Graded Assignment 3 – Leah Ross

Intro: This memo is addressed to the NYC Board of Education and examines the impact of a mentoring program tested in two NYC high schools where graduates worked with mentees on basic academic skills over the summer of 2016. The program targeted students with frequent absents and at an increased risk of dropping out. As high school dropouts are at higher risk for substance abuse, teen pregnancy, and job security problems, limiting dropout rates should be a priority for the Board.

Methods: This program sampled students from two NYC high schools who applied for the program. Following a difference-in-difference design, those who were not accepted acted as the baseline (control) group that represented how those who were selected into the program would have changed over time if they had not been in the program. Data on scores from an academic achievements test and absents was collected before and after the program. A randomized lottery determined who was accepted into the program of those who applied. Table 1 shows that the randomization process worked well, as the groups have similar characteristics, as 54.6% of applicants were not selected and 45.2% were selected. Looking at the mean values in Table 2 and Table 3, we see very similar values between those not in the program and those in the program in terms of scores at application, students' ethnicity, past mentoring experience, and school characteristics, further suggesting the randomization process worked well.

Findings: Before the program, participants and non-participants had nearly identical mean scores (90.2% and 89.5%). After the intervention, program participants had a slightly higher mean score of 92% compared to their starting point, and non-participants

saw a reduction in their scores to about 75.7% (Table 3, Figure 1). Meaning, scores increased by 16.3% among program participants compared to non-participants, who saw a decrease of 13.8% on average. As the program occurred in the summer, these findings suggest that program participation led to a more constant level of knowledge attainment over the summer months than students who were not in the program and perhaps learned less or forgot more over the summer.

Prior to the program, students in the program had fewer mean days absents than non-participants, 15 compared to 19 (Table 4). As seen in Table 4 and Figure 2, after the intervention, mean absents for program participants was reduced to 7 days. Absents for those not in the program remained relatively unchanged (18 days). Tables 5 and 6 provide more statistically advanced estimates of the program's impact. We find that being in the program reduced absents by around 11 days (Table 6). And being in the program also increased scores by 16.3% (Table 5).

The program had different impacts on different subgroups of students. Table 7 and Table 8 show the program affected students who had previously been mentored compared to those who had not in varying ways. In Table 7, the impact of the program on scores for students who were not previously mentored was an increase of 17.4%. Comparatively, students in the program who had previous mentor experience saw a smaller positive impact on their scores of 4.3% (Table 8). Turning to absences, for those in the program with low absences before the program (less than 17 days), being in the program increased their scores by 11.2% (Table 9). For students with high absences before the program (above 17 days), the program's impact increased their scores by 17.7% (Table 10).

Table 11 shows those with past mentoring experience were affected by the program statistically different than those without past mentoring. The program increased scores to a lesser extent for those with past mentoring than those without past mentoring experience. This suggests that perhaps those previously mentored had already reaped the benefits of mentoring, so additional mentoring led to minimal changes in scores. In comparison, students who had not been previously mentored saw significant improvements in their post-program scores. Table 12 shows that the program did not have different effects on students who had above or below the mean number of days absent. This finding suggests that varying levels of days absent prior to the program did not impact how the program affected students' scores.

Discussion: The program generally increased scores and reduced days absent for participants. It was especially beneficial in increasing scores among students who had not been previously mentored. And varying levels of absences at baseline did not affect the program's impact on scores. Limitations of this study include concerns with external validity, as the sample included only students who applied for the program. It is possible that students who took the initiative to apply were more concerned with their scores and absents relative to those who did not apply. This would cause the sample to not be representative of all the students in the two high schools. The study's internal validity is not a concern because a random lottery determined the assignment of who was selected/not selected into the program. A future mentoring program of interest to the Board could target students without prior mentorship, as these students' scores would benefit greatly. Further research is also needed to better understand if and how number of absents relate to students’ risk of dropping out of school.

Table 1: Randomization Process of Program Selection

Selected into Program?	Frequency	Percentage
No	131	54.6
Yes	109	45.2
Total	240	100.0

Table 2: Breakdown of Participants Not in Program

Not in Program

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
Score at Application	131	89.5	7.6	67	99
Days Absent 3 months before Application	131	18.5	12.1	0	51
Ethnic Group Black	131	0.4	0.5	0	1
Ethnic Group Hispanic	131	0.5	0.5	0	1
Ethnic Group White	131	0.1	0.3	0	1
JFK High	131	0.3	0.4	0	1
Average Class Size	131	23.7	3.1	20	30
Ever in Mentoring Before	131	0.1	0.3	0	1

Table 2: Breakdown of Participants in Program

In Program

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
Score at Application	109	90.2	6.8	68	99
Days Absent 3 months before Application	109	14.9	11.8	0	48
Ethnic Group Black	109	0.4	0.5	0	1
Ethnic Group Hispanic	109	0.6	0.5	0	1
Ethnic Group White	109	0.1	0.3	0	1
JFK High	109	0.4	0.5	0	1
Average Class Size	109	24.1	3.7	20	30
Ever in Mentoring Before	109	0.2	0.4	0	1

Table 3: Mean Score Pre and Post Program (as %)

Mean Score	Participated in Program	No Participation
Pre	90.2	89.5
Post	92	75.7

Table 4: Mean Absents Pre and Post Program

Mean Days Absent	Participated in Program	No Participation
Pre	15	19
Post	7	18

Figure 1: Mean Score (as %)

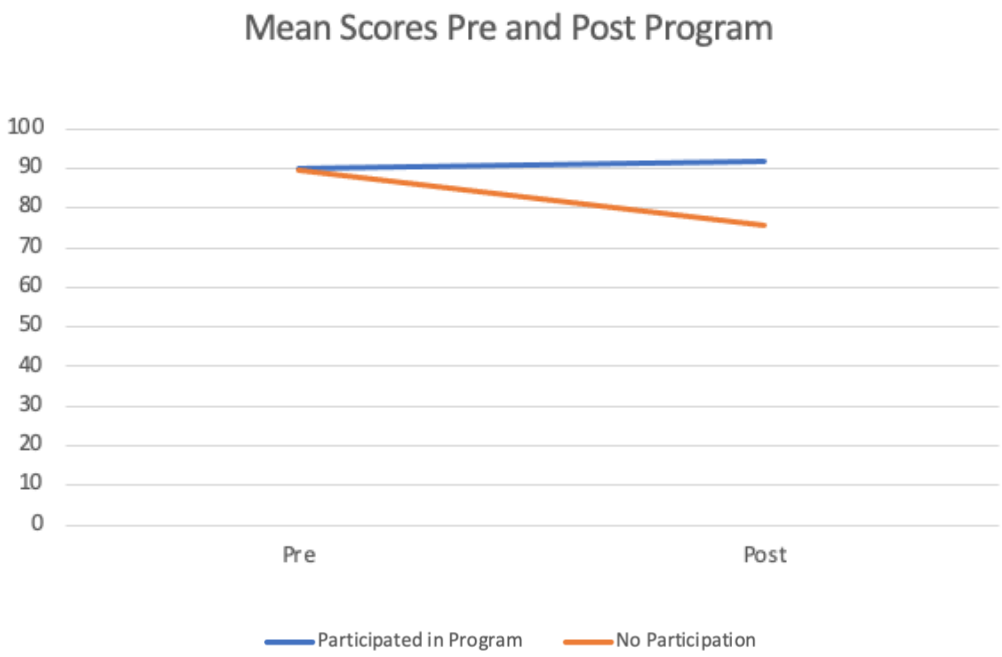

Figure 2: Mean Absents Pre and Post Program

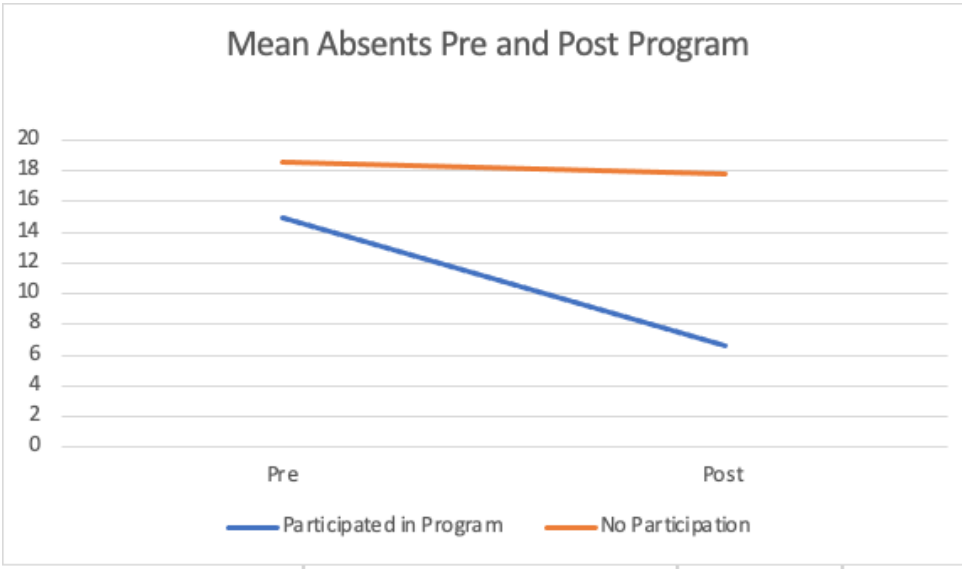

Table 5: Impact of Program Participation on Score (as %)

	OUTCOME
VARIABLES	Score
Program	16.3***
	(2.2)
Constant	75.7***
	(1.5)
Observations	240
R-squared	0.2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Impact of Program Participation on Days Absent

	OUTCOME
VARIABLES	Absent
Program	-11.2***
	(1.2)
Constant	17.9***
	(0.8)
Observations	240
R-squared	0.3

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Set A - Impact of Program on Score (as %) for Students Not Previously Mentored

VARIABLES	OUTCOME Score
Program	17.4***
	(2.5)
Constant	73.2***
	(1.6)
Observations	203
R-squared	0.2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Set A - Impact of Program on Score (as %) for Students Previously Mentored

VARIABLES	OUTCOME Score
Program	4.3***
	(1.4)
Constant	93.4***
	(1.0)
Observations	37
R-squared	0.2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Set B - Impact of Program on Score (as %) for Students with Absences Below the Mean Pre-Program

	OUTCOME
VARIABLES	Score
Program	11.2***
	(2.1)
Constant	84.5***
	(1.5)
Observations	117
R-squared	0.2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Set B - Impact of Program on Score (as %) for Students with Absences Above the Mean Pre-Program

	OUTCOME
VARIABLES	Score
Program	17.7***
	(3.7)
Constant	69.2***
	(2.3)
Observations	118
R-squared	0.2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11: Set A - Difference in the Impact of Program on Scores (as %) for Students with Past Mentorship

VARIABLES	OUTCOME Score
Program	17.4***
	(2.3)
Past Mentorship	20.2***
	(4.3)
Program x Past Mentorship	-13.1**
	(5.9)
Constant	73.2***
	(1.5)
Observations	240
R-squared	0.3

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12: Set B - Difference in the Impact of Program on Scores (as %) for Students with Absence Level Above the Mean

VARIABLES	OUTCOME Score
Program	11.2***
	(2.9)
Above Mean Absences Pre-Program	-15.2***
	(2.9)
Program x Above Mean Absences Pre-Program	6.6
	(4.207)
Constant	84.5***
	(2.2)
Observations	235
R-squared	0.3

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

```
  ___  ____  ____  ____  ____ ®
 /__    /   ____/   /   ____/      17.0
___/   /   /___/   /   /___/       BE—Basic Edition

 Statistics and Data Science       Copyright 1985-2021 StataCorp LLC
                                   StataCorp
                                   4905 Lakeway Drive
                                   College Station, Texas 77845 USA
                                   800-STATA-PC        https://www.stata.com
                                   979-696-4600        stata@stata.com

Stata license: Single-user , expiring 30 Sep 2022
Serial number: 301709040695
  Licensed to: Leah Ross


Notes:
      1. Unicode is supported; see help unicode_advice.

. do "/var/folders/48/zxs488bx2qs93wr6cxmf9fvr0000gq/T//SD16999.000000"

. clear all

. capture log close

.
. use "/Users/leah/Desktop/EIPR/mentorStata9copy.dta", clear

.
. log using "/Users/leah/Desktop/EIPR/GA 3 log file.log", replace
-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  /Users/leah/Desktop/EIPR/GA 3 log file.log
  log type:  text
 opened on:  30 Mar 2022, 12:46:14

.
. ssc install outreg2
checking outreg2 consistency and verifying not already installed...
all files already exist and are up to date.

.
.
.  * question 1 describe the sample
.
. summarize

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       black |        240    .3791667    .4861936          0          1
    hispanic |        240    .5416667    .4993022          0          1
    whiteoth |        240    .0791667     .270563          0          1
     kennedy |        240       .3125    .4644811          0          1
    classize |        240    23.91667    3.372503         20         30
-------------+---------------------------------------------------------
     program |        240    .4541667    .4989354          0          1
       score |        240    83.08333    18.78265         10        100
```

pastsc	240	89.83333	7.249104	67	99
absent	240	12.75417	11.02991	0	48
pastab	240	16.8625	12.06892	0	51
pastment	240	.1541667	.3618631	0	1

```
15 . centile
```

Variable	Obs	Percentile	Centile	Binom. interp. [95% conf.	interval]
black	240	50	0	0	0
hispanic	240	50	1	0	1
whiteoth	240	50	0	0	0
kennedy	240	50	0	0	0
classize	240	50	25	25	25
program	240	50	0	0	1
score	240	50	89	86.30566	92
pastsc	240	50	92	91	92
absent	240	50	12	10	14
pastab	240	50	17	14.30566	19
pastment	240	50	0	0	0

```
16 .
17 . *question 2 - describe the sample and show that the randomization worked
18 .
19 . *look at breakdown and n of program participants
20 . tab1 program
```

-> tabulation of program

program dummy	Freq.	Percent	Cum.
0	131	54.58	54.58
1	109	45.42	100.00
Total	240	100.00	

```
21 .          *label the table
22 .         label variable program "selected into program?"

23 .         label define yesno 0 "no" 1 "yes"

24 .         label values program yesno

25 .         codebook program
```

---

program **selected into program?**

Type: Numeric (**byte**)
Label: **yesno**

Range: [**0,1**] Units: **1**
Unique values: **2** Missing **.**: **0/240**

```
              Tabulation: Freq.   Numeric  Label
                            131         0  no
                            109         1  yes

26 .         tab1 program

   -> tabulation of program

   selected |
       into |
   program? |      Freq.     Percent        Cum.
------------+-----------------------------------
         no |        131       54.58       54.58
        yes |        109       45.42      100.00
------------+-----------------------------------
      Total |        240      100.00

27 .         *outreg2 using 5.doc
28 .
29 .                 * this is who was selected into the program vs not see dummy of who was in th
   > e program vs not, find that group not in program (0) bigger then group in program (1) 55% not
   >  in program and 45% in program of those who applied.
30 . *look at frequency of variables, make sure in percentage
31 .
32 . bysort program: sum pastsc pastab black hispanic whiteoth kennedy classize pastment

-------------------------------------------------------------------------------------------------
-> program = no

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      pastsc |        131    89.50382    7.596937        67         99
      pastab |        131    18.51145    12.07309         0         51
       black |        131    .3969466    .4911429         0          1
    hispanic |        131    .5267176    .5012023         0          1
    whiteoth |        131    .0763359    .2665541         0          1
-------------+---------------------------------------------------------
     kennedy |        131     .259542    .4400662         0          1
    classize |        131    23.74046    3.122099        20         30
    pastment |        131    .1221374    .3287014         0          1

-------------------------------------------------------------------------------------------------
-> program = yes

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      pastsc |        109    90.22936    6.821308        68         99
      pastab |        109    14.88073    11.81628         0         48
       black |        109    .3577982    .4815664         0          1
    hispanic |        109     .559633    .4987242         0          1
    whiteoth |        109    .0825688    .2765006         0          1
-------------+---------------------------------------------------------
     kennedy |        109    .3761468    .4866551         0          1
    classize |        109    24.12844    3.654274        20         30
    pastment |        109    .1926606    .3962104         0          1

33 .
```

```
34 . *outreg2 using 11.doc
35 .                 *shows the split is fairly even between groups
36 .
37 .
38 . * ccan onduct chi^2 test, test by association tells dif between two groups significant
39 .
40 . oneway program score

                        Analysis of variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      19.7954365     57    .34728836      1.59     0.0113
 Within groups      39.7003968    182   .218134048
------------------------------------------------------------------------
    Total           59.4958333    239   .248936541

Bartlett's equal-variances test: chi2(27) =   3.9008    Prob>chi2 = 1.000

note: Bartlett's test performed on cells with positive variance:
      17 single-observation cells not used
      13 multiple-observation cells not used

41 .         * this is comparison of association between program and score
42 .
43 .
44 . * question 3 - pre and post means for scores and absent - EXCEL GRAPHS
45 .
46 . *find mean values of pre and post absent and score
47 .
48 . mean pastsc if program==0

Mean estimation                           Number of obs = 131

--------------------------------------------------------------
             |       Mean   Std. err.     [95% conf. interval]
-------------+------------------------------------------------
      pastsc |   89.50382   .6637474      88.19067    90.81696
--------------------------------------------------------------

49 .         *=89.5
50 . mean pastsc if program==1

Mean estimation                           Number of obs = 109

--------------------------------------------------------------
             |       Mean   Std. err.     [95% conf. interval]
-------------+------------------------------------------------
      pastsc |   90.22936   .6533628      88.93428    91.52444
--------------------------------------------------------------

51 .         *=90.2
52 .
53 . mean score if program==0

Mean estimation                           Number of obs = 131

--------------------------------------------------------------
             |       Mean   Std. err.     [95% conf. interval]
```

score	**75.68702**	**1.801125**	**72.12371**	**79.25033**

```
54 .         *=75.7
55 . mean score if program==1
```

Mean estimation — Number of obs = **109**

	Mean	Std. err.	[95% conf.	interval]
score	**91.97248**	**1.062557**	**89.8663**	**94.07865**

```
56 .         *=92.0
57 .
58 . mean pastab if program==0
```

Mean estimation — Number of obs = **131**

	Mean	Std. err.	[95% conf.	interval]
pastab	**18.51145**	**1.054831**	**16.42459**	**20.59831**

```
59 .         *=18.5
60 . mean pastab if program==1
```

Mean estimation — Number of obs = **109**

	Mean	Std. err.	[95% conf.	interval]
pastab	**14.88073**	**1.131795**	**12.63732**	**17.12415**

```
61 .         *=14.9
62 . mean absent if program==0
```

Mean estimation — Number of obs = **131**

	Mean	Std. err.	[95% conf.	interval]
absent	**17.84733**	**.9219368**	**16.02339**	**19.67127**

```
63 .         *=17.8
64 . mean absent if program==1
```

Mean estimation — Number of obs = **109**

	Mean	Std. err.	[95% conf.	interval]
absent	**6.633028**	**.7776679**	**5.091555**	**8.1745**

```
-------------+------------------------------------------------------------

. *=6.6
. 
. 
.  *numbers put into excel table and make graphs
. 
. 
. *question 4 - estimate the program impacts
.         *formula: score=trt
. 
. reg score program
```

Source	SS	df	MS	Number of obs	=	240
				F(1, 238)	=	54.79
Model	15779.248	1	15779.248	Prob > F	=	0.0000
Residual	68537.0854	238	287.970947	R-squared	=	0.1871
				Adj R-squared	=	0.1837
Total	84316.3333	239	352.788006	Root MSE	=	16.97

score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	16.28545	2.200042	7.40	0.000	11.95141	20.6195
_cons	75.68702	1.48265	51.05	0.000	72.76623	78.60782

```
. outreg2
  outreg2 using `"4.doc"'
4.doc
dir : seeout

. *         * findings; the impact of the program on scores is 16, highly statistically significa
> nt at 1% confidence level (this what expect)
. *                 **can also do with covarites: reg score program pastsc pastab black hispanic
> whiteoth kennedy classize pastment
. 
. 
. reg absent program
```

Source	SS	df	MS	Number of obs	=	240
				F(1, 238)	=	82.46
Model	7482.22817	1	7482.22817	Prob > F	=	0.0000
Residual	21594.2677	238	90.7322171	R-squared	=	0.2573
				Adj R-squared	=	0.2542
Total	29076.4958	239	121.658978	Root MSE	=	9.5253

absent	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	-11.2143	1.234916	-9.08	0.000	-13.64706	-8.781539
_cons	17.84733	.8322333	21.45	0.000	16.20784	19.48681

```
. *outreg2
. *         * findings; the impact of program on absents is -11, highly statistically significant
>  at 1% confidence interval. Meaning when accounting for program, absents decline (this what e
> xpect)
```

```
 83 .         ** can also do with covarietes:reg absent program pastsc pastab black hispanic whiteo
   > th kennedy classize pastment
 84 .
 85 . *put findings into table and describe them
 86 .
 87 . * question 5 - compare impacts of subgroups, outcome is score
 88 .
 89 .         *Set A:  Students who were previously mentored (pastment =1), versus students who wer
   > e not previously mentored (pastment =0)
 90 .
 91 .         *(For Set A, was the program impact different for students who were and were not prev
   > iously mentored;  for Set B, same question for students with baseline low and high absences.)
 92 .
 93 .
 94 .         *make sub groups previously mentored vs not previously mentored
 95 .
 96 .         reg score program if pastment ==0
```

Source	SS	df	MS	Number of obs	=	203
				F(1, 201)	=	49.33
Model	15071.6649	1	15071.6649	Prob > F	=	0.0000
Residual	61404.9854	201	305.49744	R-squared	=	0.1971
				Adj R-squared	=	0.1931
Total	76476.6502	202	378.597278	Root MSE	=	17.478

score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	17.38755	2.475492	7.02	0.000	12.50628	22.26882
_cons	73.22609	1.629877	44.93	0.000	70.01224	76.43994

```
 97 .         outreg2 using "m.doc"
   m.doc
   dir : seeout

 98 .                         * estimating for if not past mentoring and in the program coef = 17.7
   > , highly statistically significant at the 1% confidence interval, means for kids who were not
   >  previously mentored and then in program, scores went up significantly.
 99 .
100 .         reg score program if pastment ==1
```

Source	SS	df	MS	Number of obs	=	37
				F(1, 35)	=	9.98
Model	167.259009	1	167.259009	Prob > F	=	0.0033
Residual	586.416667	35	16.7547619	R-squared	=	0.2219
				Adj R-squared	=	0.1997
Total	753.675676	36	20.9354354	Root MSE	=	4.0933

score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	4.291667	1.358315	3.16	0.003	1.534142	7.049192
_cons	93.375	1.023315	91.25	0.000	91.29756	95.45244

```
101 .         outreg2 using "y.doc"
   y.doc
```

dir : seeout

```
102 .                 * estimating for if did have past mentoring and in the program coef = 4.3 hig
    > hly statistically significant. Shows when had previous mentoring, not as much boost in score
    > when also in this new program.
103 .
104 .
105 .
106 .                 *takeaway: when never past mentored and in the program, the magnitude of the
    > effect of the program on scores is much larger (17.3) compared to the coeficient on program i
    > f have been mentored before (4.3), both cases, these coeficents are highly statistically sign
    > ificant at 1% level.
107 .
108 .
109 .         *Set B:  Students with high versus low absences at baseline (pastab, where "low" is l
    > ower than the mean, and "high" is higher than the mean.
110 .
111 .         *make sub groups above and below average absent pre-program, and how this subgroup ef
    > fected score
112 .
113 .         mean pastab

Mean estimation                           Number of obs = 240

------------------------------------------------------------------------------
             |       Mean   Std. err.     [95% conf. interval]
-------------+----------------------------------------------------------------
     pastab |    16.8625   .7790455      15.32783    18.39717
------------------------------------------------------------------------------

114 .         *mean pre-program absents round to = 17
115 .
116 .         gen lowmeanpastab = 1 if (pastab <17)
(123 missing values generated)

117 .         replace lowmeanpastab  = 0 if (pastab <17)
(117 real changes made)

118 .                 *means if lowmeanpastab = 0, then less then average absents
119 .         tab1 lowmeanpastab

-> tabulation of lowmeanpastab

lowmeanpast |
         ab |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        117      100.00      100.00
------------+-----------------------------------
      Total |        117      100.00

120 .
121 .         *check done correctly
122 .         tab2 pastab lowmeanpastab

-> tabulation of pastab by lowmeanpastab

days abs 3 | lowmeanpas
mo. before |    tab
```

app.	0	Total
0	**31**	**31**
1	**2**	**2**
2	**3**	**3**
3	**6**	**6**
4	**4**	**4**
5	**7**	**7**
6	**6**	**6**
7	**5**	**5**
8	**5**	**5**
9	**9**	**9**
10	**11**	**11**
11	**4**	**4**
12	**4**	**4**
13	**4**	**4**
14	**4**	**4**
15	**6**	**6**
16	**6**	**6**
Total	**117**	**117**

```
123 .                *see assiged to 0 when absents are below 17 and assigned to 1 when values abo
    > ve 17
124 .
125 .         sort lowmeanpastab

126 .         by lowmeanpastab: summarize pastab
```

-> lowmeanpastab = 0

Variable	Obs	Mean	Std. dev.	Min	Max
pastab	**117**	**6.538462**	**5.350698**	**0**	**16**

-> lowmeanpastab = .

Variable	Obs	Mean	Std. dev.	Min	Max
pastab	**123**	**26.68293**	**7.660113**	**17**	**51**

```
127 .                * we see here because doing >,< miss 5 observations that =17
128 .
129 .         gen highmeanpastab = 0 if (pastab <17)
    (123 missing values generated)

130 .         replace highmeanpastab  = 1 if (pastab >17)
    (118 real changes made)

131 .         tab1 highmeanpastab
```

-> tabulation of highmeanpastab

highmeanpastab	Freq.	Percent	Cum.

0	117	49.79	49.79
1	118	50.21	100.00
Total	235	100.00	

```
132 .                 * tells how many pre treatment absenses were above the mean = 118 or 50%
133 .
134 .         *check done correctly:
135 .         sort highmeanpastab

136 .         by highmeanpastab: summarize pastab
```

-> highmeanpastab = 0

Variable	Obs	Mean	Std. dev.	Min	Max
pastab	117	6.538462	5.350698	0	16

-> highmeanpastab = 1

Variable	Obs	Mean	Std. dev.	Min	Max
pastab	118	27.09322	7.550386	18	51

-> highmeanpastab = .

Variable	Obs	Mean	Std. dev.	Min	Max
pastab	5	17	0	17	17

```
137 .
138 .
139 . reg score program if lowmeanpastab == 0
```

Source	SS	df	MS		
Model	3618.86223	1	3618.86223	Number of obs	= 117
Residual	14224.2831	115	123.689418	F(1, 115)	= 29.26
				Prob > F	= 0.0000
Total	17843.1453	116	153.820218	R-squared	= 0.2028
				Adj R-squared	= 0.1959
				Root MSE	= 11.122

score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	11.15608	2.062491	5.41	0.000	7.070687	15.24148
_cons	84.46296	1.513455	55.81	0.000	81.4651	87.46083

```
140 .         *program coef is 11.2, highly statistically significant at 1% level
141 .         outreg2 using "1.doc"
    1.doc
    dir : seeout
```

```
142 . reg score program if highmeanpastab== 1
```

Source	SS	df	MS	Number of obs	=	118
				F(1, 116)	=	23.05
Model	8730.52725	1	8730.52725	Prob > F	=	0.0000
Residual	43933.4728	116	378.736834	R-squared	=	0.1658
				Adj R-squared	=	0.1586
Total	52664	117	450.119658	Root MSE	=	19.461

score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	17.70898	3.688435	4.80	0.000	10.40357	25.01439
_cons	69.24658	2.277757	30.40	0.000	64.73519	73.75796

```
143 .         * program coef 17.7 highly statistically significant at 1% level
144 .         outreg2 using "2.doc"
    2.doc
    dir : seeout

145 .
146 . *Then, for each of the two sets individually, create an interaction variable to test whether
    > the program impact was significantly different within the two subgroups in the set.(For Set A
    > , was the program impact different for students who were and were not previously mentored;  f
    > or Set B, same question for students with baseline low and high absences.)
147 .
148 . * Formula: Score = trt subgroup trt*subgroup
149 .
150 . *Set A:
151 .
152 .         gen yespastmentprogram = (pastment*program)

153 .         reg score program pastment yespastmentprogram
```

Source	SS	df	MS	Number of obs	=	240
				F(3, 236)	=	28.33
Model	22324.9313	3	7441.64376	Prob > F	=	0.0000
Residual	61991.402	236	262.675432	R-squared	=	0.2648
				Adj R-squared	=	0.2554
Total	84316.3333	239	352.788006	Root MSE	=	16.207

score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	17.38755	2.295448	7.57	0.000	12.86536	21.90974
pastment	20.14891	4.324506	4.66	0.000	11.62935	28.66848
yespastmentprogram	-13.09588	5.847619	-2.24	0.026	-24.61608	-1.575682
_cons	73.22609	1.511335	48.45	0.000	70.24866	76.20352

```
154 .         outreg2 using "3.doc"
    3.doc
    dir : seeout

155 .                 *interaction coefficent = -13,
156 .                 *for those with past mentoring vs those not past mentoring there is different
    >  impact when program is applied
```

```
157 .                         *- Those with pastmentoring in program had less of an affect on score
    >  than those without past mentoring in the program.
158 .
159 . *Set B:
160 .
161 . gen highmeanpastabprogram = (highmeanpastab*program)
    (5 missing values generated)

162 . reg score program highmeanpastab highmeanpastabprogram
```

Source	SS	df	MS	Number of obs	=	235
				F(3, 231)	=	32.64
Model	24650.4399	3	8216.81331	Prob > F	=	0.0000
Residual	58157.7558	231	251.765177	R-squared	=	0.2977
				Adj R-squared	=	0.2886
Total	82808.1957	234	353.881178	Root MSE	=	15.867

score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	11.15608	2.942548	3.79	0.000	5.358421	16.95375
highmeanpastab	-15.21639	2.848009	-5.34	0.000	-20.82778	-9.604993
highmeanpastabprogram	6.552896	4.207399	1.56	0.121	-1.736888	14.84268
_cons	84.46296	2.15924	39.12	0.000	80.20864	88.71729

```
163 .         outreg2 using "4.doc"
    4.doc
    dir : seeout

164 .         * 6.6 = interaction term but not statistically significant
165 .                 *remember interaction term tells Difference in the impact of TRT in the subgr
    > oups SUBGRP = 0 and SUBGRP = 1
166 .                 * no difference in how the treatment affected those with high vs low meanpast
    > ab
167 .
    end of do-file

168 .
```