FIN 550: Big Data Analytics

Problem Set #1- Case Executive Summary

■ Group Submission. List group member names: Hsiang Lee, Chu-Chun Ku, Yi-Ting Lee

(I) a concise overview of the case and objectives

- Overview: This case explores the impact of workplace wellness programs on employee healthcare spending and other outcomes. Through random assignment, employees were divided into treatment and control groups to assess the causal effect of participating in a wellness program.
- Objectives: The objective is to estimate the impact of the wellness program on healthcare spending by comparing the outcomes between the treatment and control groups. Additionally, the study examines the differences between participants and non-participants within the treatment group, accounting for demographic factors such as gender, race, and age to reduce selection bias.

(III) a conclusion that draws lessons for employers who may be considering adopting a wellness program in their workplace.

The results show that wellness programs may not greatly reduce healthcare spending in the short term. Therefore, employers should have realistic expectations when adopting these programs. However, wellness programs can still improve employee health and well-being, possibly leading to larger benefits in the long term. To get the most out of these programs, employers should offer ongoing participation opportunities and resources so that all employees can keep engaging in these activities.

(II) answers to each questions asked in the Case Description above

1. To make this comparison reflect a causal effect, we must ensure that, aside from the wellness program, there are no other differences between companies with wellness programs and those without that could impact healthcare spending. For example, these companies should be similar in most factors, such as employee composition, job functions, healthcare policies, company size, funding, industry type, etc., so that any changes in healthcare spending can be attributed to the workplace wellness program rather than

other factors. However, it’s often hard to fully meet this condition because companies with wellness programs usually offer more employee benefits, like gym access and mental health support. Employees who work at these companies may also care more about health, which could mean their health status and healthcare spending are already different from those at companies without wellness programs. Therefore, just comparing companies with and without wellness programs may lead to inaccurate results due to selection bias.

2. To make this comparison reflect a causal effect, we must ensure that, aside from the wellness program, there are no other differences that could impact the results. However, this condition is often difficult to fully satisfy because eligible employees may have higher positions, have been with the company longer, or care more about health. Therefore, eligible employees may already be different from ineligible employees in terms of health status, lifestyle, and other factors. This could make the results look more positive for the wellness program, but the effect might be due to selection bias rather than the program itself.
3. The "treatment group" means employees who were randomly chosen to be eligible for the wellness program, while the "control group" includes employees who were randomly chosen not to join the wellness program. This is different from "participant" and "non-participant," which are based on personal choice and can lead to bias. In this study, the treatment group had 3,300 employees, and the control group had 1,534 employees. Out of those in the treatment group, 1,848 employees participated in the initial (screening) segment of the wellness program in the first year.
4.

Description: df [4 × 3]

	Control <dbl>	Treatment <dbl>	p_value <dbl>
spend_0715_0716	505.58	464.81	0.31
spendRx_0715_0716	103.37	101.27	0.91
spendOff_0715_0716	66.71	57.98	0.38
spendHosp_0715_0716	283.36	259.33	0.39

The p-value results show no significant difference in spending between the control and treatment groups before randomization. This results suggest that randomization effectively balanced the groups and reduced potential selection bias.

5.

Description: df [4 × 3]

Outcome <chr>	No.controls <chr>	Demographic.controls <chr>
spend_0816_0717	-31.18 (54.26)	-24.29 (53.77)
spendRx_0816_0717	-10.42 (24.63)	-9.09 (24.63)
spendOff_0816_0717	-7.84 (8.83)	-7.61 (8.8)
spendHosp_0816_0717	-10.3 (40.26)	-5.73 (39.96)

4 rows

The similarity in first-year medical spending between the treatment and control groups suggests that wellness programs did not have a significant impact on these costs within the year. Since the treatment and control groups were already similar in baseline demographic variables, adding demographic controls is unlikely to substantially change the main estimates. However, including these controls may make the estimates slightly more precise by further accounting for any minor differences and reducing the margin of error.

6.

Description: df [4 × 3]

Outcome <chr>	No.controls <chr>	Demographic.controls <chr>
spend_0816_0717	-117.16 (59.13)	-135.32 (58.56)
spendRx_0816_0717	-26.63 (26.84)	-26.31 (26.94)
spendOff_0816_0717	14.96 (7.37)	11.47 (7.34)
spendHosp_0816_0717	-103.05 (45.87)	-113.97 (45.64)

4 rows

We use the variable "hra_c_yr1" to distinguish between participants and non-participants, and we control for demographic variables in our regression analysis. From the results, we observe that the estimated treatment effect does not change much after adding demographic controls, which indicates that the influence of variables such as sex, race, and age on the outcomes is relatively small. Therefore, we should not expect a significant difference between column 2 and column 3.