task2

Keyu Mao

2023/6/8

```r
data <- read.csv("final_data.csv", header = T)
data <- na.omit(data)
head(data)
```

```
##   X Country AQI.Value AQI.Category CO.AQI.Value CO.AQI.Category Ozone.AQI.Value
## 1 0     123        51            1            1               0              36
## 2 1      71        41            0            1               0               5
## 3 2     108        66            1            1               0              39
## 4 3       0        34            0            1               0              34
## 5 4      37        22            0            0               0              22
## 6 5     164        54            1            1               0              14
##   Ozone.AQI.Category NO2.AQI.Value NO2.AQI.Category PM2.5.AQI.Value
## 1                  2             0                0              51
## 2                  0             1                0              41
## 3                  2             2                1              66
## 4                  2             0                0              20
## 5                  1             0                0               6
## 6                  0            11                2              54
##   PM2.5.AQI.Category
## 1                  1
## 2                  0
## 3                  1
## 4                  0
## 5                  0
## 6                  1
```

```r
# Without confounders
model_formula_with_2 = 'PM2.5.AQI.Value ~ Country+CO.AQI.Category+NO2.AQI.Value+I(NO2.AQI.Value^2)+NO2.A

# With confounders
model_lr = lm(model_formula_with_2, data=data)
summary(model_lr)
```

```
##
## Call:
## lm(formula = model_formula_with_2, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -147.88  -24.82   -8.34    9.87  444.32
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)            32.8166875  1.1848904  27.696  < 2e-16 ***
## Country               -0.0799932  0.0067169 -11.909  < 2e-16 ***
## CO.AQI.Category        40.3845562  0.9452686  42.723  < 2e-16 ***
## NO2.AQI.Value           1.9800556  0.2110941   9.380  < 2e-16 ***
## I(NO2.AQI.Value^2)      0.0142559  0.0048085   2.965  0.00303 **
## NO2.AQI.Category       -2.2120017  0.7747038  -2.855  0.00430 **
## Ozone.AQI.Value         0.8304036  0.0653855  12.700  < 2e-16 ***
## I(Ozone.AQI.Value^2)   -0.0018996  0.0003135  -6.060 1.38e-09 ***
## Ozone.AQI.Category      2.5944170  0.6402305   4.052 5.09e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46.45 on 23026 degrees of freedom
## Multiple R-squared:  0.2884, Adjusted R-squared:  0.2882
## F-statistic:  1167 on 8 and 23026 DF,  p-value: < 2.2e-16
```

```r
# Without confounders
model_formula_with = 'PM2.5.AQI.Value ~ Country+CO.AQI.Category+NO2.AQI.Value+NO2.AQI.Category+Ozone.AQI

# With confounders
model_formula_without = 'PM2.5.AQI.Value ~ CO.AQI.Category'
# Without confounders
model_without = lm(model_formula_without, data=data)

# With confounders
model_with = lm(model_formula_with, data=data)
```

```r
# Without confounders
summary(model_without)
```

```
##
## Call:
## lm(formula = model_formula_without, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -95.81 -28.32  -9.32   9.68 444.68
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      55.3228     0.3720  148.72   <2e-16 ***
## CO.AQI.Category  57.4846     0.7659   75.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 49.35 on 23033 degrees of freedom
## Multiple R-squared:  0.1965, Adjusted R-squared:  0.1965
## F-statistic:  5633 on 1 and 23033 DF,  p-value: < 2.2e-16
```

```r
# With confounders
summary(model_with)
```

```
##
## Call:
## lm(formula = model_formula_with, data = data)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -111.12  -24.74   -8.50    9.85  444.17
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       37.849653   0.821305  46.085  < 2e-16 ***
## Country           -0.077664   0.006688 -11.613  < 2e-16 ***
## CO.AQI.Category    39.874352   0.934067  42.689  < 2e-16 ***
## NO2.AQI.Value       2.473239   0.091386  27.064  < 2e-16 ***
## NO2.AQI.Category   -3.754427   0.584406  -6.424 1.35e-10 ***
## Ozone.AQI.Value     0.448912   0.016388  27.393  < 2e-16 ***
## Ozone.AQI.Category  5.649331   0.403636  13.996  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46.49 on 23028 degrees of freedom
## Multiple R-squared:  0.287,  Adjusted R-squared:  0.2869
## F-statistic:  1545 on 6 and 23028 DF,  p-value: < 2.2e-16
```

```r
model_formula_without = 'PM2.5.AQI.Value ~  CO.AQI.Category+NO2.AQI.Value+NO2.AQI.Category+Ozone.AQI.Val
# Without confounders
model_without = lm(model_formula_without, data=data)
summary(model_without)
```

```
##
## Call:
## lm(formula = model_formula_without, data = data)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -98.86  -24.92   -8.53   10.17  441.87
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        32.01311    0.65145  49.141  < 2e-16 ***
## CO.AQI.Category    40.78226    0.93349  43.688  < 2e-16 ***
## NO2.AQI.Value       2.35456    0.09108  25.853  < 2e-16 ***
## NO2.AQI.Category   -4.66858    0.58076  -8.039 9.51e-16 ***
## Ozone.AQI.Value     0.44529    0.01643  27.099  < 2e-16 ***
## Ozone.AQI.Category  5.04500    0.40143  12.568  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46.63 on 23029 degrees of freedom
## Multiple R-squared:  0.2829, Adjusted R-squared:  0.2827
## F-statistic:  1817 on 5 and 23029 DF,  p-value: < 2.2e-16
```

```r
model_formula_without = 'PM2.5.AQI.Value ~ CO.AQI.Category+ NO2.AQI.Category+Ozone.AQI.Category'
# Without confounders
model_without = lm(model_formula_without, data=data)
summary(model_without)
```

```
##
## Call:
```

```
## lm(formula = model_formula_without, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -102.29  -26.26   -8.89   11.74  440.43
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         36.9455     0.6516  56.696  < 2e-16 ***
## CO.AQI.Category     56.4135     0.8384  67.283  < 2e-16 ***
## NO2.AQI.Category     3.3179     0.4680   7.089 1.39e-12 ***
## Ozone.AQI.Category  11.3109     0.2962  38.190  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.85 on 23031 degrees of freedom
## Multiple R-squared:  0.2449, Adjusted R-squared:  0.2448
## F-statistic:  2490 on 3 and 23031 DF,  p-value: < 2.2e-16
```
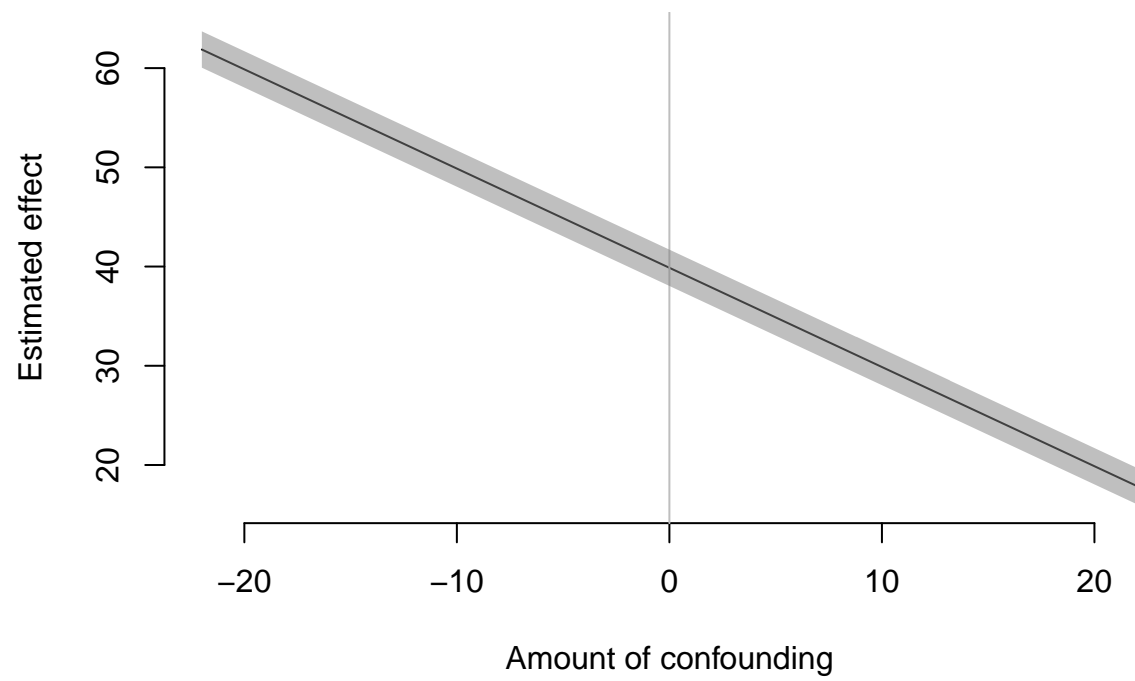
sen

```r
library(causalsens)              # Load MatchIt package
library(tidyverse)           # For data manipulation
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
model <- lm(model_formula_with, data = data)
p_model <- glm(CO.AQI.Category~NO2.AQI.Category+Ozone.AQI.Category+Country, data=data)
```

```r
alpha <- seq(-4500, 4500, by = 250)
ll.sens <- causalsens(model, p_model, ~ NO2.AQI.Category+Ozone.AQI.Category, data = data,
 confound = one.sided.att)
```

```r
plot(ll.sens, type = "raw", bty = "n")
```

```
plot(ll.sens, type = "r.squared", bty = "n")
```