Natural Experiments in Use

Vinish Shrestha

9/15/2020

Some Drawbacks of Randomized Control Trials

- As early as 1920s, Fisher (1928) argued that the only satisfactory way of looking at the treatment effects was to assign treatment and control groups "wholly at random"
- RCT has been widely utilized in fields such as medicine, agriculture, pharma industry
- Costly
- Unethical
- External Validity problems

Natural Experiments to trace the causal effects

- $y = \alpha + \beta x + \gamma temp + \epsilon$

- Here, variables in ϵ affects both x and y, which creates difficulty in tracing the causal effect

- Remember the example of the bottles of sun screen and glasses of lemonade

Natural Experiments to trace the causal effects

▶ Now, say you are interested in finding whether increase in glasses of lemonade does increase the purchase of sun screen. (A dull question, but still lets see if we can identify the causal effects.)

▶ To do so, we need external force (also known as shocks in economics), that affects the quantity of lemonade purchased.

   ▶ This shock should only affect quantity of lemonade and should not directly affect the demand for sun screen.
   ▶ can you think of something that resemble such shocks?

▶ Some thing in the natural, that affects the purchase of lemonade but does not directly affect the purchase of sun screen. So, any effect on sun screen will be driven through the changes in purchase of lemonade.

Natural Experiments ..

- ▶ These types of shocks are also known as "instruments" in economics.

- ▶ Shocks in lemon farming due to increasing Botrytis fungus around the area.

- ▶ increases the price of lemons, which drives the price of lemonade up

- ▶ Utilize the variation in price to identify, whether increase in purchase of lemonades also makes people buy more of sun scree

- ▶ $y_t = \alpha + \beta price_t(lemons) + \gamma temp_t + \epsilon_t$, where t is the time subscript.

  - ▶ $\beta_{hat} > 0$, that price of lemons is causally and positively related to demand of suncreen lotion.
  - ▶ price of lemons increases, reduces the quantity demand of lemons

Natural Experiments ..

```
set.seed(1)
#simulate quantity demanded of lemonades
lemonade = rnorm(100, 50, 10)
error = rnorm(100, 0, 1)
sunscreen = 1/2*lemonade + 8*error
temp = 1.2*lemonade + 0.4*sunscreen
data <- data.frame(cbind(lemonade, sunscreen, temp))
reg1 <- lm(sunscreen ~ lemonade, data)
reg2 <- lm(lemonade ~ sunscreen, data)
reg_control <- lm(lemonade ~ sunscreen + temp, data)
```

Natural Experiments ..

```
summary(reg1)
```

```
##
## Call:
## lm(formula = sunscreen ~ lemonade, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.015  -4.910  -1.116   4.315  18.769
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.25913    4.46978  -0.058    0.954
## lemonade     0.49915    0.08618   5.792 8.42e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 7.702 on 98 degrees of freedom
## Multiple R-squared:  0.255,  Adjusted R-squared:  0.2474
```
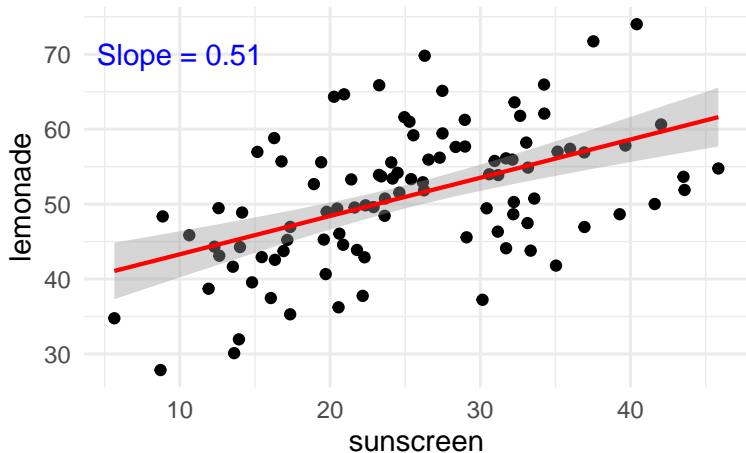
Natural Experiments ..

```
library(ggplot2)
ggplot(data, aes(x=sunscreen, y = lemonade)) + geom_point()
            label = paste("Slope =", 0.51))
```
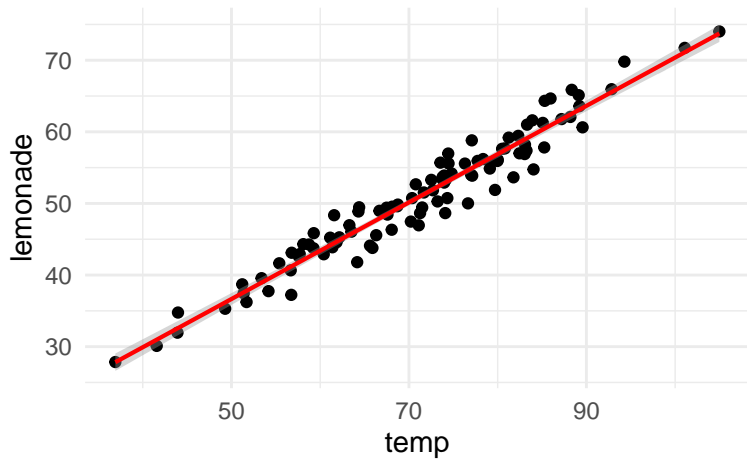
```
## `geom_smooth()` using formula 'y ~ x'
```

Natural Experiments ..

```r
library(ggplot2)
ggplot(data, aes(x=temp, y = lemonade)) + geom_point() + th
            label = paste())
```
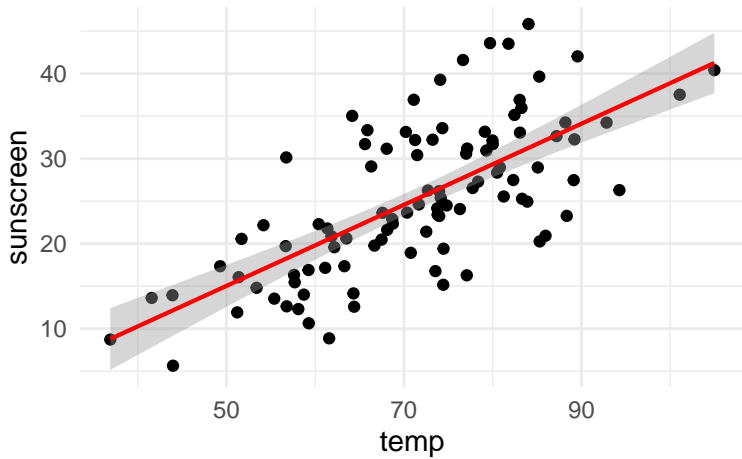
```
## `geom_smooth()` using formula 'y ~ x'
```

Natural Experiments ..

```
library(ggplot2)
ggplot(data, aes(x=temp, y = sunscreen)) + geom_point() + t
            label = paste())
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Natural Experiments ..

```
summary(reg2)
```

```
##
## Call:
## lm(formula = lemonade ~ sunscreen, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.360  -5.680   0.226   4.716  18.174
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.19302    2.35896  16.191  < 2e-16 ***
## sunscreen    0.51089    0.08821   5.792 8.42e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 7.792 on 98 degrees of freedom
## Multiple R-squared:  0.255,  Adjusted R-squared:  0.2474
```

Natural Experiments ..

```
summary(reg_control)
```

```
## Warning in summary.lm(reg_control): essentially perfect
## unreliable
##
## Call:
## lm(formula = lemonade ~ sunscreen + temp, data = data)
##
## Residuals:
##        Min        1Q    Median        3Q       Max
## -5.475e-14 -2.091e-15  2.200e-17  3.537e-15  1.081e-14
##
## Coefficients:
##               Estimate Std. Error   t value Pr(>|t|)
## (Intercept) -2.274e-14  4.225e-15 -5.381e+00 5.13e-07 **
## sunscreen   -3.333e-01  1.146e-16 -2.908e+15  < 2e-16 **
## temp         8.333e-01  7.865e-17  1.059e+16  < 2e-16 **
## ---
```
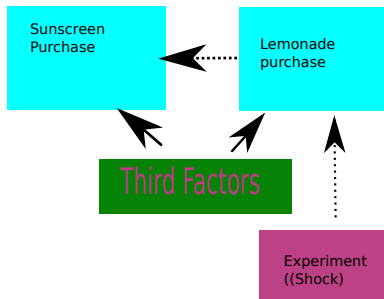
Natural Experiments (simplified)

Figure 1: Using Botrytis fungus as a Natural Experiment

Note: The third factor variables can be driving consumption of both sunscreen and lemonade. What we want is a 'shock' to lemonade production, that affects the purchase of lemonade and traces the dotted link. Note that the 'shock' does not affect the suncreen

Natural Experiment (Shrestha 2016)

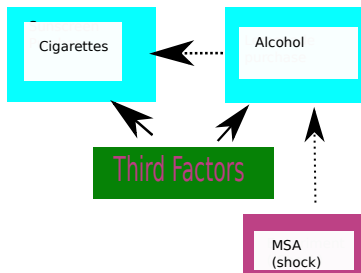▶ Do young adults substitute cigarettes for alcohol? Learning from the Master Settlement Agreement

Figure 2: Using MSA as a Natural Experiment.

Natural Experiments using the Regression Discontinuity Approach

- ▶ Now lets get back to estimating whether demand for health care and services are downward sloping. To understand this phenomenon, Card, Dobkin and Maestas (2008) used eligibility criteria as a 'shock' (if you will), to determine whether an average individual uses more health care when eligible for Medicare.

- ▶ Note that a US citizen, after turning 65, is eligible for Medicare. Age determines the eligibility criteria. However, you cannot really control age (there are some other limitations however).

- ▶ Medicare, in a sense, resembles a form of universal health insurance in US.

- ▶ The authors compare people around the eligibility criteria to identify the effects of change in health insurance coverage through Medicare coverage on utilization of health care and services.

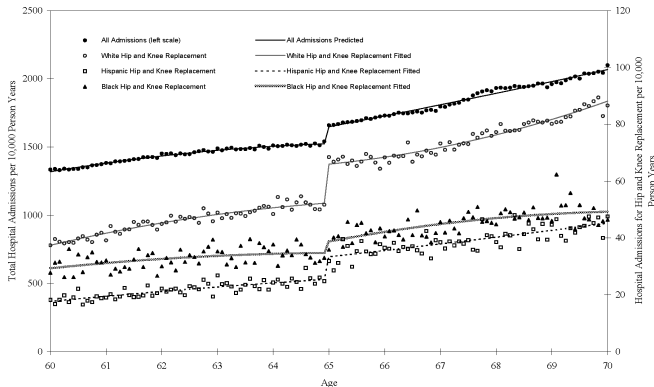# Card, Dobkin and Maestas (2008) RD Approach

Figure 3: Health Care Usage

The Oregon Health Insurance Experiment (Finkelstein et al. 2012)

- ▶ In 2008, Oregon did a lottery, to randomly assign health insurance to a group of low-income adults
  - ▶ 90,000 people who signed up received it.
- ▶ The lottery provides a unique randomized experiment through policy.
  - ▶ confounding or third factors are avoided due to random assignment
- ▶ Use administrative data from hospital discharge, credit report and mortality records
  - ▶ Higher utilization of health care (2.1 percentage points increase in probability of hospital visits, 8.8 pp increase in probability of taking any prescription drug, 21 pp increase in probability of any outpatient visit)
  - ▶ lower out of pocket medical expenses
  - ▶ better self reported health

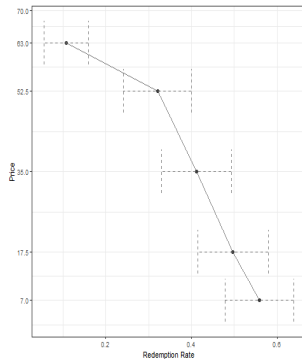# A Randomized Experiment Shrestha and Shrestha (2020)

Figure 4: Figure 2a. Demand for Sanitary Pads
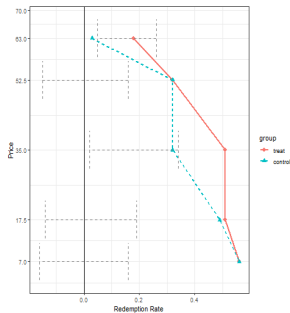
A Randomized Experiment Shrestha and Shrestha (2020)

Figure 5: Figure 2b. Demand for Sanitary Pads by Treatment Criteria

Concluding Remarks

► So the empirical evidence do point out that the demand for health care and services are downward sloping

► However, the effects are homogeneous. For instance, although inelastic, outpatient visits are relatively more elastic than inpatient visits.

► Nevertheless, it should be realized that there are other forces at play that economists generally tend to ignore.

  ► For instance, Alsan and Wanamaker (2017) finds that the disclosure of Tuskegee experiment in 1972 is correlated with medical mistrust among older black men.
    ► reduced life expectancy
  ► In presence of medical mistrust, do people still respond to lower prices?

Feedback to Homework 2

1. $q = -\frac{1}{4} \times p + 200$.

▶ point elasticity $= \frac{\Delta Q}{\Delta P} \times \frac{P}{Q}$ .. eq (1)
▶ $\frac{\Delta Q}{\Delta P} = -\frac{1}{4}$
▶ Find Q when $p = \$5$
▶ replace eq. and solve
▶ You need to note the - signs for the elasticity estimates
▶ Also, make sure that you realize the tope part of the demand curve is relatively more elastic, compared to the bottom part. This should be evidence from e.

Question 2

doc visits = *g*(*Insurance*, *Education*, *Income*, *Demographics*, ϵ) a. It is difficult to say, because there are factors in ϵ that confounds the relationship. For example, risk preference, which is in ϵ, can be driving both purchase of insurance and doctor visits. b. Nope. See a. c. This will not say anything about the demand curve as those who are insured and uninsured cannot be compared.

- ▶ At this point it is important to understand that correlation is not causal. Even in this powerpoint, there are ways that can help build causal interpretations, including randomized control trials and natural experiments.
- ▶ For example, Question 3 on PS2 deals with a setting of a randomized control trial. Here, the assignment mechanism of who gets the insurance verus who does not is random. Please revise question 3 on yourself.