Biostatistics 208 HW #2, Solutions

Question 1

What is the crude or unadjusted association of regular exercise with fasting glucose at baseline?

```
. reg glucose0 exercise

      Source |       SS       df       MS              Number of obs =    2032
-------------+------------------------------           F(  1,  2030) =   14.97
       Model |  1412.50418     1  1412.50418           Prob > F      =  0.0001
    Residual |  191605.195  2030  94.3867954           R-squared     =  0.0073
-------------+------------------------------           Adj R-squared =  0.0068
       Total |  193017.699  2031  95.0357946           Root MSE      =  9.7153
------------------------------------------------------------------------------
    glucose0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    exercise |  -1.692789   .4375862    -3.87   0.000    -2.550954   -.8346242
       _cons |   97.36104   .2815138   345.85   0.000     96.80895    97.91313
------------------------------------------------------------------------------
```

In an unadjusted comparison among post-menopausal women with CHD, average fasting glucose was 1.7 mg/dL lower (95% CI 0.8-2.6, $p < 0.0005$) among those who exercised at least 3 times per week, as compared to women who exercised less.

Question 2

Use a multi-predictor linear model to estimate the adjusted association of exercise with `glucose0`*, adjusting for age, race/ethnicity, and smoking. Is there substantial confounding of the association between exercise and baseline glucose by the other factors? Is exercise a significant predictor after adjustment?*

```
. reg glucose0 exercise age i.raceth smoker
      Source |       SS       df       MS              Number of obs =    2032
-------------+------------------------------           F(  5,  2026) =    4.70
       Model |  2213.72401     5  442.744803           Prob > F      =  0.0003
    Residual |  190803.975  2026  94.1776776           R-squared     =  0.0115
-------------+------------------------------           Adj R-squared =  0.0090
       Total |  193017.699  2031  95.0357946           Root MSE      =  9.7045
------------------------------------------------------------------------------
    glucose0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    exercise |  -1.821726   .4441249    -4.10   0.000    -2.692715    -.950737
         age |  -.0043868   .0328457    -0.13   0.894    -.0688017     .060028
             |
      raceth |
          2  |   .1559008   .9040724     0.17   0.863    -1.617108    1.928909
          3  |   3.093502   1.366503     2.26   0.024     .4136038    5.773399
             |
      smoker |  -1.120567   .6323409    -1.77   0.077    -2.360673    .1195392
       _cons |   97.78406   2.252041    43.42   0.000      93.3675    102.2006
------------------------------------------------------------------------------
```

- After adjustment for age, race/ethnicity, and current smoking, average glucose levels were 1.8 mg/dL lower (95% CI 1.0-2.7, $p < 0.0005$) among women who exercised at least 3 times per week. So clearly exercise is a significant predictor after adjustment.
- Since the coefficient for `exercise` is larger in absolute value after adjustment, there is slight *negative* confounding, by smoking, and exercise remains highly statistically significant after adjustment. The negative confounding is explained by the fact that smoking is much less prevalent among women who exercise (8 vs 20%), so they are inversely correlated, but both are associated with lower glucose levels, possibly because women who smoke are thinner on average.

Question 3

In practice, it can be difficult to distinguish confounding from mediation. It is not possible on statistical grounds. Instead, it is judgment call based on plausible models of causation. In particular, do you think BMI is a confounder or a mediator of the association between exercise and glucose levels? Describe a plausible mechanism for the effect of exercise on glucose under which BMI is a mediator.

- Long-term regular exercise, for which current exercise could be a decent surrogate, may have helped prevent weight gain
- Increased exercise may have resulted in recent weight loss

Come up with one under which BMI is a confounder.

- Higher BMI makes it more difficult to exercise
- BMI is a marker for other lifestyle factors such as diet that more directly increase glucose levels

If we do add BMI to the model, exercise still appears protective, but the effect is only half as large:

```
. reg glucose0 exercise age i.raceth smoker bmi0

      Source |       SS       df       MS              Number of obs =    2030
-------------+------------------------------           F(  6,  2023) =   26.34
       Model |  13979.4852     6   2329.9142           Prob > F      =  0.0000
    Residual |  178976.732  2023  88.4709498           R-squared     =  0.0724
-------------+------------------------------           Adj R-squared =  0.0697
       Total |  192956.217  2029  95.0991704           Root MSE      =  9.4059

------------------------------------------------------------------------------
    glucose0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    exercise |  -.9372157   .4376237    -2.14   0.032    -1.795456   -.0789754
         age |   .0572837   .0323737     1.77   0.077    -.0062056    .120773
             |
      raceth |
          2  |  -.5751108   .8785541    -0.65   0.513    -2.298076    1.147854
          3  |   3.089222   1.324498     2.33   0.020     .4916996    5.686745
             |
      smoker |  -.1087472   .6195892    -0.18   0.861    -1.323847    1.106352
        bmi0 |   .4854092   .0421249    11.52   0.000     .4027964     .568022
       _cons |   79.76392   2.695955    29.59   0.000     74.47678    85.05106
------------------------------------------------------------------------------
```

If you primarily view BMI as a mediator of long-term exercise patterns, then the result for `exercise` in this model would represent the effects of exercise by pathways other than BMI. If you view it primarily as a confounder, this might be seen as representing the overall effect of exercise.

Question 4

A previous HERS paper has shown that hormone therapy prevented progression to diabetes. Does change in glucose from baseline to the first annual visit (`glucchange`*) differ in the hormone therapy and placebo groups? What is the average change in each treatment group?*

```
. reg glucchange ht
      Source |       SS       df       MS              Number of obs =    1933
-------------+------------------------------           F(  1,  1931) =   20.41
       Model |  2768.22866     1  2768.22866           Prob > F      =  0.0000
    Residual |  261844.603  1931   135.60052           R-squared     =  0.0105
-------------+------------------------------           Adj R-squared =  0.0099
       Total |  264612.832  1932  136.963163           Root MSE      =  11.645
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ht |  -2.393937   .5298372    -4.52   0.000     -3.43305   -1.354824
       _cons |   1.983789   .3706569     5.35   0.000     1.256859    2.710719
------------------------------------------------------------------------------

. lincom _cons + ht
 ( 1)  ht + _cons = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   -.410148   .3786039    -1.08   0.279    -1.152663    .3323675
------------------------------------------------------------------------------
```

So HT did produce a net average reduction of 2.4 mg/dL (95% CI 1.4-3.4, $p < 0.0005$) in glucose levels over the first year of treatment. Average changes by group were

- an increase of 2.0 mg/dL (95% CI 1.3 to 2.7) in the placebo group
- a decrease of 0.4 mg/dL (95% CI -0.3 to 1.2) in the HT group.

The treatment effect is statistically significant even though the CI for the average change in the HT group includes zero, because we estimate the treatment effect by the *difference* between the average changes in the treated and placebo groups, not just by the change in the treated group. This could have been done with equal validity using a t-test.

```
. ttest glucchange, by(ht)
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 placebo |     987    1.983789     .427926    13.44396    1.144039     2.82354
     E+P |     946    -.410148    .3058667    9.407577   -1.010405    .1901085
---------+--------------------------------------------------------------------
combined |    1933     .812209    .2661865    11.70313    .2901661    1.334252
---------+--------------------------------------------------------------------
    diff |            2.393937    .5298372                1.354824     3.43305
------------------------------------------------------------------------------
    diff = mean(placebo) - mean(E+P)                              t =   4.5183
Ho: diff = 0                                     degrees of freedom =     1931

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Question 5

Is there an interaction between the predictors 1) hormone therapy and 2) baseline glucose on the outcome change in glucose? Give and interpret the statistical test of interaction.

A simple solution is to treat baseline glucose as a continuous variable, and let Stata set up cross-product with the treatment indicator.

```
. reg glucchange i.ht##c.glucose0

      Source |       SS       df       MS              Number of obs =    1933
-------------+------------------------------           F(  3,  1929) =   24.53
       Model |  9723.53001     3  3241.17667           Prob > F      =  0.0000
    Residual |  254889.302  1929   132.13546           R-squared     =  0.0367
-------------+------------------------------           Adj R-squared =  0.0352
       Total |  264612.832  1932  136.963163           Root MSE      =  11.495

------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        1.ht |   2.424923   5.264795     0.46   0.645    -7.900365    12.75021
    glucose0 |  -.1701347   .0376219    -4.52   0.000    -.2439186   -.0963508
             |
          ht#|
  c.glucose0 |
          1  |   -.051057   .0541798    -0.94   0.346    -.1573141    .0552001
             |
       _cons |   18.48772   3.667813     5.04   0.000     11.29442    25.68101
------------------------------------------------------------------------------
```

- The effects on HT on first-year change in glucose appeared similar across levels of baseline glucose ($p = 0.35$), so there is essentially no evidence for interaction.
- Nonetheless, there is an association between baseline levels and subsequent change, possibly due to regression to the mean. This also arises from the fact that the outcome is calculated as the difference between the year 1 and baseline values, so the latter is effectively included – with different signs – on both sides of the regression equation, producing the strong negative association. From the previous analysis with no interaction, we know that glucose levels increased on average in the placebo group. However, the coefficient for `glucose0` in this analysis suggests that the average placebo-group increase in glucose was smaller by 0.17 mg/dL (95% 0.10-0.24, $p < 0.0005$) for each mg/dL increase in baseline glucose. The big p-value for interaction suggests that the association between baseline and change is the same in the HT and placebo groups.
- Centering `glucose0` (and by extension the cross-product term) would make the estimate of the effect of `ht` in the `regress` output more interpretable, but would not change the results for participants with given values of baseline glucose, computed using `lincom` statements below.

Based on the linear model you fit for interaction, estimate the effect of hormone therapy on glucose change (`glucchange`*), with a 95% CI, among patients with baseline glucose of 80, 90, 100, and 110 mg/dL.*

Here is the table we need to figure out the right `lincom` statement:

HT	Baseline Glucose	`_cons`	`ht`	`glucose0`	`1.ht#c.glucose0`
yes	k	1	1	k	k
no	k	1	0	k	0
	Diff.	0	1	0	k

So we can use `lincom ht +` k`*1.ht#c.glucose0`. This can be implemented using a `forvalues` loop (though it is certainly fine to type the commands one at a time):

```
. forvalues glucose0 = 80(10)110 {
  2.         dis _newline as result "Effect of HT when baseline glucose = `glucose0':"
  3.         lincom 1.ht + `glucose0'*1.ht#c.glucose0
  4.         }

Effect of HT when baseline glucose = 80:
 ( 1)  1.ht + 80*1.ht#c.glucose0 = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -1.659638   1.044825    -1.59   0.112    -3.708744     .389467
------------------------------------------------------------------------------

Effect of HT when baseline glucose = 90:
 ( 1)  1.ht + 90*1.ht#c.glucose0 = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -2.170209   .6366019    -3.41   0.001    -3.418709   -.9217084
------------------------------------------------------------------------------

Effect of HT when baseline glucose = 100:
 ( 1)  1.ht + 100*1.ht#c.glucose0 = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -2.680779    .553131    -4.85   0.000    -3.765576   -1.595981
------------------------------------------------------------------------------

Effect of HT when baseline glucose = 110:
 ( 1)  1.ht + 110*1.ht#c.glucose0 = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -3.191349   .8909184    -3.58   0.000    -4.938613   -1.444085
------------------------------------------------------------------------------
```

Here is a summary of the results:

Baseline Glucose	Estimated effects of HT		
(mg/dL)	Net reduction (mg/dL)	95% CI	p-value
80	1.7	-0.4, 3.7	0.112
90	2.2	0.9, 3.4	0.001
100	2.7	1.6, 3.8	< 0.0005
110	3.2	1.4, 4.9	< 0.0005

Note that this analysis assumes a linear interaction between HT and baseline glucose, with the estimated HT effects increasing by a constant 0.5 mg/dL for each 10 mg/dL increase in baseline glucose. Of course, the interaction is clearly not statistically significant, and this shows up in these relatively minor differences by baseline glucose level, as compared to the average net reduction of 2.4 mg/dL estimated for HT in the unadjusted analysis for Question 4.

This problem could also be examined using categorization of baseline glucose levels. I split them at 85, 95, and 105 mg/dL, on the argument that 90 and 100 at least are thus more or less in the middle of their respective categories, though this is not the only possible way to do it.

```
. recode glucose0 min/85 = 1 86/95=2 96/105=3 106/max=4, gen(grp)
(2032 differences between glucose0 and grp)

. reg glucchange i.grp##i.ht

      Source |       SS       df       MS              Number of obs =    1933
-------------+------------------------------           F(  7,  1925) =   12.56
       Model |  11560.9659     7  1651.56655           Prob > F      =  0.0000
    Residual |  253051.866  1925  131.455515           R-squared     =  0.0437
-------------+------------------------------           Adj R-squared =  0.0402
       Total |  264612.832  1932  136.963163           Root MSE      =  11.465

------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         grp |
          2  |  -3.816878    1.28931    -2.96   0.003     -6.34547   -1.288287
          3  |  -5.745294   1.312074    -4.38   0.000     -8.31853   -3.172058
          4  |  -6.284194   1.421726    -4.42   0.000    -9.072479   -3.495908
             |
       1.ht  |  -1.681171   1.580772    -1.06   0.288    -4.781376    1.419034
             |
     grp#ht  |
       2 1   |  -.4230616   1.792138    -0.24   0.813    -3.937798    3.091675
       3 1   |  -.9857429   1.822226    -0.54   0.589    -4.559488    2.588002
       4 1   |  -1.918053   2.009375    -0.95   0.340    -5.858833    2.022727
             |
       _cons |       6.51   1.146541     5.68   0.000     4.261408    8.758592
------------------------------------------------------------------------------
```

Here are the tables we need to figure out how to compute treatment effects for the four participants of interest. First, the effect of HT in a participant with baseline glucose of 80 mg/dL (subgroup 1):

Tx	FPG	2.grp	3.grp	4.grp	1.ht	2.grp#1.ht	3.grp#1.ht	4.grp#1.ht
HT	80	0	0	0	1	0	0	0
Pbo	80	0	0	0	0	0	0	0
	Diff.	0	0	0	1	0	0	0

So we could simply use the coefficient for `ht` from `regress` output. Now for the effect of HT for a participant with baseline glucose of 90 mg/dL (subgroup 2):

Tx	FPG	2.grp	3.grp	4.grp	1.ht	2.grp#1.ht	3.grp#1.ht	4.grp#1.ht
HT	90	1	0	0	1	1	0	0
Pbo	90	1	0	0	0	0	0	0
	Diff.	0	0	0	1	1	0	0

So for this participant we would use `lincom ht + 2.grp#1.ht`, and similarly for the participants with baseline glucose of 100 mg/dL (subgroup 3) and 110 mg/dL (subgroup 4). Again we can use a `forvalues` loop to compute the effects. Note that we can include the baseline level, for which there is no apparent interaction term.

```
. forvalues grp = 1/4 {
  2.         di _newline as result "Effect of HT in baseline glucose group `grp':"
  3.         lincom 1.ht + `grp'.grp#1.ht
  4.         }

Effect of HT in baseline glucose group 1:
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -1.681171   1.580772    -1.06   0.288    -4.781376    1.419034
------------------------------------------------------------------------------

Effect of HT in baseline glucose group 2:
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -2.104233    .844346    -2.49   0.013    -3.760162   -.4483039
------------------------------------------------------------------------------

Effect of HT in baseline glucose group 3:
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -2.666914   .9064597    -2.94   0.003     -4.44466    -.889168
------------------------------------------------------------------------------

Effect of HT in baseline glucose group 4:
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -3.599224   1.240462    -2.90   0.004    -6.032015   -1.166433
------------------------------------------------------------------------------
```

The tests for interaction are of interest with this solution. We could do it one of two ways:

```
. * test for heterogeneity in treatment effects
. testparm i.grp#i.ht
 ( 1)  2.grp#1.ht = 0
 ( 2)  3.grp#1.ht = 0
 ( 3)  4.grp#1.ht = 0

       F(  3,  1925) =    0.44
            Prob > F =    0.7275

. * tests for trend in treatment effects
. test 2.grp#1.ht = 3.grp#1.ht + 3*4.grp#1.ht
 ( 1)  2.grp#1.ht - 3.grp#1.ht - 3*4.grp#1.ht = 0
       F(  1,  1925) =    1.05
            Prob > F =    0.3048

. contrast q(1).grp#ht
Contrasts of marginal linear predictions
Margins      : asbalanced
------------------------------------------------
             |         df           F        P>F
-------------+----------------------------------
      grp#ht |          1        1.05     0.3048
------------------------------------------------
```

Note that the second and third tests, for *trend* rather than *heterogeneity* in the treatment effects across the four categories, are equivalent and more nearly analogous to the test in the first model, where we kept baseline glucose level as continuous and modeled the interaction by a product term, which was also continuous. The `test` statement for the test for trend is based on Table 4.5 on page 82 in VGSM.

Question 6

Here are the code and results for the regression using centered glucose:

```
. sum glucose0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    glucose0 |      2032    96.66043     9.74863         47        125

. local mglucose0 = r(mean)
. gen cglucose0 = glucose0 - `mglucose0'
```

STATA notes: After running the `sum glucose0` command to get the mean of baseline glucose, the command `local mglucose0 = r(mean)` stores the mean returned by the `sum` command in a so-called local macro variable called `mglucose0` that we can then use, first to created the centered values. Note that when we call it later we have to enclose `mglucose0` in single quotes (using the key below escape for the leading single quote as usual). To see what else is returned by the `sum` command, type `return list` just after running the command `sum glucose0`.

```
. reg glucchange i.ht##c.cglucose0

      Source |       SS       df       MS              Number of obs =    1933
-------------+------------------------------           F(  3,  1929) =   24.53
       Model |  9723.52997     3  3241.17666           Prob > F      =  0.0000
    Residual |  254889.302  1929   132.13546           R-squared     =  0.0367
-------------+------------------------------           Adj R-squared =  0.0352
       Total |  264612.832  1932  136.963163           Root MSE      =  11.495
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.ht |   -2.51027   .5232834    -4.80   0.000    -3.536531    -1.48401
   cglucose0 |  -.1701347   .0376219    -4.52   0.000    -.2439186   -.0963508
             |
          ht#|
 c.cglucose0 |
          1  |   -.051057   .0541798    -0.94   0.346    -.1573141    .0552001
             |
       _cons |   2.042423   .3661201     5.58   0.000      1.32439    2.760456
------------------------------------------------------------------------------
```

STATA notes: After running the `sum glucose0` command to get the mean of baseline glucose, the command `local mglucose0 = r(mean)` stores the mean returned by the `sum` command in a so-called local macro variable called `mglucose0` that we can then use, first to created the centered values. Note that when we call it later we have to enclose `mglucose0` in single quotes (using the key below escape for the leading single quote as usual). To see what else is returned by the `sum` command, type `return list` just after running the command `sum glucose0`.

Here is the table we need to figure out the right `lincom` statement:

HT	Baseline Glucose	_cons	1.ht	cglucose0	1.ht#c.cglucose0
yes	k	1	1	k-96.7	k-96.7
no	k	1	0	k-96.7	0
	Diff.	0	1	0	k-96.7

So we can use `lincom ht +` $(k-96.7)$`*htglucose0`. This can also be implemented using a `forvalues` loop (though it is certainly fine to type the commands one at a time); we can again use the local variable `mglucose0` to avoid having to type the mean value repeatedly.

```
. forvalues glucose0 = 80(10)110 {
  2.         dis _newline as result "Effect of HT when baseline glucose = ‘glucose0’:"
  3.         lincom 1.ht + (‘glucose0’-‘mglucose0’)*1.ht#c.cglucose0
  4.         }

Effect of HT when baseline glucose = 80:
 ( 1)  1.ht - 16.66043*1.ht#c.cglucose0 = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -1.659638   1.044825    -1.59   0.112    -3.708744     .389467
------------------------------------------------------------------------------

Effect of HT when baseline glucose = 90:
```

```
 ( 1)  1.ht - 6.660433*1.ht#c.cglucose0 = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -2.170209   .6366019    -3.41   0.001    -3.418709   -.9217084
------------------------------------------------------------------------------

Effect of HT when baseline glucose = 100:
 ( 1)  1.ht + 3.339567*1.ht#c.cglucose0 = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -2.680779    .553131    -4.85   0.000    -3.765576   -1.595981
------------------------------------------------------------------------------

Effect of HT when baseline glucose = 110:
 ( 1)  1.ht + 13.33957*1.ht#c.cglucose0 = 0
------------------------------------------------------------------------------
  glucchange |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -3.191349   .8909184    -3.58   0.000    -4.938613   -1.444085
------------------------------------------------------------------------------
```

You can see that the `lincom` results are unchanged after centering. But with this model, the coefficient for `ht` shown in the regression output estimates the effect of HT when baseline glucose is at the mean value of 96.7 mg/dL rather than the impossible value of 0.

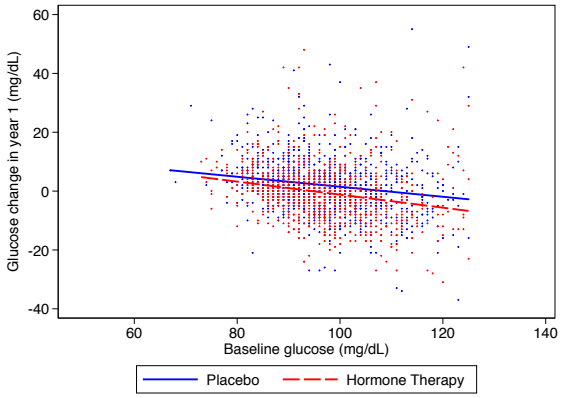
Question 7

Produce a single graph for the fitted regression model showing the association of baseline glucose with change in glucose for each of the treatment groups.

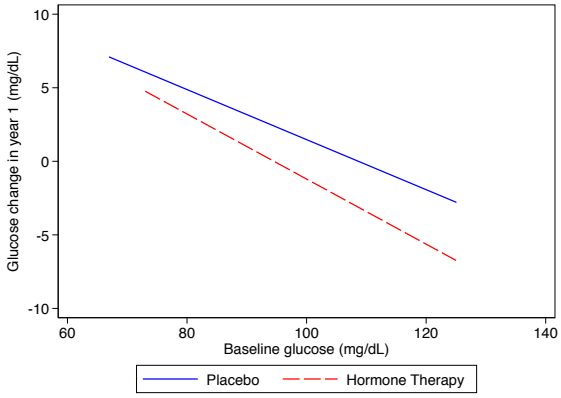
Here is code for the plot. The line separators (`///`) will only work in a do-file.

```
twoway ///
    (scatter glucchange glucose0 if ht==0 & glucchange<75, mcol(blue) msize(vtiny)) ///
    (lfit glucchange glucose0 if ht==0, lpattern(solid) lw(medthick) lcolor(blue)) ///
    (scatter glucchange glucose0 if ht==1, mcolor(red) msize(vtiny)) ///
    (lfit glucchange glucose0 if ht==1, lpattern(longdash) lw(medthick) lcolor(red)), ///
    legend(order(2 "Placebo" 4 "Hormone Therapy")) ///
    ylabel(, angle(horizontal)) ytitle("Glucose change in year 1 (mg/dL)") ///
    xscale(range(60 140)) xlabel(60(20)140) xtitle("Baseline glucose (mg/dL)") ///
    scheme(s1color) name(plota, replace)
graph export hwk2plota.pdf, replace
```

Here is the resulting figure:



I had to limit the range of the data points in the scatter plot for the placebo group to avoid making the two regression lines practically invisible. Omitting the two `scatter` sub-commands gives this plot, which makes the lines easier to see, at the expense of exaggerating the difference.



Another way to make this plot is to fit the interaction model, then plot the fitted values.

```
qui reg glucchange i.ht##c.glucose0
predict fitted, xb
replace fitted = . if missing(glucchange)
twoway ///
    (line fitted glucose0 if ht == 0, sort lpattern(solid) lcolor(blue)) ///
    (line fitted glucose0 if ht == 1, sort lpattern(longdash) lcolor(red)), ///
    legend(order(1 "Placebo" 2 "Hormone Therapy")) ///
    yscale(range(-10 10)) ylabel(-10(5)10, angle(horizontal)) ///
    ytitle("Glucose change in year 1 (mg/dL)") ///
    xscale(range(60 140)) xlabel(60(20)140) xtitle("Baseline glucose (mg/dL)") ///
    scheme(s1color) name(plotc, replace)
graph export hwk2plotc.pdf, replace
```

I replace one fitted value with a missing outcome that is way off to the left; without this point, the resulting figure looks just like the one above.

Note that a plot does not make sense for the model treating baseline glucose as categorical, which just leads to four point treatment estimates, one for each baseline glucose category, rather than two regression lines.