Assignment 2b

STEP 1: Setting my working directory and loading required packages in R

```
rm(list = ls())         #clears environment
directory <- setwd("~/Documents/SFU courses/Spring 2022/Econ836 - A.
Econometrics/Assignm") #sets working directory

library(haven)          #needed to load the .dta stata data file
library(tidyverse)      #needed for datacleaning

library(quantreg)

library(labelled)       #needed for labels
library(modelsummary)   #needed for tables
library(data.table)

library(stargazer)
```

STEP 2: Loading the dataset

Here I will load the *2016 Census Public Use Microdata File (PUMF) needed for the project*

```
d.s <- read_dta("~/Documents/SFU courses/Spring 2022/Econ836 - A.
Econometrics/Assignm/Census_2016_Individual_PUMF.dta")
```

STEP 3: Cleaning the dataset

Here I will clean all 8 variables that are needed in my analysis which have missing data and require a reference level for the dummies:

1. EmpIn: this is the variable that defines the "income" and has data that is not available is labelled as "8,888,888" and data that is not applicable is labelled as "99,999,999" so here I will replace those observations with NA

2. agegrp: this is the variable that defines the "agegroup" and has data that is not available is labelled as "88", so here I will replace those observations with NA.

   The "left out" category for agegroup is "25-29 years" which is coded as 9 in the data, so I put this as my reference level

3. MatStH: this is the variable that defines "martial status" and There are no unavailable data in this category.

   The "left out" category for "martial status" is the never-married individuals which is coded as 1 in the data, so I put this as my reference level

4. hdegree: this is the variable that defines "education" and and has data that is not available is labelled as "88" and is not applicable is labelled as "99", so here I will replace those observations with NA.

The “left out” category for “education” are individuals with a highschool degree which is coded as 2 in the data, so I put this as my reference level

5. hhsize:this is the variable that defines “household size” and and has data that is not available is labelled as “8”, so here I will replace those observations with NA.

6. kol: this is the variable that defines “knowledge language” and this has data that is not available labelled as “8”, so here I replace those observations with NA.

   The “left out” category for “knowledge language” are individuals with “English Only” which is coded as 1 in the data, so I put this as my reference level

7. cma: this defines “*census metropolitan areas”* and here the left out catagory are the”non-cmas” which is coded as 999, so I put this as the reference level

8. abor: I create this variable to correctly capture the aboriginal people

```
#filtering out "Income" (income)----------------------------------------------------
--------------------
class(d.s$EmpIn) #checking the class of the variable: numeric
d.s <- d.s %>% rename(income=EmpIn) %>%
  mutate(income = case_when(
    income %in% c(88888888,99999999) ~ as.numeric(NA),
    TRUE ~ as.numeric(income)
  ))

#filtering out "Agegroup" (agegrp) ------------------------------------------
----------------------
class(d.s$agegrp) #checking the class of the variable: numeric
d.s <- d.s %>%
  mutate(agegrp = case_when(
    agegrp == 88 ~ as.double(NA),
    TRUE ~ as.double(agegrp)
  ))
d.s$agegrp <- factor(d.s$agegrp) #converting agegroup to factor/catagorical
variable

d.s$agegrp <- relevel(d.s$agegrp, ref=9) #taking the correct reference level
(25-29)
d.s <- d.s %>% rename(age=agegrp) #renaming agegrp to age


#converting MarStH to a factor and releveling it (marital)-------------------
-------------------
class(d.s$MarStH)
d.s$MarStH <- factor(d.s$MarStH)
d.s$MarStH <- relevel(d.s$MarStH, ref=1) #taking the correct reference level
(single people)
d.s <- d.s %>% rename(marital=MarStH)      #renaming MarStH to marital
```

```
#filtering out education (school)----------------------------------------------------
------------------

class(d.s$hdgree)
d.s <- d.s %>%
  mutate(hdgree = case_when(
    hdgree %in% c(88,99) ~ as.double(NA),
    TRUE ~ as.double(hdgree)
  ))
d.s$hdgree <- factor(d.s$hdgree)
d.s$hdgree <- relevel(d.s$hdgree, ref=2) #taking the correct reference level
(highschool)
d.s <- d.s %>% rename(school=hdgree)      #renaming hdgree to school

#filtering out household size (hhsize)---------------------------------------------
---------------------

class(d.s$hhsize)

d.s <- d.s %>%
  mutate(hhsize = case_when(
    hhsize ==8 ~ as.double(NA),
    TRUE ~ as.double(hhsize)
  ))
d.s$hhsize <- factor(d.s$hhsize)
levels(d.s$hhsize) #checking the level

#filtering the language catagory (lang) --------------------------------------------
------------------

class(d.s$kol) #checking the class: double
d.s <- d.s %>%
  mutate(kol = case_when(
    kol ==8 ~ as.double(NA),
    TRUE ~ as.double(kol)
  ))


d.s$kol <- factor(d.s$kol)
d.s$kol <- relevel(d.s$kol, ref=1) #setting the correct reference level
(English only)
d.s <- d.s %>% rename(lang=kol)  #renaming kol to lang

#releveling CMAs ------------------------------------------------------------
-------

class(d.s$cma)
```

```
d.s$cma <- factor(d.s$cma)
d.s$cma <- relevel(d.s$cma, ref='999')


#making the abor variable to include all catagories of indians --------------
----------------

#ethder=1 are the 'North American Aboriginal origins'
#regind=1 are the 'registered Indians'
d.s <-
  d.s %>%
  filter(ethder != 88) %>%
  mutate(
    abor = case_when(
      ethder != 1 ~ ethder,
      ethder == 1 & regind == 1 ~ as.double(999), # 999 is Registered
Indians,
      ethder == 1 & regind == 0 & aboid != 6 ~ aboid,
      ethder == 1 & regind == 0 & aboid == 6 ~ as.double(998) # 998 is
Ancestry
    )
  )

# Now give them labels

d.s <-
  d.s %>%
  mutate(abor =
    case_when(
      abor == 999 ~ 'Registered Indian',
      abor == 100 ~ 'North American Indian',
      abor == 200 ~ 'Métis',
      abor == 300  ~ 'Inuit',
      abor == 400  ~ 'Multiple',
      abor == 500  ~ 'Other aboriginal response',
      abor == 600 ~ 'Ancestry',
      TRUE ~ as.character(ethder),
    ),
    abor = as.factor(abor) %>% relevel('4')
  )

#Get the accurate sample size by filtering out only individuals who work for
someone else

d.s <-
  d.s %>%
  filter(income>0, cow==1)
```

Step 4: Get the correct samples for the regressions

```
#WHOLE OF CANADA

men_can <-
  d.s %>%
  filter(Sex == 2)

women_can <-
  d.s %>%
  filter(Sex ==1)

# MONTREAL

men_montreal <-
  d.s %>%
  filter(cma == 462,
         Sex == 2)

women_montreal <-
  d.s %>%
  filter(cma == 462,
         Sex == 1)

# TORONTO

men_toronto <-
  d.s %>%
  filter(cma == 535,
         Sex == 2)

women_toronto <-
  d.s %>%
  filter(cma == 535,
         Sex == 1)

# WINNIPEG

men_winnipeg <-
  d.s %>%
  filter(cma == 535,
         Sex == 2)

women_winnipeg <-
  d.s %>%
  filter(cma == 602,
         Sex == 1)

#CALGARY

men_calgary <-
```

```
  d.s %>%
  filter(cma == 825,
         Sex == 2)

women_calgary <-
  d.s %>%
  filter(cma == 825,
         Sex == 1)

# EDMONTON

men_edmonton <-
  d.s %>%
  filter(cma == 835,
         Sex == 2)

women_edmonton <-
  d.s %>%
  filter(cma == 835,
         Sex == 1)

# VANCOUVER

men_vancouver <-
  d.s %>%
  filter(cma == 933,
         Sex == 1)

women_vancouver <-
  d.s %>%
  filter(cma == 933,
         Sex == 2)
```

Step 5: Run the regressions

```
#Canada

reg_canada_men <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = men_can,
     tau = c(0.2,0.5,0.8,0.9))

reg_canada_women <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = women_can,
     tau = c(0.2,0.5,0.8,0.9))

#Men pseudo R-squared
```

```r
fit0 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = men_can)
fit1 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = men_can)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R1 <- 1 - fit1$rho/fit0$rho


#Women pseudo R-squared
fit2 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = women_can)

fit3 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = women_can)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R2 <- 1 - fit3$rho/fit2$rho

#Montreal

reg_men_montreal <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = men_montreal,
     tau = c(0.2,0.5,0.8,0.9))

reg_women_montreal <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = women_montreal,
     tau = c(0.2,0.5,0.8,0.9))

#men pseudo R-squared
fit4 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = men_montreal)
fit5 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = men_montreal)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R3 <- 1 - fit5$rho/fit4$rho

#women pseudo R-squared

fit6 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = women_montreal)
fit7 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = women_montreal)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R4 <- 1 - fit7$rho/fit6$rho

#Toronto

reg_men_toronto <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = men_toronto,
     tau = c(0.2,0.5,0.8,0.9))

reg_women_toronto <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
```

```
     data = women_toronto,
     tau = c(0.2,0.5,0.8,0.9))

#men pseudo R-squared
fit8 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = men_toronto)
fit9 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = men_toronto)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R5 <- 1 - fit9$rho/fit8$rho

#women pseudo R-squared

fit10 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = women_toronto)

fit11 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = women_toronto)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R6 <- 1 - fit7$rho/fit6$rho

#Winnipeg

reg_men_winnipeg <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = men_winnipeg,
     tau = c(0.2,0.5,0.8,0.9))

reg_women_winnipeg <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = women_winnipeg,
     tau = c(0.2,0.5,0.8,0.9))

#men pseudo R-squared
fit12 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = men_winnipeg)
fit13 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = men_winnipeg)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R7 <- 1 - fit13$rho/fit12$rho

#women pseudo R-squared

fit14 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = women_winnipeg)

fit15 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = women_winnipeg)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R8 <- 1 - fit15$rho/fit14$rho
#Calgary

reg_men_calgary <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
```

```r
     data = men_calgary,
     tau = c(0.2,0.5,0.8,0.9))

reg_women_calgary <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = women_calgary,
     tau = c(0.2,0.5,0.8,0.9))

#men pseudo R-squared
fit16 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = men_calgary)

fit17 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = men_calgary)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R9 <- 1 - fit17$rho/fit16$rho

#women pseudo R-squared

fit18 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = women_calgary)
fit19 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = women_calgary)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R10 <- 1 - fit19$rho/fit18$rho

#Edmonton

reg_men_edmonton <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = men_edmonton,
     tau = c(0.2,0.5,0.8,0.9))

reg_women_edmonton <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = women_edmonton,
     tau = c(0.2,0.5,0.8,0.9))

#men pseudo R-squared
fit20 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = men_edmonton)

fit21 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = men_edmonton)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R11 <- 1 - fit21$rho/fit20$rho

#women pseudo R-squared

fit22 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = women_edmonton)

fit23 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = women_edmonton)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R12 <- 1 - fit23$rho/fit22$rho
```

```r
#Vancouver

reg_men_vancouver <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = men_vancouver,
     tau = c(0.2,0.5,0.8,0.9))

reg_women_vancouver <-
  rq(log(income) ~ abor + marital + age + hhsize + school + lang,
     data = women_vancouver,
     tau = c(0.2,0.5,0.8,0.9))

#men pseudo R-squared
fit24 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = men_vancouver)
fit25 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = men_vancouver)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R13 <- 1 - fit25$rho/fit24$rho

#women pseudo R-squared

fit26 <- rq(log(income)~1,tau=c(0.2,0.5,0.8,0.9),data = women_vancouver)

fit27 <- rq(log(income)~abor,tau=c(0.2,0.5,0.8,0.9),data = women_vancouver)

rho <- function(u,tau=.5)u*(tau - (u < 0))
R14 <- 1 - fit27$rho/fit26$rho
```

Step 5: Creating output to extract relevant coefficients for each table

```r
#when youre taking the coefs it is already a data.frame

#canada

write.csv(reg_canada_men$coef, "1.reg_canada_men.csv")

write.csv(reg_canada_women$coef, "2.reg_canada_women.csv")

#montreal

write.csv(reg_men_montreal$coef, "3.reg_men_montreal.csv")

write.csv(reg_women_montreal$coef, "4.reg_women_montreal.csv")

#toronto

write.csv(reg_men_toronto$coef, "5.reg_men_toronto.csv")

write.csv(reg_women_toronto$coef, "6.reg_women_toronto.csv")
```

```
#winnipeg reg_men_winnipeg

write.csv(reg_men_winnipeg$coef, "7.reg_men_winnipeg.csv")

write.csv(reg_women_winnipeg$coef, "8.reg_women_winnipeg.csv")

#calgary

write.csv(reg_men_calgary$coef, "9.reg_men_calgary.csv")
write.csv(reg_women_calgary$coef, "10.reg_women_calgary.csv")

#edmonton reg_men_edmonton

write.csv(reg_men_edmonton$coef, "11.reg_men_edmonton.csv")
write.csv(reg_women_edmonton$coef, "12.reg_women_edmonton.csv")

#vancouver reg_men_vancouver


write.csv(reg_men_vancouver$coef, "13.reg_men_vancouver.csv")
write.csv(reg_women_vancouver$coef, "14.reg_women_vancouver.csv")
```

Step 6: Finally, I use excel to create the completed summary table

Comment: From the data table generated, we can see that for men in Canada in the bottom quintile the estimated disparities range from none (for those reporting single-origin Aboriginal ancestry) to 77% (for those who are registered Indians). Similar patterns for men broken int provinces. For those in the top quintile the estimated disparities range from none (for those reporting single-origin Aboriginal ancestry) to 24% (for those who are registered Indians). Similar patterns for men broken into provinces. There are however several outliers.
Similarly, for women in Canada in the bottom quintile the estimated disparities range from none (for those reporting single-origin Aboriginal ancestry) to 78% (for those who are registered Indians). Similar patterns for men broken int provinces. For those in the top quintile the estimated disparities range from none (for those reporting single-origin Aboriginal ancestry) to 37% (for those who are multiple-origin Aboriginal Ancestry). Similar patterns for women broken into provinces. There are however several outliers.

In summary, we can say for both men and women, the disparity at the bottom quintiles are worse and twice as severe as that at the top.

	Male					Female				
		Q20	Q50	Q80	Q90		Q20	Q50	Q80	Q90
	Pseudo R-squared	0.013	0.013	0.015	0.015	**Pseudo R-squared**	0.01	0.01	0.01	0.01
Canada	**Aborginal Ancestry (single)**	0.00	-0.08	-0.03	-0.03	**Aborginal Ancestry (single)**	-0.20	-0.12	-0.09	-0.08
	Metis	-0.15	-0.01	0.02	-0.01	**Metis**	-0.01	-0.01	-0.01	-0.09
	Aboriginal Ancestry (multiple)	-0.08	-0.17	-0.11	-0.23	**Aboriginal Ancestry (multiple)**	-0.28	-0.46	-0.62	-0.37
	North American Indian	-0.10	-0.10	-0.05	-0.11	**North American Indian**	-0.05	-0.05	-0.03	-0.11
	Registered Indians	-0.77	-0.43	-0.27	-0.24	**Registered Indians**	-0.78	-0.18	-0.12	-0.12
	Pseudo R-squared	0.015	0.020	0.021	0.020	**Pseudo R-squared**	0.019	0.018	0.016	0.012
Montreal	**Aborginal Ancestry (single)**	-0.03	-0.05	-0.06	-0.10	**Aborginal Ancestry (single)**	-0.10	-0.07	-0.28	-0.25
	Metis	-0.31	-0.12	-0.18	0.07	**Metis**	0.20	0.04	-0.21	-0.42
	Aboriginal Ancestry (multiple)	0.94	0.31	-0.13	-0.28	**Aboriginal Ancestry (multiple)**				
	North American Indian	-0.97	0.06	-0.12	-0.34	**North American Indian**	0.02	0.10	0.25	0.11
	Registered Indians	-0.21	-0.43	-0.48	-0.60	**Registered Indians**	-0.77	-0.46	-0.29	-0.41
	Pseudo R-squared	0.017	0.022	0.026	0.032	**Pseudo R-squared**	0.019	0.018	0.016	0.012
Toronto	**Aborginal Ancestry (single)**	-0.04	0.18	-0.24	-0.55	**Aborginal Ancestry (single)**	-1.02	-0.88	-1.14	-1.30
	Metis	0.46	0.08	-0.22	-0.22	**Metis**	1.34	0.61	0.18	-0.03
	Aboriginal Ancestry (multiple)	0.67	0.24	-0.16	-0.41	**Aboriginal Ancestry (multiple)**				
	North American Indian	0.14	-0.07	-0.23	-0.30	**North American Indian**	0.24	0.20	-0.07	-0.27
	Registered Indians	-0.34	-0.31	-0.21	-0.28	**Registered Indians**	-1.40	-0.26	-0.22	-0.05
	Pseudo R-squared	0.017	0.022	0.026	0.032	**Pseudo R-squared**	0.029	0.027	0.032	0.031
Winnipeg	**Aborginal Ancestry (single)**	-0.84	-0.26	0.58	0.40	**Aborginal Ancestry (single)**	-0.45	-0.84	-0.03	-0.06
	Metis					**Metis**	-0.95	-0.85	-0.21	-0.38
	Aboriginal Ancestry (multiple)	0.40	0.11	-0.16	-0.28	**Aboriginal Ancestry (multiple)**				
	North American Indian	-0.19	-0.16	-0.15	-0.16	**North American Indian**	0.17	0.48	-0.03	0.46
	Registered Indians	-0.34	-0.31	-0.21	-0.28	**Registered Indians**	-0.70	-0.48	-0.41	-0.38
	Pseudo R-squared	0.020	0.022	0.025	0.030	**Pseudo R-squared**	0.018	0.018	0.022	0.021
Calgary	**Aborginal Ancestry (single)**	0.24	0.01	-0.97	-0.96	**Aborginal Ancestry (single)**	0.66	0.14	-0.38	-0.31
	Metis	1.40	0.09	-0.38	-0.18	**Metis**	-0.38	0.10	0.03	0.29
	Aboriginal Ancestry (multiple)	2.31	0.38	-0.60	-0.60	**Aboriginal Ancestry (multiple)**				
	North American Indian	0.44	0.03	-0.48	-0.36	**North American Indian**	-0.44	-0.02	0.04	0.20
	Registered Indians	-0.22	-0.08	-0.11	-0.06	**Registered Indians**	-0.79	-0.43	-0.43	-0.49
	Pseudo R-squared	0.029	0.027	0.026	0.027	**Pseudo R-squared**	0.017	0.017	0.018	0.017
Edmonton	**Aborginal Ancestry (single)**	0.89	-0.09	0.05	0.14	**Aborginal Ancestry (single)**	0.99	0.52	0.64	0.73
	Metis	0.36	0.01	0.00	0.27	**Metis**	0.35	0.39	0.18	0.44
	Aboriginal Ancestry (multiple)	0.51	-0.11	-0.04	-0.02	**Aboriginal Ancestry (multiple)**	0.47	0.20	-0.19	-0.16
	North American Indian					**North American Indian**				
	Registered Indians	-0.66	-0.44	-0.34	-0.24	**Registered Indians**	-0.47	-0.22	-0.13	-0.08
	Pseudo R-squared	0.014	0.018	0.020	0.019	**Pseudo R-squared**	0.019	0.024	0.021	0.025
Vancouver	**Aborginal Ancestry (single)**	-0.17	-0.54	-0.43	-0.70	**Aborginal Ancestry (single)**	1.10	1.00	0.60	0.70
	Metis					**Metis**	1.65	1.90	1.22	1.37
	Aboriginal Ancestry (multiple)					**Aboriginal Ancestry (multiple)**				
	North American Indian	-0.01	0.14	0.05	-0.01	**North American Indian**	1.87	1.23	1.33	1.47
	Registered Indians	-0.94	-0.27	-0.16	-0.18	**Registered Indians**	-1.25	-0.31	0.02	-0.12