Assignment 2

YOUR NAME HERE

Invalid Date

Questions

1. Suppose that we are interested in the effect of a medical treatment on an outcome y. The code below creates potential outcome with treatment y_1 and the potential outcome y_0 without treatment for 1000 fictional people.

Plot the kernel density estimate of y_0 and y_1 on the same plot. On your plot, ensure that the two density estimates have different colours, the axes are labelled intuitively, and the plot has a title.

```
potential <- tibble(
  y0 = 6 + rnorm(1000,0,3),
  y1 = 8 + rnorm(1000,0,1)
)
```

2. Pretend that the doctor who assigns people to treatment is evil, and puts people in the treatment group only if they don't benefit from it. Add the treatment variable $w = 1$ for the treatment group and $w = 0$ for the control to the dataset based on this scenario, and compute the Average Treatment Effect on the Treated (ATT), the Average Treatment Effect on the Non-Treated (ATNT), and the Average Treatment Effect (ATE). Comment on the differences between them an explain why you think they are different.

INSERT COMMENTS HERE

3. Compute the observed outcome $y = y_0 + (y_1 - y_0)w$ based on the actual treatment status. Estimate the regression of y on w, and report the results in a professional-looking output table (for example stargazer, but you can do it any way you like). Explain why it does not match any of the average treatment effects computed above.

INSERT COMMENTS HERE

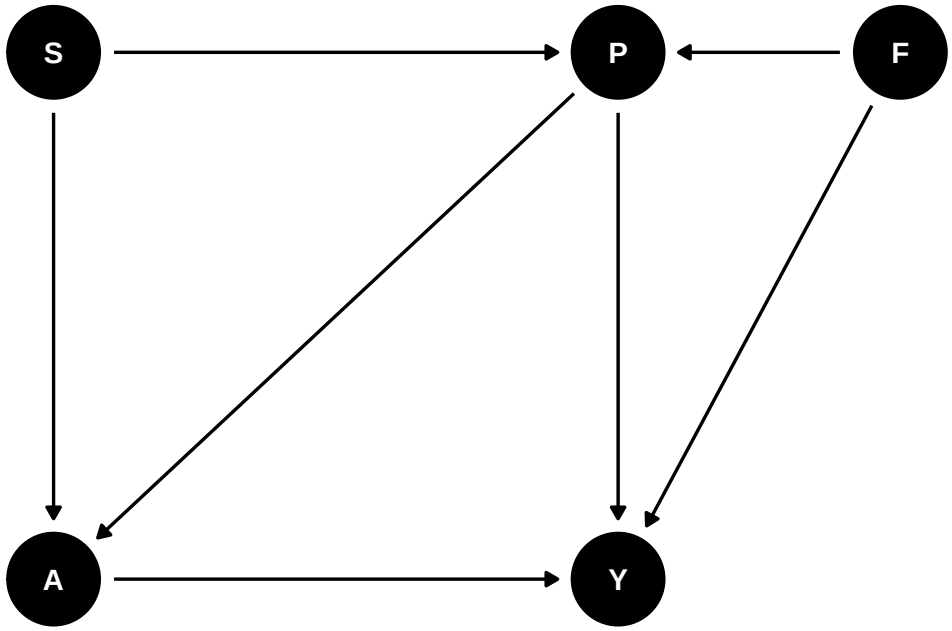
4. You are a research assistant helping an instructor on a study that tries to measure the effect of Parental Involvement (P) on Student Test Scores (Y). The DAG below relates these two

variables with other factors including Socioeconomic Status (S), Extracurricular Activities (A), and Family Stress (F) [NOTE: compile the document to see the DAG]. List all of the paths that connect P to Y, and indicate whether they are front door or back door paths.

INSERT COMMENTS HERE

```
coord_dag<-list(x = c(A = 0, S = 0, Y = 1, P = 1, F = 1.5), y = c(A = 0, S = 1, Y = 0, P =
dag <- dagify(Y~P + A + F, P ~ F + S, A~ P + S, coords = coord_dag) %>%
  tidy_dagitty()

ggplot(dag,aes(x = x, y = y, xend = xend, yend = yend)) +
  geom_dag_point() +
  geom_dag_edges() +
  geom_dag_text() +
  theme_dag()
```



5. Assume that the relationships above are all linear. Write down a linear regression model that would allow you estimate the direct effect of P on Y, and explain why the model identifies the causal effect.

INSERT COMMENTS HERE

6. You are considering having children and you want to know whether it will affect your sleep. You decide to use your econometric skills to estimate the relationship with some data. The dataset `sleep75` loaded below has information on sleep time in minutes for a set of 706

individuals, in addition to demographic and other information [for variable descriptions, load the `wooldridge` package and type `?sleep75` in the console]. Create a single professional-looking table for variables *minutes of sleep at night per week, age, gender, spousal pay, health status*, where the table has the mean of each variable separately by whether or not the person has young kids. Ensure your table has an intuitive title and variable names. Are there any notable differences in these variables?

INSERT COMMENTS HERE

```
data <- sleep75
```

7. Estimate the relationship between minutes of sleep at night per week and the presence of young kids, controlling for the remaining variables listed in the previous question. Is there any evidence that children affect sleep? When drawing these conclusions, comments on the coefficient estimate and perform an appropriate hypothesis test using robust standard errors.

INSERT COMMENTS HERE

8. Pretend that you now think the effect of young kids on sleep depends on your age. Alter your model to incorporate this new information, report the results in a professional-looking table, and test whether kids affect sleep using an appropriate hypothesis test with robust standard errors. Comment on the results.

INSERT COMMENTS HERE