12265092

12265092

10/02/2022

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(estimatr)
```

```
## Warning: package 'estimatr' was built under R version 4.1.2
```

```
library(Rcpp)
library(readxl)
library(haven)
library(boot)


data <- read.csv("cpsmar12_chi.csv")
```

```
reg_wages <- lm(wages ~ age, data = data)

summary(reg_wages)
```

```
##
## Call:
## lm(formula = wages ~ age, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -54505  -29596  -14002   11265 1063809
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   6884.8     3891.8   1.769    0.077 .
## age            732.6       91.1   8.042 1.26e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 66330 on 2975 degrees of freedom
## Multiple R-squared:  0.02128,    Adjusted R-squared:  0.02095
## F-statistic: 64.68 on 1 and 2975 DF,  p-value: 1.26e-15
```

```
##  COEFF_1 of age observed in the summary = 732.6
#For 1 year increase in age, the total wages of an individual
##would increase at an estimate of 732.6 dollars
```

```
##Transform ages so that its measured in cents

data <- data %>% mutate(wages_cents = wages *100)

reg_wages_cents <- lm(wages_cents ~ age, data = data)
summary(reg_wages_cents)
```

```
##
## Call:
## lm(formula = wages_cents ~ age, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
##  -5450497  -2959594  -1400237   1126501 106380950
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   688476     389180   1.769    0.077 .
## age            73262       9110   8.042 1.26e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6633000 on 2975 degrees of freedom
## Multiple R-squared:  0.02128,    Adjusted R-squared:  0.02095
## F-statistic: 64.68 on 1 and 2975 DF,  p-value: 1.26e-15
```

```
##COEFF_2 of age observed in the summary = 73262
###For 1 year increase in age, the total wages of an
##individual would increase at an estimate of 73262 cents
```

```
##Transform age so that it is now measured in months

data <- data %>% mutate(age_months = age*12)

reg_age_months <- lm(wages_cents ~ age_months , data = data)

summary(reg_age_months)
```

```
##
```

```
## Call:
## lm(formula = wages_cents ~ age_months, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
##  -5450497  -2959594  -1400237   1126501 106380950
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 688475.9   389179.8   1.769    0.077 .
## age_months    6105.2      759.1   8.042 1.26e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6633000 on 2975 degrees of freedom
## Multiple R-squared:  0.02128,    Adjusted R-squared:  0.02095
## F-statistic: 64.68 on 1 and 2975 DF,  p-value: 1.26e-15
```

```
##Coefficient of age_monts observed = 6105.2
##COEFF_3 = 6105.2

##For 1 month increase in age of an individual
##the wages will increase by an estimate of 6105.2 cents

##Relationship between COEFF_1, COEFF_2
##COEFF_2 = COEFF_1 * 100
##As we multiplied the independent variable wages in dollars
##to transform it into cents, the estimated coefficient of the
##resultant variable is multiplied by 100. Hence we get COEFF_2 as 100 * COEFF_1

##Relationship between COEFF_2 and COEFF_3
##COEFF_3 = COEFF_2/12 => (COEFF_1*100/12)
##As we multiplied the dependant variable age in years
##to transform it into months, the estimated coefficient of the
##resultant variable is divided by 12
```

```
data <- data %>% mutate(male_collgrad = male * collgrad)


reg_b1 <- lm(wages ~ male + collgrad + male_collgrad, data = data)

summary(reg_b1)
```

```
##
## Call:
## lm(formula = wages ~ male + collgrad + male_collgrad, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -85113  -25574  -10765   13426 1072365
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      16265       1958   8.305  < 2e-16 ***
## male             11369       2839   4.005 6.36e-05 ***
## collgrad         30309       3346   9.058  < 2e-16 ***
## male_collgrad    27170       4837   5.617 2.12e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 62640 on 2973 degrees of freedom
## Multiple R-squared:  0.1278, Adjusted R-squared:  0.1269
## F-statistic: 145.2 on 3 and 2973 DF,  p-value: < 2.2e-16
```

```
##Coefficient of the interaction term male_collgrad = 27170
## Males who have a colelge degree earn 27170 dollars
##more in wages as compared to females without a college degree
##The interaction term is basically a difference between wages of males
##with college degrees and others without college degrees
```

```
reg_b2 <- lm(formula = wages ~ collgrad + male_collgrad, data = data)

summary(reg_b2)
```

```
##
## Call:
## lm(formula = wages ~ collgrad + male_collgrad, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -85113  -21676  -11676   13324 1078323
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      21676       1421  15.250  < 2e-16 ***
## collgrad         24898       3069   8.113 7.16e-16 ***
## male_collgrad    38539       3926   9.816  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 62800 on 2974 degrees of freedom
## Multiple R-squared:  0.1231, Adjusted R-squared:  0.1225
## F-statistic: 208.7 on 2 and 2974 DF,  p-value: < 2.2e-16
```

```
##Coefficient of the interaction term male_collgrad = 38539
## Males who have a colelge degree earn 38539 dollars
##more in wages as compared to females without a college degree
##The interaction term is basically a difference between wages of males
##with college degrees and others without college degrees


##As we observed in both the regressions , all the coefficients are
##statistically significant
##It is obvious the postivie correlation between males and male_collgrad
##If male is 0, then male_collgrad = 0, because male_collgrad = male * collgrad
##As we have ommitted the males variable in the second regression,
```

```
##and it is positively correlated with another variable in regression
##we observe an Ommitted variable bias in the coefficients
##of the regression in second regression.
##Now to determine the sign of bias, look at the sign of coefficient
##as we already know the sign of correlation.
##From first regression, we say the estimated coefficient is positive
##Hence it is positive bias, as coeeficient and correlation are positive
##Thus a positive bias is induced in the coefficients of second regression
##as we ommitted a variable males that is correlated with another variable
##in the regression equation.
```