Data Translation Challenge

Presented by Rachel Nguyen, Khanh Quach, Nghi La, Tiffany Song, and Anna Smalley.

1. How has COVID affected the health of the retail industry, as measured by employment?

library(lubridate)

library(tidyverse)

library(vtable)

library(fixest)  
library(readr)  
library(dplyr)  
library(rio)  
library(ipumsr)  
library(lubridate)

#Load Data  
ddi <- read\_ipums\_ddi("cps\_00002.xml")  
data\_econ <- read\_ipums\_micro(ddi)

Use of data from IPUMS CPS is subject to conditions including that users should  
cite the data appropriately. Use command `ipums\_conditions()` for more details.

# merge two files by industry   
indnames <- import('indnames.csv')  
  
data\_econ <- merge(data\_econ, indnames, by.x ="IND", by.y ="ind")

# filter Retail industry only

data\_econ <- data\_econ %>%

mutate(retail = as.character(indname == 'Retail Trade'))

# Year as numeric   
data\_econ$YEAR <- as.numeric(data\_econ$YEAR)

# Join year and month   
data\_econ <- data\_econ %>% mutate(year.month = ym(paste0(YEAR, '/',MONTH)))

# Cutoff point is March 2020   
data\_econ <- data\_econ %>% mutate(covid\_cutoff = case\_when(  
 year.month <= ym('2020/03') ~ "Before",  
 year.month > ym('2020/03') ~ "After"  
))

# center is March 2020  
data\_econ <- data\_econ %>%  
 mutate(Center.Date = year.month - as.Date("2020-03-01"))

# EMPSTAT as binary   
data\_econ <- data\_econ %>% mutate(Employ\_numeric = case\_when(  
 EMPSTAT == 10 ~ 1,  
 EMPSTAT == 21 ~ 0  
))

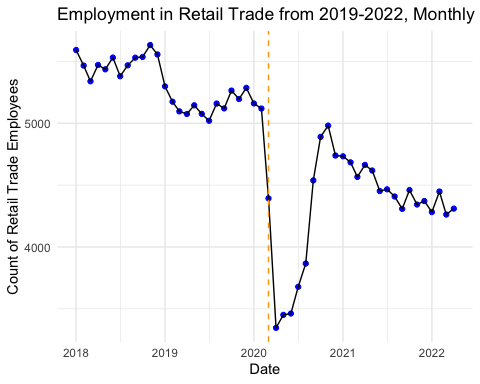
# Create new data frame with only individuals employed in retail

Retail.data <- filter(data\_econ, retail == "TRUE")

Retail.data <- data\_econ %>%   
 filter(retail == "TRUE")

#   
Retail.df <- data\_econ %>% filter(indname == "Retail Trade" & EMPSTAT %in% c(10, 21))  
  
# Create a new column that contains a varible that counts the number of retail employees, per month.   
Retail.df <- Retail.df %>% group\_by(year.month) %>%  
 mutate(count.EMPSTAT.10 = sum(EMPSTAT == 10),  
 count.EMPSTAT.21 = sum(EMPSTAT == 21),  
 )  
  
Retail.df <- Retail.df %>% distinct(year.month, covid\_cutoff, .keep\_all = TRUE)  
  
# Graph data  
Retail\_plot <- ggplot(Retail.df, aes(x = year.month, y = count.EMPSTAT.10)) +  
 geom\_point(color = "blue") +  
 geom\_line() +  
 geom\_vline(xintercept = ym('2020/03'), color = "orange", linetype = "dashed") +  
 theme\_minimal() +  
 labs(title = 'Employment in Retail Trade from 2019-2022, Monthly', x = 'Date', y = 'Count of Retail Trade Employees')

Retail\_plot

  
data\_econ <- data\_econ %>%  
 mutate(covid\_cutoff = factor(covid\_cutoff, levels = c('Before','After')))

# Regression  
reg1 <- data\_econ %>%  
feols(Employ\_numeric ~ covid\_cutoff\*I(indname == 'Retail Trade')\*Center.Date)

etable(reg1)

Retail.reg1  
Dependent Var.: Employment.Status.numeric  
   
Constant 0.9685\*\*\* (0.0004)  
covid\_cutoffAfter -0.0690\*\*\* (0.0006)  
I(indname=="RetailTrade")TRUE -0.0099\*\*\* (0.0011)  
Center.Date 4.17e-6\*\*\* (7.76e-7)  
covid\_cutoffAfter x I(indname=="RetailTrade")TRUE -0.0025 (0.0018)  
covid\_cutoffAfter x Center.Date 0.0001\*\*\* (1.23e-6)  
I(indname=="RetailTrade")TRUE x Center.Date 6.29e-6\*\* (2.4e-6)  
covid\_cutoffAfter x I(indname=="RetailTrade")TRUE x Center.Date -5.54e-6 (3.84e-6)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type IID  
Observations 2,551,165  
R2 0.00917  
Adj. R2 0.00917  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# (covid.pre.post)\*RETAILTRUE is the RDD effect of covid.pre.post on the Retail industry.

# Results:

The research question is ‘How has COVID affected the health of the retail industry, as measured by employment?’

First, a range of data is selected from IPUMS, including EMPSTAT(employment status), Year, Month, Ind(industry). Importing the csv file of the industry.

Then combine data selection with the csv file by ind, and then filter the Retail Trade industry only.

Covid started in March 2020 in the states, so we decided to set March 2020 as the ‘cutoff’ point. To do so, first, combine the month and year in the data set, and set any month before March 2020 as Pre-Covid, and after March 2020 as Post-Covid. From the data set, EMPSTAT, 10 is employed, and 21 is unemployed, we set this as the binary dependent variable (0,1). To graph, set the x-axis as time, and the y-axis as the count of employed. From the output of the graph we can see that the discontinuity, or the estimated effect of the intervention, is the difference between pre and after-Covid retail trade employment.

The model, includes covid\_after cutoff, I (retail, binary), center date(March 2020), and the interaction term. From the output, The average employment status numeric value is approximately 0.9685 when all the independent variables are zero. After the pandemic started, the average employment decreased by 0.069, holding everything else constant. The effect of Covid on the retail industry is an additional decrease in employment at 0.0025. After the pandemic began, for every unit increase in the centered date, the employment status numeric value increased by an additional 0.0001. The triple interaction term suggests that after the pandemic started, in the retail trade industry, for each unit increase in the centered date, the employment status numeric value decreased by an additional 5.54e-6.

To sum up, the analysis suggests that the COVID-19 pandemic has had a detrimental effect on employment, with a sharp drop observed in the retail industry right before the designated 'cutoff' point. This decline in employment persisted immediately after the 'cutoff' point as well. Throughout the ongoing pandemic, there was a gradual increase in retail employment, though there were certain periods of decrease, potentially influenced by preceding policy decisions or concerns related to public safety. However, by April 2022, the level of employment in the retail sector had not yet returned to its pre-pandemic levels.

1. **How has retail fared relative to other industries?**

library(rio)  
library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':  
  
 filter, lag

The following objects are masked from 'package:base':  
  
 intersect, setdiff, setequal, union

library(ggplot2)  
library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ forcats 1.0.0 ✔ stringr 1.5.0  
✔ lubridate 1.9.2 ✔ tibble 3.2.1  
✔ purrr 1.0.1 ✔ tidyr 1.3.0  
✔ readr 2.1.4

── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lubridate)  
library(fixest)  
library(readxl)  
library(vtable)

Loading required package: kableExtra  
  
Attaching package: 'kableExtra'  
  
The following object is masked from 'package:dplyr':  
  
 group\_rows

library(ipumsr)  
library(rdrobust)  
library(stringr)

ddi <- read\_ipums\_ddi("cps\_00005.xml")  
data <- read\_ipums\_micro(ddi)

Use of data from IPUMS CPS is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

indnames <- read.csv('indnames.csv')  
data <- merge(data, indnames, by.x ="IND", by.y ="ind")

data$YEAR <- as.numeric(data$YEAR)  
  
data <- data %>%   
 filter(is.na(ASECFLAG)) %>%   
 filter(EMPSTAT %in% c(10, 21))  
  
data <- data %>%  
 mutate(  
 Date = ymd(paste0(YEAR, '/', MONTH, '/01')),  
 COVID.PERIOD = if\_else(Date <= ymd('2020/03/15'), "Before COVID", "After COVID"),  
 Employment\_Status = factor(EMPSTAT, levels = c(10, 21), labels = c("Employed", "Unemployed")),  
 Employment\_Binary = if\_else(EMPSTAT == 10, 1, 0)  
 )

data <- data %>%  
 mutate(  
 industry = if\_else(str\_detect(indname, "Retail"), "Retail", "Other industries")  
 )  
  
data$COVID.PERIOD = factor(data$COVID.PERIOD, levels = c("Before COVID", "After COVID"))  
  
data$industry = factor(data$industry, levels = c("Other industries", "Retail"))

# Run the regression model with weights   
rdd\_model <- feols(Employment\_Binary ~ COVID.PERIOD \* industry, weights = data$WTFINL, data = data)  
  
# Show the summary of the model  
summary(rdd\_model)

OLS estimation, Dep. Var.: Employment\_Binary  
Observations: 2,309,487   
Weights: data$WTFINL   
Standard-errors: IID   
 Estimate Std. Error t value  
(Intercept) 0.965936 0.000212 4555.73797  
COVID.PERIODAfter COVID -0.031405 0.000305 -103.09870  
industryRetail -0.012838 0.000653 -19.64590  
COVID.PERIODAfter COVID:industryRetail -0.001586 0.000938 -1.69111  
 Pr(>|t|)   
(Intercept) < 2.2e-16 \*\*\*  
COVID.PERIODAfter COVID < 2.2e-16 \*\*\*  
industryRetail < 2.2e-16 \*\*\*  
COVID.PERIODAfter COVID:industryRetail 0.090816 .   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
RMSE: 11.9 Adj. R2: 0.005535

etable(rdd\_model)

rdd\_model  
Dependent Var.: Employment\_Binary  
   
Constant 0.9659\*\*\* (0.0002)  
COVID.PERIODAfterCOVID -0.0314\*\*\* (0.0003)  
industryRetail -0.0128\*\*\* (0.0007)  
COVID.PERIODAfterCOVID x industryRetail -0.0016. (0.0009)  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type IID  
Observations 2,309,487  
R2 0.00554  
Adj. R2 0.00554  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

data\_grouped <- data %>%  
 group\_by(industry, COVID.PERIOD, Date) %>%  
 summarise(employment\_percentage = mean(Employment\_Binary)\*100)

`summarise()` has grouped output by 'industry', 'COVID.PERIOD'. You can  
override using the `.groups` argument.

ggplot(data\_grouped, aes(x = Date, y = employment\_percentage, color = industry)) + geom\_line() + geom\_vline(xintercept = ym('2020/03')) +  
 labs(title = 'Employment rates in Retail vs Other Industries',  
 x = 'Date',  
 y = 'Employment Rate',  
 color = 'Industry') +  
 theme\_minimal()

A picture containing text, plot, line, diagram

Description automatically generated

**Results:**

In this analytical model, our primary objective is to compare the changes in employment rates between the ‘retail trade industry’ and other industries, due to the onset of the COVID-19 pandemic. Our choice of methodology, the Regression Discontinuity Design (RDD), was a strategic decision driven by the nature of the research question at hand.

Given the categorical nature of the industries we’re examining—specifically the retail trade industry versus others—we found RDD to be an effective analytical tool. RDD shines in circumstances where we can clearly delineate a cutoff or threshold variable that triggers a different treatment or condition. In our case, this threshold variable is the start of the COVID-19 outbreak in the U.S., which we have chosen to be March 15th.

Our choice of March as the COVID cutoff is a pragmatic one, reflective of the reality of the situation in the U.S. around that time. As the pandemic began to grip the country, various aspects of the economy were significantly affected, not the least of which was employment. It’s important to note that our data does not specify the actual employment commencement or termination dates; therefore, we made a uniform assumption that the COVID impact commenced on this date.

The implementation of the RDD model provides us with fascinating insights into the impact of the COVID period on employment. The ‘COVID.PERIODAfterCOVID’ coefficient is both negative and statistically significant. This suggests that there was a decrease in employment probability across all sectors by approximately 3.14 percentage points on average in the post-COVID period. Holding industry constant, we find that the employment probability declined to about 93.45% (=96.59% - 3.14%) following the arrival of COVID-19. This statistically significant result highlights the pervasive adverse effect of COVID-19 on employment across all sectors.

Interestingly, our analysis also reveals that the retail trade industry faced a unique set of challenges. The ‘industryRetail’ coefficient is negative and significant, indicating that being in the retail trade industry, irrespective of the period (pre- or post-COVID), is associated with a further decrease in employment probability by 1.28 percentage points as compared to non-retail industries. This suggests that even before the advent of COVID-19, the retail sector had a slightly lower employment probability, around 95.31% (=96.59% - 1.28%).

Finally, when we consider the differential impact of the COVID period for the retail industry, the interaction term ‘COVID.PERIODAfterCOVID x industryRetail’ unveils an additional decrease in employment probability during the post-COVID period for this sector.

Taken together, these results indicate that both the retail trade and non-retail industries experienced a decrease in employment probability from the pre-COVID to the post-COVID period, with the retail trade industry already starting at a lower point of employment probability in the pre-COVID period. Importantly, the post-COVID period’s impact on employment does not seem to differ significantly between the retail trade and non-retail industries, as evidenced by the non-significant interaction term.

This analysis underscores the critical role of external macroeconomic factors like the COVID-19 pandemic in shaping industry employment trends and highlights the need for further research to fully understand these dynamics.

3. What has changed about who is working and earning money?

library(rio)  
library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':  
  
 filter, lag

The following objects are masked from 'package:base':  
  
 intersect, setdiff, setequal, union

library(ggplot2)  
library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ forcats 1.0.0 ✔ stringr 1.5.0  
✔ lubridate 1.9.2 ✔ tibble 3.2.1  
✔ purrr 1.0.1 ✔ tidyr 1.3.0  
✔ readr 2.1.4

── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lubridate)  
library(fixest)  
library(readxl)  
library(vtable)

Loading required package: kableExtra  
  
Attaching package: 'kableExtra'  
  
The following object is masked from 'package:dplyr':  
  
 group\_rows

library(ipumsr)  
library(rdrobust)  
library(stringr)

ddi <- read\_ipums\_ddi("cps\_00004.xml")  
data <- read\_ipums\_micro(ddi)

Use of data from IPUMS CPS is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

indnames <- read.csv('indnames.csv')  
# merge the data together to get the new table   
data <- merge(data, indnames, by.x ="IND", by.y ="ind")

data$YEAR <- as.numeric(data$YEAR)  
  
data <- data %>%   
 filter(is.na(ASECFLAG)) %>%   
 filter(EMPSTAT %in% c(10, 21))

data <- data %>%  
 mutate(  
 Date = ymd(paste0(YEAR, '/', MONTH, '/01')),  
 COVID.PERIOD = if\_else(Date <= ymd('2020/03/15'), "Before COVID", "After COVID"),  
 Employment\_Status = factor(EMPSTAT, levels = c(10, 21), labels = c("Employed", "Unemployed")),  
 Employment\_Binary = if\_else(EMPSTAT == 10, 1, 0)  
 )

data <- data %>%   
 mutate(Race = case\_when(  
 RACE == 100 ~ "White",  
 RACE == 200 ~ "Black",  
 RACE <= 651 ~ "Asian"  
))  
  
data <- data %>%  
 mutate(AgeGroup = case\_when(  
 AGE >= 18 & AGE <= 34 ~ "Young Adults",  
 AGE >= 35 & AGE <= 54 ~ "Middle Aged Adults",  
 AGE >= 55 & AGE <= 65 ~ "Older Adults",  
 TRUE ~ "Outside Working Age"  
))  
data <- data %>%   
 mutate(Gender = case\_when(  
 SEX == 1 ~ "Male",  
 SEX == 2 ~ "Female"  
))

data$COVID.PERIOD = factor(data$COVID.PERIOD, levels = c("Before COVID", "After COVID"))  
vtable(data)

Warning in vtable(data): Some labelled variables have unlabeled values.  
Treating these as numeric variables and ignoring labels.

data

| Name | Class | Label | Values |
| --- | --- | --- | --- |
| IND | numeric | NULL | Num: 170 to 9890 |
| YEAR | numeric | NULL | Num: 2018 to 2022 |
| SERIAL | numeric | NULL | Num: 1 to 72367 |
| MONTH | haven\_labelled | Month | '1: January' '2: February' '3: March' '4: April' '5: May' and more |
| HWTFINL | numeric | NULL | Num: 153.427 to 26194.133 |
| CPSID | numeric | NULL | Num: 20161000001300 to 20220406880900 |
| ASECFLAG | haven\_labelled | Flag for ASEC | '1: ASEC' '2: March Basic' |
| PERNUM | numeric | NULL | Num: 1 to 16 |
| WTFINL | numeric | NULL | Num: 135.759 to 26194.133 |
| CPSIDP | numeric | NULL | Num: 20161000001301 to 20220406880902 |
| AGE | haven\_labelled | Age | Num: 15 to 65 |
| SEX | haven\_labelled | Sex | '1: Male' '2: Female' '9: NIU' |
| RACE | haven\_labelled | Race | '100: White' '200: Black' '300: American Indian/Aleut/Eskimo' '650: Asian or Pacific Islander' '651: Asian only' and more |
| EMPSTAT | haven\_labelled | Employment status | '0: NIU' '1: Armed Forces' '10: At work' '12: Has job, not at work last week' '20: Unemployed' and more |
| CLASSWKR | haven\_labelled | Class of worker | '0: NIU' '10: Self-employed' '13: Self-employed, not incorporated' '14: Self-employed, incorporated' '20: Works for wages or salary' and more |
| COVIDUNAW | haven\_labelled | Unable to work due to COVID-19 pandemic | '1: No' '2: Yes' '99: NIU' |
| indname | character | NULL |  |
| Date | Date | NULL | Time: 2018-01-01 to 2022-04-01 |
| COVID.PERIOD | factor | NULL | 'Before COVID' 'After COVID' |
| Employment\_Status | factor | NULL | 'Employed' 'Unemployed' |
| Employment\_Binary | numeric | NULL | Num: 0 to 1 |
| Race | character | NULL |  |
| AgeGroup | character | NULL |  |
| Gender | character | NULL |  |

model\_race <- feols(Employment\_Binary ~ COVID.PERIOD\*Race | AgeGroup + Gender, data = data, se = "hetero", weights = data$WTFINL)

NOTE: 53,837 observations removed because of NA values (RHS: 53,837).

etable(model\_race)

model\_race  
Dependent Var.: Employment\_Binary  
   
COVID.PERIODAfterCOVID -0.0405\*\*\* (0.0012)  
RaceBlack -0.0278\*\*\* (0.0010)  
RaceWhite -0.0017\* (0.0007)  
COVID.PERIODAfterCOVID x RaceBlack -0.0008 (0.0018)  
COVID.PERIODAfterCOVID x RaceWhite 0.0120\*\*\* (0.0013)  
Fixed-Effects: -------------------  
AgeGroup Yes  
Gender Yes  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 2,291,882  
R2 0.01041  
Within R2 0.00765  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Results:

In this model, we ran a fixed effects regression to examine how employment rates changed from before COVID to after COVID periods for different racial groups (White, Black, and Asian), while controlling for age group and gender. We used the “Before COVID” period as the reference period and considered Asian as the reference group.

As in all of these models, we used the assumption that COVID-19 started in March 15th 2020, which is only the case for the US. Another assumption we made to answer this third question we only used white, Black, and Asian races. Similarly, we used the data variable “sex”to indicate gender however we know that these are not always the same. Also in all below models we accounted for heteroskedastity.

The findings of the analysis show a decrease in employment rates during the after COVID period. Specifically, the coefficient for COVID.PERIODAfterCOVID (-0.0405) suggests a decrease in employment rates in the after COVID period compared to the before COVID period.

When considering the racial groups, the coefficients for RaceBlack (-0.0278) and RaceWhite (-0.0017) suggest that both Black and White individuals had lower employment rates compared to the reference group. However, White experienced a relatively better recovery in terms of employment rates (COVID.PERIODAfterCOVID x RaceWhite (0.0120) ) for the after COVID period.

Overall, these results indicate that the COVID period had a significant impact on employment rates, with both Black and White individuals having lower rates than the Asian reference group. However, the impact appears to be worse for Black individuals, whereas Whites recover relatively more quickly. This points out the need for Retail to consider shifting demographics and employment trends when determining who has money to spend.

# Filter out the "Outside Working Age" group  
data\_working\_age <- data %>%  
 filter(AgeGroup != "Outside Working Age")  
  
data\_working\_age <- data\_working\_age %>%  
 mutate(COVID.PERIOD = factor(COVID.PERIOD, levels = c("Before COVID","After COVID")))  
  
model\_age\_group <- feols(Employment\_Binary ~ COVID.PERIOD\*AgeGroup | Race + Gender, data = data\_working\_age, se = "hetero", weights = data\_working\_age$WTFINL)

NOTE: 52,271 observations removed because of NA values (Fixed-effects: 52,271).

etable(model\_age\_group)

model\_age\_group  
Dependent Var.: Employment\_Binary  
   
COVID.PERIODAfterCOVID -0.0271\*\*\* (0.0005)  
AgeGroupOlderAdults -0.0003 (0.0004)  
AgeGroupYoungAdults -0.0182\*\*\* (0.0005)  
COVID.PERIODAfterCOVID x AgeGroupOlderAdults -0.0024\*\* (0.0008)  
COVID.PERIODAfterCOVID x AgeGroupYoungAdults -0.0102\*\*\* (0.0008)  
Fixed-Effects: -------------------  
Race Yes  
Gender Yes  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 2,257,216  
R2 0.01030  
Within R2 0.00771  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Results:

In this model, we examined the changes in employment rates before and after the COVID period for different age groups, while controlling for race and gender. The “Before COVID” period was used as the reference category for the COVID period.

The coefficient -0.0271 suggests a significant decline in employment rates in the after COVID period compared to the before COVID period.

When looking at different age groups, both Older Adults and Young Adults show lower employment rates. The coefficients for AgeGroupOlderAdults (-0.0003) and AgeGroupYoungAdults (-0.0182) means that Young Adults had a bigger decrease in employment rates compared to Older Adults.

Similarly, the coefficient (-0.0102) indicates a significant additional decrease in employment rates for Young Adults in the after COVID period while -0.0024 shows another small decrease in employment rates for Older Adults during the after COVID period. However, it is smaller compared to the Young adult group.

In conclusion, young adults were more negatively impacted by the COVID era in terms of employment. Compared to Older Adults, they saw a greater decline in employment opportunities. This might be the case because Young Adults, who are just starting their careers, had more difficulty during the COVID era finding employment or securing stable jobs. Older adults, on the other hand, experienced a relatively smaller decline in employment rates, perhaps as a result of the fact that they already had established careers or were less impacted by the pandemic’s changes to the labor market.

model\_gender <- feols(Employment\_Binary ~ COVID.PERIOD\*Gender | Race + AgeGroup,   
data = data, se = "hetero", weights = data$WTFINL)

NOTE: 53,837 observations removed because of NA values (Fixed-effects: 53,837).

etable(model\_gender)

model\_gender  
Dependent Var.: Employment\_Binary  
   
COVID.PERIODAfterCOVID -0.0330\*\*\* (0.0005)  
GenderMale -0.0018\*\*\* (0.0004)  
COVID.PERIODAfterCOVID x GenderMale 0.0035\*\*\* (0.0007)  
Fixed-Effects: -------------------  
Race Yes  
AgeGroup Yes  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
S.E. type Heteroskedast.-rob.  
Observations 2,291,882  
R2 0.01029  
Within R2 0.00508  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Results:**

In this model, we examined the changes in employment rates before and after the COVID period, specifically focusing on the differences between gender groups while controlling for race and age group. The “Before COVID” period was used as the reference category.

The coefficient for COVID.PERIODAfterCOVID (-0.0330) shows decrease in employment rates during the after COVID period compared to the before COVID period. This suggests that the pandemic had a negative impact on employment rate.

The coefficient for GenderMale (-0.0018) indicates that male employment rates were lower than those of females when gender differences were taken into account. However, the COVID interaction term coefficient.Male employment rates declined relatively less than female employment rates during the post-COVID period, according to PERIODAfterCOVID x GenderMale (0.0035).

Overall, these findings suggest that although employment rates dropped for both genders during the COVID period, the decrease for men was slightly smaller. In conclusion, the COVID period significantly affected employment rates, with drops seen in both genders. However, compared to females, the decrease in males was relatively less.