

LaborEcon PS2

Boyu Chen

2023-05-25

1 Research Question and Read Data

1. Write a paragraph to introduce the research question in your term paper

The topic of my research is to examine whether distance workers have lower well-being than commuters, and whether there are significant differences in time allocation between distance workers and commuters.

2. read data

```
timeuse <- read_dta("/Users/bychen/Documents/LaborEcon/term/work/tu.dta")
```

2 Examine Data

1. Use summary() to display summary statistics for your dataset

```
summary(timeuse)
```

```
##      year      statefip      age      race      marst
## Min.   :2020   Min.    : 1.00   Min.   :16.00   Min.    :100.0   Min.    :1.000
## 1st Qu.:2020   1st Qu.:13.00   1st Qu.:34.00   1st Qu.:100.0   1st Qu.:1.000
## Median :2021   Median :28.00   Median :43.00   Median :100.0   Median :1.000
## Mean   :2021   Mean    :28.45   Mean    :44.46   Mean    :104.6   Mean    :2.922
## 3rd Qu.:2021   3rd Qu.:42.00   3rd Qu.:55.00   3rd Qu.:100.0   3rd Qu.:6.000
## Max.    :2021   Max.     :56.00   Max.     :85.00   Max.     :320.0   Max.     :6.000
##
##      hh_numkids      bls_pcare_sleep      wbladder      bls_leis_sport
## Min.   : 0.0000   Min.    :  0.0   Min.    : 0.000   Min.    :  0.0
## 1st Qu.: 0.0000   1st Qu.: 450.0   1st Qu.:  6.000   1st Qu.:  0.0
## Median : 0.0000   Median : 515.0   Median :  7.000   Median :  0.0
## Mean    : 0.7813   Mean     :527.2   Mean     : 7.286   Mean     :21.19
## 3rd Qu.: 1.0000   3rd Qu.: 600.0   3rd Qu.:  8.000   3rd Qu.:  0.0
## Max.    :12.0000   Max.     :1270.0   Max.     :10.000   Max.     :685.00
##
##      NA's      :1827
##      bls_leis_soc      empstat      occ2      ind2      earnweek
## Min.    :  0.0   Min.    :1.000   Min.    :110   Min.    :100.0   Min.    :  0
```

```

## 1st Qu.: 110.0    1st Qu.:1.000    1st Qu.:120    1st Qu.:160.0    1st Qu.: 660
## Median : 210.0    Median :1.000    Median :127    Median :220.0    Median :1058
## Mean   : 248.4    Mean   :1.042    Mean   :137    Mean   :210.1    Mean   :1266
## 3rd Qu.: 360.0    3rd Qu.:1.000    3rd Qu.:150    3rd Qu.:250.0    3rd Qu.:1731
## Max.   :1360.0    Max.   :2.000    Max.   :200    Max.   :300.0    Max.   :2885
##
##      famincome      hh_size      hh_child      citizen
## Min.   : 1.00    Min.   : 1.000    Min.   :0.0000    Min.   :1.000
## 1st Qu.:12.00    1st Qu.: 2.000    1st Qu.:0.0000    1st Qu.:1.000
## Median :14.00    Median : 2.000    Median :0.0000    Median :1.000
## Mean   :13.03    Mean   : 2.757    Mean   :0.4295    Mean   :1.579
## 3rd Qu.:15.00    3rd Qu.: 4.000    3rd Qu.:1.0000    3rd Qu.:1.000
## Max.   :16.00    Max.   :14.000    Max.   :1.0000    Max.   :5.000
##
##      clwkr      fullpart      bls_carehh      bls_educ
## Min.   :1.000    Min.   :1.000    Min.   : 0.00    Min.   : 0.000
## 1st Qu.:4.000    1st Qu.:1.000    1st Qu.: 0.00    1st Qu.: 0.000
## Median :4.000    Median :1.000    Median : 0.00    Median : 0.000
## Mean   :3.761    Mean   :1.136    Mean   : 38.32    Mean   : 7.477
## 3rd Qu.:4.000    3rd Qu.:1.000    3rd Qu.: 25.00    3rd Qu.: 0.000
## Max.   :5.000    Max.   :2.000    Max.   :799.00    Max.   :900.000
##
##      bls_food      bls_hhact      bls_leis_attend      bls_leis_relax
## Min.   : 0.00    Min.   : 0.0    Min.   : 0.000    Min.   : 0.0
## 1st Qu.: 30.00    1st Qu.: 15.0    1st Qu.: 0.000    1st Qu.: 75.0
## Median : 60.00    Median : 65.0    Median : 0.000    Median : 170.0
## Mean   : 72.42    Mean   : 116.9    Mean   : 4.504    Mean   : 207.9
## 3rd Qu.: 90.00    3rd Qu.: 170.0    3rd Qu.: 0.000    3rd Qu.: 295.0
## Max.   :780.00    Max.   :1080.0    Max.   :590.000    Max.   :1360.0
##
##      bls_leis_soccom      bls_leis_soccomex      bls_pcare_act      bls_pcare_health
## Min.   : 0.0    Min.   : 0.00    Min.   : 0.0000    Min.   : 0.000
## 1st Qu.: 0.0    1st Qu.: 0.00    1st Qu.: 0.0000    1st Qu.: 0.000
## Median : 0.0    Median : 0.00    Median : 0.0000    Median : 0.000
## Mean   : 37.3    Mean   : 32.79    Mean   : 0.7292    Mean   : 1.575
## 3rd Qu.: 30.0    3rd Qu.: 15.00    3rd Qu.: 0.0000    3rd Qu.: 0.000
## Max.   :1105.0    Max.   :1105.00    Max.   :640.0000    Max.   :1201.000
##
##      bls_purch      bls_work_travel      bls_work_working      bls_work_workrel
## Min.   : 0.00    Min.   : 0.00    Min.   : 0.0    Min.   : 0.0000
## 1st Qu.: 0.00    1st Qu.: 0.00    1st Qu.: 0.0    1st Qu.: 0.0000
## Median : 0.00    Median : 0.00    Median : 164.5    Median : 0.0000
## Mean   : 41.02    Mean   : 14.92    Mean   : 255.7    Mean   : 0.1625
## 3rd Qu.: 60.00    3rd Qu.: 20.00    3rd Qu.: 495.0    3rd Qu.: 0.0000
## Max.   :770.00    Max.   :320.00    Max.   :1311.0    Max.   :465.0000
##
##      female      distance_work
## Min.   :0.0000    Min.   :0.0000
## 1st Qu.:0.0000    1st Qu.:0.0000
## Median :0.0000    Median :0.0000
## Mean   :0.4836    Mean   :0.2874
## 3rd Qu.:1.0000    3rd Qu.:1.0000
## Max.   :1.0000    Max.   :1.0000
##

```

3 Create Sample for Analysis

1. Briefly discuss how you construct estimation sample in your term paper

I selected the IPUMS Time use data in 2020 and 2021 because, in these two years, the subjects were asked: “Whether you work remotely because of COVID-19?” and “How satisfied are you with your current life?” For this reason, I excluded those observations that did not answer these two question, as well as excluded those observations with no job and no income.

2. Use `mutate()` to generate a variable for your empirical analysis

Create interaction terms for subgroup analysis.

```
timeuse_sup <- timeuse %>%  
  mutate(  
    D.female = distance_work * female,  
    D.hh_numkids = distance_work * hh_numkids,  
    D.age = distance_work * age)
```

3. Use `summarise()` to create a new dataset for your empirical analysis

I created a new dataset called `wb_grpby_marst` here, which is the average well-being index for different marital statuses, and we can see that for those who never married, their well-being index is lower than other groups.

```
wb_grpby_marst <- timeuse %>%  
  group_by(marst) %>%  
  summarize(wb.average.grpby.marst = mean(wbladder, na.rm = T))  
wb_grpby_marst
```

```
## # A tibble: 6 x 2  
##   marst                wb.average.grpby.marst  
##   <dbl+lbl>                <dbl>  
## 1 1 [married - spouse present]                7.63  
## 2 2 [married - spouse absent]                7.32  
## 3 3 [widowed]                                7.07  
## 4 4 [divorced]                                7.07  
## 5 5 [separated]                                7.26  
## 6 6 [never married]                            6.75
```

4. Use `full_join()`, `inner_join()`, `right_join()` or `left_join()` to merge the dataset you have created for question 3-2 (previous question)

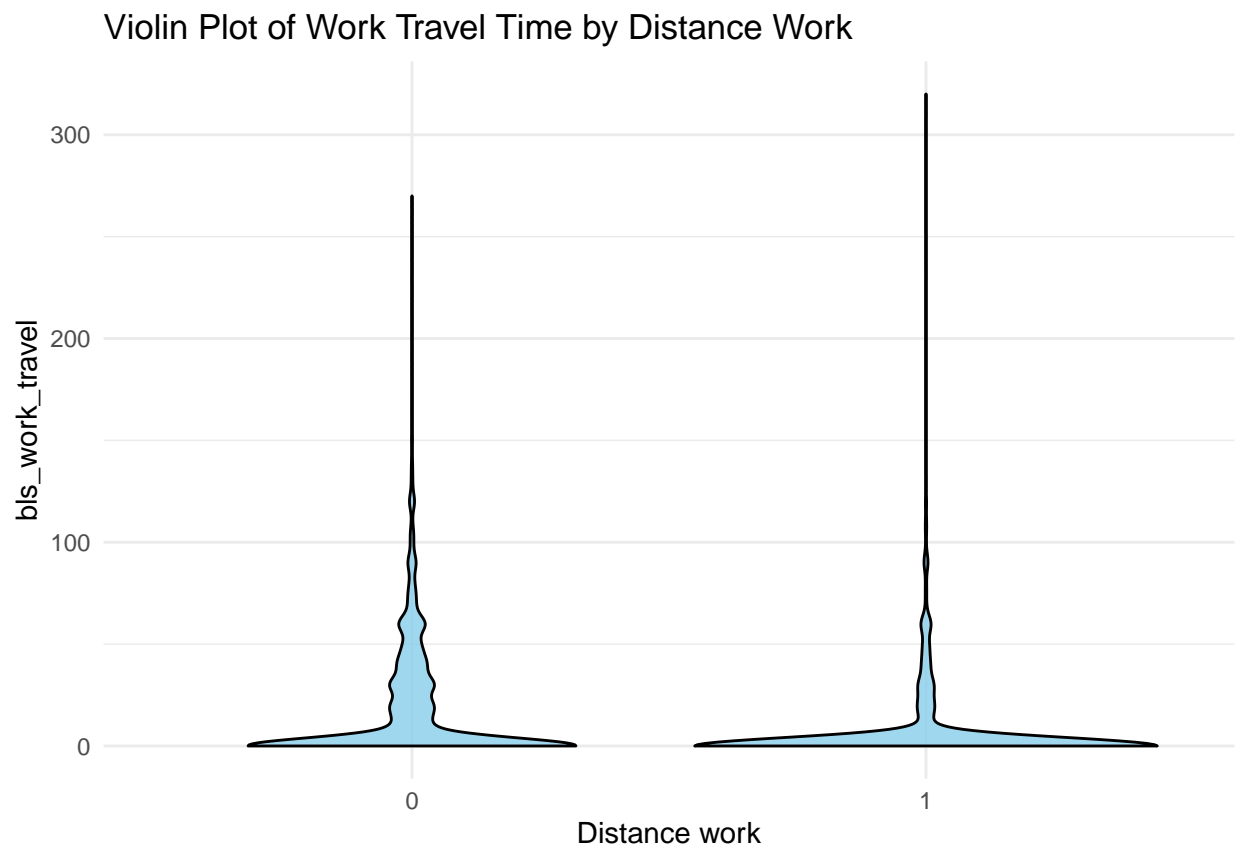
```
combine.data <- full_join(timeuse, wb_grpby_marst)
```

```
## Joining with 'by = join_by(marst)'
```

4 Visualize Data

1. Use any command that in R's graphical packages, such as `ggplot2` or others, to create a graph that can represent one of the findings in your term paper

```
ggplot(timeuse, aes(x = factor(distance_work), y = bls_work_travel)) +  
  geom_violin(fill = "skyblue", color = "black", alpha = 0.8) +  
  labs(x = "Distance work", y = "bls_work_travel") +  
  ggtitle("Violin Plot of Work Travel Time by Distance Work") +  
  theme_minimal()
```



2. Please explain the main findings of this graph

From the violin plot we can see the density of distance workers and commuters. We can see that distance workers are more likely to spend less time on work-related travel compared to commuters. Nevertheless, it is important to note that there are several outliers within both groups. This is because the variable `bls_work_travel` contains not only commuting time but also the time spent on business trips.

5 Empirical Analysis

1. Write a paragraph to explain the empirical specification you want to implement in your term paper.

We use **propensity score matching** to control the attributes of the experimental group (Distance workers) and the control group (Commuters), including age, sex, industry, occupation, marital status, how many children, etc., and then see if there is a significant difference in well-being between the treatment group and the control group.

2. Use any command in R's statistical/econometrics packages, such as `lm` or others, to analyze causal relationship in your term paper

For Y = well-being, method: `pscore`

```
y <- timeuse %>% select(starts_with("bls_"), wbladder)
x <- timeuse %>% select(-starts_with("bls_"), -wbladder)
cata <- c("statefip", "race", "marst", "empstat", "occ2", "ind2", "famincome",
         "citizen", "clwkr", "fullpart")
catagorical.data <- x %>% select(all_of(cata)) %>% to_label()
numeric.data <- x %>% select(-all_of(cata))
covariate <- bind_cols(catagorical.data, numeric.data)
covariate <- covariate %>%
  relocate(c(distance_work, female),
           .before = statefip)

idx <- 18 # wbladder
regdata <- bind_cols(y[idx], covariate) %>% drop_na()
fmla1 <- as.formula(paste("distance_work", ".", sep=" ~ "))
fmla2 <- as.formula(paste(colnames(y[idx]), "distance_work", sep=" ~ "))
m.out1 <- matchit(fmla1, data = regdata, method = "nearest", distance = "logit",
                 replace=FALSE)
m.data1 <- match.data(m.out1, distance = "pscore")
print(summary(lm(fmla2, data = m.data1)))

##
## Call:
## lm(formula = fmla2, data = m.data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.309 -1.222 -0.222  0.778  2.778
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.30881    0.05629  129.839  <2e-16 ***
## distance_work -0.08685    0.07961  -1.091    0.275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1.621 on 1656 degrees of freedom
## Multiple R-squared:  0.0007182, Adjusted R-squared:  0.0001148
## F-statistic: 1.19 on 1 and 1656 DF, p-value: 0.2754
```

For Y = well-being, method: Double Selection

```
catdummy = dummy_cols(catagorical.data) %>% select(-(statefip:fullpart))
idx <- 18

dat = bind_cols(y[, idx] ,numeric.data, catdummy) %>% drop_na()
YY = dat[,1]
DD = dat[,9]
XX = as.matrix(dat[, -c(1,9)])
doublesel.effect = rlassoEffect(x = XX, y = YY, d =DD, method = "double selection")
summary(doublesel.effect)
```

```
## [1] "Estimates and significance testing of the effect of target variables"
##           Estimate. Std. Error t value Pr(>|t|)
## distance_work -0.11548    0.07531  -1.533    0.125
```

3. Please explain your preliminary results

Here I did **pscore matching** and **Double Selection**. The coefficient of `distance_working` estimated by these two methods are both insignificant, which show that there is no difference in well-being between distance workers and commuters.