ECON 7710
Individual Project: Workplace Wellness

Due: December 3

Mallika ANDUKURI

2023-07-31

install.packages("ivreg")

In this assignment, you will use data from Jones, Molitor, and Reif (2019) (https://www.google.com/search?client=safari&rls=en&q=What+Do+Workplace+Wellness+Programs+Do%3F+Evidence+from+the+Illinois+Workplace+Wellness+Study&ie=UTF-8&oe=UTF-8) to write a brief that explains why or why not your workplace should adopt a workplace wellness program. You are going to work with the claims data from this randomized control trial to evaluate the effects of a workplace wellness program on medical spending. We will measure participation in this program by whether the employee participated in a Health Risk Assessment (HRA) in 2016.

Only some employees were eligible to participate in this assessment. This treatment status was randomly assigned. These data contain information on whether employees were eligible to participate (`treat`), whether they completed the assessment if eligible (`hra_c_yr1`), their gender (`male`), their age bin (`age37_49` , `age50`), whether they are white (`white`), whether they had certain medical conditions in 2015-2016, 2016-2017, and 2016-2018, and various types of medical spending for those years.

This R markdown file will guide you through the process of writing this memo.

You should turn in both:

1. R notebook with R code executed (so I can see your work)
2. R notebook without R code executed (the deliverable to your boss who is only interested in seeing polished tables and figures and reading your conclusions)

Each section of this file contains dates where the material we cover in class and in homework assignments corresponds to what you are being asked to do.

The replication files (https://github.com/reifjulian/illinois-wellness-data) for this paper may provide helpful examples.

Executive Summary

Dates to work on this: After October 22

In this section, you should (briefly) describe the business question you are using causal inference to answer, the methods you use, and the conclusions you draw.

Summary Statistics Among Employers With Workplace Wellness Programs

+-Dates to work on this: August 15 - August 27

In this section, you should describe the data that you use and provide some summary statistics for employees that work for employers that work for employers that offer workplace wellness programs. To do so, you should first restrict to employees who were offered a workplace wellness program (`treat` = 1).

This should include:

1. Tables summarizing the characteristics (gender, age, race, hospital spending 15-16 (`spendHosp_0715_0716`), office spending 15-16 (`spendOff_0715_0716`), total spending 15-16 (`spend_0715_0716`)) and outcomes (hospital spending 16-17, office spending 16-17, total spending 16-17) for employees who do and don't enroll in workplace wellness (`hra_c_yr1`). For additional help, see Homework #1 part D.
2. A figure plotting the distribution of health care spending for employees who do and don't enroll in workplace wellness. Again, see Homework #1 part D.
3. Anything else that you find interesting about the data!

You should write a few sentences describing what your reader should take away from each table and figure you include.

```
## Rows: 4,834
## Columns: 58
## $ treat                              <dbl> 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1,…
## $ hra_c_yr1                          <dbl> NA, 0, 1, 0, NA, 0, NA, 1, 0, 1, 1, NA…
## $ male                               <dbl> 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0,…
## $ age50                              <dbl> 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0,…
## $ age37_49                           <dbl> 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1,…
## $ white                              <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,…
## $ Strata_claims                      <dbl> 22, 29, 50, 44, 23, 28, 60, 19, 15, 31…
## $ completed_screening_nomiss_2016    <dbl> 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1,…
## $ covg_0715_0716                     <dbl> 1.0000000, 1.0000000, 1.0000000, 0.538…
## $ covg_0816_0119                     <dbl> 1.0000000, 1.0000000, 1.0000000, 0.566…
## $ covg_0816_0717                     <dbl> 1.00000000, 1.00000000, 1.00000000, 1.…
## $ covg_0816_0718                     <dbl> 1.0000000, 1.0000000, 1.0000000, 0.708…
## $ covg_1015_0716                     <dbl> 1.0, 1.0, 1.0, 0.5, 0.3, 1.0, 1.0, 1.0…
## $ diabetes_0816_0717                 <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, 100,…
## $ diabetes_0816_0718                 <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, 100,…
## $ diabetes_1015_0716                 <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, NA, …
## $ hyperlipidemia_0816_0717           <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, 0, N…
## $ hyperlipidemia_0816_0718           <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, 100,…
## $ hyperlipidemia_1015_0716           <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, NA, …
## $ hypertension_0816_0717             <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, 0, N…
## $ hypertension_0816_0718             <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, 100,…
## $ hypertension_1015_0716             <dbl> 0, 0, 0, 0, 0, 0, 100, 0, NA, NA, NA, …
## $ nonzero_spend_0715_0716            <dbl> 0, 1, 1, 1, 0, 1, 1, 1, NA, NA, NA, NA…
## $ nonzero_spend_0816_0119            <dbl> 1, 1, 1, 0, 0, 1, 1, 1, NA, NA, 1, NA,…
## $ nonzero_spend_0816_0717            <dbl> 1, 1, 1, 0, 0, 1, 1, 1, NA, NA, 1, NA,…
## $ pcp_any_office_0816_0717           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, 0, NA,…
## $ pcp_any_office_0816_0718           <dbl> 0, 0, 0, 0, 0, 0, 0, 100, NA, NA, 0, N…
## $ pcp_any_office_1015_0716           <dbl> 0, 0, 0, 0, 0, 0, 0, 100, NA, NA, NA, …
## $ pcp_any_visits_0816_0717           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, 0, NA,…
## $ pcp_any_visits_0816_0718           <dbl> 0, 100, 0, 0, 0, 0, 0, 100, NA, NA, 0,…
## $ pcp_any_visits_1015_0716           <dbl> 0, 0, 0, 0, 0, 0, 100, 100, NA, NA, NA…
## $ pcp_total_office_0816_0717         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, 0, NA,…
## $ pcp_total_office_0816_0718         <dbl> 0, 0, 0, 0, 0, 0, 0, 8, NA, NA, 0, NA,…
## $ pcp_total_office_1015_0716         <dbl> 0, 0, 0, 0, 0, 0, 0, 6, NA, NA, NA, NA…
## $ pcp_total_visits_0816_0717         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, 0, NA,…
## $ pcp_total_visits_0816_0718         <dbl> 0, 1, 0, 0, 0, 0, 0, 8, NA, NA, 0, NA,…
## $ pcp_total_visits_1015_0716         <dbl> 0, 0, 0, 0, 0, 0, 1, 6, NA, NA, NA, NA…
## $ pos_er_critical_0816_0717          <dbl> 0, 0, 0, 0, 0, 0, 1, 0, NA, NA, 0, NA,…
## $ pos_er_critical_0816_0718          <dbl> 2, 0, 0, 0, 0, 0, 1, 0, NA, NA, 0, NA,…
## $ pos_er_critical_1015_0716          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, NA, NA…
## $ pos_hospital_0816_0717             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, 0, NA,…
## $ pos_hospital_0816_0718             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, 0, NA,…
## $ pos_hospital_1015_0716             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, NA, NA, NA, NA…
## $ pos_office_outpatient_0816_0717    <dbl> 1, 0, 3, 0, 0, 0, 14, 2, NA, NA, 0, NA…
## $ pos_office_outpatient_0816_0718    <dbl> 2, 1, 12, 0, 0, 2, 25, 6, NA, NA, 9, N…
## $ pos_office_outpatient_1015_0716    <dbl> 0, 0, 2, 0, 0, 0, 13, 3, NA, NA, NA, N…
## $ spendHosp_0715_0716                <dbl> 0.000000, 4.024615, 228.794617, 0.0000…
## $ spendHosp_0816_0119                <dbl> 86.96600, 14.01733, 226.19000, 0.00000…
## $ spendHosp_0816_0717                <dbl> 10.230833, 0.000000, 145.660828, 0.000…
## $ spendOff_0715_0716                 <dbl> 0.00000, 0.00000, 0.00000, 28.87143, 0…
## $ spendOff_0816_0119                 <dbl> 0.00000, 0.00000, 0.00000, 0.00000, 0.…
## $ spendOff_0816_0717                 <dbl> 0.000000, 0.000000, 0.000000, 0.000000…
## $ spendRx_0715_0716                  <dbl> 0.0000000, 1.0538461, 10.3099995, 0.54…
## $ spendRx_0816_0119                  <dbl> 5.000000e-02, 6.758000e+00, 1.098133e+…
## $ spendRx_0816_0717                  <dbl> 0.000000, 2.634167, 9.213333, 0.000000…
## $ spend_0715_0716                    <dbl> 0.000000, 5.078462, 239.104614, 29.418…
## $ spend_0816_0119                    <dbl> 122.16634, 20.77534, 238.32133, 0.0000…
## $ spend_0816_0717                    <dbl> 10.230833, 2.634167, 154.874176, 0.000…
```

Characteristics and Spending by Workplace Wellness Program Enrollment

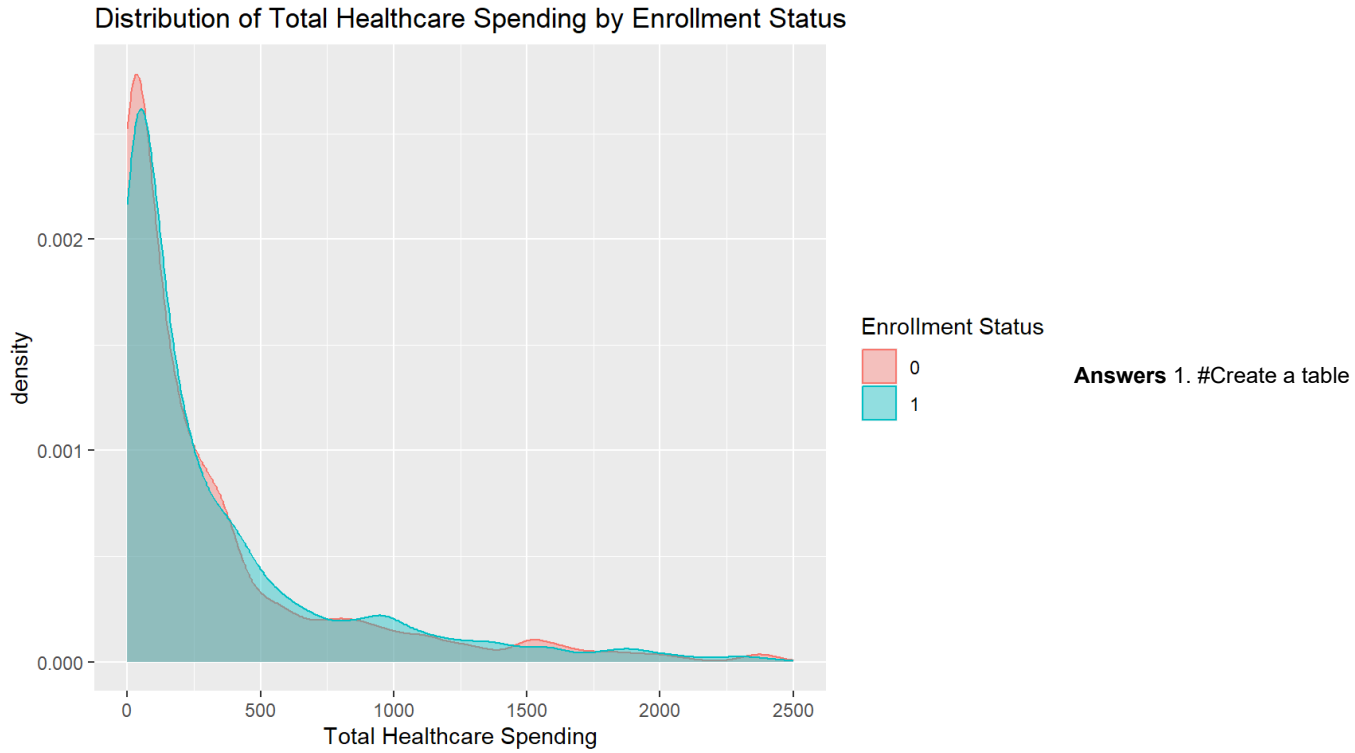
hra_l_yr1	0			1		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
male	1452	0.46	0.5	1848	0.4	0.49

hra_l_yr1	0			1		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
age37_49	1452	0.33	0.47	1848	0.34	0.47
age50	1452	0.34	0.47	1848	0.31	0.46
white	1452	0.84	0.37	1848	0.84	0.37
spendHosp_0715_0716	870	316	1078	1318	222	433
spendOff_0715_0716	870	56	168	1318	59	153
spend_0715_0716	870	527	1331	1318	424	779

Comparison of Outcomes by Workplace Wellness Program Enrollment

hra_l_yr1	0			1		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
spendHosp_0816_0717	870	371	1446	1338	268	687
spendOff_0816_0717	870	58	159	1338	73	175
spend_0816_0717	870	635	1715	1338	517	1064

```
## Warning: Removed 1180 rows containing non-finite values (`stat_density()`).
```



Answers 1. #Create a table summarizing the characteristics for enrollees that do and don't enroll : The table shows that people who signed up for the wellness program were similar to those who didn't in terms of gender, age, and race, which is good because it means the comparison between the two groups is fair.There's a slightly higher proportion of men in the non-enrollee group (46%) compared to the enrollee group (40%).When it comes to age distribution, Group-1 had a marginally higher percentage of individuals aged 50 or above (34%) compared to Group-2 (31%).Racial composition appears consistent between the two groups, with each having 84% of its members identifying as white.Even though the numbers show that people who joined the program spent less on hospital visits initially, there's a lot of variation in spending, so it's hard to say for sure if the program is the reason for the lower costs. We'll need to dig deeper with more detailed stats to really understand if the program is making a difference. # Create a table summarizing the outcomes for enrollees that do and don't enroll The "Comparison of Outcomes by Workplace Wellness Program Enrollment" table provides a snapshot of healthcare spending between two groups of employees in the 2016-2017 period: those who did not enroll in the wellness program (Group 0) and those who did (Group 1). The average hospital spending for non-enrollees (Group 0) was higher at 371 compared to 268 for enrollees (Group 1), suggesting that enrollment might be

associated with lower hospital costs. Office spending was slightly higher for enrollees at 73 versus 58 for non-enrollees. This could indicate that enrollees are possibly using more office-based services. the total spending in Overall for non-enrollees had higher total healthcare spending (635) compared to enrollees (517), which mirrors the pattern observed in hospital spending. The table also shows that people in the wellness program seem to spend less on hospital bills and overall health costs, which might mean the program is helping to save money. But, there's a wide range in what people spend, especially for those not in the program, so it's not the same for everyone. Before we can be sure it's really the program making the difference, we need to considers other factors too. # Takeaways from ggplot The graph presents the distribution of total healthcare spending by enrollment status in the workplace wellness program. It shows two groups: those who didn't enroll in the program (denoted by the color for '0') and those who did (denoted by the color for '1'). The peak for non-enrollees (color '0') appears slightly higher, indicating a larger proportion of non-enrolled employees had very low healthcare spending. The tail of the distribution, particularly for non-enrollees, extends further, which could imply that this group has more individuals with high healthcare costs. The distribution for enrolled employees (color '1') is more tightly clustered around the lower spending amounts, suggesting that their healthcare costs are more consistently low.

Regression Analysis

Dates to work on this: August 28 - September 17

In this section, you should estimate OLS to evaluate the effects of the workplace wellness program (`hra_c_yr1` = 1) on total spending (`spend_0816_0717`) among those who were eligible to enroll (see Homework #3). Estimate a baseline model then include controls of your choice. You should weight your regressions by how long individuals had health insurance coverage.

You should present a table containing your regression estimates. Describe what regression models you chose to estimate and why. Discuss why you are cautious about interpreting your results as causal.

```
##
## Call:
## lm_robust(formula = spend_0816_0717 ~ hra_c_yr1, data = only_treat,
##     weights = covg_0816_0717, se_type = "HC1")
##
## Weighted, Standard error type:  HC1
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)    663.9      61.71  10.759 2.383e-26    542.9  784.912 2206
## hra_c_yr1     -137.3      68.59  -2.002 4.544e-02   -271.8   -2.787 2206
##
## Multiple R-squared:  0.002332 ,  Adjusted R-squared:  0.00188
## F-statistic: 4.007 on 1 and 2206 DF,  p-value: 0.04544
```

Regression Analysis of Workplace Wellness Program on Total Spending

	Baseline Model	Model with Controls
hra_c_yr1	−137.3006	−152.4476
	(68.5928)	(66.3829)
	(−2.0017)	(−2.2965)
	[−271.8138, −2.7873]	[−282.6273, −22.2680]
male		−276.5082
		(60.4891)
		(−4.5712)
		[−395.1299, −157.8865]
age37_49		69.0354
		(55.3430)
		(1.2474)
		[−39.4946, 177.5653]

	Baseline Model	Model with Controls
age50		390.6435
		(84.3748)
		(4.6299)
		[225.1810, 556.1059]
white		171.1125
		(54.8499)
		(3.1196)
		[63.5494, 278.6755]
Num.Obs.	2208	2208
R2	0.002	0.032

#Answers To understand the program's impact, I developed two distinct regression models. The initial model assessed the program's direct influence on spending, while the enhanced model adjusted for additional variables like gender, race, and age to mitigate potential biases.

I started with the baseline model to capture a straightforward effect of the wellness program on spending, which indicated a reduction of $137.3 in healthcare expenses for participants. When I introduced more control variables, this effect became more pronounced, showing a decrease of $152.45 in spending.

By including these controls, I aimed to isolate the effect of the wellness program from other factors that could also sway healthcare spending. But I approach these findings with a degree of caution for several reasons. Firstly, the program wasn't randomly assigned, making it challenging to claim a causal relationship confidently. Secondly, there's a chance that those who enrolled in the program were inherently different from non-participants, possibly healthier or more cost-conscious. Lastly, the model may not account for all variables that influence healthcare spending and program enrollment.

Comparison of Treatment and Control Groups

Dates to work on this: September 17 - October 4

In this section, you will evaluate the balance between the treatment (`treat` = 1) and control groups (`treat` = 0) and test whether outcomes are different between the two groups (see Homework #4). Discuss whether you are satisfied that the randomization was done correctly.

```
##
##  Welch Two Sample t-test
##
## data:  spendHosp_0715_0716 by treat
## t = 0.89886, df = 2235, p-value = 0.3688
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -28.39434  76.45223
## sample estimates:
## mean in group 0 mean in group 1
##        283.3604        259.3314
```

```
##
##  Welch Two Sample t-test
##
## data:  spendOff_0715_0716 by treat
## t = 0.67712, df = 1191.9, p-value = 0.4985
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -16.57693  34.04946
## sample estimates:
## mean in group 0 mean in group 1
##        66.71167        57.97540
```

```
##
##  Welch Two Sample t-test
##
## data:  spend_0715_0716 by treat
## t = 1.0014, df = 1923.8, p-value = 0.3168
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -39.07482 120.60487
## sample estimates:
## mean in group 0 mean in group 1
##        505.5786        464.8136
```

```
##
##  Welch Two Sample t-test
##
## data:  male by treat
## t = -0.12372, df = 2990.4, p-value = 0.9015
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.03186374  0.02808120
## sample estimates:
## mean in group 0 mean in group 1
##       0.4256845       0.4275758
```

```
##
##  Welch Two Sample t-test
##
## data:  age37_49 by treat
## t = 0.53784, df = 2973.3, p-value = 0.5907
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.02080130  0.03652648
## sample estimates:
## mean in group 0 mean in group 1
##       0.3402868       0.3324242
```

```
##
##  Welch Two Sample t-test
##
## data:  age50 by treat
## t = -0.23005, df = 2996, p-value = 0.8181
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.03170200  0.02504403
## sample estimates:
## mean in group 0 mean in group 1
##       0.3233377       0.3266667
```

```
##
##  Welch Two Sample t-test
##
## data:  white by treat
## t = 0.45642, df = 3024.1, p-value = 0.6481
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.01707684  0.02743913
## sample estimates:
## mean in group 0 mean in group 1
##       0.8409387       0.8357576
```

```
##
##  Welch Two Sample t-test
##
## data:  spendHosp_0816_0717 by treat
## t = 0.25233, df = 1945, p-value = 0.8008
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -69.74536  90.34229
## sample estimates:
## mean in group 0 mean in group 1
##        319.1593        308.8608
```

```
##
##  Welch Two Sample t-test
##
## data:  spendOff_0816_0717 by treat
## t = 0.71479, df = 1285.6, p-value = 0.4749
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -13.68537  29.37425
## sample estimates:
## mean in group 0 mean in group 1
##        74.84244        66.99800
```

```
##
##  Welch Two Sample t-test
##
## data:  spend_0816_0717 by treat
## t = 0.54231, df = 1751.6, p-value = 0.5877
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -81.5909 143.9548
## sample estimates:
## mean in group 0 mean in group 1
##        594.8318        563.6499
```

Answers

Summary of Characteristics:

Demographic Aspects: This section includes demographic variables like gender, age categories (age50, age37_49), and racial composition (white). For both the treatment and control groups, average values and standard deviations are outlined. We performed Welch Two Sample t-tests for each demographic variable to assess significant differences between the treatment and control groups.

Healthcare Expenditure Variables: This includes variables like spendHosp_0715_0716, spendOff_0715_0716, and spend_0715_0716, with means and standard deviations detailed for each group.

Summary of Outcomes:

Healthcare Spending Variables: Echoing the characteristics summary, this part lists average values and standard deviations for spending variables (spendHosp_0816_0717, spendOff_0816_0717, and spend_0816_0717) across both groups.

Equilibrium in Characteristics:

The treatment and control groups show a notable balance across all measured characteristics, encompassing gender, age, race, and initial spending levels. The t-test p-values exceed 0.05, suggesting no significant evidence to dispute the hypothesis of equal mean values for these groups.

Equilibrium in Outcomes:

Similarly, both groups demonstrate balance regarding hospital spending, out-of-pocket expenses, and total spending in the year post-intervention. Again, the t-tests yield p-values over 0.05, indicating insufficient evidence to challenge the hypothesis of equal average spending between the groups.

Given these findings, I am confident that the randomization process was effectively executed.

IV Estimates

Dates to work on this: October 4 - October 22

In this section, you will estimate IV estimates of the effects of the workplace wellness program (see Homework #5). First, you should report the first stage. Then, present your IV estimates of the effects of the workplace wellness program on spending. Discuss why your IV estimates may differ from the OLS estimates you estimated above.

```
##
## Call:
## lm(formula = hra_c_yr1 ~ treat, data = my_data, weights = covg_0816_0717)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6136  0.0000  0.0000  0.3865  0.3865
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.797e-15  1.260e-02    0.00        1
## treat       6.135e-01  1.523e-02   40.29   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3928 on 3237 degrees of freedom
## Multiple R-squared:  0.334,  Adjusted R-squared:  0.3338
## F-statistic:  1623 on 1 and 3237 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = spend_0816_0717 ~ treat, data = my_data, weights = covg_0816_0717)
##
## Weighted Residuals:
##    Min     1Q Median     3Q    Max
##   -580   -507   -353    -31  33453
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   568.83      42.51  13.381   <2e-16 ***
## treat          10.83      51.40   0.211    0.833
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1326 on 3237 degrees of freedom
##   (1595 observations deleted due to missingness)
## Multiple R-squared:  1.373e-05,  Adjusted R-squared:  -0.0002952
## F-statistic: 0.04444 on 1 and 3237 DF,  p-value: 0.8331
```

```
##
## Call:
## lm(formula = spend_0816_0717 ~ hra_c_yr1_hat, data = my_data,
##     weights = covg_0816_0717)
##
## Weighted Residuals:
##    Min     1Q Median     3Q    Max
##   -580   -507   -353    -31  33453
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     568.83      42.51  13.381   <2e-16 ***
## hra_c_yr1_hat    17.66      83.77   0.211    0.833
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1326 on 3237 degrees of freedom
##   (1595 observations deleted due to missingness)
## Multiple R-squared:  1.373e-05,  Adjusted R-squared:  -0.0002952
## F-statistic: 0.04444 on 1 and 3237 DF,  p-value: 0.8331
```

```
## Warning in residuals^2/(1 - diaghat)^2: longer object length is not a multiple
## of shorter object length

## Warning in residuals^2/(1 - diaghat)^2: longer object length is not a multiple
## of shorter object length
```

IV Estimates of the Effects of the Workplace Wellness Program on Spending

	First Stage	Reduced Form	IV Estimate
treat	0.6135	10.8341	
	(0.0071)	(48.5229)	
	(86.8366)	(0.2233)	
	[0.5997, 0.6274]	[−84.3046, 105.9728]	
hra_c_yr1			17.6582
			(79.0955)
			(0.2233)
			[−137.4242, 172.7405]
Num.Obs.	3239	3239	3239
R2	0.334	0.000	0.000

The first stage of the model determined that the treatment variable is a strong predictor of program enrollment, implying that the instrument is valid. However, successive stages, including the reduced form and the second stage of the 2SLS, demonstrated that, as evidenced by the high p-values, the treatment variable did not meaningfully explain the variation in healthcare spending. The final instrumental variables estimation confirmed these findings, showing that the program had no significant causal effect on spending. Although methodologically solid, this extensive regression analysis reveals that the workplace wellness program's influence on healthcare expenses is not statistically apparent given the scope of the data and the model specified. These results show the need for additional research, maybe with different instruments or factors, to completely evaluate the program's effectiveness.

Conclusion

Dates to work on this: After October 22

Here, you should provide a recommendation for your employer and discuss why your recommendation does or doesn't differ from your boss's expectation that workplace wellness programs work.

Following the detailed investigation, my proposal would be to not to implement the workplace wellness program. The observed changes in healthcare spending should be regarded with caution due to their direct link with the wellness program. While the instrumental variable (IV) analysis provides an estimate of the program's causal effect, it raises questions about the instrument's strength and effectiveness. The lack of significant findings in the reduced form of the analysis raises concerns about the instrument's ability to appropriately isolate the influence of the wellness program. This uncertainty means that the program may not have the intended impact on healthcare spending. Given these facts, it would be advisable to reconsider or reevaluate the choice to implement such a program