Biostatistics 208 Lab #5 2/4/21

The purpose of this lab is to give you practice in using the `regress` and `lincom` commands in Stata to look at interaction. Download the dataset `lab5.dta` from the website. Please note: the optional sections are likely of most interest to those who would like to learn a bit more Stata programming.

Interaction of hormone therapy and statin use

In this section we will examine whether assignment to hormone therapy (HT) interacts with baseline statin use in its effect on change in LDL cholesterol from baseline to the first annual visit in the HERS data. First, we will compute and tabulate the mean changes in LDL by treatment group, stratified by statin use, using the following commands:

```
gen ldlch = ldl1 - ldl0
table ht statins, contents(mean ldlch)
```

This may help to interpret some of the regression results. Now run the regression, letting Stata calculate the product term.

```
regress ldlch i.statins##i.ht
```

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

Use `lincom` statements as necessary to estimate the effect of HT in the two subgroups defined by baseline statin use. Here is the regression equation:

$E[\texttt{ldlch} \mid \texttt{ht, statins}] = \beta_0 + \beta_1 \texttt{1.statins} + \beta_2 \texttt{1.ht} + \beta_3 \texttt{1.statins\#1.ht}$

Here is a table providing the desired effect for non-users of statins:

Statins	HT	_cons(β_0)	1.statins(β_1)	1.ht(β_2)	1.statins#1.ht (β_3)
No	Yes	1	0	1	0
No	No	1	0	0	0
	Difference	0	0	1	0

So the effect of HT in non-users of statins is given by the coefficient (β_2) for `ht`, which we can read from the regress output. Now here is the table for statin users:

Statins	HT	_cons(β_0)	1.statins(β_1)	1.ht(β_2)	1.statins#1.ht (β_3)
Yes	Yes	1	1	1	1
Yes	No	1	1	0	0
	Difference	0	0	1	1

So, to evaluate the effect of HT among users of statins, use the command:

```
lincom 1.ht + 1.statins#1.ht
```

Question 2. Summarize the HT effects by baseline statin use.

Now use the following table to compute the associations of being a statin user with change in LDL, in the HT group.

Statins	HT	_cons(β_0)	1.statins(β_1)	1.ht(β_2)	1.statins#1.ht (β_3)
Yes	Yes	1	1	1	1
No	Yes	1	0	1	0
	Difference	0	1	0	1

Use `lincom` if necessary to compute the difference in LDL changes within the HT group, and interpret. Now do the computation for the placebo group. Again, use `lincom` if necessary to compute the statin effect.

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.

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`.

Interaction between glucose metabolism and BMI

Using the tools you have developed here, we will check whether the association of BMI with systolic blood pressure (SBP) is modified by glucose metabolism group, a three-level variable with categories normal (fasting glucose (FG) < 100), impaired fasting glucose (IFG; FG 100-125), and diabetes (by self-report or FG ≥ 126). The code to make the variable and run the model is given below. Note that to get an overall test for interaction, we need to use an *F*-test, since we are in effect comparing the effect of BMI across three, not just two, subgroups (i.e. the *F*-test evaluates the hypothesis that the difference between BMI effects across glucose groups is zero by comparing a model including the two interaction terms to a reduced model excluding these terms). As in the lecture, we will center BMI to make the estimates for metabolic group in the regression output more interpretable.

```
recode glucose min/99=1 100/125=2 126/max=3, gen(dmgrp)
replace dmgrp = 3 if diabetes==1
label define dmgrpvals 1 "normal" 2 "IFG" 3 "diabetes"
label values dmgrp dmgrpvals
gen cbmi = bmi-28.6
regress sbp i.dmgrp##c.cbmi
* overall test for interaction
testparm dmgrp#c.cbmi
```

Stata notes: The new variables automatically made by Stata are keyed to the codes of the underlying `dmgrp` variable. Here is a table giving the generated variable names:

Variable definition	*Variable names*
Indicator for IFG	`2.dmgrp`
Indicator for Diabetes	`3.dmgrp`
IFG-cbmi product term	`2.dmgrp#c.cbmi`
Diabetes-cbmi product term	`3.dmgrp#c.cbmi`

The reference group is normal individuals, because they have the lowest code, and no indicator is made for them. The indicator `2.dmgrp = 1` for the IFG group and 0 for the others, while `3.dmgrp`

= 1 for the diabetes group and 0 for the others. Within the `regress` command, the ## operator instructs Stata to set up the interaction product terms automatically as discussed in lecture. Note that these variables are not appended to the working dataset, but are available for use in post-estimation commands like `testparm` and `lincom`, until another regression model is fitted.

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

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

To further interpret these results, we will first evaluate the effects of a 5-kg/m^2 increase BMI in the different glucose metabolism groups. Here is the regression equation

$$E[\texttt{ldlch} \mid \texttt{dmgrp, cbmi}] = \beta_0 + \beta_1 \texttt{2.dmgrp} + \beta_2 \texttt{3.dmgrp} + \beta_3 \texttt{cbmi} + \beta_4 \texttt{2.dmgrp\#c.cbmi} + \beta_5 \texttt{3.dmgrp\#c.cbmi}$$

Recall that `2.dmgrp#c.cbmi = cbmi` for the IFG group (`2.dmgrp = 1`) and 0 for the other two groups; likewise `3.dmgrp#c.cbmi = cbmi` for the diabetes group and 0 for others. Now we can set up our table, first for the normal group:

Group	BMI	_cons	2.dmgrp	3.dmgrp	cbmi	2.dmgrp#c.cbmi	3.dmgrp#c.cbmi
Normal	*k*+5	1	0	0	*k* – 28.6 + 5	0	0
Normal	*k*	1	0	0	*k* – 28.6	0	0
	Diff.	0	0	0	5	0	0

So we can use the following command to estimate what we want:

```
lincom 5*cbmi
```

Now for the IFG group:

Group	BMI	_cons	2.dmgrp	3.dmgrp	cbmi	2.dmgrp#c.cbmi	3.dmgrp#c.cbmi
IFG	*k*+5	1	1	0	*k* – 28.6 + 5	*k* – 28.6 + 5	0
IFG	*k*	1	1	0	*k* – 28.6	*k* – 28.6	0
	Diff.	0	0	0	5	5	0

So in this case, we can use:

```
lincom 5*(cbmi+2.dmgrp#c.cbmi)
```

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					
Diab.	*k*	1					
	Diff.	0					

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					
IFG	30	1					
	Diff.	0					

Optional: interaction between BMI and glucose

Now consider the interaction between BMI and glucose in a model for SBP adjusting for age and race-ethnicity. For interpretability, we will center glucose on 100 (we already did this for BMI). We will do checks for departures from linearity and influential points in next week's lab.

```
gen cglucose = glucose-100
regress sbp age i.raceth c.cbmi##c.cglucose
```

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

Now use the following code to compute the effect of 5 kg/m^2 increases in BMI and 10 mg/dL increases in glucose at selected levels of the other predictor.

```
foreach g of numlist 80 90 100 125 {
      dis _newline as result "Effect of a 5-unit increase in BMI when glucose = `g'"
      lincom 5*(c.cbmi + (`g'-100)*c.cbmi#c.cglucose)
      }
foreach b of numlist 18.5 25 28.6 30 35 {
      dis _newline as result "Effect of a 10-unit increase in glucose when BMI = `b'"
      lincom 10*(c.cglucose + (`b'-28.6)*c.cbmi#c.cglucose)
      }
```

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