Week 3

Bhagya Gunawardena

20 March 2019

import data

```
# clear environment
rm(list = ls())
# load libaries
library(tidyverse)
library(lubridate)

#import data
data_01 <- read.csv("data_01.csv", header = TRUE, sep = ",", quote = "\"",
                    dec = ".", fill = TRUE, comment.char = "")
```

Organize data

```
#select males
data_02<-data_01%>%
  filter(sex==1)

#create dummies for education level
data_02<-data_02%>%
  mutate(edlvl = ifelse(deduc_1==1,1,ifelse(deduc_2==1,3,ifelse(deduc_3==1,4,ifelse(deduc_4==1,5,2)))))

attach(data_02)

#calculate the supply
sy = 1964
y=2017-sy+1


wagein=which( colnames(data_02)=="lrwage" ) #to find the colum of wages

hs_data=array(0,dim = c(y,3))
col_data=array(0,dim=c(y,3))
colnames(hs_data)<-c("year","hs_supp","hs_wage")
colnames(col_data)<-c("year","col_supp","col_wage")


#assign data
for (i in 1:y) {
  hs_data[i,1]=sy+i-1
  hs_data[i,2]=sum(edlvl==2 & year==(sy+i-1))
  hs_data[i,3]=sum(data_02[which(edlvl==2&year==(sy+i-1)),wagein])


  col_data[i,1]=sy+i-1
```

```
  col_data[i,2]=sum((edlvl==4|edlvl==5) & year==(sy+i-1))
  col_data[i,3]=sum(data_02[which((edlvl==4|edlvl==5)&year==(sy+i-1)),wagein])
}

hs_coldata<-merge(x=hs_data,y=col_data,by=c("year"))

#relative supply and relative wage in each year
hs_coldata<-hs_coldata%>%
  mutate(colavgwage = col_wage/col_supp)%>%
  mutate(hsavgwage = hs_wage/hs_supp)%>%
  mutate(relsupp=col_supp/hs_supp)%>%
 # mutate(relwage=col_wage/hs_wage)%>%
  mutate(relwage=colavgwage/hsavgwage)%>%
  mutate(ttime = year-1963)
```

"hs_coldata" has the required data for the analysis. Data is from 1964 as there is no data in 1963

Regression 1: From 1964-2017

```
summary(reg1<-lm(log(relwage) ~ log(relsupp)+ttime,data=hs_coldata))
```

```
##
## Call:
## lm(formula = log(relwage) ~ log(relsupp) + ttime, data = hs_coldata)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0247263 -0.0044574 -0.0000009  0.0052219  0.0236116
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.895e-02  3.103e-03   19.00  < 2e-16 ***
## log(relsupp) 1.204e-02  1.758e-03    6.85 9.38e-09 ***
## ttime        8.092e-04  8.067e-05   10.03 1.16e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009186 on 51 degrees of freedom
## Multiple R-squared:  0.7248, Adjusted R-squared:  0.714
## F-statistic: 67.16 on 2 and 51 DF,  p-value: 5.145e-15
```

Interpretation

In this regression, all three regressors (intercept, log relative supply and tim) are highly significant small positive values. However, the results are different to the Kats and Murpgy results. One main reason will be the considering time period as they consider from 1963-1987 and we considerfrom 1964-2017.

Regression 2: 1964:1987

```
summary(reg2<-lm(log(relwage) ~ log(relsupp)+ttime,data=subset(hs_coldata,year>=1964 & year<=1987)))
```

```
##
```

```
## Call:
## lm(formula = log(relwage) ~ log(relsupp) + ttime, data = subset(hs_coldata,
##     year >= 1964 & year <= 1987))
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0077615 -0.0034267  0.0000179  0.0027682  0.0078967
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0537744  0.0025154  21.378 9.85e-16 ***
## log(relsupp) 0.0038699  0.0021510   1.799   0.0864 .
## ttime        0.0006260  0.0001297   4.826 9.05e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004397 on 21 degrees of freedom
## Multiple R-squared:  0.5538, Adjusted R-squared:  0.5113
## F-statistic: 13.03 on 2 and 21 DF,  p-value: 0.0002088
```

Interpretation

In this regression, only intercept and time are highly significant.Again, the results are different to the Kats and Murpgy results. One main reason will be the considering time period as they consider from 1963-1987 and we considerfrom 1964-2017, as we do not have data for year 1963.

Regression 3: 1988-2017

```
hs_coldata<-hs_coldata%>%
  mutate(tttime=ttime-24)

summary(reg3<-lm(log(relwage) ~ log(relsupp)+tttime,data=subset(hs_coldata,year>=1988 & year<=2017)))
```

```
##
## Call:
## lm(formula = log(relwage) ~ log(relsupp) + tttime, data = subset(hs_coldata,
##     year >= 1988 & year <= 2017))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.029491 -0.001276  0.001430  0.003987  0.028029
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.064e-01  1.035e-02  10.280 7.83e-11 ***
## log(relsupp)  2.192e-02  3.811e-03   5.752 4.08e-06 ***
## tttime       -4.286e-05  3.446e-04  -0.124    0.902
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01019 on 27 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7345
## F-statistic: 41.11 on 2 and 27 DF,  p-value: 6.41e-09
```

Interpretation

In this regression, only intercept and log relative supply are highly significant.However, the results are different to the Kats and Murpgy results. One main reason will be the considering time period as they consider from 1963-1987 and we considerfrom 1988-2017.

delete below... experiment..

```
hs_coldata<-hs_coldata%>%
  # mutate(colavgwage = col_wage/col_supp)%>%
  # mutate(hsavgwage = hs_wage/hs_supp)%>%
  mutate(relsupp1=hs_supp/col_supp)%>%
 # mutate(relwage=col_wage/hs_wage)%>%
  mutate(relwage1=hsavgwage/colavgwage)#%>%
  #mutate(ttime = year-1963)
```

```
summary(reg11<-lm(log(relwage1) ~ log(relsupp1)+ttime,data=hs_coldata))
```

```
##
## Call:
## lm(formula = log(relwage1) ~ log(relsupp1) + ttime, data = hs_coldata)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0236116 -0.0052219  0.0000009  0.0044574  0.0247263
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -5.895e-02  3.103e-03  -19.00  < 2e-16 ***
## log(relsupp1)  1.204e-02  1.758e-03    6.85 9.38e-09 ***
## ttime         -8.092e-04  8.067e-05  -10.03 1.16e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009186 on 51 degrees of freedom
## Multiple R-squared:  0.7248, Adjusted R-squared:  0.714
## F-statistic: 67.16 on 2 and 51 DF,  p-value: 5.145e-15
```

```
summary(reg21<-lm(log(relwage1) ~ log(relsupp1)+ttime,data=subset(hs_coldata,year>=1964 & year<=1987)))
```

```
##
## Call:
## lm(formula = log(relwage1) ~ log(relsupp1) + ttime, data = subset(hs_coldata,
##     year >= 1964 & year <= 1987))
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0078967 -0.0027682 -0.0000179  0.0034267  0.0077615
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.0537744  0.0025154 -21.378 9.85e-16 ***
## log(relsupp1)  0.0038699  0.0021510   1.799   0.0864 .
## ttime         -0.0006260  0.0001297  -4.826 9.05e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004397 on 21 degrees of freedom
## Multiple R-squared:  0.5538, Adjusted R-squared:  0.5113
## F-statistic: 13.03 on 2 and 21 DF,  p-value: 0.0002088
```

```
summary(reg31<-lm(log(relwage1) ~ log(relsupp1)+ttime,data=subset(hs_coldata,year>=1988 & year<=2017)))
```

```
##
## Call:
## lm(formula = log(relwage1) ~ log(relsupp1) + ttime, data = subset(hs_coldata,
##     year >= 1988 & year <= 2017))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.028029 -0.003987 -0.001430  0.001276  0.029491
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.074e-01  1.828e-02  -5.874 2.95e-06 ***
## log(relsupp1)  2.192e-02  3.811e-03   5.752 4.08e-06 ***
## ttime          4.286e-05  3.446e-04   0.124    0.902
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01019 on 27 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7345
## F-statistic: 41.11 on 2 and 27 DF,  p-value: 6.41e-09
```