Comments on Biostat 208 Lab #5, 2/04/10

1 Interaction of hormone therapy and statin use

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 `htstat`, which estimates the difference between the HT effects in the two groups defined by use of statins at baseline, is only 0.003.

2. *Summarize the HT effects by baseline statin use.* 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 ht + htstat` 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 `htstat`. It also implies that the average causal effect (ACE) of HT on LDL levels is a weighted average of these two effects.

3. *Summarize the differences between users and non-users of statins in each treatment group.* First for the HT group:

Statins	HT	`_cons`	`statins`	`ht`	`htstat`
Yes	Yes	1	1	1	1
No	Yes	1	0	1	0
	Difference	0	1	0	1

```
. lincom statins + htstat
 ( 1)  statins + htstat = 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 HT group, using statins at baseline on average was associated with reductions in LDL during the first year that were smaller by 11.1 mg/dL than the reductions among women in the HT group who were not using statins, again reflecting less scope for improvement due to HT. Now for the placebo group:

Statins	HT	`_cons`	`statins`	`ht`	`htstat`
Yes	No	1	1	0	0
No	No	1	0	0	0
	Difference	0	1	0	0

From the `regress` output, the coefficient for `statins` shows that among women assigned to placebo, statin use at baseline was associated with a reduction in LDL that was smaller by 3 mg/dL ($p = 0.12$) than the reduction among non-users, possibly due to chance, non-adherence in baseline users, or new use in non-users, which happened quite a bit. The two estimates of the statin effects again differ by the estimate for `htstat`.

4. *How do these estimates of the statin effects stratified by assignment to HT differ from the estimated effects of HT stratified by statin use?* The stratified HT effect estimates are almost surely unconfounded, because treatment assignment was uncorrelated with statin use, and there were enough users and non-users for us to expect good balance between the HT and placebo groups.

In contrast, 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.

5. *Relate the numbers in the table 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 ht + htstat` 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 `htstat` is equal to the difference between these differences.
- Finally, the coefficient estimate for `statins` and the `lincom statins + hstat` result can be computed as the differences between the right-hand 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 because there are no adjustment variables in the model. If we had included adjustment variables, this would no longer hold.

2 Interaction between glucose metabolism and BMI

1. *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 clearcut evidence for heterogeneity of the associations of BMI across the three metabolism groups.

2. *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`
Diabetes	$k+5$	1	0	1	$k-28.6+5$	0	$k-28.6+5$
Diabetes	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
------------------------------------------------------------------------------
```

So BMI that is higher by 5 kg/m^2 is associated with 1.5 $mmHg$ *lower* SBP (95% CI 0.3-2.7, $p = 0.012$) among women with diabetes.

3. *Interpret the coefficient estimates for* `2.dmgrp` *and* `3.dmgrp`. 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 DM and normal groups.

4. *Now compute the difference between the diabetic and IFG groups if BMI = 30 (i.e., centered BMI = 1.4).*

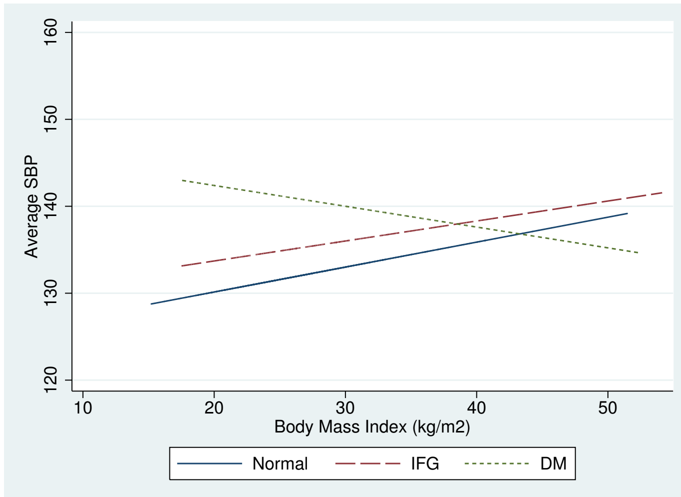
Group	BMI	`_cons`	`2.dmgrp`	`3.dmgrp`	`cbmi`	`2.dmgrp#c.cbmi`	`3.dmgrp#c.cbmi`
Diabetic	30	1	0	1	1.4	0	1.4
IFG	30	1	1	0	1.4	1.4	0
	Diff.	0	-1	1	0	-1.4	1.4

```
. 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
------------------------------------------------------------------------------
```

So among women with BMI = 30 kg/m^2, average SBP is 4.1 $mmHg$ (95% CI 2.1-6.1, $p < 0.0005$) higher among those with diabetes than those with IFG.

5. *Are the differences between the glucose metabolism groups interpretable as average causal effects?* We did center BMI, so that factor would drop out in calculating ACEs. However, the three groups almost surely differ in ways not captured in our simple unadjusted model. So the between-group differences are not interpretable as ACEs.

3 Optional: plotting adjusted regression lines



4 Optional: alternative coding of interaction

1. *Interpret the coefficient estimates for* `cbmidmgrp_1`*,* `cbmidmgrp_2`*, and* `cbmidmgrp_3`*.* The coefficients estimate the increase in SBP per unit increase in BMI within each of the three groups. In the earlier version of the model, the coefficient for `cbmi` had this interpretation for the normal glucose metabolism group, but `2.dmgrp#c.cbmi` and `3.dmgrp#c.cbmi` estimated between-group differences in the BMI effects (see next item).

2. *Is the interaction test result the same? Why do we need the* `equal` *option in the* `testparm` *statement? This was not part of the* `testparm` *command used with the other version of the model to evaluate interaction.* The interaction test result is exactly the same; the models are completely equivalent. To understand the different `testparm` commands, note that while `2.dmgrp#c.cbmi` = `cbmidmgrp_2` and `3.dmgrp#c.cbmi` = `cbmidmgrp_3` for all observations, their meaning in this model is changed. In the first version of this model, the coefficients for `2.dmgrp#c.cbmi` and `3.dmgrp#c.cbmi` estimated the *differential* effects of BMI in the IFG and diabetic groups as compared to the normal reference group. Thus in the first model, the null hypothesis for the overall 2 degree-of-freedom test for interaction or heterogeneity is that the coefficients for `2.dmgrp#c.cbmi` and `3.dmgrp#c.cbmi` are both zero, which we got using the `testparm` command without the `equal` option.

   In contrast, in this version of the model, the coefficients for `cbmidmgrp_1`, `cbmidmgrp_2`, and `cbmidmgrp_3` directly capture the within-group BMI effects. There is no interaction if those three coefficients are *equal to each other*; we get this test using the `equal` option. The 3-df test of whether they are all zero that we would get without the `equal` option would represent a global test for an effect of BMI in any of the three groups, not a test for interaction.

3. *To see why this coding is convenient, use the following to evaluate the effects of 5-unit increase in BMI in each group.* With this coding, you get all three within-group estimates of the effects of a 1-unit increase in BMI from the `regress` output directly, or, for differences of $k \neq 1$ units, using simpler `lincom` statements not involving product terms.

4. *Figure out the* `lincom` *command to compare diabetes to IFG when BMI = 30.*

Group	BMI	_cons	_Idmgrp_2	_Idmgrp_3	cbmidmgrp_1	cbmidmgrp_2	cbmidmgrp_3
Diabetic	30	1	0	1	0	0	1.4
IFG	30	1	1	0	0	1.4	0
	Diff.	0	-1	1	0	-1.4	1.4

So the `lincom` statement and result are essentially the same as with the other coding.

```
. lincom 3.dmgrp - 2.dmgrp + 1.4*(cbmidmgrp_3 - cbmidmgrp_2)
 ( 1)  - 2.dmgrp + 3.dmgrp - 1.4*cbmidmgrp_2 + 1.4*cbmidmgrp_3 = 0
------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   4.122757   1.007944     4.09   0.000     2.146354     6.09916
------------------------------------------------------------------------------
```