

Econometria 2

Miguel Sallum

23/06/2021

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.4      v purrr 0.3.4
## v tibble 3.1.2       v dplyr 1.0.7
## v tidyr 1.1.3        v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.1

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

Regressão Linear e Estimação

A regressão linear é um modelo de relação entre variáveis, e pode ser chamado também de função de esperança condicional. Ela é tradicionalmente estimada com método dos mínimos quadrados ordinários, mas é equivalente por método dos momentos e (se não me engano) por máxima verossimilhança.

Caso tenhamos somente um regressor, estamos estimando a esperança condicional da forma

$$E[Y|X] = \beta_0 + \beta_1 X$$

Tradicionalmente, no entanto, representamos o modelo como

$$Y = \beta_0 + \beta_1 X + \mu$$

Onde μ são os fatores não observados. Para estimarmos as regressões, são necessárias algumas hipóteses: 1. É importante lembrar que a regressão *não é* um modelo causal. Para ser causal, são necessárias algumas hipóteses a mais. Sendo flexível com notação, em geral nosso interesse é estimar o modelo causal

$$E[Y|do(X)] = \beta_0 + \beta_1 X$$

Queremos encontrar então o valor adequado de β_1 , que seria o efeito médio de X sobre Y. para isso, precisamos da hipótese (**não-observável**) 4. $Corr(X, \mu) = 0$

Estimação matricial

```
ols <- function(Y, X){
  X <- as.matrix(X)
  Y <- as.matrix(Y)
  Xt <- t(X)
  XtX <- Xt %*% X
  XtX_inv <- solve(XtX)
```

```

  XtX_inv %*% Xt %*% Y
}

SST <- function(Y) {Y - mean(Y) %>% t(.) %*% .}

res <- function(Y, X) {Y - (X %*% ols(Y, X))}

SSR <- function(Y, X) {t(res(Y, X)) %*% res(Y, X)}

R2 <- function(Y, X){
  X <- as.matrix(X)
  Y <- as.matrix(Y)

  1 - SSR(Y,X) /SST(Y)
}

vcov<- function(Y, X, robust = FALSE){

  k <- nrow(ols(Y, X))
  n <- length(Y)
  if (robust) {
    sigma2

  } else {
    sigma2 <- as.numeric(SSR(Y, X)/(n-k))*diag(nrow = k)
  }
  var_cov <- sigma2 %*% XtX_inv
}

```

Métodos do R e Bibliotecas

Para rodar as regressões, podemos usar a função nativa `lm()`, em geral já em conjunção com `summary()` para termos algumas características importantes, como R^2 , erros-padrão e significância assumindo homoscedasticidade. Caso estejamos usando tidyverse, podemos usar as funções assim:

```

#lm(dependente ~ regressores, data = dados)%>%
# summary()

```

No entanto, na maior parte dos casos, a hipótese da homoscedasticidade é muito forte

Potencial Outcomes

Potencial Outcomes é uma forma de pensar sobre causalidade usando de contrafactuais. A ideia é que teríamos a informação do resultado de cada indivíduo para cada nível de intervenção de X. Os exemplos do tema em geral são binários, **mas o método não se restringe a isso**. Em casos binários, podemos representar o resultado do indivíduo i caso ele receba o tratamento ($X = 1$) como Y_i^1 , e como Y_i^0 caso ele não seja tratado.

Trabalhando com esses dados hipotéticos surgem alguns conceitos interessantes:

$$\begin{aligned}
 \text{Treatment Effect}_i (TE_i) &= Y_i^1 - Y_i^0 \\
 \text{Average Treatment Effect (ATE)} &= E[Y^1 - Y^0] = E[Y^1] - E[Y^0] \\
 \text{Treatment Effect on the Treated (ATT)} &= E[Y^1 - Y^0 | \text{Treatment} = 1] \\
 \text{Treatment Effect on the Untreated (ATU)} &= E[Y^1 - Y^0 | \text{Treatment} = 0]
 \end{aligned}$$

Nenhum desses efeitos pode ser extraído diretamente dos dados, mas eles ajudam a entender quais são as hipóteses necessárias para nossas estimativas valerem em cada caso. Caso quisermos estimar o efeito a partir da diferença simples de médias, por exemplo, chegamos que:

$$\begin{aligned} SDO &= E[Y^1|D=1] - E[Y^0|D=0] \\ &= (E[Y^1] - E[Y^0]) + (E[Y^0|D=1] - E[Y^0|D=0]) + (1 - \pi)(ATT - ATU) \\ &= ATE + \text{viés de seleção} + (1 - \pi)(\text{viés de efeito heterogêneo}) \end{aligned}$$

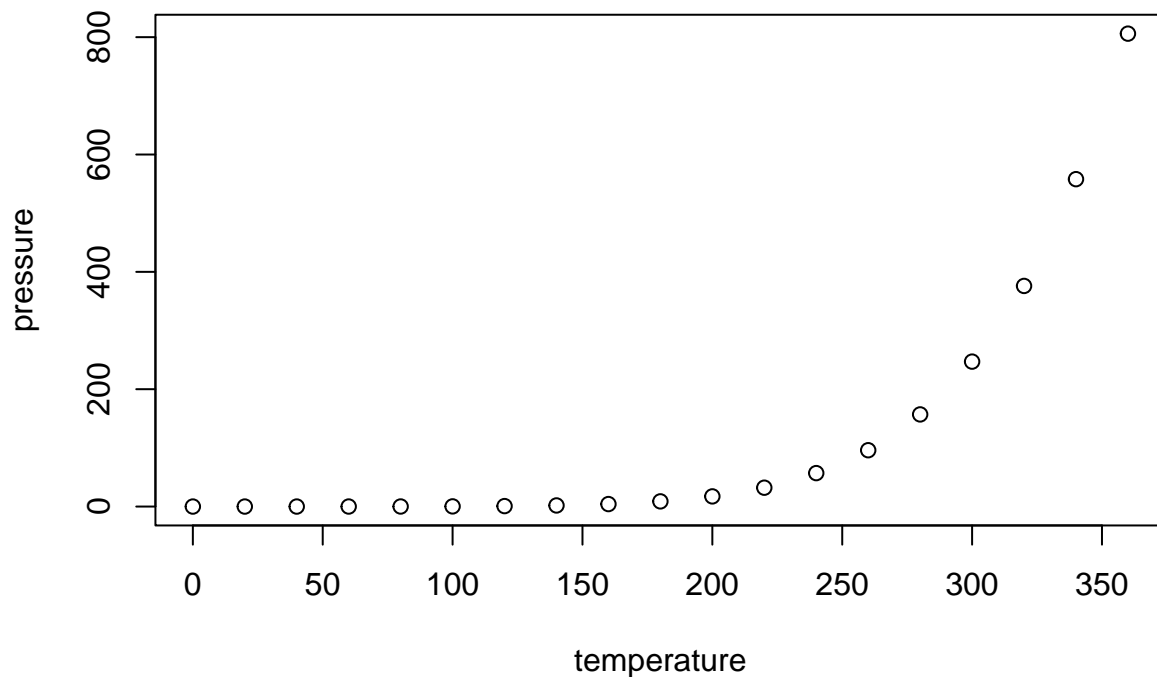
Considerando que o que nós buscamos conhecer são os efeitos de fato, ATE, então devemos nos preocupar com métodos para que os vieses sejam nulos. A forma mais simples é a randomização, fazendo com que as esperanças de ambos os grupos sejam as mesmas.

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.:26.00
## Median :15.0    Median :36.00
## Mean   :15.4    Mean   :42.98
## 3rd Qu.:19.0    3rd Qu.:56.00
## Max.   :25.0    Max.   :120.00
```

Matching e Estratificação

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Regression Discontinuity Design

Variável Instrumental

Panel Data

Diferenças-em-Diferenças

Event Study, Two-Way Fixed Effects e Generalização de DiD