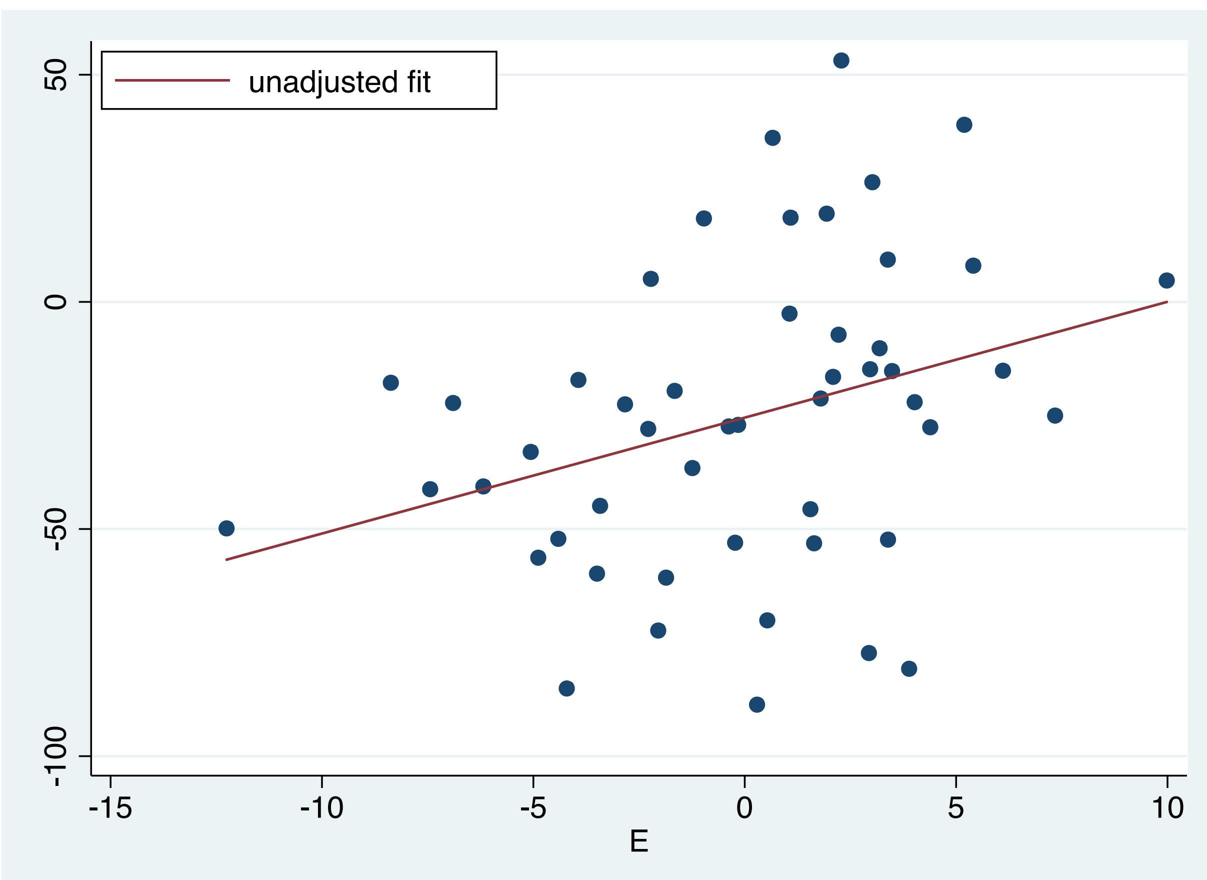
Adjusted



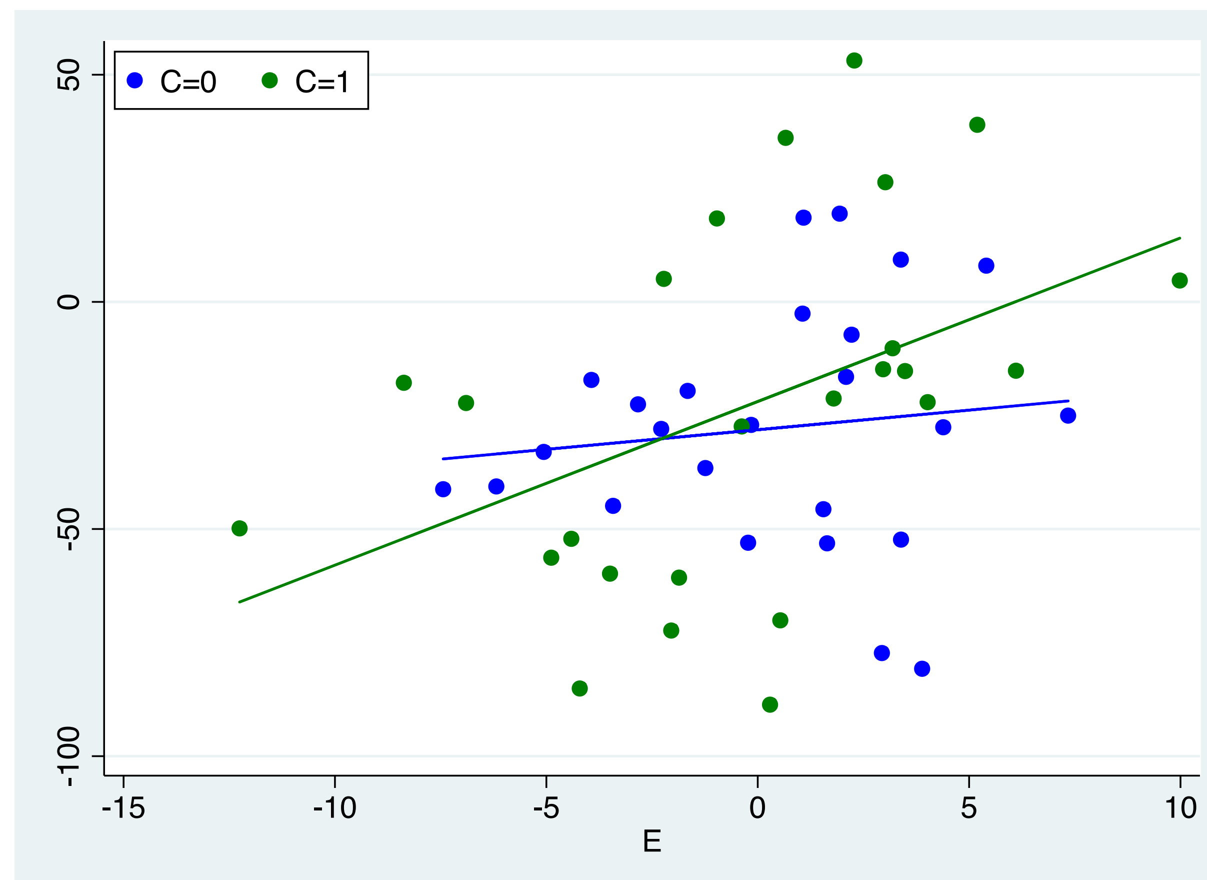
Example based on simulated data including interaction

- continuous exposure E
- binary predictor C - interaction with E
- continuous outcome $y = \beta_0 + \beta_1 E + \beta_2 C + \beta_3 CE + \varepsilon$

Unadjusted



Adjusted



Linear regression models including interaction

To allow interaction between two predictors x_1 & x_2 in a regression model for a continuous outcome y, include both predictors individually, as *main effects* along with their product (for *interaction*):

$$\mathrm{E}[y|x_1, x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2$$

Resulting model allows separate linear effects for x_1 & x_2 :

- Outcome mean as a function of x_1, with x_2 fixed at a constant value c:

$$\begin{aligned}\mathrm{E}[y|x_1, x_2 = c] &= \beta_0 + \beta_1 x_1 + \beta_2 c + \beta_3 x_1 \times c \\ &= (\beta_0 + \beta_2 c) + (\beta_1 + \beta_3 c)\, x_1\end{aligned}$$

- Outcome mean as a function x_2, with x_1 fixed at a constant value c:

$$\begin{aligned}\mathrm{E}[y|x_1 = c, x_2] &= \beta_0 + \beta_1 c + \beta_2 x_2 + \beta_3 c \times x_2 \\ &= (\beta_0 + \beta_1 c) + (\beta_2 + \beta_3 c)\, x_2\end{aligned}$$

effect of x_1:

effect of x_2:

intercept:

- Allows different intercepts & slopes as c varies
- Reduces to standard linear model with no interaction ($\beta_3 = 0$)
- $\beta_3 \neq 0$ represents *interaction* between x_1 & x_2

Example: interaction of BMI and physical activity (PA) effects on SBP in HERS

Motivating questions:

- Excess weight increases SBP, but PA may blunt this effect
- Does higher BMI lead to higher SBP only among inactive?
- Does protective effect of PA on SBP increase with BMI?

- Predictors:

```
egen mb = mean(BMI)
gen BMIc = BMI - mb
recode physact (1/2=0) (3/5=1), g(active)
```

  - Centered BMI: `BMIc = BMI - 28.6` (centered on sample mean for interpretability)
  - Physically active (indicator for activity generated by including `i.active` in model):

    `1.active` (0 = inactive, 1 = active)

- Interaction (product) term `1.active#c.BMIc` (# denotes product term in model):

$$
\begin{aligned}
\texttt{1.active\#c.BMIC} &= \texttt{1.active} \times \texttt{c.BMIc} \\
&= 0 \text{ if } \texttt{active = 0} \\
&= \texttt{BMIc} \text{ if } \texttt{active} = 1
\end{aligned}
$$

Note: "`c.`" syntax denotes a continuous predictor in a regression model - used for including continuous variables in interaction terms

Effects of PA - dependence on BMI

- Model for mean SBP:

$$E[\texttt{SBP}|\ \texttt{active, BMIc}] = \beta_0 + \beta_1 \texttt{1.active} + \beta_2 \texttt{c.BMIc} + \beta_3 \texttt{1.active\#c.BMIc}$$

- Estimated mean SBP for each level of PA:

$$E[\texttt{SBP}|\ \texttt{active=0, BMIc}] = \beta_0 + \beta_2\, \texttt{c.BMIc}$$

$$E[\texttt{SBP}|\ \texttt{active=1, BMIc}] = \beta_0 + \beta_1 + (\beta_2 + \beta_3)\, \texttt{c.BMIc}$$

- Difference is $\beta_1 + \beta_3\, \texttt{c.BMIc}$

- Effect of PA depends on BMI

Effects of BMI among active & inactive participants

$\mathrm{E}[\texttt{SBP}|\ \texttt{active},\ \texttt{BMIc}] = \beta_0 + \beta_1 \texttt{1.active} + \beta_2 \texttt{c.BMIc} + \beta_3 \texttt{1.active\#c.BMIc}$

For inactive:

$\mathrm{E}[\texttt{SBP}|\ \texttt{active=0},\ \texttt{BMIc}] = \beta_0 + \beta_2\, \texttt{c.BMIc}$

- β_2 gives change in mean SBP (*mmHg*) for unit (*kg/m*^{2}) increase in BMI

For active:

$\mathrm{E}[\texttt{SBP}|\ \texttt{active=1},\ \texttt{BMIc}] = (\beta_0 + \beta_1) + (\beta_2 + \beta_3)\, \texttt{c.BMIc}$

- $\beta_2 + \beta_3$ gives change in mean SBP for unit increase in BMI

- Effect of BMI depends on PA

Interaction model for SBP in baseline HERS dataset

```
. regress SBP i.active##c.BMIc

      Source |       SS       df       MS              Number of obs =    2758
-------------+------------------------------           F(  3,  2754) =    4.24
       Model |  4579.65484     3  1526.55161           Prob > F      =  0.0054
    Residual |  992671.256  2754  360.447079           R-squared     =  0.0046
-------------+------------------------------           Adj R-squared =  0.0035
       Total |  997250.911  2757  361.715963           Root MSE      =  18.985

------------------------------------------------------------------------------
         SBP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.active |  -.5468102 β1 .8611566    -0.63   0.525    -2.235388    1.141768
        BMIc |   .0512451 β2 .1138269     0.45   0.653    -.1719496    .2744397
             |
active#c.BMIc |
          1  |   .2293768 β3 .1406966     1.63   0.103    -.0465047    .5052584
             |
       _cons |   135.5883 β0 .7508309   180.58   0.000      134.116    137.0605
-------------+----------------------------------------------------------------
```

Note: syntax "`i.active##c.BMIc`" is shorthand for including: `i.active`, `c.BMIc` & `i.active#c.BMIc`.

Equivalent command from earlier Stata versions ($\leq$10):

```
xi: regress sbp i.active*BMIc
```

Equivalent specification of previous model

```
. regress SBP i.active c.BMIc i.active#c.BMIc

      Source |       SS       df       MS              Number of obs =    2758
-------------+------------------------------           F(  3,  2754) =    4.24
       Model |  4579.65484     3  1526.55161           Prob > F      =  0.0054
    Residual |  992671.256  2754  360.447079           R-squared     =  0.0046
-------------+------------------------------           Adj R-squared =  0.0035
       Total |  997250.911  2757  361.715963           Root MSE      =  18.985

------------------------------------------------------------------------------
          SBP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     1.active |  -.5468102   .8611566    -0.63   0.525    -2.235388    1.141768
         BMIc |   .0512451   .1138269     0.45   0.653    -.1719496    .2744397
              |
active#c.BMIc |
            1 |   .2293768   .1406966     1.63   0.103    -.0465047    .5052584
              |
        _cons |   135.5883   .7508309   180.58   0.000      134.116    137.0605
------------------------------------------------------------------------------
```

Interpretation of coefficients: `_cons` (β_0)

E[SBP| active, BMIc] = $\beta_0 + \beta_1$1.active + β_2c.BMIc + $\beta_3$1.active#c.BMIc

- Interpretation: average SBP in inactive participants with BMIc= 0

  (i.e., BMI set at mean value of 28.6 *kg*/*m*^{2}):

E[SBP|active=0, BMIc=0] = β_0

- Average SBP among inactive with BMI = 28.6 *kg*/*m*^{2} is 135.6 *mmHg* (95% CI: 134.1, 137.1)

Interpretation of coefficients: `1.active` (β_1)

$$E[\texttt{SBP}|\ \texttt{active},\ \texttt{BMIc}] = \beta_0 + \beta_1 \texttt{1.active} + \beta_2 \texttt{c.BMIc} + \beta_3 \texttt{1.active\#c.BMIc}$$

- Interpretation: effect of being active for participants with $\texttt{BMIc} = 0$:

$$E[\texttt{SBP}|\texttt{active} = 1,\ \texttt{BMIc} = 0] = \beta_0 + \beta_1$$

$$E[\texttt{SBP}|\texttt{active} = 0,\ \texttt{BMIc} = 0] = \beta_0$$

- Difference is β_1 (i.e. effect of being active for those with average BMI)
  - At BMI of 28.6 *kg/m*^{2} (mean), PA reduces SBP by 0.55 *mmHg* (95% CI: -2.2, 1.1)
- For other values of BMI the effect of PA on SBP involves the coefficient β_3

Interpretation of coefficients: $\texttt{BMIc}$ (β_2)

$$E[\texttt{SBP}|\ \texttt{active}, \texttt{BMIc}] = \beta_0 + \beta_1 \texttt{1.active} + \beta_2 \texttt{BMIc} + \beta_3 \texttt{1.active\#c.BMIc}$$

- $\texttt{BMIc}$ (β_2): effect of a 1-*kg/m*^{2} increase in BMI, if inactive

$$E[\texttt{SBP}|\texttt{active} = 0,\ \texttt{BMIc} = k+1] = \beta_0 + \beta_2 (k+1)$$

$$E[\texttt{SBP}|\texttt{active} = 0,\ \texttt{BMIc} = k] = \beta_0 + \beta_2 k$$

- Difference is β_2 for any value of BMI (k)

- Among inactive, a 1-*kg/m*^{2} increase in BMI increases SBP by 0.05 *mmHg* (95% CI: -0.2, 0.3)

Interpretation of coefficients: `1.active#c.BMIc` (β_3)

$$E[\texttt{SBP}|\ \texttt{active}, \texttt{BMIc}] = \beta_0 + \beta_1 \texttt{1.active} + \beta_2 \texttt{BMIc} + \beta_3 \texttt{1.active\#c.BMIc}$$

1. Difference between BMI effects in active and inactive

$$E[\texttt{SBP}|\texttt{active} = 1,\ \texttt{BMIc} = k{+}1] - E[\texttt{SBP}|\texttt{active} = 1,\ \texttt{BMIc} = k] -$$
$$E[\texttt{SBP}|\texttt{active} = 0,\ \texttt{BMIc} = k{+}1] - E[\texttt{SBP}|\texttt{active} = 0,\ \texttt{BMIc} = k]$$
$$= \{(\ \beta_0 + \beta_1 + (\beta_2 + \beta_3)(k{+}1)\ ) - (\ \beta_0 + \beta_1 + (\beta_2 + \beta_3)k\ )\ \} -$$
$$\{(\ \beta_0 + \beta_2(k{+}1)\ ) - (\ \beta_0 + \beta_2 k\ )\ \} = \beta_3$$

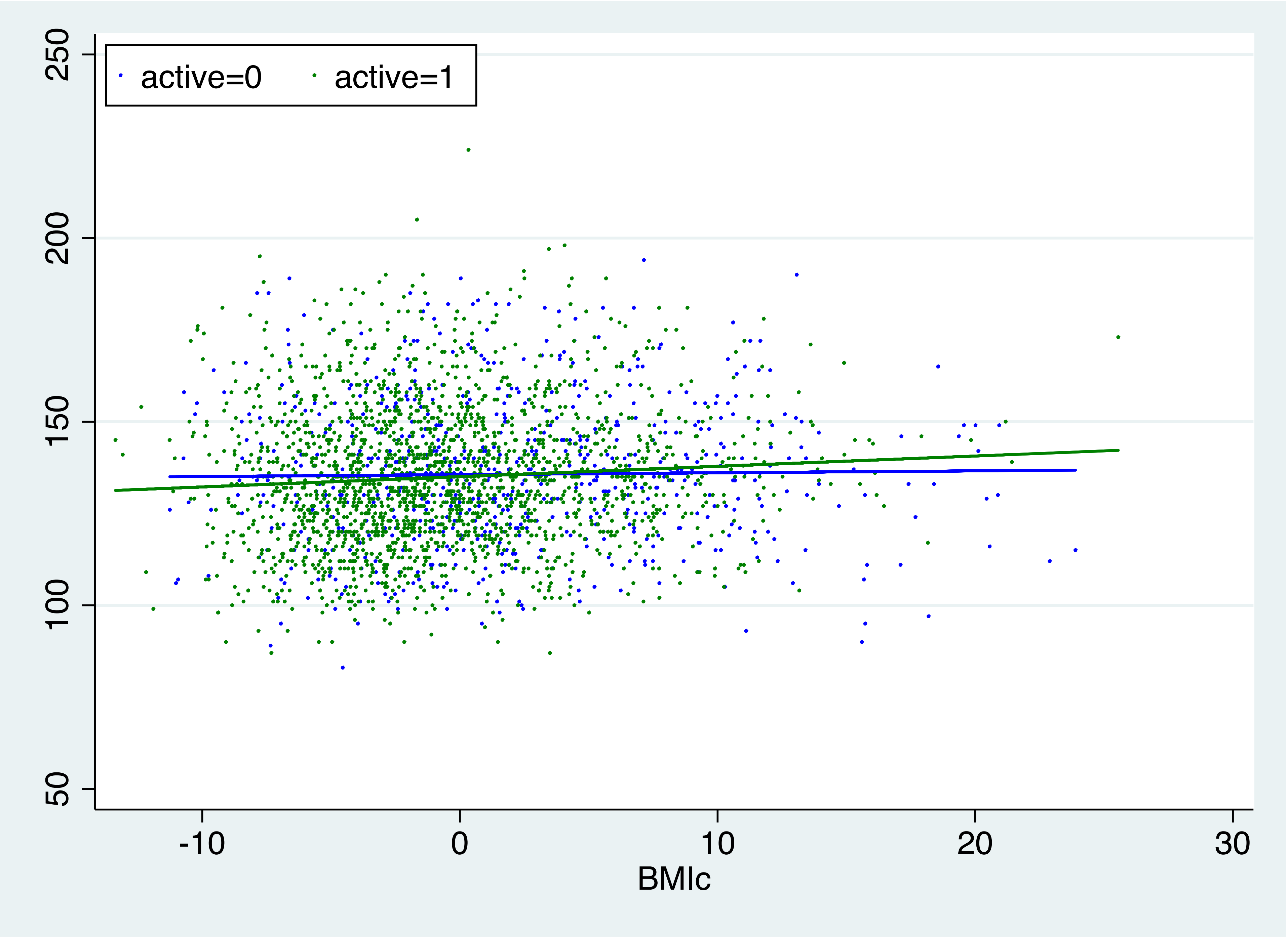
- among active, a unit *kg/m*^{2} increase in BMI predicts a 0.23 *mmHg greater* increase in SBP than among inactive (95% CI: -0.05, 0.51)

2. Change in effect of PA for every *kg/m*^{2} increase in BMI

- for every *kg/m*^{2} increase in BMI, the beneficial effect of being active on SBP *decreases* by 0.23 *mmHg* (95% CI: -0.05, 0.51)

3. *p*-value of 0.10 for `1.active#c.BMIc` shows that evidence for interaction is weak

Illustrating BMI - activity interaction in HERS




```
. regress SBP i.active##c.BMIc
. predict p
. graph twoway (scatter SBP BMIc if active==0, mcolor(blue) msize(vtiny))
(scatter SBP BMIc if active==1, mcolor(green) msize(vtiny))  (line p BMIc if
active==0, clcolor(blue)) (line p BMIc if active==1, clcolor(green)),
legend( order(1 "active=0" 2 "active=1") ring(0) pos(10) )
```

Use of `lincom` for summarizing interaction effects

- Effects of PA & BMI depend on interaction coefficient `1.active#c.BMIc` (β_3)
  - `lincom` statements can be used to illustrate effects (preferred to presenting raw coefficients)
  - Example - two `lincom` statements:

```
. lincom 1.active + 1.4*1.active#c.BMIc
```
($\beta_1 + 1.4\beta_3$)

```
( 1)  1.active + 1.4*1.active#c.BMIc = 0

------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.2256828   .8505544    -0.27   0.791    -1.893472    1.442106
------------------------------------------------------------------------------
```

```
. lincom BMIc + 1.active#c.BMIc
```
($\beta_2 + \beta_3$)

```
 ( 1)  BMIc + 1.active#c.BMIc = 0

------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .2806219   .0826981     3.39   0.001     .1184653    .4427785
------------------------------------------------------------------------------
```

Interpretation of first `lincom` result:

`lincom 1.active + 1.4*1.active#c.BMIc` $(\beta_1 + 1.4\beta_3)$

- Interpretation: effect of PA for an individual with BMI of 30 *kg/m*^{2}
  - for `BMI=30`, `BMIc`=(30-28.6) = 1.4 *kg/m*^{2}

$$E[\texttt{SBP}|x] = \beta_0 + \beta_1\texttt{1.active} + \beta_2\texttt{BMIc} + \beta_3\texttt{1.active\#c.BMIc}$$

Active: $E[\texttt{SBP}|\texttt{active} = 1, \texttt{BMIc} = 1.4] = \beta_0 + \beta_1 + 1.4(\beta_2 + \beta_3)$

Inactive: $E[\texttt{SBP}|\texttt{active} = 0, \texttt{BMIc} = 1.4] = \beta_0 + 1.4\beta_2$

- Difference is $\beta_1 + 1.4\beta_3 = -0.23$
- At BMI of 30 *kg/m*^{2}, PA reduces mean SBP by 0.23 *mmHg* (95% CI: -1.9, 1.4)

Interpretation of second `lincom` result:

`lincom c.BMIc + 1.active#c.BMIc` ($\beta_2 + \beta_3$)

- Interpretation: effect of a 1-kg/m^2 increase in BMI, for an active individual

$$E[\texttt{SBP}|x] = \beta_0 + \beta_1\texttt{1.active} + \beta_2\texttt{BMIc} + \beta_3\texttt{1.active\#c.BMIc}$$

$$E[\texttt{SBP}|\texttt{active} = 1, \texttt{BMIc} = k+1] = \beta_0 + \beta_1 + (k+1)(\beta_2 + \beta_3)$$

$$E[\texttt{SBP}|\texttt{active} = 1, \texttt{BMIc} = k] \quad = \beta_0 + \beta_1 + k(\beta_2 + \beta_3)$$

- Difference is $\beta_2 + \beta_3 = 0.28$, for any value of k
- Among active, a 1- kg/m^2 increase in BMI increases SBP by 0.28 *mmHg*

  (95% CI: -0.12, 0.44)

Figuring out what `lincom` statement to use:

- Example 1: Estimate effect of PA at BMI of 35 (i.e. `BMIc`= 35-28.6 = 6.4 *kg/m*^{2}):

$$E[\texttt{SBP}|\texttt{active}, \texttt{BMIc}] = \beta_0 + \beta_1 \texttt{1.active} + \beta_2 \texttt{BMIc} + \beta_3 \texttt{1.active\#c.BMIc}$$

Active	BMIc	_cons	1.active	BMIc	1.active#BMIc
yes	6.4	1	1	6.4	6.4
no	6.4	1	0	6.4	0
	Diff.	0	1	0	6.4

- Use `lincom 1.active + 6.4 * 1.active#c.BMIc` (output on next slide)

Figuring out what `lincom` statement to use:

```
* Example 1
. lincom 1.active + 6.4 * 1.active#c.BMIc

 ( 1)  1.active + 6.4*1.active#c.BMIc = 0

------------------------------------------------------------------------------
         SBP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .9212014     1.1367     0.81   0.418    -1.307668    3.150071
-------------+----------------------------------------------------------------
```

Interpretation: The estimated difference in average SBP between active and inactive individuals with a BMI of 35 kg/m^2 is 0.92 *mmHg* (95% CI: -1.37, 3.15).

Using `margins` to get the group means:

```
. margins active, at(BMIc=6.4)
Adjusted predictions                              Number of obs     =      2758
Model VCE    : OLS

Expression   : Linear prediction, predict()
at           : BMIc            =         6.4

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      active |
          0  |   135.9163    .8817122   154.15   0.000     134.1874    137.6451
          1  |   136.8375    .7174047   190.74   0.000     135.4307    138.2442
------------------------------------------------------------------------------
```

Figuring out what `lincom` statement to use:

- Example 2: Estimate effect of additional 5 kg/m^2 of BMI among active individuals:

$$E[\texttt{SBP}|\texttt{active}, \texttt{BMIc}] = \beta_0 + \beta_1 \texttt{1.active} + \beta_2 \texttt{BMIc} + \beta_3 \texttt{1.active\#c.BMIc}$$

Active	BMIc	_cons	1.active	BMIc	1.active#BMIc
yes	*k*+5	1	1	*k*+5	*k*+5
yes	*k*	1	1	*k*	*k*
	Diff.	0	0	5	5

- Use `lincom 5 * (BMIc + 1.active#c.BMIc)`

Figuring out what `lincom` statement to use:

```
* Example 2
. lincom 5 * (BMIc + 1.active#c.BMIc)

 ( 1)  5*BMIc + 5*1.active#c.BMIc = 0

------------------------------------------------------------------------------
         SBP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |    1.40311   .4134907     3.39   0.001     .5923264    2.213893
------------------------------------------------------------------------------
```

Interpretation: The estimated effect on average SBP of an increase of 5 *kg/m*^{2} in BMI among active individuals is 1.40 *mmHg* (95% CI: $0.59, 2.21$)

Figuring out what `lincom` statement to use:

- Example 3: Estimate effect of additional 5 *kg/m*^{2} among the *inactive*:

$$E[\texttt{SBP}|\texttt{active}, \texttt{BMIc}] = \beta_0 + \beta_1 \texttt{1.active} + \beta_2 \texttt{BMIc} + \beta_3 \texttt{1.active\#c.BMIc}$$

Active	BMIc	_cons	1.active	BMIc	1.active#BMIc
no	*k+5*	1	0	k+5	0
no	*k*	1	0	k	0
	Diff.	0	0	5	0

- Use `lincom 5*BMIc`

Figuring out what `lincom` statement to use:

```
* Example 3
. lincom 5*BMIc

 ( 1)  5*BMIc = 0

------------------------------------------------------------------------------
         SBP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .2562254   .5691343     0.45   0.653    -.8597478    1.372199
------------------------------------------------------------------------------
```

Interpretation: The estimated effect on average SBP of an increase of 5 *kg/m*^{2} in BMI among inactive individuals is 0.26 *mmHg* (95% CI: -0.86, 1.37)

Interactions between continuous predictors

- Be careful with interactions between continuous variables
  - non-linearity, influential points can be a problem
  - easier to interpret if one or both dichotomized, but information gets lost
- Restricted cubic splines work (next week), but results only interpretable by plotting
- An example will be covered in this week's lab

Interactions between multi-category predictors

- Hard to detect and interpret
- Example - predictors with 3 and 4 categories:
  - $(3 - 1) \times (4 - 1) = 6$ interaction parameters
- Manually-constructed product terms often don't work; best to use Stata syntax
  - newer versions (≥ 11): `regress y i.catvar1##i.catvar2`
  - earlier versions (≤ 10): `xi: regress y i.catvar1*i.catvar2`

Additive models, interactions and outcome scale

- Outcome scale:
  - as measured vs log-transformed; absolute vs percent change
  - logistic, Poisson, negative binomial models: $\mathrm{E}[y \mid x]$ linked to predictors via logit or log transformations
- Additive models excluding interaction terms hold exactly on at most one outcome scale (may hold approximately on > 1)
- Changes of scale can eliminate need for interaction term

Effects of hormone therapy (`ht`) on change in LDL (`ldlch`)

```
. tab ht
  assignment to |
hormone therapy |      Freq.     Percent        Cum.
----------------+-----------------------------------
        placebo |      1,383       50.05       50.05
hormone therapy |      1,380       49.95      100.00
----------------+-----------------------------------
          Total |      2,763      100.00
```

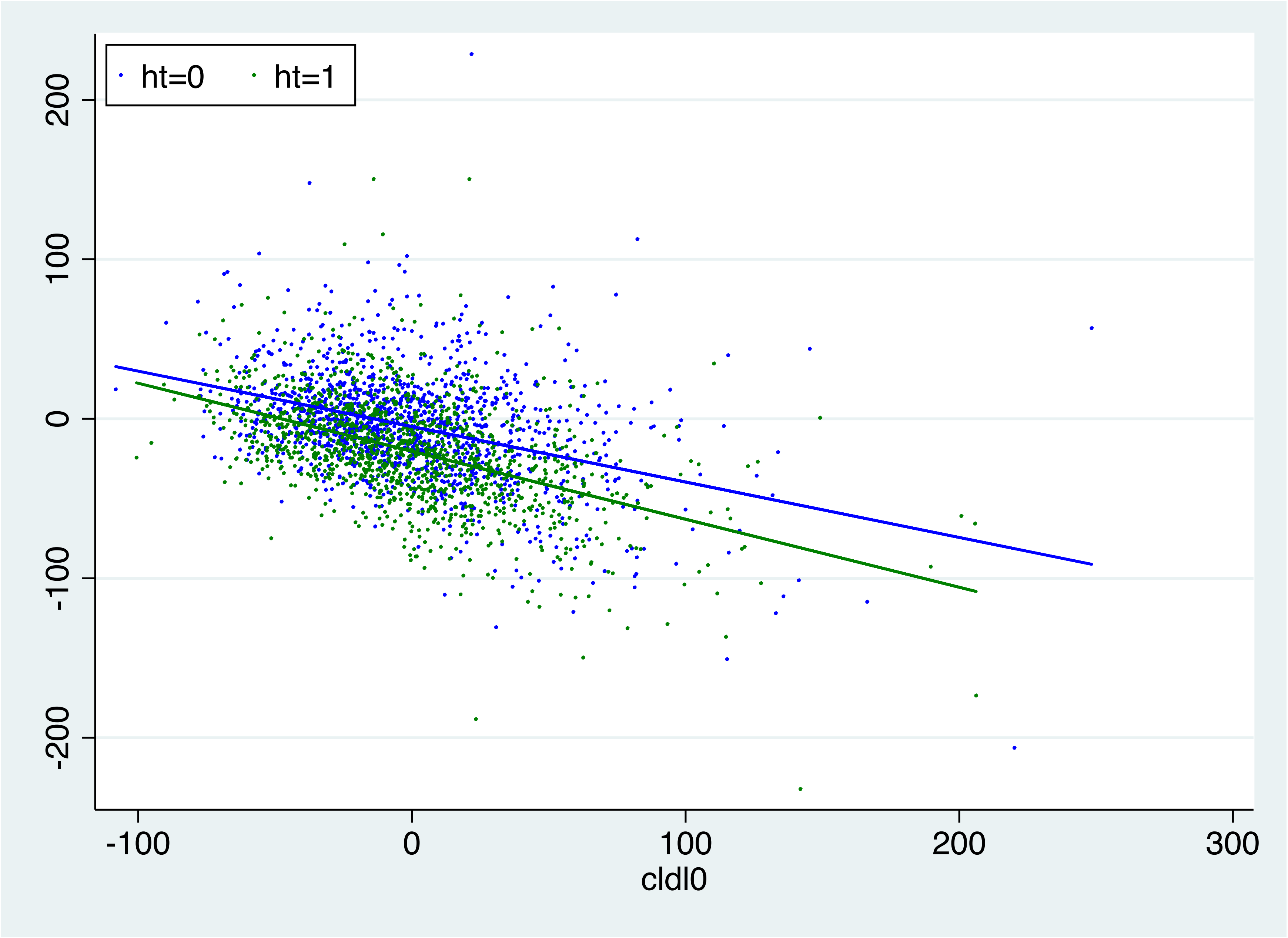
```
. regress ldlch ht##c.cldl0
```
(`ldlch` is change from baseline in LDL level, cldl0 is centered baseline level)

```
      Source |       SS       df       MS              Number of obs =    2597
-------------+------------------------------           F(  3,  2593) =  258.81
       Model |  721218.969     3  240406.323           Prob > F      =  0.0000
    Residual |  2408575.51  2593  928.876015           R-squared     =  0.2304
-------------+------------------------------           Adj R-squared =  0.2295
       Total |  3129794.48  2596  1205.62191           Root MSE      =  30.477

------------------------------------------------------------------------------
           ldlch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
              ht |
 hormone therapy |  -15.47703   1.196246   -12.94   0.000    -17.82273   -13.13134
           cldl0 |  -.3477064   .0225169   -15.44   0.000    -.3918593   -.3035534
                 |
      ht#c.cldl0 |
 hormone therapy |  -.0786871   .0316365    -2.49   0.013    -.1407226   -.0166517
                 |
           _cons |  -4.888739   .8408392    -5.81   0.000    -6.537523   -3.239955
------------------------------------------------------------------------------
```

Interpretation: among post-menopausal women with CHD, HT induces greater absolute reductions in LDL among women with higher baseline levels

Effects of hormone therapy (`ht`) on change in LDL (`ldlch`)



Effects of hormone therapy (`ht`) on change in LDL (`ldlch`)

```
* Use of margins to estimate effect of HT at diff. levels of (centered) baseline LDL
* Effect among indiv. at mean, and at 25 and 50 units more than the mean base. level
. margins ht, at(cldl0=(0 25 50))
Adjusted predictions                              Number of obs   =       2597
Model VCE    : OLS

Expression   : Linear prediction, predict()

1._at        : cldl0           =           0
2._at        : cldl0           =          25
3._at        : cldl0           =          50
```

	Margin	Delta-method Std. Err.	t	P>|t|	[95% Conf.	Interval]
_at#ht						
1#placebo	-4.888739	.8408392	-5.81	0.000	-6.537523	-3.239955
1#hormone therapy	-20.36577	.8508778	-23.94	0.000	-22.03424	-18.6973
2#placebo	-13.5814	1.017519	-13.35	0.000	-15.57663	-11.58617
2#hormone therapy	-31.02561	1.017472	-30.49	0.000	-33.02075	-29.03047
3#placebo	-22.27406	1.41331	-15.76	0.000	-25.04539	-19.50273
3#hormone therapy	-41.68545	1.401366	-29.75	0.000	-44.43336	-38.93754

Interpretation: estimated effect of HT increases with higher baseline levels

HT and percent change LDL

```
. gen ldlpctch = (ldlch / ldl0) * 100
. regress ldlpctch ht##c.cldl0
```
(ldlpctch is % change from baseline in LDL level)

```
      Source |       SS       df       MS              Number of obs =    2597
-------------+------------------------------           F(  3,  2593) =  165.33
       Model |  233394.163     3  77798.0545           Prob > F      =  0.0000
    Residual |  1220171.82  2593  470.563756           R-squared     =  0.1606
-------------+------------------------------           Adj R-squared =  0.1596
       Total |  1453565.98  2596  559.925263           Root MSE      =  21.692

------------------------------------------------------------------------------
    ldlpctch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ht |
hormone therapy |  -10.79035   .8514335   -12.67   0.000    -12.45991   -9.120789
       cldl0 |  -.2162437   .0160265   -13.49   0.000    -.2476697   -.1848176
             |
 ht#c.cldl0 |
hormone therapy |   .0218767   .0225175     0.97   0.331    -.0222773    .0660307
             |
       _cons |  -1.284977   .5984713    -2.15   0.032    -2.458507   -.1114465
------------------------------------------------------------------------------
```

Interpretation: *percent* reductions in LDL induced by HT *do not* clearly vary by baseline level

Data dredging for interaction in multi-predictor models

- With 10 predictors
  - 45 possible 2-way interactions – 120 possible 3-way interactions
- Considering them all inflates type-I error, uncovers trivial interactions in big datasets
- If goal is outcome prediction, criteria for considering interactions differs (lecture 7)

Focused data dredging

- Focus on plausible interactions (Harrell et al., Stat Med, 1996;15:361-87)
  - treatment and severity of disease
  - treatment and calendar time
  - age and risk factors
  - measurement and state during measurement (e.g. l. ventric. function and stress level)
  - *with primary predictor*

If we hadn't centered BMI ...

- In BMI and PA example, if we did not center BMI
  - β_0: average SBP if inactive, BMI $= 0$
  - β_1: effect of PA if BMI $= 0$
  - estimated effect of PA at any plausible BMI requires `lincom`
- With centered BMI:
  - β_0 and β_1 refer to participants with average BMI

If we hadn't centered BMI ...

```
. regress SBP i.active##c.BMI

      Source |       SS       df       MS              Number of obs =    2758
-------------+------------------------------           F(  3,  2754) =    4.24
       Model |  4579.65484     3  1526.55161           Prob > F      =  0.0054
    Residual |  992671.256  2754  360.447079           R-squared     =  0.0046
-------------+------------------------------           Adj R-squared =  0.0035
       Total |  997250.911  2757  361.715963           Root MSE      =  18.985

------------------------------------------------------------------------------
         SBP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.active |  -7.102228   4.251184    -1.67   0.095    -15.43806    1.233603
         BMI |   .0512451   .1138269     0.45   0.653    -.1719496    .2744397
             |
active#c.BMI |
          1  |   .2293768   .1406966     1.63   0.103    -.0465047    .5052584
             |
       _cons |   134.1237   3.544294    37.84   0.000      127.174    141.0735
------------------------------------------------------------------------------
```

- Average SBP among inactive with BMI = 0 *kg/m*^{2} is 134.1 *mmHg*
- At BMI of 0 *kg/m*^{2}, PA reduces SBP by 7.1 *mmHg*

Jeff's (Epi 203) criteria for reporting interaction

- *Substantively trivial* (e.g., *RR*s of 3.2 and 3.5 for two levels of a binary interacting factor): ignore, regardless of *p*-value for interaction
- *Substantively important* (e.g., *RR*s of 2.0 and 3.5): report if interaction *p*-value < 0.10
- *Qualitative interaction* (e.g., RRs of -2.0 and 2.0): report if interaction *p*-value < 0.15

Possible additional criteria for reporting

- Trying to prove something (e.g., show that BiDil should be approved for heart failure in African American patients):
  - need strong evidence, especially for ad hoc finding, possibly penalized for multiple comparisons
- Weak evidence for interaction that undermines your controversial hypothesis:
  - good to report as a sensitivity analysis

Further recommendations[1]

- Not enough to show a within-subgroup effect; need to show that effects significantly differ by subgroup
- Focus on a priori interactions, those with primary predictor
- Evaluate interactions for strength, biological plausibility
- Identify post hoc interactions
- In ambiguous cases, present both overall and stratified results

1 Wang et al., *NEJM*, 2008;357(21):2189-94

Summary: Interaction

- Summary: Interaction
- Interaction: association of one predictor with outcome modified by another predictor
- Modeled using product terms, made automatically by Stata ## operator
- Subgroup-specific regression lines for continuous predictor not parallel
- Centering makes “main effects” easier to interpret
- Power not always low, but can be

Confounding, mediation, and interaction

- Confounders are a cause of both outcome and primary predictor, or surrogates for such a cause
- Potential confounders must be associated with predictor and independently with the outcome
- Confounding can account for the some or all of the unadjusted association between a predictor and an outcome

Mediation vs confounding

- Mediators are on causal pathway from predictor to outcome
- In multi-predictor models, mediation and confounding behave alike, must be distinguished on substantive grounds
- Some problems too complex for this paradigm:
  - LDL both confounds, mediates statin effects on CHD
- You have the substantive expertise to draw the DAG

Interaction vs confounding/mediation

- Interaction vs confounding/mediation
- Important to distinguish between:
  - change in the coefficient for a primary predictor in two models *with and without adjustment for a confounder or mediator*, which we sometimes use to assess confounding and/or mediation
  - differences in the estimated effect of a primary predictor *across strata defined by an effect modifier* in a single model that includes interaction term(s)

Confounding, mediation, or interaction?

1. Is alendronate effect on BMD reduced in vitamin-D deficient patients?
2. Is alendronate effect on vertebral fracture explained by changes in BMD?
3. Is HT effect on CHD explained by a healthier lifestyle?
4. Are beta-blockers less effective among African Americans?
5. Is diabetes a stronger risk factor for CHD events in women than in men?