

ECON 753 Homework 3

Jesús Lara Jáuregui

10/27/2020

Part 1: Exercises from Berndt, chapter 5

Exercise 1

Part (a)

In this exercise I work with the Current Population Survey (CPS) for the years 1978 and 1985, that shows demographic, social and economic information of 550 and 534 individuals, respectively.

In this part I calculate the arithmetic and geometric means of the the wage rate and annual wages for 1978. The results are shown in Table 1.

Table 1: Geometric and Arithmetic Mean for 1978 Wages

	Geometric Mean	Arithmetic Mean
Hourly	5.37	6.06
Annual	10741.87	12125.53

Table 2: Mean and Standard Deviation of Education and Experience

	Mean	Standard Deviation
Education	12.54	2.77
Experience	18.72	13.35

We have data for the log of the wage rate, so to calculate the arithmetic mean we exponentiate that variable and then get the average. To get the mean of the annual wage we just multiply that number by 2000, which are the assumed work hours of a year. In contrast, the geometric mean is calculated by getting the average of the log wage and then exponentiating it. Hence, it is no surprise that the two results are different. The arithmetic mean of the wage rate and annual wages is greater than the geometric mean because the latter reduces large values more than it increases small values (of the wage-rate). Given that logarithm and exponentiation are not linear transformations, the arithmetic and geometric means will, in general, be different.

In Table 2 we show the mean and standard deviation of education and experience. As we would expect, the standard deviation of education is much smaller (2.8) than that of experience (13.4). The former means that we will have a very large share of the population within the range 9.5-15.5 of years of education. This is so, because, in general, almost all workers have a minimum years of education which is a number greater than zero. In contrast, in the labor market there is always a greater dispersion of experience (new people join the labor force all the time), which is captured by the standard deviation of 13.4

Table 4: Years of Education and Mean Wages by Race

	Education		lnWage		
	Mean	SD	Mean Hourly	SD	Mean Annual
Gender					
Male	12.40	3.05	6.13	3.42	12256.64
Female	12.76	2.22	4.32	2.50	8632.72
Race					
Whites	12.81	2.49	5.55	3.38	11100.34
NonWhites	11.72	3.44	4.54	2.55	9084.33
Hispanic	10.31	3.66	4.62	1.96	9233.10

Part (b)

In this part we get the frequencies of non-whites, females and hispanics. The results are shown in Table 3. The share of each category is given by the arithmetic mean, whereas the total count is given by the mean times the sample size or, equivalently, by the sum over each of these dummy variables.

Table 3: Means and Number of Individuals in Demographic Groups

	Mean	Count
nonwh	0.10	57
hisp	0.07	36
fe	0.38	207

Our results show that 10% of the individuals in the sample are nonwhite, 7% hispanic and 38% women.

Part (c)

In this part we sub-divide the sample by gender and race and analyze the mean and standard deviation of years of education and the wage rate. The results are in Table 4

Regarding education and gender, men and women have on average almost the same years of education, although the standard deviation for men is greater. Regarding race, the results show that whites in the sample have on average one more year of education than non-whites, and the latter have, on average, one more year of education than hispanics.

Regarding the wage rate, the average in the sample is greater for men than for women and it is greater for whites than for nonwhites and hispanics. The same is true for the mean annual wages.

Part (d)

In part d we repeat the comparisons of shares, means and standard deviations of wage and education but for 1985. With respect to shares, the results are in Table 5

There we can see that share of non-whites in the sample increased and that of hispanics decreased, although the changes are small. In contrast, the share of females in the sample increased from 38% in 1978 to 46% in 1985

In Table 6 we show the results for Education and Real Wages, incorporating the percentage change that took place from 1978 to 1985. Real wages were calculated with a price index which is equal to one for 1978 and 1.649 for 1985

Table 5: Comparison of demographics between 1978 and 1985

	1978		1985	
	Mean	Count	Mean	Count
nonwh	0.10	57	0.13	67
hisp	0.07	36	0.05	27
fe	0.38	207	0.46	245

Table 6: Comparison of education and real wages between 1978 and 1985

	Education					lnWage				
	1978		1985		%Change	1978		1985		%Change
	Mean	SD	Mean	SD		Mean Hourly	SD	Mean Hourly	SD	
Gender										
Male	12.40	3.05	13.01	2.77	0.05	6.13	3.42	5.29	3.21	-0.14
Female	12.76	2.22	13.02	2.43	0.02	4.32	2.50	4.19	2.86	-0.03
Race										
Whites	12.81	2.49	13.17	2.48	0.03	5.55	3.38	4.89	3.17	-0.12
NonWhites	11.72	3.44	12.64	2.60	0.08	4.54	2.55	4.33	2.47	-0.05
Hispanic	10.31	3.66	11.52	4.05	0.12	4.62	1.96	3.74	3.35	-0.19
Total										
Whole Sample	12.54	2.77	13.02	2.62	0.04	5.37	3.26	4.75	3.12	-0.11

Just by looking at the signs, we can see the main trends regarding these two variables: the average years of education increased for all groups, whereas the average wage rate fell also for all groups. For the whole sample, the average years of education increased by 4% and the real wage rate fell by 11%.

However, this trends, although general, are not exactly the same for all groups. The average years of education increased more for men than for women. But the average real wage rate for men fell by 14% whereas for women the decline was 3%. Regarding race, hispanics had the greatest increase in the average years of education. However, they also experienced the largest decline in real wage (19%).

Part (e)

In this part we want to analyze the distribution of wages and log wages. In particular, we want to test whether they are normally distributed. To do that, we follow two approaches. In the first one, we calculate the proportion of observations that are within 6 different groups defined in terms of the sample mean and sample standard deviation. We compare them with the proportions of a normal distribution for these six groups. The second approach, more formal, consists of chi-squared goodness-of-fit tests for normal distribution.

From Table 7 we can observe that the proportions of log wages match almost perfectly with those of the normal distributions, whereas the proportions of Wage for each group are substantially different. Now we conduct the chi-squared test. The null-hypothesis H_0 is that wages and log wages are normally distributed. The results of the test are shown below.

As we would expect, for log wages the p-value is 0.59 so we failed to reject H_0 . In contrast, the p-value for wage is close to zero, so we reject the null-hypothesis H_0 at the 1% significance that wages are normally distributed.

Table 7: Counts and Proportions of lnwage and wage Distributions

	lnwage		wage		Normal Distribution Proportions
	Counts	Proportion	Counts	Proportion	
wi ≤ w-2s	9	0.02	0	0.00	0.02
w-2s < wi ≤ w-s	80	0.15	45	0.08	0.14
w-s < wi ≤ w	194	0.35	278	0.51	0.34
w < wi ≤ w+s	193	0.35	164	0.30	0.34
w+s < wi ≤ w+2s	63	0.11	45	0.08	0.14
w+2s < wi	11	0.02	18	0.03	0.02

Table 8: Chi-Squared Goodness-of-Fit Tests for Normal Distribution

	Statistic	pValue
lnwage	3.74	0.59
wage	84.99	0.00

Problem 6

Part (a)

In this part we make use of a linear regression model to test for the existence of wage premiums and penalties associated to gender and race. The results are presented in table 9. The interpretation of the coefficients of our relevant dummy variables is the percentage difference in wage associated to belonging to a race/gender group, controlling for other factors (union, experience and education)

The results point to a 30% wage penalty for women, a 16% penalty for nonwhites and a 3% penalty for hispanics, although the latter is not statistically significant at any reasonable level.

We are asked to test the hypothesis that the coefficients of nonwh and hisp are the same, that is that both groups experience the same wage penalty. That is done with a Chow test where $H_0 : \beta_{hisp} - \beta_{nonwh} = 0$. The results of the test are provided in table 10. The F-statistic is 2.46 and the associated p-value is 0.12. Hence we do not have statistical evidence to reject the hypothesis that the wage penalty for hispanics is equal to that of nonwhites.

Part (b)

For this part we will analyze our data by race (non-whites, hispanics and others). We first analyze shares and frequencies, which are shown in Table 11. The most salient fact is that 83% of the individuals in the sample belong to the category “others”, which are mainly whites, whereas hispanics and nonwhites have almost the same share.

We then analyze education, experience, gender, unionization and wages by race and conduct a comparison with whites. The results are shown in Table 12.

The most important aspects from Table 12 are that whites have 1 and 2 and a half years more of education than non-whites and hispanics, respectively. Whites have a wage rate approximately 20% higher than hispanics and non-whites. In contrast, on average, non-whites and hispanics have more years of experience and, very interestingly for me, non-whites are more likely to belong to a union than whites.

Table 9: Simple Regression of log Wages to Determine Wage Discrimination

	<i>Dependent variable:</i>
	lnwage
Constant	0.488*** (0.098)
fe	-0.306*** (0.034)
union	0.207*** (0.037)
nonwh	-0.157*** (0.055)
hisp	-0.027 (0.069)
ed	0.075*** (0.007)
ex	0.026*** (0.005)
exsq	-0.0003*** (0.0001)
Observations	550
R ²	0.392
Adjusted R ²	0.385
Residual Std. Error	0.385 (df = 542)
F Statistic	50.012*** (df = 7; 542)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Table 10: Chow test

F statistic	P-value
2.46	0.12

Table 11: Means and Number of Individuals in Demographic Groups

	Mean	Count
other	0.83	457
nonwh	0.10	57
hisp	0.07	36

Table 12: Comparison of Demographics Variables

	Education		Experience		Female		Unionized		Wages	
	ed	diff	ex	diff.1	fe	diff.2	union	diff.3	lnwage	diff.4
Others	12.81	0.00	18.11	0.00	0.37	0.00	0.30	0.00	1.71	0.00
NonWhites	11.72	-1.09	22.16	4.05	0.49	0.13	0.39	0.08	1.51	-0.20
Hispanic	10.31	-2.51	21.06	2.95	0.33	-0.03	0.19	-0.11	1.53	-0.18

Part (c)

In this part we split the sample by race to run three separate linear regressions with log wage as the dependent variable. This allow us to see the effect that experience, education, unionization and gender exert on wages across race.

Table 13: Regressions of lnwage by Racial Category

	<i>Dependent variable:</i>		
	Others	lnwage NonWhites	Hispanics
	(1)	(2)	(3)
Constant	0.456*** (0.111)	0.755* (0.389)	0.682*** (0.234)
fe	-0.338*** (0.037)	-0.059 (0.144)	-0.235** (0.104)
union	0.197*** (0.040)	0.205 (0.139)	0.417*** (0.123)
ed	0.077*** (0.008)	0.054** (0.026)	0.055*** (0.015)
ex	0.028*** (0.005)	0.003 (0.020)	0.021 (0.014)
exsq	-0.0003*** (0.0001)	-0.00001 (0.0004)	-0.0003 (0.0003)
Observations	457	57	36
R ²	0.420	0.149	0.489
Adjusted R ²	0.414	0.066	0.404
Residual Std. Error	0.376 (df = 451)	0.484 (df = 51)	0.287 (df = 30)
F Statistic	65.423*** (df = 5; 451)	1.789 (df = 5; 51)	5.752*** (df = 5; 30)

Note:

*p<0.1; **p<0.05; ***p<0.01

The results are shown in Table 13. Among the most interesting things, we can see that the wage penalty for women is greater within whites, (it is close to 40% and is statistically significant). In contrast, the coefficient of female is not statistically significant for non-whites. Regarding union, it is positive and significant for whites and hispanics but not for non-whites. The coefficient of education is positive and significant, whereas that of experience is only significant for whites (positive in its linear form and negative in quadratic form).

Table 14: RSS comparison

Sample	RSS
Complete	81.35
White	63.83
Hisp	2.46
Non-white	11.94

Table 15: Chow test information

DF 1	DF 2	F statistic	P-Value
5	65	0.52	0.24

These last facts may reflect the characteristics of the the jobs of non-whites and hispanics, where education and experience may not be very relevant.

Part (d)

In this part we are asked to test whether the coefficients of the 3 models that we estimated in part c are in reality the same. To do so, we first estimate the same equation than in part (c) but with the complete sample (results not shown here), and perform a Chow test in which we make use of the Sum of Squares Residuals (RSS). These are shown in Table 14.

What I see from this table is that, we are getting a low RSS for the models with a very few number of observations. We use this information and the number of variables ($k = 5$) to construct our F-statistic and evaluate it in the F-distribution function with the correct degrees of freedom. The results are shown in Table 15

The P-value associated to this test is 0.24. Hence, we fail to reject the null hypothesis that the coefficients are the same for the 3 models estimated in part c.

Part 2: Replication of Card & Krueger (1994)

In this part I replicate the first three rows of Table 3 of Card & Krueger(1994), in which they present the basic Diff-in-Diff results of the effect of the increase in minimum wage in New Jersey on employment in the fast-food sector. My replication is Table 16 (Evan and I spent so many hours figuring out how to put standard errors below means in parenthesis and feel very proud about how nice that table is).

The variable analyzed in all columns is Full-Time Equivalent employment (FTE), which, at the store level is defined by the equation:

$$FTE_i = FullTimeEmployees_i + Managers_i + \frac{1}{2}PartTimeEmployees_i$$

I create that variable and then group the data by state and period (after and before the treatment, that happened in 1992) to build Panel 1 (Stores by state) of Table 3.

That panel shows the main results of the study, where can be seen that:

- (1) Before the treatment, stores in PA were, on average, bigger in terms of FTE than those of New Jersey
- (2) The difference became smaller after the treatment
- (3) FTE fell in PA after the treatment, while in NJ it remained constant or slightly increased

Table 16: Replication of Rows 1-3 of Table 3 of Card and Krueger (1994)

Variable	Stores By State			Stores in NJ			Differences in NJ	
	PA	NJ	NJ-PA	Wage=4.25	Wage=4.26-4.99	Wage>4.99	Low-high	Midrange-high
FTE before	23.33 (1.35)	20.44 (0.51)	-2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	-2.69 (1.37)	-2.17 (1.41)
FTE after	21.17 (0.94)	21.03 (0.52)	-0.14 (1.08)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.66 (1.44)	0.74 (1.27)
Change in mean FTE	-2.17 (1.65)	0.59 (0.73)	2.75 (1.34)	1.32 (1.13)	0.87 (1.53)	-2.04 (1.27)	3.36 (1.3)	2.91 (1.22)

- (4) Most importantly, the diff-in-diff estimate, which under the Conditional Independence Assumption (or parallel trends) is a estimation of the causal effect of minimum wage on FTE, is non-negative (2.75 in our results)

Then I group the NJ data by wage category before the treatment: Low (Minimum wage=\$4.25), Midrange (\$4.26-\$4.99), and High (>\$4.99). I use this categories to replicate panel 2 (Stores in NJ) and panel 3 (Differences in NJ) of Table 3

In these panels we can see that, in general, stores with a high starting wage are bigger in terms of FTE than the rest before the treatment (that is seen in the negative numbers of panel 3). However, this changes after the increase in minimum wage since stores with a high starting wage before the treatment are the only ones whose average FTE falls after the increase in minimum wage (this can be seen in the positive numbers of panel 3). In row 3 of panel two we can see that average FTE did not fall neither for Low nor for Midrange wage stores, and only did for high-starting wage stores, (although the difference is not statistically significant).

By comparing with the original, you can see that both means and standard errors match except by the standard errors in the bottom row. These are the standard errors of the difference in means within each state. I got them by extracting the standard error associated to the t-test that the difference in means is equal to zero. The fact that I get different standard errors for that row is surprising as I am getting the correct ones for column 3 of Panel 1, which tests the difference in FTE means between NJ and PA. In any case, the standard errors don't change the main results, which are that both the difference in means within and between state and our diff-in-diff estimate are not significantly different zero, which is empirical evidence against the claim of Walrasian economics that an increase in minimum wage necessarily reduces employment.

Part 3: Advances in replication

The data consists of four main variables: trade flows, tariffs, rules origin and input-output (IO) information. As I said in problem set 2, IO is only used for robustness checks and I will replicate that part if I time allows me. I already gathered all the data and am in the cleaning process. Up to this point, I successfully cleaned the data on trade flows, which consists of the imports of Mexico from NAFTA partners and non-NAFTA countries. Below I show a replication of panels 1 and 2 of Table 2 of Conconi et al. (2018)

There we see some very well-known facts about the Mexican economy in the last decades, which are that since trade liberalization started un the 80s, trade volumes increased substantially in general but exponentially with NAFTA partners, specially with the US.

I am almost done with the cleaning of data on tariffs and rules of origin. Once I have them I will be able to replicate Tables 1 and 2 of descriptive statistics. With that, the replication of the main regressions should be very straightforward.

R code

```
library(dplyr)
library(tidyr)
```


Table 17: Trade: Panels 1 and 2 of Table 2 of Conconi et al. (2018)

HS Code and Description	Mexican imports non-NAFTA		Mexican imports NAFTA	
	1991	2003	1991	2003
01-05: animal products	105015.80	396708.2	974826.06	2300023.24
06-15: vegetables	163746.87	245054.6	1470589.97	4893730.51
16-24: foodstuffs	81746.57	133052.3	630233.92	2241186.71
25-27: mineral products	122199.61	718771.1	1295175.35	4932581.91
28-38: chemicals	166600.88	1194191.4	1844428.29	7416817.92
39-40: plastics/rubber	164092.39	1365552.9	1364731.88	11502199.59
41-43: raw hides, skins, leathers	22956.68	222964.6	183434.99	796171.69
44-49: wood products	39232.65	359856.6	1291545.32	4276956.29
50-63: textiles	325965.40	1468048.9	662750.51	6778142.79
64-67: footwear/headgear	82539.92	260928.9	72100.04	60278.68
68-71: stone/glass	39363.39	525586.3	244871.21	1492556.97
72-84: metals	192869.11	1585282.5	2045470.12	9408639.86
84-85: machinery/electrical	1224048.48	21999538.6	6540453.92	36770636.94
86-89: transportation	135931.48	1444667.6	1230502.75	11583126.23
90-97: miscellaneous	324508.53	1839537.2	1361375.09	5104762.24

```

library(data.table)
library(ggplot2)
library("rio")
library(matlib)
library(gdata)
library(tinytex)
library(car)
library(scales)
library(ggplot2)
library(foreign)
library(rmarkdown)
library(fastDummies)
library(haven)
library(pmdplyr)
library(plotrix)
library(foreign)

options(scipen=10000)
options(digits=4)

rm(list=ls())

setwd("C:/Users/User/Documents/GitHub/Problem-Sets--753/PS3")

```

```

#####
##### PART 1 #####

```

```
#####
## Question 1 ##
## PART a ##
#####

## Extract data
cps <- read.dta("http://courses.umass.edu/econ753/berndt/stata/chap5-cps.dta")

## Filter for years 1978 and 1985
cps78 <- filter(cps, year==1978)
cps85 <- filter(cps, year==1985)

## Exponentiate wage
cps78 <- mutate(cps78, wage=exp(lnwage))

## Get geometric and arithmetic mean
cps78.geometricMeanlnwage <- exp(mean(cps78$lnwage))
cps78.arithmeticMeanlnwage <- mean(cps78$wage)

## Annual means
cps78.annualGeometricMeanlnwage <- 2000 * cps78.geometricMeanlnwage
cps78.annualArithmeticMeanlnwage <- 2000 * cps78.arithmeticMeanlnwage

## Mean and standard deviation of ED and EX
cps78.meanEd <- mean(cps78$ed)
cps78.sdEd <- sd(cps78$ed)
cps78.meanEx <- mean(cps78$ex)
cps78.sdEx <- sd(cps78$ex)

## Make tables with info
cps78.lnwageMeanTable <- data.frame("Geometric Mean" = c(cps78.geometricMeanlnwage, cps78.annualGeometricMeanlnwage),
                                   "Arithmetic Mean" = c(cps78.arithmeticMeanlnwage, cps78.annualArithmeticMeanlnwage))
rownames(cps78.lnwageMeanTable) <- c("Hourly", "Annual")
colnames(cps78.lnwageMeanTable) <- c("Geometric Mean", "Arithmetic Mean")
save(cps78.lnwageMeanTable, file="cps78.lnwageMeanTable.Rdata")

cps78.edExTable <- data.frame(Mean = c(cps78.meanEd, cps78.meanEx),
                             "Standard Deviation" = c(cps78.sdEd, cps78.sdEx))
rownames(cps78.edExTable) <- c("Education", "Experience")
colnames(cps78.edExTable) <- c("Mean", "Standard Deviation")
save(cps78.edExTable, file="cps78.edExTable.Rdata")

#####
## Question 1 ##
## PART b ##
#####

## Sample size
cps78.sampleSize <- 550

## Means for demographic dummy variables
cps78.nonwh.mean <- mean(cps78$nonwh)
cps78.hisp.mean <- mean(cps78$hisp)
```

```

cps78.fe.mean <- mean(cps78$fe)

## Counts for demographic dummy variables
cps78.nonwh.count <- cps78.sampleSize * cps78.nonwh.mean
cps78.hisp.count <- cps78.sampleSize * cps78.hisp.mean
cps78.fe.count <- cps78.sampleSize * cps78.fe.mean

## Make table with info
cps78.demographicsTable <- data.frame(Mean = c(cps78.nonwh.mean, cps78.hisp.mean, cps78.fe.mean),
                                     Count = c(cps78.nonwh.count, cps78.hisp.count, cps78.fe.count))
rownames(cps78.demographicsTable) <- c("nonwh", "hisp", "fe")
save(cps78.demographicsTable, file="cps78.demographicsTable.Rdata")

#####
## Question 1 ##
## PART c ##
#####

## Filter to subgroups
cps78.male <- filter(cps78, fe==0)
cps78.female <- filter(cps78, fe==1)
cps78.wh <- filter(cps78, nonwh==0 & hisp==0)
cps78.nonwh <- filter(cps78, nonwh==1)
cps78.hisp <- filter(cps78, hisp==1)

## Table for Gender
cps78.genderTable <- data.frame("Mean ed" = c(mean(cps78.male$ed), mean(cps78.female$ed)),
                              "sd ed" = c(sd(cps78.male$ed), sd(cps78.female$ed)),
                              "Mean Hourly Wage" = c(exp(mean(cps78.male$lnwage)), exp(mean(cps78.female$lnwage))),
                              "sd Wage" = c(sd(cps78.male$wage), sd(cps78.female$wage)),
                              "Mean Annual Wage" = c(2000*exp(mean(cps78.male$lnwage)),
                                                    2000*exp(mean(cps78.female$lnwage))))
rownames(cps78.genderTable) <- c("Male", "Female")
colnames(cps78.genderTable) <- c("Mean", "SD", "Mean Hourly", "SD", "Mean Annual")
save(cps78.genderTable, file="cps78.genderTable.Rdata")

## Table for Race
cps78.raceTable <- data.frame("Mean ed" = c(mean(cps78.wh$ed), mean(cps78.nonwh$ed), mean(cps78.hisp$ed)),
                              "sd ed" = c(sd(cps78.wh$ed), sd(cps78.nonwh$ed), sd(cps78.hisp$ed)),
                              "Mean Hourly Wage" = c(exp(mean(cps78.wh$lnwage)), exp(mean(cps78.nonwh$lnwage)),
                                                    exp(mean(cps78.hisp$lnwage))),
                              "sd Wage" = c(sd(cps78.wh$wage), sd(cps78.nonwh$wage), sd(cps78.hisp$wage)),
                              "Mean Annual Wage" = c(2000*exp(mean(cps78.wh$lnwage)),
                                                    2000*exp(mean(cps78.nonwh$lnwage)),
                                                    2000*exp(mean(cps78.hisp$lnwage))))
rownames(cps78.raceTable) <- c("Whites", "NonWhites", "Hispanic")
colnames(cps78.raceTable) <- c("Mean", "SD", "Mean Hourly", "SD", "Mean Annual")
save(cps78.raceTable, file="cps78.raceTable.Rdata")

#####
## Question 1 ##
## PART d ##
#####

```

```

## Exponentiate wage
cps85 <- mutate(cps85, wage=exp(lnwage))

## Deflate wages by deflator
cps85.deflator <- 1.649
cps85 <- mutate(cps85, wage=wage/cps85.deflator)
cps85 <- mutate(cps85, lnwage=log(wage))

## Get geometric mean and standard deviation
cps85.geometricMeanlnwage <- exp(mean(cps85$lnwage))
cps85.sdwage <- sd(cps85$wage)

## Annual mean
cps85.annualGeometricMeanlnwage <- 2000 * cps85.geometricMeanlnwage

## Mean and standard deviation of ED
cps85.meanEd <- mean(cps85$ed)
cps85.sdEd <- sd(cps85$ed)

## Sample size
cps85.sampleSize <- 534

## Means for demographic dummy variables
cps85.nonwh.mean <- mean(cps85$nonwh)
cps85.hisp.mean <- mean(cps85$hisp)
cps85.fe.mean <- mean(cps85$fe)

## Counts for demographic dummy variables
cps85.nonwh.count <- cps85.sampleSize * cps85.nonwh.mean
cps85.hisp.count <- cps85.sampleSize * cps85.hisp.mean
cps85.fe.count <- cps85.sampleSize * cps85.fe.mean

## Make table with info
cps85.demographicsTable <- data.frame(Mean = c(cps85.nonwh.mean, cps85.hisp.mean, cps85.fe.mean),
                                     Count = c(cps85.nonwh.count, cps85.hisp.count, cps85.fe.count))
rownames(cps85.demographicsTable) <- c("nonwh", "hisp", "fe")
save(cps85.demographicsTable, file="cps85.demographicsTable.Rdata")

## Filter to subgroups
cps85.male <- filter(cps85, fe==0)
cps85.female <- filter(cps85, fe==1)
cps85.wh <- filter(cps85, nonwh==0 & hisp==0)
cps85.nonwh <- filter(cps85, nonwh==1)
cps85.hisp <- filter(cps85, hisp==1)

## Table for Gender
cps85.genderTable <- data.frame("Mean ed" = c(mean(cps85.male$ed), mean(cps85.female$ed)),
                                "sd ed" = c(sd(cps85.male$ed), sd(cps85.female$ed)),
                                "Mean Hourly Wage" = c(exp(mean(cps85.male$lnwage)), exp(mean(cps85.female$lnwage))),
                                "sd Wage" = c(sd(cps85.male$wage), sd(cps85.female$wage)),
                                "Mean Annual Wage" = c(2000*exp(mean(cps85.male$lnwage)),
                                                         2000*exp(mean(cps85.female$lnwage))))
rownames(cps85.genderTable) <- c("Male", "Female")

```

```

colnames(cps85.genderTable) <- c("Mean", "SD", "Mean Hourly", "SD", "Mean Annual")
save(cps85.genderTable, file="cps85.genderTable.Rdata")

## Table for Race
cps85.raceTable <- data.frame("Mean ed" = c(mean(cps85.wh$ed), mean(cps85.nonwh$ed), mean(cps85.hisp$ed)),
                             "sd ed" = c(sd(cps85.wh$ed), sd(cps85.nonwh$ed), sd(cps85.hisp$ed)),
                             "Mean Hourly Wage" = c(exp(mean(cps85.wh$lnwage)), exp(mean(cps85.nonwh$lnwage)),
                                                       exp(mean(cps85.hisp$lnwage))),
                             "sd Wage" = c(sd(cps85.wh$wage), sd(cps85.nonwh$wage), sd(cps85.hisp$wage)),
                             "Mean Annual Wage" = c(2000*exp(mean(cps85.wh$lnwage)),
                                                       2000*exp(mean(cps85.nonwh$lnwage)),
                                                       2000*exp(mean(cps85.hisp$lnwage))))

rownames(cps85.raceTable) <- c("Whites", "NonWhites", "Hispanic")
colnames(cps85.raceTable) <- c("Mean", "SD", "Mean Hourly", "SD", "Mean Annual")
save(cps85.raceTable, file="cps85.raceTable.Rdata")

## Table with totals for both years
cps78.totalTable <- data.frame("Mean ed" = cps78.meanEd,
                              "sd ed" = cps78.sdEd,
                              "Mean Hourly Wage" = cps78.geometricMeanlnwage,
                              "sd Wage" = sd(cps78$wage),
                              "Mean Annual Wage" = cps78.annualGeometricMeanlnwage)
rownames(cps78.totalTable) <- "Whole Sample"
colnames(cps78.totalTable) <- c("Mean", "SD", "Mean Hourly", "SD", "Mean Annual")
save(cps78.totalTable, file="cps78.totalTable.Rdata")
cps85.totalTable <- data.frame("Mean ed" = cps85.meanEd,
                              "sd ed" = cps85.sdEd,
                              "Mean Hourly Wage" = cps85.geometricMeanlnwage,
                              "sd Wage" = cps85.sd wage,
                              "Mean Annual Wage" = cps85.annualGeometricMeanlnwage)
rownames(cps85.totalTable) <- "Whole Sample"
colnames(cps85.totalTable) <- c("Mean", "SD", "Mean Hourly", "SD", "Mean Annual")
save(cps85.totalTable, file="cps85.totalTable.Rdata")

#####
## Question 1 ##
## PART e ##
#####

#####
## First for lnwage ##

## Calculate mean and standard deviation of lnwage
cps78.w <- mean(cps78$lnwage)
cps78.s <- sd(cps78$lnwage)

## Separate wage data into categories
cps78$lnwagecat.subgroup <- cut(cps78$lnwage,
                                breaks = c(-Inf,
                                             cps78.w - 2 * cps78.s,
                                             cps78.w - cps78.s,
                                             cps78.w,
                                             cps78.w + cps78.s,

```

```

                                cps78.w + 2 * cps78.s,
                                Inf))
cps78$lnwagecat <- factor(cps78$lnwagecat.subgroup,
                          labels = c("wi<=w-2s",
                                      "w-2s<wi<=w-s",
                                      "w-s<wi<=w",
                                      "w<wi<=w+s",
                                      "w+s<wi<=w+2s",
                                      "w+2s<wi"))

## Create table with counts and proportion of sample size
cps78.logNormalTable <- t(rbind(with(cps78, table(lnwagecat)),
                                with(cps78, table(lnwagecat))/cps78.sampleSize))
colnames(cps78.logNormalTable) <- c("Counts", "Proportion")
save(cps78.logNormalTable, file="cps78.logNormalTable.Rdata")

#####
##                               Next for wage                               ##

## Calculate mean and standard deviation of wage
cps78.w <- mean(cps78$wage)
cps78.s <- sd(cps78$wage)

## Separate wage data into categories
cps78$wagecat.subgroup <- cut(cps78$wage,
                              breaks = c(-Inf,
                                          cps78.w - 2 * cps78.s,
                                          cps78.w - cps78.s,
                                          cps78.w,
                                          cps78.w + cps78.s,
                                          cps78.w + 2 * cps78.s,
                                          Inf))
cps78$wagecat <- factor(cps78$wagecat.subgroup,
                        levels = c("(-Inf,2.8]",
                                    "(-0.453,2.8]",
                                    "(2.8,6.06]",
                                    "(6.06,9.32]",
                                    "(9.32,12.6]",
                                    "(12.6, Inf]"),
                        labels = c("wi<=w-2s",
                                    "w-2s<wi<=w-s",
                                    "w-s<wi<=w",
                                    "w<wi<=w+s",
                                    "w+s<wi<=w+2s",
                                    "w+2s<wi"))

## Create table with counts and proportion of sample size
cps78.normalTable <- t(rbind(with(cps78, table(wagecat)),
                              with(cps78, table(wagecat))/cps78.sampleSize))
colnames(cps78.normalTable) <- c("Counts", "Proportion")
save(cps78.normalTable, file="cps78.normalTable.Rdata")

## Chi-squared tests

```

```

testlnwage <- chisq.test(cps78.logNormalTable[,1],
                        p=c(pnorm(-2),pnorm(-1)-pnorm(-2),pnorm(0)-pnorm(-1),
                           pnorm(1)-pnorm(0),pnorm(2)-pnorm(1),1-pnorm(2)))

testwage <- chisq.test(cps78.normalTable[,1],
                      p=c(pnorm(-2),pnorm(-1)-pnorm(-2),pnorm(0)-pnorm(-1),
                         pnorm(1)-pnorm(0),pnorm(2)-pnorm(1),1-pnorm(2)))

chiTestTable <- data.frame(Statistic=c(testlnwage$Statistic,testwage$Statistic),
                           pValue=c(testlnwage$p.value,testwage$p.value))
rownames(chiTestTable) <- c("lnwage", "wage")
save(chiTestTable, file="chiTestTable.Rdata")

## Below is just stuff I was practicing with that I want to keep

# dnorm(0)
#
# pnorm(0)-pnorm(-1)
#
# seq(cps78.w-3*cps78.s,cps78.w+3*cps78.s, by=0.01),
# dnorm(seq(cps78.w-3*cps78.s,cps78.w+3*cps78.s, by=0.01),
#       mean=cps78.w,sd=cps78.s)
#
#
#
# plot(seq(cps78.w-3*cps78.s,cps78.w+3*cps78.s, by=0.01),
#       dnorm(seq(cps78.w-3*cps78.s,cps78.w+3*cps78.s, by=0.01),
#             mean=cps78.w,sd=cps78.s))
#
# ggplot(cps78, aes(x=lnwage)) + geom_density()
#
#
# , y=densityPlot(lnwage)

#####
## Question 6 ##
## PART a ##
#####

cps78.lnwage.lm1 <- lm(lnwage ~ fe + union + nonwh + hisp + ed + ex + exsq, data=cps78)
save(cps78.lnwage.lm1, file="cps78.lnwage.lm1.Rdata")

linearHypothesis(cps78.lnwage.lm1, c("nonwh = hisp"))
str(linearHypothesis)

table6achow<-data.frame(2.46,0.12)
colnames(table6achow)<-c("F statistic", "P-value")

save(table6achow,file="table6achow.Rdata")

#####
## Question 6 ##
## PART b ##

```

```
#####
```

```
## Counts table
```

```
cps78.demographicsTable6b <- data.frame(Mean = c(1-cps78.nonwh.mean-cps78.hisp.mean,
                                                cps78.nonwh.mean,
                                                cps78.hisp.mean),
                                         Count = c((1-cps78.nonwh.mean-cps78.hisp.mean)*cps78.sampleSize,
                                                    cps78.nonwh.count,
                                                    cps78.hisp.count))
rownames(cps78.demographicsTable6b) <- c("other", "nonwh", "hisp")
save(cps78.demographicsTable6b, file="cps78.demographicsTable6b.Rdata")
```

```
cps78.raceTable6b <- rbind(with(cps78.wh, data.frame(ed=mean(ed),
                                                    diff=mean(ed)-mean(cps78.wh$ed),
                                                    ex=mean(ex),
                                                    diff=mean(ex)-mean(cps78.wh$ex),
                                                    fe=mean(fe),
                                                    diff=mean(fe)-mean(cps78.wh$fe),
                                                    union=mean(union),
                                                    diff=mean(union)-mean(cps78.wh$union),
                                                    lnwage=mean(lnwage),
                                                    diff=mean(lnwage)-mean(cps78.wh$lnwage))),
                           with(cps78.nonwh, data.frame(ed=mean(ed),
                                                         diff=mean(ed)-mean(cps78.wh$ed),
                                                         ex=mean(ex),
                                                         diff=mean(ex)-mean(cps78.wh$ex),
                                                         fe=mean(fe),
                                                         diff=mean(fe)-mean(cps78.wh$fe),
                                                         union=mean(union),
                                                         diff=mean(union)-mean(cps78.wh$union),
                                                         lnwage=mean(lnwage),
                                                         diff=mean(lnwage)-mean(cps78.wh$lnwage))),
                           with(cps78.hisp, data.frame(ed=mean(ed),
                                                         diff=mean(ed)-mean(cps78.wh$ed),
                                                         ex=mean(ex),
                                                         diff=mean(ex)-mean(cps78.wh$ex),
                                                         fe=mean(fe),
                                                         diff=mean(fe)-mean(cps78.wh$fe),
                                                         union=mean(union),
                                                         diff=mean(union)-mean(cps78.wh$union),
                                                         lnwage=mean(lnwage),
                                                         diff=mean(lnwage)-mean(cps78.wh$lnwage))))
rownames(cps78.raceTable6b) <- c("Others", "NonWhites", "Hispanic")
save(cps78.raceTable6b, file="cps78.raceTable6b.Rdata")
```

```
#####
```

```
## Question 6 ##
```

```
## PART c ##
```

```
#####
```

```
cps78.wh.lnwage.lm1 <- lm(lnwage ~ fe + union + ed + ex + exsq, data=cps78.wh)
cps78.nonwh.lnwage.lm1 <- lm(lnwage ~ fe + union + ed + ex + exsq, data=cps78.nonwh)
```



```

cps78.hisp.lnwage.lm1 <- lm(lnwage ~ fe + union + ed + ex + exsq, data=cps78.hisp)
save(cps78.wh.lnwage.lm1, file="cps78.wh.lnwage.lm1.Rdata")
save(cps78.nonwh.lnwage.lm1, file="cps78.nonwh.lnwage.lm1.Rdata")
save(cps78.hisp.lnwage.lm1, file="cps78.hisp.lnwage.lm1.Rdata")

#####
## Question 6 ##
## PART d ##
#####

cps78.lm1 <- lm(lnwage ~ fe + union + ed + ex + exsq, data=cps78)
str(cps78.lm1)

Sc<-deviance(cps78.lm1)

Swh<-deviance(cps78.wh.lnwage.lm1)
Snwh<-deviance(cps78.nonwh.lnwage.lm1)
Shisp<-deviance(cps78.hisp.lnwage.lm1)
k<-5

Nwh<-length(cps78.wh)
Nnwh<-length(cps78.nonwh)
Nhisps<-length(cps78.hisp)

F_stat<-((Sc-(Swh+Snwh+Shisp))/k)/((Swh+Snwh+Shisp)/(Nwh+Nnwh+Nhisps-2*k))

f_pvalue<-pf(F_stat, k, Nwh+Nnwh+Nhisps-2*k)

#Make table RSS

table6dRSS<- data.frame(c("Complete", "White", "Hisps", "Non-white"), c(Sc, Swh, Shisp, Snwh))
colnames(table6dRSS)<-c("Sample", "RSS")

#Make table Chow test

table6dchow<-data.frame(k, Nwh+Nnwh+Nhisps-2*k, F_stat, f_pvalue)
colnames(table6dchow)<-c("DF 1", "DF 2", "F statistic", "P-Value")

save(table6dRSS, file="table6dRSS.Rdata")
save(table6dchow, file="table6dchow.Rdata")

#####
#####
#####

##### PART2 #####
#####
#####
#####

```

```
#####

rm(list=ls())

#Replicate rows 1-3 of the panel "Stores by State" of Table 3 in Card and Krueger, "Minimum Wages and Employment:
#A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania" (AER 84(4) 1994). The material
#git provides the Card and Krueger data in several formats. Discuss your method and findings in relation

fastfood<-read_dta("fastfood.dta")
fastfood<- fastfood %>% mutate(state=ifelse(state==1,"NJ","PA"))

fastfood_empft<-fastfood %>% select(sheet, state, empft, empft2 )
fastfood_emppt<-fastfood %>% select(sheet, state, emppt, emppt2)

fastfood_wage<-fastfood %>% select(sheet, state, wage_st, wage_st2)
fastfood_wage<-fastfood_wage %>%mutate(wage_cat= ifelse(wage_st==4.25,"L",
  ifelse(wage_st<5,"M",
  ifelse(wage_st>=5,"H",NA))))

fastfood_man<-fastfood %>% select(sheet, state, nmgrs, nmgrs2)

#Pivot Longer empft

fastfood_empft<-fastfood_empft %>% pivot_longer(c(empft,empft2), values_to= "empft", names_to="post")
fastfood_empft<-fastfood_empft %>% mutate(post=ifelse(post=="empft", 0,1))

#Pivot Longer emppt
fastfood_emppt<-fastfood_emppt %>% pivot_longer(c(emppt,emppt2), values_to= "emppt", names_to="post")
fastfood_emppt<-fastfood_emppt %>% mutate(post=ifelse(post=="emppt", 0,1))

#Pivot Longer wage

fastfood_wage<-fastfood_wage %>% pivot_longer(c(wage_st,wage_st2), values_to= "wage_st", names_to="post")
fastfood_wage<-fastfood_wage %>% mutate(post=ifelse(post=="wage_st", 0,1))

#Pivot Longer managers and assistants

fastfood_man<-fastfood_man %>% pivot_longer(c(nmgrs,nmgrs2), values_to= "nmgrs", names_to="post")
fastfood_man<-fastfood_man %>% mutate(post=ifelse(post=="nmgrs", 0,1))

# merge everything
fastfood2<-merge(merge(fastfood_empft,fastfood_emppt),merge(fastfood_wage,fastfood_man))

#Panel Means 1
fastfood2<-fastfood2 %>% mutate(fte1=empft+0.5*emppt,
  fte2=empft+nmgrs+0.5*emppt)
fte<-fastfood2 %>% group_by(post,state) %>% summarise(across(fte2, mean ,na.rm=TRUE)) %>% ungroup()
fte<-fte %>% pivot_wider(names_from=state,values_from=fte2) %>% mutate(NJminusPA=NJ-PA)
```

```

fte<-fte %>% rbind(fte[2,]-fte[1,])

#Panel 2 Means
w<-fastfood2 %>% group_by(post, state, wage_cat) %>% summarise(across(fte2, mean, na.rm=TRUE)) %>%
  ungroup() %>% filter(state=="NJ")

w<-w %>% pivot_wider(names_from=wage_cat, values_from=fte2)
w<-w %>% select(-c(post,state,"NA"))
w<-w %>% rbind(w[2,]-w[1,])

#Panel 3 Means
w.between<-w %>% mutate(LminusH=L-H, MminusH=M-H)
w.between<-w.between %>% select(LminusH, MminusH)

#### Standard errors!
### Generate individual variables

fteNJ0<-filter(fastfood2, state=="NJ",post==0)$fte2
fteNJ1<-filter(fastfood2, state=="NJ",post==1)$fte2
ftePA0<-filter(fastfood2, state=="PA",post==0)$fte2
ftePA1<-filter(fastfood2, state=="PA",post==1)$fte2

fteNJ0L<-filter(fastfood2, state=="NJ",post==0, wage_cat=="L")$fte2
fteNJ1L<-filter(fastfood2, state=="NJ",post==1, wage_cat=="L")$fte2

fteNJ0M<-filter(fastfood2, state=="NJ",post==0, wage_cat=="M")$fte2
fteNJ1M<-filter(fastfood2, state=="NJ",post==1, wage_cat=="M")$fte2

fteNJ0H<-filter(fastfood2, state=="NJ",post==0, wage_cat=="H")$fte2
fteNJ1H<-filter(fastfood2, state=="NJ",post==1, wage_cat=="H")$fte2

fte.std<-fastfood2 %>% group_by(state,post) %>% summarise(across(fte2, .fns=list(std=std.error),.names=
w.std<-fastfood2 %>% group_by(post, state, wage_cat) %>%
  summarise(across(fte2, .fns=list(std=std.error),.names="{fn}.{col}", na.rm=TRUE)) %>%
  ungroup() %>% filter(state=="NJ")

fte.std<-fte.std %>% pivot_wider(names_from=state,values_from=std.fte2)
w.std<-w.std%>% pivot_wider(names_from=wage_cat, values_from=std.fte2)

### Add the standard error of the differences

#Panel 1 differences standard errors
diff_within_state<-c(1, t.test(fteNJ0,fteNJ1)$stderr,t.test(ftePA0,ftePA1)$stderr, t.test(fteNJ0-fteNJ
diff_between_state<-c(t.test(fteNJ0,ftePA0)$stderr,t.test(fteNJ1,ftePA1)$stderr)

```

```

#Panel 2 differences standard errors

diff_within_wagecat<-c(t.test(fteNJOL,fteNJ1L)$stderr,t.test(fteNJOM,fteNJ1M)$stderr,t.test(fteNJOH,fteNJ1H)$stderr)

#colnames(diff_within_state)<-colnames(fte.std)
#fte.std<-fte.std %>% mutate(NJminusPA=diff_between_state)
#fte.std<-fte.std %>% rbind(diff_within_state)

#Panel 3 standard errors

diff_between_wagecat0<-c(t.test(fteNJOL,fteNJOH)$stderr,t.test(fteNJOM,fteNJOH)$stderr)
diff_between_wagecat1<-c(t.test(fteNJ1L,fteNJ1H)$stderr,t.test(fteNJ1M,fteNJ1H)$stderr)
dd_time_wagecat<-c(t.test(fteNJ1L-fteNJOL,fteNJ1H-fteNJOH)$stderr, t.test(fteNJ1M-fteNJOM,fteNJ1H-fteNJ1M)$stderr)

### Putting everything together
#Panel1
fte.std<-fastfood2 %>% group_by(state,post) %>% summarise(across(fte2, .fns=list(std=std.error),.names=
fte.std<-fte.std %>% pivot_wider(names_from=state,values_from=std.fte2)

fte.std<-fte.std %>% mutate(NJminusPA=diff_between_state)
fte.std<-fte.std %>% rbind(diff_within_state)

###Panel 2

w.std<-fastfood2 %>% group_by(post, state, wage_cat) %>%
  summarise(across(fte2, .fns=list(std=std.error),.names="{fn}.{col}", na.rm=TRUE)) %>%
  ungroup() %>% filter(state=="NJ")

w.std<-w.std %>% pivot_wider(names_from=wage_cat, values_from=std.fte2)
w.std<-w.std %>% select(-c(post, state,"NA"))
w.std<-w.std %>% rbind(diff_within_wagecat)

###Panel 3

w.std.between<-data.frame(t(diff_between_wagecat0))
w.std.between<-w.std.between%>% rbind(diff_between_wagecat1,dd_time_wagecat)

## Putting everything together

## Format Panel 1
## Pasting \n in between means and standard deviations for LaTeX
x1<-as.data.frame(do.call(cbind,lapply(1:ncol(fte),
function(i) paste0(round(fte[1,i],digits=2),
                    "\n",
                    "(",
                    round(fte.std[1,i],digits=2),
                    ")"))))

```

```

x2<-as.data.frame(do.call(cbind,lapply(1:ncol(fte),
                                     function(i) paste0(round(fte[2,i],digits=2),
                                                         "\n",
                                                         "(",
                                                         round(fte.std[2,i],digits=2),
                                                         ")"))))

x3<-as.data.frame(do.call(cbind,lapply(1:ncol(fte),
                                     function(i) paste0(round(fte[3,i],digits=2),
                                                         "\n",
                                                         "(",
                                                         round(fte.std[3,i],digits=2),
                                                         ")"))))

panel_1 <-x1 %>% rbind(x2,x3)
rm(x1,x2,x3)

## Format panel2
## Pasting \n in between means and standard deviations for LaTeX
y1<-as.data.frame(do.call(cbind,lapply(1:ncol(w),
                                     function(i) paste0(round(w[1,i],digits=2),
                                                         "\n",
                                                         "(",
                                                         round(w.std[1,i],digits=2),
                                                         ")"))))

y2<-as.data.frame(do.call(cbind,lapply(1:ncol(w),
                                     function(i) paste0(round(w[2,i],digits=2),
                                                         "\n",
                                                         "(",
                                                         round(w.std[2,i],digits=2),
                                                         ")"))))

y3<-as.data.frame(do.call(cbind,lapply(1:ncol(w),
                                     function(i) paste0(round(w[3,i],digits=2),
                                                         "\n",
                                                         "(",
                                                         round(w.std[3,i],digits=2),
                                                         ")"))))

panel_2<-y1 %>% rbind(y2,y3)
rm(y1,y2,y3)

## Format panel 3
## Pasting \n in between means and standard deviations for LaTeX
z1<-as.data.frame(do.call(cbind,lapply(1:ncol(w.between),
                                     function(i) paste0(round(w.between[1,i],digits=2),
                                                         "\n",
                                                         "(",
                                                         round(w.std.between[1,i],
                                                         digits=2),
                                                         ")"))))

z2<-as.data.frame(do.call(cbind,lapply(1:ncol(w.between),
                                     function(i) paste0(round(w.between[2,i],digits=2),
                                                         "\n",
                                                         "(",
                                                         round(w.std.between[2,i],digits=2),

```

```

        "))))))
z3<-as.data.frame(do.call(cbind,lapply(1:ncol(w.between),
                                     function(i) paste0(round(w.between[3,i],digits=2),
                                                           "\n",
                                                           "(",
                                                           round(w.std.between[3,i],digits=2),
                                                           ")"))))

panel_3<-z1 %>% rbind(z2,z3)
rm(z1,z2,z3)

## Panel 1 final
colnames(panel_1)<-c("Variable","NJ","PA","NJ-PA")
panel_1<-panel_1 %>% relocate(Variable,PA) %>% mutate(Variable=c("FTE before","FTE after", "Change in m

## Panel 2 final
colnames(panel_2)<-c("Wage>4.99", "Wage=4.25","Wage=4.26-4.99")
panel_2<-panel_2 %>% mutate(Variable=c("FTE before","FTE after", "Change in mean FTE"))
panel_2<-panel_2 %>% relocate(Variable, `Wage=4.25`, `Wage=4.26-4.99`, `Wage>4.99`)

## Panel 3 final
colnames(panel_3)<-c("Low-high","Midrange-high")
panel_3<-panel_3 %>% mutate(Variable=c("FTE before","FTE after", "Change in mean FTE"))
panel_3<-panel_3 %>% relocate(Variable)

#####
##### REPLICATION #####

#Jesús Lara
#Critical Replication of Conconi et al (2018)

#Trade: IMPORTS
library(dplyr)
library(tidyr)
library(data.table)
library(ggplot2)
library("rio")
library(matlib)
library(gdata)
library(tinytex)
library(scales)
library(ggplot2)
library(foreign)
library(rmarkdown)
library(fastDummies)
library(haven)
library(pmdplyr)

memory.limit(size=10000)
options(scipen=10000)

```

```

options(digits=3)

rm(list=ls())

setwd("C:/Users/User/Documents/Fall 2020 UMass/753/Replication/data")

Mex_imports<-read_dta("./trade/raw-data/DTA_Mexico19912003-WTO-countries.dta")
HS92conversion<-read_dta("./trade/raw-data/HS92conversion.dta")

Mex_imports<-Mex_imports %>% filter(PartnerName!="Switzerland"&
                                   PartnerName!="Iceland"&
                                   PartnerName!="Norway"&
                                   PartnerName!="WTO All members (123) --- WTO-ALL")

Mex_imports<-Mex_imports %>%
  select(-c(SelectedNomen,TradeSource,DutyType,SimpleAverage,
            WeightedAverage,MinimumRate,MaximumRate,TariffYear,
            ProductName))

Mex_imports<-Mex_imports %>%
  rename("year"="TradeYear",
         "imports"="ImportsValuein1000USD")
### In the author's version this is TEMP_regression.dta

#Convert HS2002 to HS1992

Mex_imports2<-Mex_imports %>%
  filter(year==2003) %>%
  rename("HS2002"="Product")

Mex_imports2<-merge(Mex_imports2,HS92conversion, all.x = TRUE)
Mex_imports2<-Mex_imports2 %>%
  rename("Product"="HS1992")

#Duplicates SUM total import flows

Mex_imports2 <-Mex_imports2 %>%
  group_by(Product, PartnerName, Partner, year) %>%
  summarise(across(imports,sum)) %>% ungroup()

#Drop 2003 of Mex_imports
Mex_imports<-Mex_imports %>% filter(year!=2003)

Mex_imports_F<-bind_rows(Mex_imports2,Mex_imports)

Mex_imports_F <- Mex_imports_F %>% mutate(NAFTA=ifelse(PartnerName=="United States"|
                                                       PartnerName=="Canada",1,0))

```

```

Mex_imports_F <- Mex_imports_F %>% mutate(RTAmex=ifelse(PartnerName=="Chile" |
                                                         PartnerName=="Israel" |
                                                         PartnerName=="Austria" |
                                                         PartnerName=="Belgium" |
                                                         PartnerName=="Denmark" |
                                                         PartnerName=="Finland" |
                                                         PartnerName=="France" |
                                                         PartnerName=="Germany" |
                                                         PartnerName=="Ireland" |
                                                         PartnerName=="Italy" |
                                                         PartnerName=="Luxembourg" |
                                                         PartnerName=="Netherlands" |
                                                         PartnerName=="Portugal" |
                                                         PartnerName=="Spain" |
                                                         PartnerName=="Sweden" |
                                                         PartnerName=="United Kingdom" |
                                                         PartnerName=="Greece" |
                                                         PartnerName=="Brazil" |
                                                         PartnerName=="Argentina" |
                                                         PartnerName=="Uruguay" |
                                                         PartnerName=="Paraguay" |
                                                         PartnerName=="Cuba" |
                                                         PartnerName=="Bolivia" |
                                                         PartnerName=="Ecuador" |
                                                         PartnerName=="Colombia" |
                                                         PartnerName=="Peru" |
                                                         PartnerName=="Guatemala" |
                                                         PartnerName=="Panama" |
                                                         PartnerName=="Venezuela"
                                                         ,1,0))

Mex_imports_F<-Mex_imports_F %>%
  select(-c(NativeNomen,Reporter,ReporterName))

Mex_imports_F<-Mex_imports_F %>%
  rename("product"="Product")

rm(Mex_imports,Mex_imports2)

###Make Panel A of Tab 2
Mex_imports_F<-Mex_imports_F%>%
  mutate(product_2_digits= ifelse(product<60000,"01-05: animal products",
                                   ifelse(product<160000,"06-15: vegetables",
                                   ifelse(product<250000,"16-24: foodstuffs",
                                   ifelse(product<280000,"25-27: mineral products",
                                   ifelse(product<390000,"28-38: chemicals",
                                   ifelse(product<410000,"39-40: plastics/rubber",
                                   ifelse(product<440000,"41-43: raw hides, skins, leathers",
                                   ifelse(product<500000,"44-49: wood products",
                                   ifelse(product<640000,"50-63: textiles",
                                   ifelse(product<680000,"64-67: footwear/headgear",
                                   ifelse(product<720000,"68-71: stone/glass",
                                   ifelse(product<840000,"72-84: metals",
                                   ifelse(product<860000,"84-85: machinery/electrical",

```



```

        ifelse(product<900000,"86-89: transportation",
        ifelse(product<990000,"90-97: miscellaneous",NA))))))
    )))))))

```

Descriptive Statistics:

NAFTA COUNTRIES

```

s_Mex_imports_NAFTA<-Mex_imports_F %>% filter(NAFTA==1) %>%
  group_by(year, product_2_digits) %>%
  summarise(across(imports, sum))

```

```

s_Mex_imports_NAFTA <- s_Mex_imports_NAFTA %>%
  pivot_wider(names_from = year, values_from=imports)

```

No NAFTA COUNTRIES

```

s_Mex_imports_NoNAFTA<-Mex_imports_F %>% filter(RTAmex==0 & NAFTA==0) %>%
  group_by(year, product_2_digits) %>%
  summarise(across(imports, sum))

```

```

s_Mex_imports_NoNAFTA <- s_Mex_imports_NoNAFTA %>%
  pivot_wider(names_from = year, values_from=imports)

```

###Preliminary Table 2

```

Table2_rep<-cbind(s_Mex_imports_NoNAFTA,(select(s_Mex_imports_NAFTA,-product_2_digits)))
colnames(Table2_rep)<-c("HS Code and Description","1991","2003","1991","2003")

```

```

save(Table2_rep,file="Table2_rep.Rdata")

```

Save all tables

```

save(panel_1,file="panel_1.Rdata")
save(panel_2,file="panel_2.Rdata")
save(panel_3,file="panel_3.Rdata")

```