Comments on Biostatistics 208 Lab #5, 2/6/20

Interaction of hormone therapy and statin use

Question 1. Is there evidence for differential effects of HT on change in LDL according to use of statins at baseline?

Yes, clearly; the *p*-value for `1.ht#1.statins`, which estimates the difference between the HT effects in the two groups defined by use of statins at baseline, is only 0.003.

Question 2. Summarize the HT effects by baseline statin use.
Now use the following two tables to compute the associations of being a statin user with change in LDL.

From the `regress` output, the coefficient for `ht` shows that HT causes an average reduction of 18.5 *mg/dL* in LDL (95% CI 15.2-21.8, $p < 0.0005$) among women not using statins at baseline. The result for `lincom 1.ht + 1.ht#1.statins` shows that HT caused an average reduction of 10.4 *mg/dL* (95% CI 6.2-14.6, $p < 0.0005$) among those who were using statins, possibly indicating less scope for improvement in LDL levels in this group. Note that these two estimates differ by the estimate for `1.ht#1.statins`.

Question 3. Summarize the differences between users and non-users of statins in each treatment group, and comment on the results for the placebo group.

First in the HT group:

Statins	HT	_cons	1.statins	1.ht	1.statins#1.ht
Yes	Yes	1	1	1	1
No	Yes	1	0	1	0
	Diff.	0	1	0	1

```
. lincom 1.statins + 1.ht#1.statins
  ( 1)  1.statins + 1.ht#1.statins = 0
------------------------------------------------------------------------------
       ldlch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   11.14062   1.945967     5.72   0.000     7.324816    14.95643
------------------------------------------------------------------------------
```

This means that within the group assigned HT, the change in LDL during the first year is on average 11.1 *mg/dL* greater (i.e., less negative) among women using statins at baseline, compared to those not using statins. This reflects less scope for improvement due to HT.

Now for the placebo group:

Statins	HT	_cons	1.statins	1.ht	1.statins#1.ht
Yes	No	1	1	0	0
No	No	1	0	0	0
	Diff.	0	1	0	0

```
. lincom 1.statins
------------------------------------------------------------------------------
       ldlch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   2.978842   1.901594     1.57   0.117    -.7499543    6.707639
------------------------------------------------------------------------------
```

From the `lincom` (and `regress`) output, the coefficient for statins shows that among women assigned to placebo, changes in LDL were greater (i.e., less negative) by 3 *mg/dL* ($p = 0.12$) among statin users, compared to user baseline non-users. This could be due to chance, non-adherence in baseline users, or new use of statins in non-users, which happened quite a bit. The two estimates of the statin effects again differ by the estimate for `1.ht#1.statins`.

In contrast to the results for HT stratified by statin use, comparisons of statin users to non-users within the HT and placebo groups (or overall) are *observational*, so these unadjusted analyses should be taken as descriptive. Furthermore, while we know that statins do in fact lower LDL, these estimates do not measure that effect, for two reasons. One is the usual potential for confounding, but the other is that we are not comparing changes after the initiation of use, as with HT; rather we are comparing post-baseline changes in LDL among women who were already taking statins at baseline to changes among women who were not. So the changes by statin use do not estimate the effect of new use of statins on LDL within either group.

Question 4. Relate the numbers in the table of LDL changes by treatment assignment and baseline statin use (produced above) to estimates you got from `regress` and `lincom`.

```
. table ht statins, contents(mean ldlch)
--------------------------------------
random          |
assignment to   |    use of statins
hormone therapy |        no        yes
----------------+---------------------
        placebo | -5.894632   -2.91579
hormone therapy | -24.42815  -13.28753
--------------------------------------
```

- The intercept from the regression is equal to the mean change among non-users of statins assigned to placebo, -5.9.

- The estimated effect of HT among non-users, shown as the coefficient for `ht` in the `regress` output as -18.5, can be computed from the table as the difference between the mean changes in LDL among statin non-users assigned to HT or placebo, or –24.4 – (–5.9) = –18.5.

- The estimated effect of HT in statin-users, shown in the `lincom 1.ht + 1.ht#1.statins` result as -10.3, can be computed from the table as the difference between the mean changes among statin users by treatment assignment, or –13.3 – (–2.9) = –10.4. In both cases, we estimate the effect of treatment by subtracting off the change in the placebo group, a measure of what would have happened in the absence of treatment. The interaction coefficient `1.ht#1.statins` is equal to the difference between these differences.

- Finally, the estimate for `1.statins` and the `lincom 1.statins + 1.h#1.statins` result can be computed as the differences between the right- and left-hand numbers in the first and second rows respectively. The last difference, equivalently our last lincom result, was positive because among women assigned to HT, the average change in statin users was less negative (i.e., a smaller reduction) than in non-users of statins, presumably because statin use modifies the effect of HT on LDL.

A final point: the numbers match up exactly with the regression results because there are no adjustment variables in the model. If we had included adjustment variables, this would no longer hold.

Interaction between glucose metabolism and BMI

Question 5. Interpret the coefficient estimates for `2.dmgrp` and `3.dmgrp`.

```
. regress sbp i.dmgrp##c.cbmi
      Source |       SS       df       MS              Number of obs =    2758
-------------+------------------------------           F(  5,  2752) =   18.11
       Model |  31769.2393     5  6353.84785           Prob > F      =  0.0000
    Residual |  965481.672  2752  350.829096           R-squared     =  0.0319
-------------+------------------------------           Adj R-squared =  0.0301
       Total |  997250.911  2757  361.715963           Root MSE      =   18.73
------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dmgrp |
         IFG |    3.31619   .8950856     3.70   0.000     1.561083    5.071297
    diabetes |   8.081379   .9335075     8.66   0.000     6.250933    9.911825
             |
        cbmi |   .1896834   .1060169     1.79   0.074    -.0181973     .397564
             |
dmgrp#c.cbmi |
         IFG |  -.0342154   .1690438    -0.20   0.840    -.3656809    .2972501
    diabetes |  -.4930951   .1610166    -3.06   0.002    -.8088208   -.1773694
             |
       _cons |   132.4312     .54383   243.52   0.000     131.3648    133.4975
------------------------------------------------------------------------------

. testparm dmgrp#c.cbmi
 ( 1)  2.dmgrp#c.cbmi = 0
 ( 2)  3.dmgrp#c.cbmi = 0

       F(  2,  2752) =    5.37
            Prob > F =    0.0047
```

The coefficient for `2.dmgrp` estimates the difference in mean SBP between the IFG and normal groups if `cbmi` = 0 (i.e., BMI = 28.6 kg/m^2), and similarly for `3.dmgrp` and the difference between the diabetes and normal groups.

Question 6. Is there statistically significant variability in the associations of BMI with SBP across the three groups?

The `testparm` result ($F = 5.37$, $p = 0.0047$) shows clear-cut evidence for heterogeneity of the associations of BMI across the three metabolism groups.

Question 7. Do this for the women with diabetes on your own.

Group	BMI	`_cons`	`2.dmgrp`	`3.dmgrp`	`cbmi`	`2.dmgrp#c.cbmi`	`3.dmgrp#c.cbmi`
Diab.	*k*+5	1	0	1	*k*-28.6+5	0	*k*-28.6+5
Diab.	*k*	1	0	1	*k*-28.6	0	*k*-28.6
	Diff.	0	0	0	5	0	5

```
. lincom 5*(cbmi+3.dmgrp#c.cbmi)

 ( 1)  5*cbmi + 5*3.dmgrp#c.cbmi = 0
------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -1.517058   .6059452    -2.50   0.012    -2.705212   -.3289051
------------------------------------------------------------------------------
```

Question 8. Now compute the difference between the diabetes and IFG groups if BMI = 30 (i.e., centered BMI = 1.4).

Group	BMI	_cons	2.dmgrp	3.dmgrp	cbmi	2.dmgrp#c.cbmi	3.dmgrp#c.cbmi
Diab.	30	1	0	1	1.4	0	1.4
Diab.	30	1	1	0	1.4	1.4	0
	Diff.	0	-1	1	0	-1.4	1.4

```
. lincom 3.dmgrp - 2.dmgrp + 1.4*(3.dmgrp#c.cbmi - 2.dmgrp#c.cbmi)

 ( 1)  - 2.dmgrp + 3.dmgrp - 1.4*2.dmgrp#c.cbmi + 1.4*3.dmgrp#c.cbmi = 0
------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   4.122757   1.007944     4.09   0.000     2.146354     6.09916
------------------------------------------------------------------------------
```

Optional: interaction between BMI and glucose

Question 9. What is the interpretation of the coefficients for `cbmi`, `cglucose`, and the interaction term?

- The coefficient for `cbmi` estimates the effect of a 1 kg/m^2 increase in BMI when `cglucose`=0 (equivalently when `glucose` = 100 mg/dL). Specifically, average SBP increases in 0.31 *mmHg* (95% CI 0.16-0.45) for each kg/m^2 increase in BMI, when `glucose` = 100 *mg/dL*.

- The coefficient for `cglucose` estimates the effect of a one mg/dL increase in `glucose` when `cbmi`=0 : Average SBP increases in 0.07 *mmHg* (95% CI 0.05-0.09) for each *mg/dL* increase in glucose, when BMI = 28.6 kg/m^2. (Also, these interpretations assume the adjustment implicit in including age and race/ethnicity in the model.)

- The coefficient for the interaction term `c.cbmi#c.glucose` has a dual interpretation:

  - The increment in the effect of a one kg/m^2 increase in BMI when glucose increases by one mg/dL: so the effect of a 1-kg/m2 increase in BMI decreases by 0.007 *mmHg* (95% CI 0.004-0.01) for each increase of 1 *mg/dL* in glucose.

  - The increment in the effect of a one mg/dL increase in glucose when BMI increases by one kg/m^2: the effect of a 1-*mg/dL* increase in glucose decreases by 0.007 *mmHg* (95% CI 0.004-0.01) for each increase of 1 kg/m^2 increase in BMI.

Question 10. What is the rationale for the `lincom` statements? Are the results consistent with the model output?

Here is the table we could figure out the `lincom` statement for the effect of a 5 kg/m^2 increase in BMI when glucose is 125.

BMI	glucose	_cons	cbmi	cglucose	c.cbmi#c.cglucose
k+5	125	1	*k*-28.6+5	125-100	(*k*-28.6+5)*(125-100)
k	125	1	*k*-28.6	125-100	(*k*-28.6)*(125-100)
	Diff.	0	5	0	5*(125-100)

Try to fill in the table below for a 10 *mg/dL* increase in glucose when BMI = 35.

BMI	glucose	_cons	cbmi	cglucose	c.cbmi#c.cglucose
35	*k*+10	1			
35	k	1			
	Diff.	0			

These results are consistent with the model output. When glucose = 100 *mg/dL* (so `cglucose` = `c.cbmi#c.glucose` = 0), the `lincom` results shows that the effect of a 5 kg/m^2 increase in BMI is 1.527793, exactly 5 times the coefficient for `cbmi` in the model output. Similarly, when BMI = 28.6 kg/m^2 (so `c.cbmi` = `c.cbmi#c.cglucose` = 0), the effect of a 10 *mg/dL* increase in glucose is .6634521, or 10 times the coefficient for `cglucose`.

As discussed in lecture, the model assumes that the effects of glucose & BMI are linear functions of the other variable. This is clearly a very strong assumption. Next week's lab goes over tools for checking linearity.