Basic Findings

As Graph 1 and Graph 2 show, students in the control group had a mean score of 89.5 at the baseline and a mean score of 75.7 after the program period. Their scores, on average, decreased by 13.8 points. Students in the treatment group had a mean score of 90.2 at the baseline and a mean score of 92.0 after the program. Their scores, on average, increased by 1.8 points. The difference in their change in scores was 15.6 points. Additionally, students in the control group had an average of 18.5 days of absence at the baseline and 17.8 days after the program period, with their days of absence on average decreasing by 0.7 days. Students in the treatment group had an average of 14.9 days of absence at the baseline and a mean of 6.6 days after the program, with their days of absence on average decreasing by 14.3 days. The difference in their change in days of absence was 7.6 points.

Impact Estimations

This study applies a quasi-experimental difference-in-difference design to estimate the mentor program's impact on students in each group. As Graph 3 shows, when controlling for race/ethnicity, school, past mentor experience, and class size, being in the mentor program is associated with getting 15.2 more scores on average for a student. As Graph 4 shows, when controlling for the same variates, being in the mentor program is associated with getting -7.2 fewer days of absence from school on average for a student.

Subgroup Analysis

I restructured the data to a difference-in-difference dataset structure and compared the programs' impact on two different student groups. As shown in Table 6, The program has a greater impact on students who were not previously mentored than those who were previously mentored. For

students who were not previously mentored, being in the program increased the score by 16 points. For students who were not previously mentored, being in the program increased the score by 3.6 points. The program has a greater impact on students with high absences at baseline than on students with low absences at baseline. For students with high absences at baseline, being in the program increased the score by 9.5 points. For students with low absences at baseline, being in the program increased the score by 17.5 points.

Conclusions

Based on this study, for the students in JFK High and WIH High, being in the mentor program is associated with enhancing their test scores and reducing days of absence. The program's impact was significantly larger for students who were previously mentored than who weren't; and the program's impact was greater for those who previously had high absences at baseline than those who previously had low absences.

Limitations

This study does not conduct a check on parallel trends because we don't have pre-data, which are the scores and absences of students in treatment and control groups absence further before the program takes place. There is the possibility of the composition of the treatment group changing as a result of the mentorship program. Also, there might be maturity issues between the two groups. As students grew, students in treatment and control groups may develop different understandings of studying, which may affect the students' scores and absences. Further, there might be selection bias that those who registered for the lottery are more likely to improve their grades and reduce absences than those who didn't want to register for the lottery. And the sample size is relatively small, and the two schools in NYC are not representative of other schools with different profiles or in different geographic locations. Thus, the external validity is limited.

TABLE 1
Distribution of mentoring program
By race/ethnicity, mentor experiences, and school characteristics

	Not in Program n = 131	In Program n = 109
Race/ethnicity (%)		
White	7.6	8.3
Black	39.7	35.8
Hispanic	52.7	60.0
School (%)		
Washington Irving High	74.1	62.4
John F. Kennedy High	26.0	37.6
Class Size (%)		
20	35.1	36.7
25	55.0	44.0
30	10.0	19.3
Past Mentoring Experience (%)		
no	87.8	80.7
yes	12.2	19.3

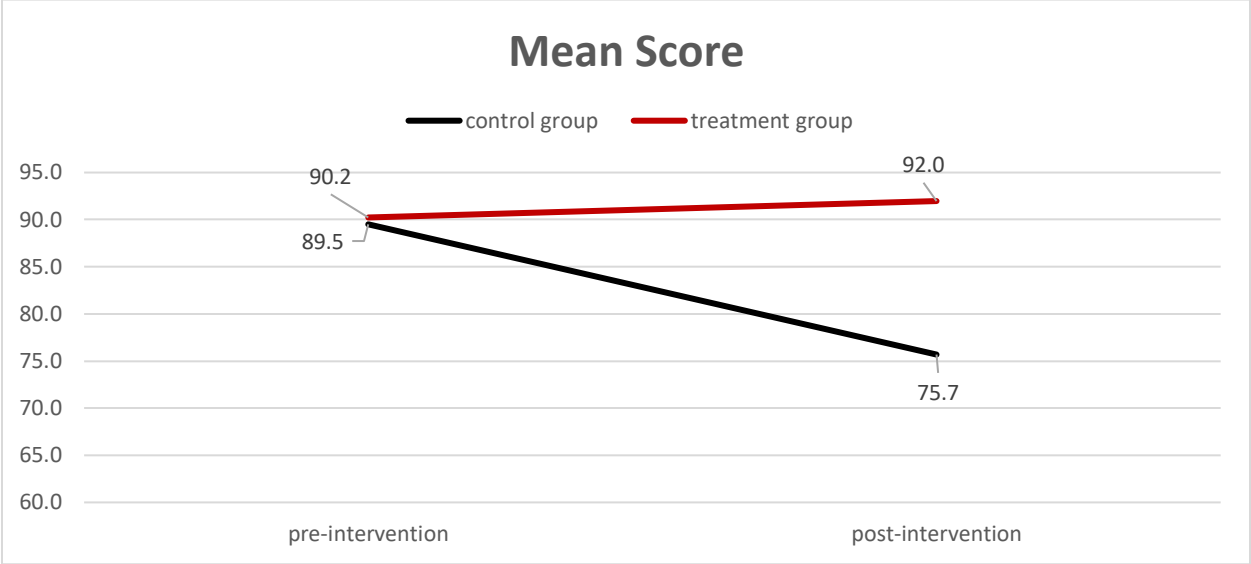
* p <.05 for a chi-square test of no association between supplemental insurance status and this characteristic.

TABLE 2
Score and days of absence at the baseline
By program status

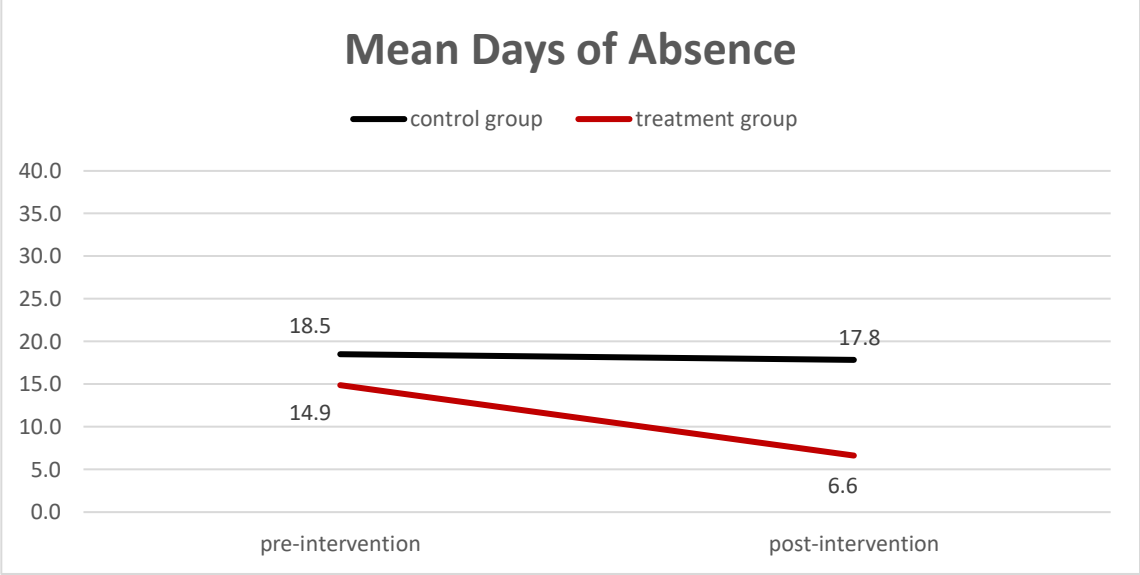
	Baseline Mean Score	Baseline Mean Days of Absence
Program Status		
Not in the program	89.5	18.5
In the program	90.2	14.9

* p < .05 for an ANOVA test of no difference between mean expenditures across groups defined by this characteristic.

GRAPH 1

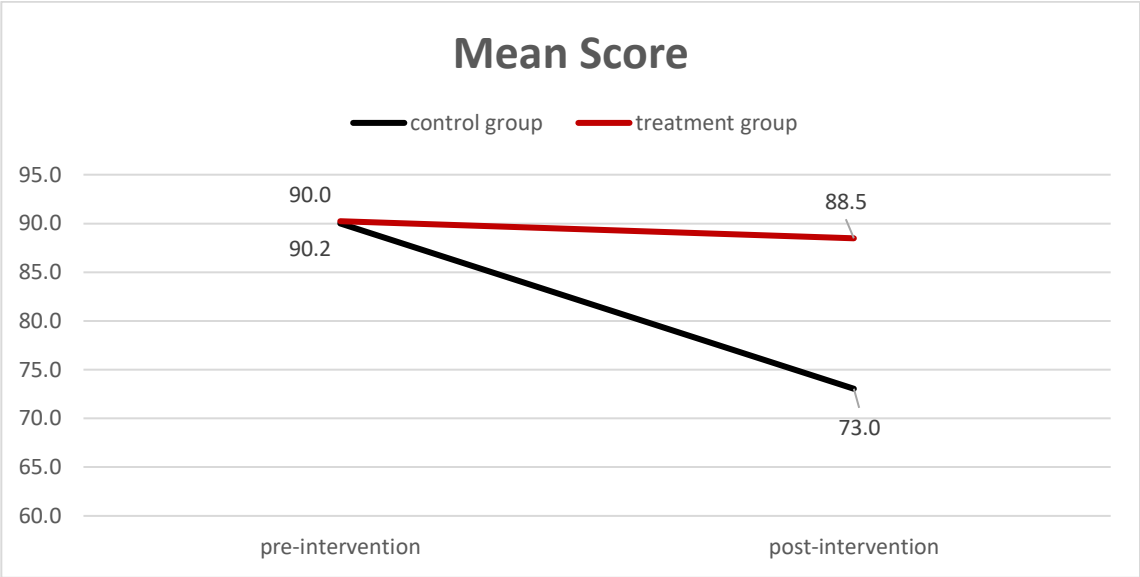


GRAPH 2



GRAPH 3

(Controlling for race/ethnicity, school, class size, and mentor experience)



GRAPH 4

(Controlling for race/ethnicity, school, class size, and mentor experience)

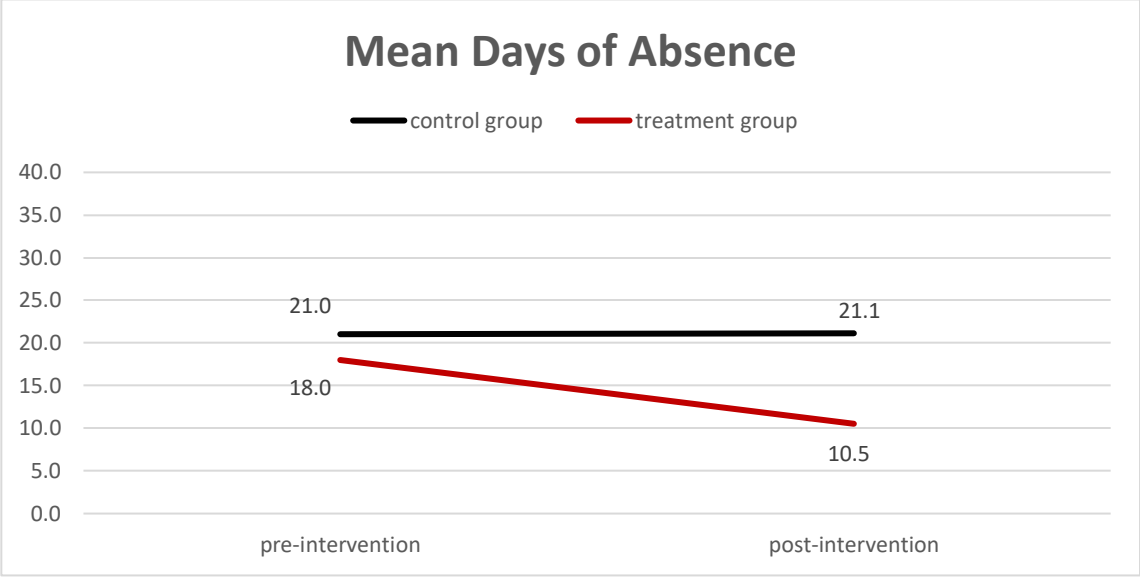


TABLE 3

Average program Impact on Student's Score

Controlling for race/ethnicity, school, class size, mentor experience

	control group	treatment group	dif-in-dif (program impact)
pre-intervention	21.0	18.0	
post-intervention	21.1	10.5	
difference	0.1	-7.5	-7.6

TABLE 4

Average program Impact on Student's Days of Absence

Controlling for race/ethnicity, school, class size, mentor experience

	control group	treatment group	dif-in-dif (program impact)
pre-intervention	90.0	90.2	
post-intervention	73.0	88.5	
difference	-17.0	-1.8	15.2

TABLE 5

Program Effect on Student Subgroups
(controlling for Controlling for race/ethnicity, school, class size, mentor experience)

	no past mentor experience	have past mentor experience	low absences at baseline	high absences at baseline
VARIABLES	program	program	program	program
after	-16.44***	5.062***	-4.278**	-20.51***
	(1.780)	(1.810)	(1.768)	(2.358)
program	0.574	1.411	1.810	-0.329
	(1.940)	(1.756)	(1.724)	(2.769)
after_program	16.55***	3.557	9.484***	17.51***
	(2.704)	(2.403)	(2.409)	(3.855)
black	-0.139	-1.879	5.092*	-4.735
	(2.557)	(2.905)	(2.697)	(3.363)
hispanic	0.679	0.383	4.373	-1.644
	(2.502)	(2.752)	(2.661)	(3.266)
kennedy	1.649	2.070	2.249	2.932
	(2.022)	(1.846)	(1.866)	(2.716)
25.classize	-0.538	-0.515	1.730	-2.029
	(1.666)	(1.481)	(1.443)	(2.423)
30.classize	-1.032	1.519	3.574	-2.803
	(2.889)	(2.890)	(2.917)	(3.891)
Constant	89.38***	87.85***	82.62***	93.31***
	(2.750)	(3.176)	(3.016)	(3.536)
Observations	406	74	234	246
R-squared	0.249	0.431	0.203	0.301

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

```
      name:  <unnamed>
       log:  /Users/wsq/Desktop/Estimating Impact in Policy Research/Graded Assignment 3/WangSiqiAssignment3
> _log.smcl
  log type:  smcl
 opened on:   5 Apr 2023, 18:12:12

1 .
2 . * explore
3 . browse

4 . codebook

-------------------------------------------------------------------------------------------------------------------
black                                                                                    ethnic group: black
-------------------------------------------------------------------------------------------------------------------

                  Type: Numeric (byte)

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

            Tabulation: Freq.  Value
                        149  0
                         91  1

-------------------------------------------------------------------------------------------------------------------
hispanic                                                                                ethnic group: hispanic
-------------------------------------------------------------------------------------------------------------------

                  Type: Numeric (byte)

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

            Tabulation: Freq.  Value
                        110  0
                        130  1

-------------------------------------------------------------------------------------------------------------------
whiteoth                                                                                ethnic group: white-oth
-------------------------------------------------------------------------------------------------------------------

                  Type: Numeric (byte)

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

            Tabulation: Freq.  Value
                        221  0
                         19  1

-------------------------------------------------------------------------------------------------------------------
kennedy                                                                                  john f. kennedy high
-------------------------------------------------------------------------------------------------------------------
```

Type: Numeric (**byte**)

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

Tabulation: Freq. Value
165 **0**
75 **1**

---

classize **average class size**

Type: Numeric (**byte**)

Range: [**20,30**] Units: **1**
Unique values: **3** Missing **.**: **0/240**

Tabulation: Freq. Value
86 **20**
120 **25**
34 **30**

---

program **program dummy**

Type: Numeric (**byte**)

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

Tabulation: Freq. Value
131 **0**
109 **1**

---

score **score at follow-up**

Type: Numeric (**float**)

Range: [**10,100**] Units: **1**
Unique values: **58** Missing **.**: **0/240**

Mean: **83.0833**
Std. dev.: **18.7827**

Percentiles: 10% 25% 50% 75% 90%
57 **73** **89** **100** **100**

---

pastsc **score at application**

Type: Numeric (**float**)

Range: [67,99] Units: 1
Unique values: 30 Missing .: 0/240

Mean: 89.8333
Std. dev.: 7.2491

Percentiles:	10%	25%	50%	75%	90%
	79	86	92	95	98

---

absent **days abs at follow-up.**

Type: Numeric (**float**)

Range: [0,48] Units: 1
Unique values: 39 Missing .: 0/240

Mean: 12.7542
Std. dev.: 11.0299

Percentiles:	10%	25%	50%	75%	90%
	0	0	12	22	28

---

pastab **days abs 3 mo. before app.**

Type: Numeric (**float**)

Range: [0,51] Units: 1
Unique values: 46 Missing .: 0/240

Mean: 16.8625
Std. dev.: 12.0689

Percentiles:	10%	25%	50%	75%	90%
	0	7	17	25.5	32

---

pastment **ever in mentoring before**

Type: Numeric (**byte**)

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

Tabulation:	Freq.	Value
	203	0
	37	1

```
 5 .
 6 . *************************
 7 . * 2.Describe the sample *
 8 . *************************
 9 . * Check for parallel trends and attrition by comparing treatment group vs. control
10 . * How well the randomization process worked?
11 . ** control group: 131 observaions; treatment group: 109 observations
12 . bys program: sum pastab pastment kennedy classize black hispanic whiteoth

-------------------------------------------------------------------------------
-> program = 0

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

-------------------------------------------------------------------------------
-> program = 1

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


13 . * notes:
14 . ** pastsc(continuous)
15 . ** score(continuous)
16 . ** pastab(continuous)
17 . ** absent(continuous)
18 . ** classize(continuous)
19 . ** kennedy(nominal)
```

```
20 .
21 . gen race =.
   (240 missing values generated)

22 . replace race=0 if white==1
   (19 real changes made)

23 . replace race=1 if black==1
   (91 real changes made)

24 . replace race=2 if hispanic==1
   (130 real changes made)

25 . *race(nominal)
26 . oneway pastsc program, tabulate
```

program dummy	Summary of score at application: Mean	Std. dev.	Freq.
0	**89.503817**	**7.5969367**	**131**
1	**90.229358**	**6.8213079**	**109**
Total	**89.833333**	**7.2491042**	**240**

Analysis of variance

Source	SS	df	MS	F	Prob > F
Between groups	**31.3191867**	**1**	**31.3191867**	**0.59**	**0.4413**
Within groups	**12528.0141**	**238**	**52.6387149**		
Total	**12559.3333**	**239**	**52.5495119**		

Bartlett's equal-variances test: chi2(**1**) = **1.3510** Prob>chi2 = **0.245**

```
27 . oneway pastab program,tabulate
```

program dummy	Summary of days abs 3 mo. before app.: Mean	Std. dev.	Freq.
0	**18.51145**	**12.073089**	**131**
1	**14.880734**	**11.816283**	**109**
Total	**16.8625**	**12.06892**	**240**

Analysis of variance

Source	SS	df	MS	F	Prob > F
Between groups	**784.280134**	**1**	**784.280134**	**5.49**	**0.0200**
Within groups	**34028.1824**	**238**	**142.975556**		
Total	**34812.4625**	**239**	**145.658839**		

Bartlett's equal-variances test: chi2(**1**) = **0.0542** Prob>chi2 = **0.816**

28 . tabulate kennedy program,chi2 column

Key
frequency *column percentage*

john f. kennedy high	program dummy 0	1	Total
0	**97** **74.05**	**68** **62.39**	**165** **68.75**
1	**34** **25.95**	**41** **37.61**	**75** **31.25**
Total	**131** **100.00**	**109** **100.00**	**240** **100.00**

Pearson chi2(**1**) = **3.7653** Pr = **0.052**

29 . tabulate classize program,chi2 column

Key
frequency *column percentage*

average class size	program dummy 0	1	Total
20	**46** **35.11**	**40** **36.70**	**86** **35.83**
25	**72** **54.96**	**48** **44.04**	**120** **50.00**
30	**13** **9.92**	**21** **19.27**	**34** **14.17**
Total	**131** **100.00**	**109** **100.00**	**240** **100.00**

Pearson chi2(**2**) = **5.1274** Pr = **0.077**

30 . tabulate pastment program, chi2 column

Key
frequency *column percentage*

ever in mentoring before	program dummy 0	1	Total
0	**115** **87.79**	**88** **80.73**	**203** **84.58**
1	**16** **12.21**	**21** **19.27**	**37** **15.42**
Total	**131** **100.00**	**109** **100.00**	**240** **100.00**

Pearson chi2(**1**) = **2.2692** Pr = **0.132**

31 . tabulate race program, chi2 column

Key
frequency *column percentage*

race	program dummy 0	1	Total
0	**10** **7.63**	**9** **8.26**	**19** **7.92**
1	**52** **39.69**	**39** **35.78**	**91** **37.92**
2	**69** **52.67**	**61** **55.96**	**130** **54.17**
Total	**131** **100.00**	**109** **100.00**	**240** **100.00**

Pearson chi2(**2**) = **0.3887** Pr = **0.823**

```
32 .
33 . *******************************************************
34 . * 3.Basic Study Findings and Examine Program impacts   *
35 . *******************************************************
36 . * Method 1: sum
37 . bys program: sum score
```

-> program = 0

Variable	Obs	Mean	Std. dev.	Min	Max
score	**131**	**75.68702**	**20.61481**	**10**	**100**

-> program = 1

Variable	Obs	Mean	Std. dev.	Min	Max
score	**109**	**91.97248**	**11.09342**	**52**	**100**

```
38 . bys program: sum pastsc
```

-> program = 0

Variable	Obs	Mean	Std. dev.	Min	Max
pastsc	**131**	**89.50382**	**7.596937**	**67**	**99**

-> program = 1

Variable	Obs	Mean	Std. dev.	Min	Max
pastsc	**109**	**90.22936**	**6.821308**	**68**	**99**

```
39 . *effect_on_trt_group=post-pre
40 . di 91.97248-90.22936 //1.74312
   1.74312

41 . *effect_on_control_group=post-pre
42 . di  75.68702-89.50382 //-13.8168
   -13.8168
```

```
. *dif_in_dif
. di 1.74312-(-13.8168) //15.55992
15.55992

.
. bys program: sum absent

-------------------------------------------------------------------------------------------------------------------
-> program = 0

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      absent |        131    17.84733    10.55205          0         48

-------------------------------------------------------------------------------------------------------------------
-> program = 1

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      absent |        109    6.633028    8.119091          0         38


. bys program: sum pastab

-------------------------------------------------------------------------------------------------------------------
-> program = 0

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      pastab |        131    18.51145    12.07309          0         51

-------------------------------------------------------------------------------------------------------------------
-> program = 1

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      pastab |        109    14.88073    11.81628          0         48


. *effect_on_trt_group=post-pre
. di 6.633028-14.88073   //-8.247702
-8.247702

. *effect_on_control_group=post-pre
. di 17.84733-18.51145 // -.66412
-.66412
```

```
52 . *dif_in_dif
53 . di -8.247702 -(-.66412) //-7.583582
   -7.583582

54 .
55 . gen difsc=score-pastsc

56 .
57 . * Method 2: regression
58 . *difference in score
59 . * / without covariates / *
60 . *score
61 . reg pastsc program
```

Source	SS	df	MS	Number of obs	=	240
				F(1, 238)	=	0.59
Model	31.3191867	1	31.3191867	Prob > F	=	0.4413
Residual	12528.0141	238	52.6387149	R-squared	=	0.0025
				Adj R-squared	=	-0.0017
Total	12559.3333	239	52.5495119	Root MSE	=	7.2553

pastsc	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	.725541	.9406095	0.77	0.441	-1.127442	2.578524
_cons	89.50382	.6338945	141.20	0.000	88.25506	90.75258

```
62 . 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

```
63 . display 16.28545  - .725541 //15.559909
   15.559909
```

```
64 .
65 . *absent
66 . reg pastab program
```

Source	SS	df	MS	Number of obs	=	240
				F(1, 238)	=	5.49
Model	784.280134	1	784.280134	Prob > F	=	0.0200
Residual	34028.1824	238	142.975556	R-squared	=	0.0225
				Adj R-squared	=	0.0184
Total	34812.4625	239	145.658839	Root MSE	=	11.957

pastab	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	-3.630716	1.5502	-2.34	0.020	-6.684581	-.5768519
_cons	18.51145	1.044709	17.72	0.000	16.45339	20.56951

```
67 . 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

```
68 . display -11.2143-(-3.630716)  // -7.583584
-7.583584

69 .
70 . * / with covariates / *
71 . *score
72 . reg pastsc program black hispanic kennedy i.classize pastment
```

Source	SS	df	MS	Number of obs	=	240
				F(7, 232)	=	2.57
Model	902.514334	7	128.930619	Prob > F	=	0.0145
Residual	11656.819	232	50.2449095	R-squared	=	0.0719
				Adj R-squared	=	0.0439
Total	12559.3333	239	52.5495119	Root MSE	=	7.0884

pastsc	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	.2312176	.9380803	0.25	0.806	-1.617028	2.079463
black	-.5023009	1.789074	-0.28	0.779	-4.027209	3.022608
hispanic	-.2190396	1.751214	-0.13	0.901	-3.669355	3.231275
kennedy	4.309236	1.371166	3.14	0.002	1.607708	7.010765
classize						
25	-1.497213	1.128038	-1.33	0.186	-3.71972	.7252941
30	-1.00792	1.998462	-0.50	0.614	-4.945374	2.929535
pastment	-1.276426	1.283471	-0.99	0.321	-3.805173	1.252322
_cons	89.77897	1.838379	48.84	0.000	86.15691	93.40102

```
73 . reg score program black hispanic kennedy i.classize pastment
```

Source	SS	df	MS	Number of obs	=	240
				F(7, 232)	=	11.17
Model	21259.1948	7	3037.02783	Prob > F	=	0.0000
Residual	63057.1385	232	271.798011	R-squared	=	0.2521
				Adj R-squared	=	0.2296
Total	84316.3333	239	352.788006	Root MSE	=	16.486

score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	15.43402	2.181811	7.07	0.000	11.13533	19.73272
black	-.3602953	4.161074	-0.09	0.931	-8.558617	7.838026
hispanic	1.648345	4.073018	0.40	0.686	-6.376485	9.673176
kennedy	.0118907	3.189093	0.00	0.997	-6.271393	6.295175
classize						
25	.7222809	2.623619	0.28	0.783	-4.446883	5.891445
30	-.451913	4.648074	-0.10	0.923	-9.609744	8.705918
pastment	12.80169	2.985128	4.29	0.000	6.920266	18.68312
_cons	73.04304	4.275748	17.08	0.000	64.61879	81.4673

```
74 . display   15.43402  - (  .2312176 ) // 15.202802
   15.202802

75 . *absent
```

```
76 . reg pastab program black hispanic kennedy i.classize pastment
```

Source	SS	df	MS	Number of obs	=	240
				F(7, 232)	=	16.77
Model	11696.4369	7	1670.91956	Prob > F	=	0.0000
Residual	23116.0256	232	99.6380412	R-squared	=	0.3360
				Adj R-squared	=	0.3159
Total	34812.4625	239	145.658839	Root MSE	=	9.9819

pastab	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	-3.0238	1.321011	-2.29	0.023	-5.626512	-.4210876
black	-1.799219	2.519387	-0.71	0.476	-6.76302	3.164583
hispanic	-1.026203	2.466072	-0.42	0.678	-5.884961	3.832556
kennedy	1.476765	1.930886	0.76	0.445	-2.327548	5.281077
classize						
25	.0655522	1.588511	0.04	0.967	-3.064198	3.195303
30	4.47595	2.814249	1.59	0.113	-1.068801	10.0207
pastment	-17.39531	1.807393	-9.62	0.000	-20.95631	-13.83431
_cons	21.02729	2.588818	8.12	0.000	15.92669	26.12789

```
77 . reg absent program black hispanic kennedy i.classize pastment
```

Source	SS	df	MS	Number of obs	=	240
				F(7, 232)	=	29.15
Model	13605.5018	7	1943.64311	Prob > F	=	0.0000
Residual	15470.9941	232	66.6853193	R-squared	=	0.4679
				Adj R-squared	=	0.4519
Total	29076.4958	239	121.658978	Root MSE	=	8.1661

absent	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	-10.60824	1.08071	-9.82	0.000	-12.7375	-8.478977
black	-1.4681	2.061092	-0.71	0.477	-5.528949	2.59275
hispanic	-.9541625	2.017476	-0.47	0.637	-4.929077	3.020752
kennedy	.074119	1.579644	0.05	0.963	-3.038161	3.186399
classize						
25	-1.350829	1.299549	-1.04	0.300	-3.911255	1.209597
30	1.798979	2.302317	0.78	0.435	-2.737141	6.3351
pastment	-13.5618	1.478615	-9.17	0.000	-16.47503	-10.64857
_cons	21.13374	2.117894	9.98	0.000	16.96098	25.3065

```
78 . *dif_in_dif
79 . di  -10.60824 -( -3.0238 ) //-7.58444
   -7.58444

80 .
81 . * Method 3: DiD Coding regression
82 . *DiD Coding
83 . gen index=_n

84 . gen after=1

85 . append using mentorStata9

86 . replace after=0 if after==.
   (240 real changes made)

87 . replace index=_n-240 if after==0
   (240 real changes made)

88 . sort index black hispanic whiteoth kennedy classize program score pastsc absent pastab pastment after

89 . order program, after(after)

90 . gen score2=score if after==1
   (240 missing values generated)

91 . replace score2=pastsc if after==0
   (240 real changes made)

92 . gen absent2=absent if after==1
   (240 missing values generated)

93 . replace absent2=pastab if after==0
   (240 real changes made)

94 . gen after_program = after*program

95 . reg score2 after program after_program black hispanic kennedy i.classize pastment //  15.55991

      Source |       SS           df       MS      Number of obs   =       480
-------------+----------------------------------   F(9, 470)       =     15.85
       Model |  23827.0805         9  2647.45339   Prob > F        =    0.0000
    Residual |  78516.0861       470  167.055502   R-squared       =    0.2328
-------------+----------------------------------   Adj R-squared   =    0.2181
       Total |  102343.167       479  213.660056   Root MSE        =    12.925
```

score2	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
after	-13.81679	1.597018	-8.65	0.000	-16.95497	-10.67862
program	.0526626	1.693175	0.03	0.975	-3.274466	3.379791
after_program	15.55991	2.369748	6.57	0.000	10.9033	20.21652
black	-.4312981	2.306735	-0.19	0.852	-4.964088	4.101491
hispanic	.7146528	2.25792	0.32	0.752	-3.722215	5.15152
kennedy	2.160564	1.767907	1.22	0.222	-1.313416	5.634543
classize						
25	-.3874661	1.454431	-0.27	0.790	-3.245457	2.470525
30	-.7299163	2.576709	-0.28	0.777	-5.793211	4.333378
pastment	5.762633	1.654837	3.48	0.001	2.510838	9.014428
_cons	88.3194	2.501193	35.31	0.000	83.4045	93.23431

96 . reg absent2 after program after_program black hispanic kennedy i.classize pastment // -7.583584

Source	SS	df	MS	Number of obs	=	480
				F(9, 470)	=	35.82
Model	26815.9137	9	2979.54597	Prob > F	=	0.0000
Residual	39098.4529	470	83.1881977	R-squared	=	0.4068
				Adj R-squared	=	0.3955
Total	65914.3667	479	137.608281	Root MSE	=	9.1208

absent2	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
after	-.6641221	1.126964	-0.59	0.556	-2.878634	1.55039
program	-3.024226	1.194819	-2.53	0.012	-5.372075	-.6763778
after_program	-7.583584	1.672255	-4.53	0.000	-10.86961	-4.297562
black	-1.633659	1.627789	-1.00	0.316	-4.832304	1.564986
hispanic	-.9901825	1.593342	-0.62	0.535	-4.121138	2.140773
kennedy	.7754418	1.247555	0.62	0.535	-1.676035	3.226918
classize						
25	-.6426383	1.026345	-0.63	0.532	-2.659431	1.374155
30	3.137465	1.818301	1.73	0.085	-.4355404	6.71047
pastment	-15.47856	1.167766	-13.25	0.000	-17.77324	-13.18387
_cons	21.41258	1.765012	12.13	0.000	17.94429	24.88087

```
 97 .
 98 . * Graph_score
 99 . graph twoway (lfit pastsc program) (lfit score program),legend(label(1 "Control Group")label(2 "Treatment
    > Group")) title("The program's impact on students' test scores." ) xlabel(minmax) xtitle("Time") ytitle("Sc
    > ore")

100 . graph export programimpact_on_testscores.png, replace
    file /Users/wsq/Desktop/Estimating Impact in Policy Research/Graded Assignment
        3/programimpact_on_testscores.png saved as PNG format

101 .
102 . * Graph_absence
103 . graph twoway (lfit pastab program) (lfit absent program),legend(label(1 "Control Group")label(2 "Treatment
    >  Group")) title("The program's impact on students' absence." ) xlabel(minmax) xtitle("Time") ytitle("Absen
    > t")

104 . graph export programimpact_on_absence.png, replace
    file /Users/wsq/Desktop/Estimating Impact in Policy Research/Graded Assignment
        3/programimpact_on_absence.png saved as PNG format

105 .
106 . **********************************
107 . * 5.Program's impact on subgroups *
108 . **********************************
109 . * Set A: previously mentored vs. previously not mentored
110 . ** subgroup dummy
111 . codebook pastment

-------------------------------------------------------------------------------------------------------------------
pastment                                                                                ever in mentoring before
-------------------------------------------------------------------------------------------------------------------

                  Type: Numeric (byte)

                 Range: [0,1]                         Units: 1
         Unique values: 2                         Missing .: 0/480

            Tabulation: Freq.  Value
                        406  0
                         74  1

112 . ** regression
113 . reg difsc program if pastment==0 // coefficient:  16.54575, significant at 0.0000 level

      Source |       SS           df       MS      Number of obs   =       203
-------------+----------------------------------   F(1, 201)       =     48.28
       Model |  13647.6364         1  13647.6364   Prob > F        =    0.0000
    Residual |  56814.4622       201  282.659016   R-squared       =    0.1937
-------------+----------------------------------   Adj R-squared   =    0.1897
       Total |  70462.0985       202   348.82227   Root MSE        =    16.812
```

difsc	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	16.54575	2.381163	6.95	0.000	11.85049	21.24102
_cons	-16.44348	1.567771	-10.49	0.000	-19.53487	-13.35209

```
114 . reg difsc program if pastment==1 // coefficient:  3.556548, significant at  0.0042 level
```

Source	SS	df	MS	Number of obs	=	37
				F(1, 35)	=	4.46
Model	114.866876	1	114.866876	Prob > F	=	0.0420
Residual	901.889881	35	25.7682823	R-squared	=	0.1130
				Adj R-squared	=	0.0876
Total	1016.75676	36	28.2432432	Root MSE	=	5.0762

difsc	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
program	3.556548	1.684511	2.11	0.042	.1368076	6.976288
_cons	5.0625	1.269062	3.99	0.000	2.486168	7.638832

```
115 . ** DiD coding regression
116 . reg score2 after program after_program black hispanic kennedy i.classize if pastment==0 // 16.54575, signi
    > ficant at 0.0000 level
```

Source	SS	df	MS	Number of obs	=	406
				F(8, 397)	=	16.48
Model	24029.7796	8	3003.72245	Prob > F	=	0.0000
Residual	72363.58	397	182.27602	R-squared	=	0.2493
				Adj R-squared	=	0.2342
Total	96393.3596	405	238.008295	Root MSE	=	13.501

score2	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
after	-16.44348	1.780454	-9.24	0.000	-19.94378	-12.94318
program	.5741892	1.940148	0.30	0.767	-3.24006	4.388438
after_program	16.54575	2.704192	6.12	0.000	11.22942	21.86208
black	-.1393244	2.557096	-0.05	0.957	-5.166467	4.887818
hispanic	.6788636	2.50152	0.27	0.786	-4.239019	5.596746
kennedy	1.648917	2.021638	0.82	0.415	-2.325537	5.623371
classize						
25	-.5383731	1.665564	-0.32	0.747	-3.812801	2.736055
30	-1.03165	2.888734	-0.36	0.721	-6.710779	4.647479
_cons	89.3847	2.750033	32.50	0.000	83.97825	94.79114

```
117 . outreg2 using Assignment3Table5,excel ctitle(no past mentor experience,program) append
    Assignment3Table5.xml
    dir : seeout

118 . reg score2 after program after_program black hispanic kennedy i.classize if pastment==1 // 3.556548, signi
    > ficant at 0.0000 level
```

Source	SS	df	MS	Number of obs	=	74
				F(8, 65)	=	6.15
Model	1290.16313	8	161.270391	Prob > F	=	0.0000
Residual	1704.43147	65	26.2220226	R-squared	=	0.4308
				Adj R-squared	=	0.3608
Total	2994.59459	73	41.0218438	Root MSE	=	5.1207

score2	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
after	5.0625	1.810457	2.80	0.007	1.446769	8.678231
program	1.410543	1.755583	0.80	0.425	-2.095598	4.916683
after_program	3.556548	2.403141	1.48	0.144	-1.242857	8.355952
black	-1.878935	2.904909	-0.65	0.520	-7.680439	3.922568
hispanic	.383234	2.752362	0.14	0.890	-5.113613	5.880081
kennedy	2.070283	1.845531	1.12	0.266	-1.615498	5.756063
classize						
25	-.5151032	1.480516	-0.35	0.729	-3.471899	2.441692
30	1.518714	2.889953	0.53	0.601	-4.25292	7.290348
_cons	87.8543	3.176059	27.66	0.000	81.51128	94.19733

```
119 . outreg2 using Assignment3Table5,excel ctitle(have past mentor experience,program) append
    Assignment3Table5.xml
    dir : seeout

120 .
121 . * Set B: high absences at baseline vs. low absences at baseline
122 . ** generate subgroup dummy
123 . codebook pastab
```

pastab — **days abs 3 mo. before app.**

```
                Type: Numeric (float)

               Range: [0,51]                        Units: 1
      Unique values: 46                         Missing .: 0/480

                Mean: 16.8625
           Std. dev.: 12.0563

       Percentiles:     10%       25%       50%       75%       90%
                          0         7        17      25.5        32
```

```
124 . gen pastablvl=.
    (480 missing values generated)

125 . replace pastablvl= 0 if pastab <= 16.8625
    (234 real changes made)

126 . replace pastablvl= 1 if pastab > 16.8625
    (246 real changes made)

127 . label variable pastablvl "absence level at baseline"

128 . ** regression
129 . reg difsc program if pastablvl==0 // coefficient: 9.484127, significant at 0.0000 level

      Source |       SS           df       MS      Number of obs   =       117
-------------+----------------------------------   F(1, 115)       =     23.03
       Model |  2615.4304         1   2615.4304   Prob > F        =     0.0000
    Residual |  13061.1508       115  113.575224   R-squared       =     0.1668
-------------+----------------------------------   Adj R-squared   =     0.1596
       Total |  15676.5812       116  135.142941   Root MSE        =     10.657

------------------------------------------------------------------------------
       difsc | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     program |   9.484127   1.976367     4.80   0.000     5.569324    13.39893
       _cons |  -4.277778   1.450257    -2.95   0.004    -7.150457   -1.405098
------------------------------------------------------------------------------

130 . reg difsc program if pastablvl==1 // coefficient: 17.50649, significant at 0.0000 level

      Source |       SS           df       MS      Number of obs   =       123
-------------+----------------------------------   F(1, 121)       =     25.29
       Model |  8825.54999         1  8825.54999   Prob > F        =     0.0000
    Residual |  42231.2468       121  349.018568   R-squared       =     0.1729
-------------+----------------------------------   Adj R-squared   =     0.1660
       Total |  51056.7967       122  418.498334   Root MSE        =     18.682

------------------------------------------------------------------------------
       difsc | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     program |   17.50649   3.481389     5.03   0.000     10.61417    24.39882
       _cons |  -20.50649   2.129016    -9.63   0.000    -24.72144   -16.29154
------------------------------------------------------------------------------

131 . reg score2 after program after_program black hispanic kennedy i.classize if pastablvl==0 // 9.484127

      Source |       SS           df       MS      Number of obs   =       234
-------------+----------------------------------   F(8, 225)       =      7.19
       Model |  4848.70515         8  606.088144   Prob > F        =     0.0000
    Residual |  18979.5726       225  84.3536561   R-squared       =     0.2035
-------------+----------------------------------   Adj R-squared   =     0.1752
       Total |  23828.2778       233  102.267287   Root MSE        =     9.1844
```

score2	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
after	-4.277778	1.767543	-2.42	0.016	-7.760834	-.7947216
program	1.810125	1.724238	1.05	0.295	-1.587596	5.207846
after_program	9.484127	2.408756	3.94	0.000	4.737521	14.23073
black	5.091842	2.696523	1.89	0.060	-.2218281	10.40551
hispanic	4.37278	2.660869	1.64	0.102	-.8706314	9.616191
kennedy	2.248808	1.865961	1.21	0.229	-1.428186	5.925802
classize						
25	1.729567	1.442837	1.20	0.232	-1.113634	4.572769
30	3.573678	2.91731	1.22	0.222	-2.175066	9.322423
_cons	82.61947	3.015974	27.39	0.000	76.6763	88.56263

132 . outreg2 using Assignment3Table5,excel ctitle(low absences at baseline,program) append
Assignment3Table5.xml
dir : seeout

133 . reg score2 after program after_program black hispanic kennedy i.classize if pastablvl==1 // 17.50649

Source	SS	df	MS	Number of obs	=	246
				F(8, 237)	=	12.78
Model	21887.4823	8	2735.93529	Prob > F	=	0.0000
Residual	50719.1884	237	214.005015	R-squared	=	0.3015
				Adj R-squared	=	0.2779
Total	72606.6707	245	296.353758	Root MSE	=	14.629

score2	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
after	-20.50649	2.357662	-8.70	0.000	-25.15114	-15.86184
program	-.3287982	2.769203	-0.12	0.906	-5.784195	5.126598
after_program	17.50649	3.855274	4.54	0.000	9.911511	25.10148
black	-4.734601	3.362698	-1.41	0.160	-11.3592	1.889996
hispanic	-1.644041	3.265726	-0.50	0.615	-8.077601	4.789518
kennedy	2.931612	2.715961	1.08	0.282	-2.418897	8.282121
classize						
25	-2.029109	2.422615	-0.84	0.403	-6.801718	2.743501
30	-2.802771	3.891287	-0.72	0.472	-10.4687	4.863158
_cons	93.31141	3.535883	26.39	0.000	86.34564	100.2772

134 . outreg2 using Assignment3Table5,excel ctitle(high absences at baseline,program) append
Assignment3Table5.xml
dir : seeout

135 . help label

136 . capture log close