Multivariate relationships

Week 6

22 February, 2016

Prof. Andrew Eggers

19 November 2012 Last updated at 18:19

Share

Does chocolate make you clever?

By Charlotte Pritchard
BBC News

Eating more chocolate improves a nation's chances of producing Nobel Prize winners - or at least that's what a recent study appears to suggest. But how much chocolate do Nobel laureates eat, and how could any such link be explained?

In today's Magazine

The Swiss children

Top Stories

Penalties 'do not stop' drug use

Sickness benefit cuts 'considered'

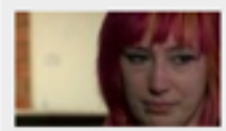
Care plan 'to ease hospital pressure'

Child sex exploitation 'social norm'

MP Jim Murphy joins Labour contest

Features

Magical masterpiece
The Leonardo hidden from Hitler in case it gave him special powers



'GamerGate'
The developer forced to leave her home due to threats

Wake up
Is eating sage better for your alertness than coffee?

Rumble revisited
Forty years since Ali took on Foreman in Kinshasa BBC SPORT

Armageddon file
The nuclear attack on the UK that

OCCASIONAL NOTES

Chocolate Consumption, Cognitive Function, and Nobel Laureates

Franz H. Messerli, M.D.

Dietary flavonoids, abundant in plant-based foods, have been shown to improve cognitive function. Specifically, a reduction in the risk of dementia, enhanced performance on some cognitive tests, and improved cognitive function in elderly patients with mild impairment have been associated with a regular intake of flavonoids.[1,2] A subclass of flavonoids called flavanols, which are widely present in cocoa, green tea, red wine, and some fruits, seems to be effective in slowing down or even reversing the reductions in cognitive performance that occur with aging. Dietary flavanols have also been shown to improve endothelial cause the population of a country is substantially higher than its number of Nobel laureates, the numbers had to be multiplied by 10 million. Thus, the numbers must be read as the number of Nobel laureates for every 10 million persons in a given country.

All Nobel Prizes that were awarded through October 10, 2011, were included. Data on per capita yearly chocolate consumption in 22 countries was obtained from Chocosuisse (www.chocosuisse.ch/web/chocosuisse/en/home), Theobroma-cacao (www.theobroma-cacao.de/wissen/wirtschaft/international/konsum), and

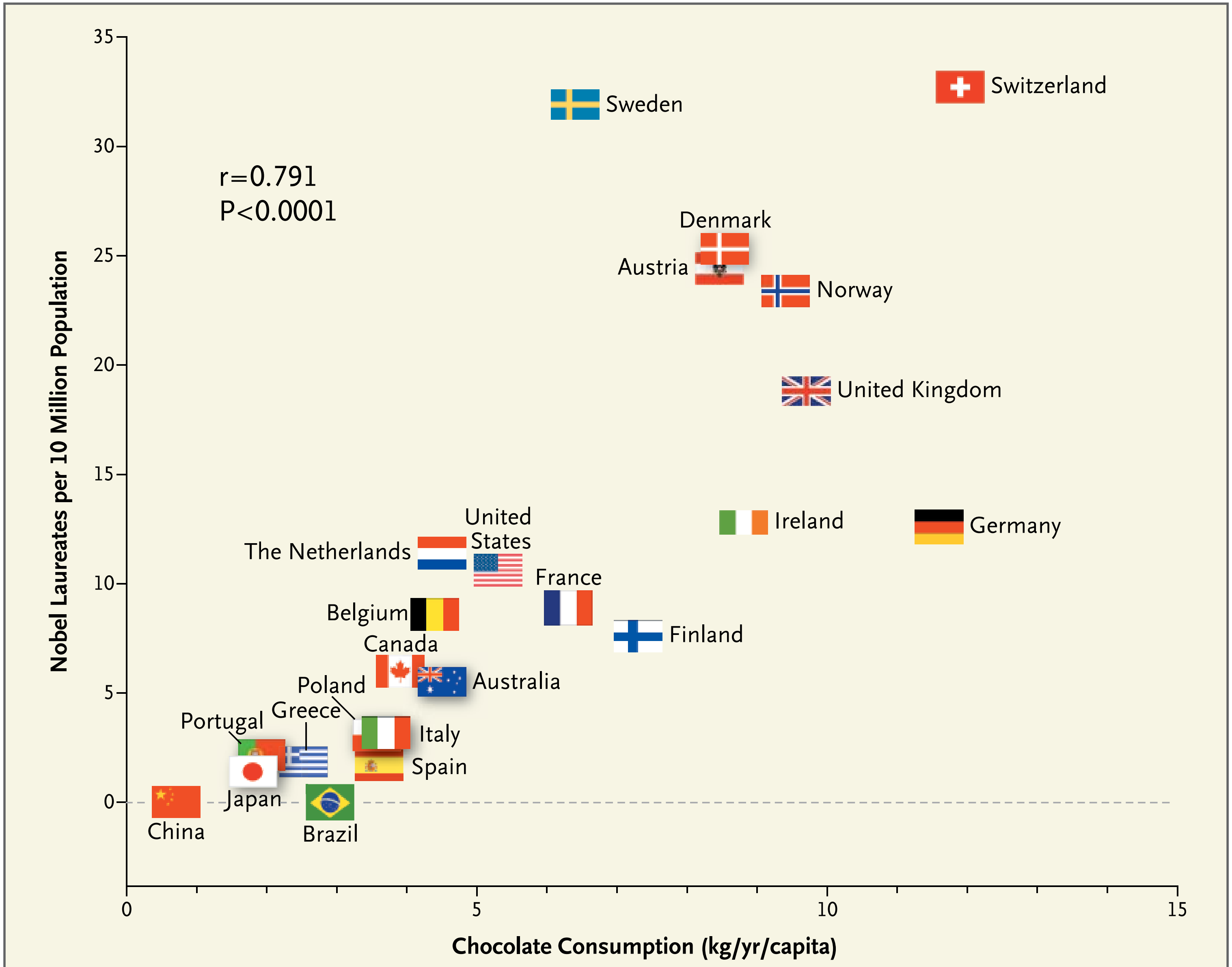
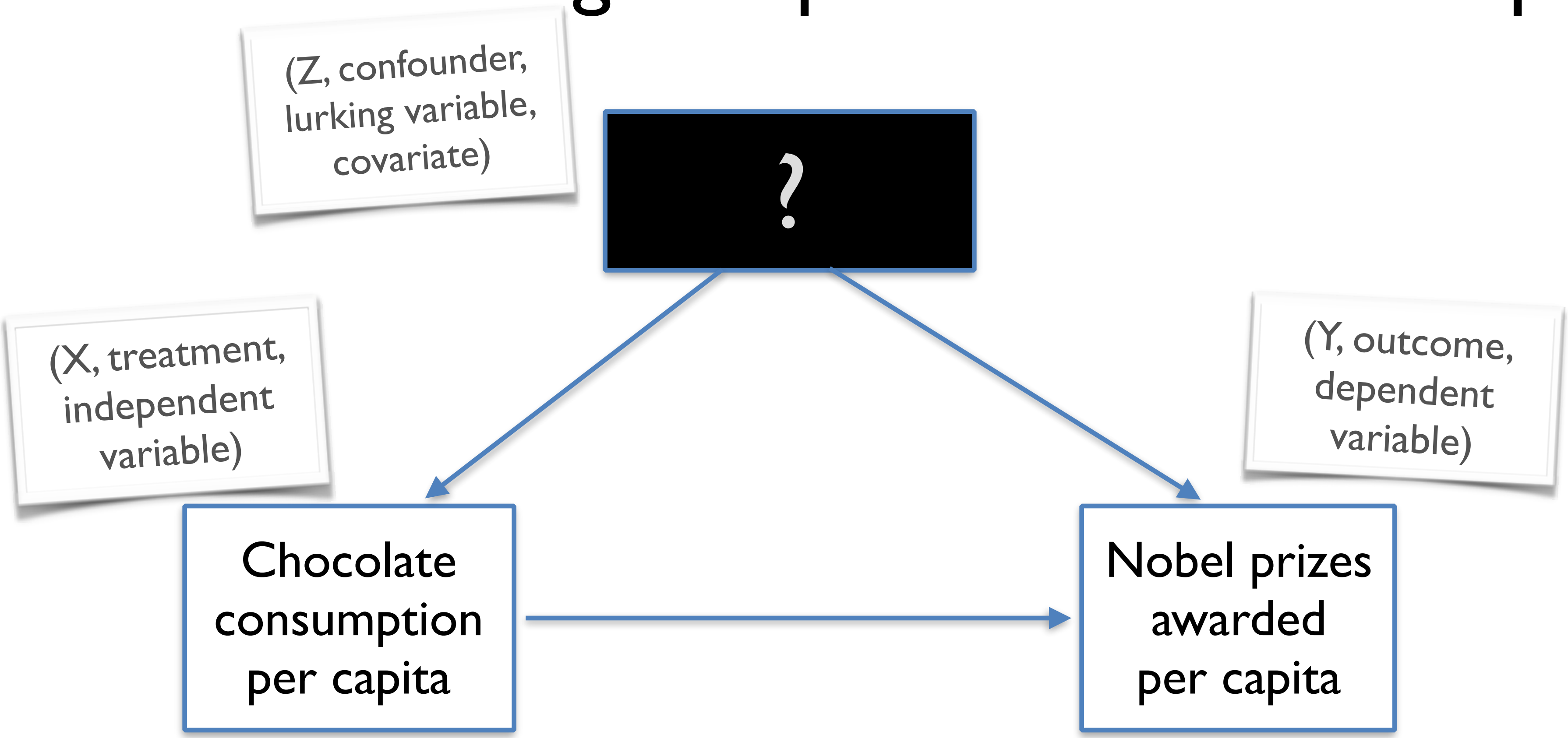


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

What else might explain this relationship?



Spurious?

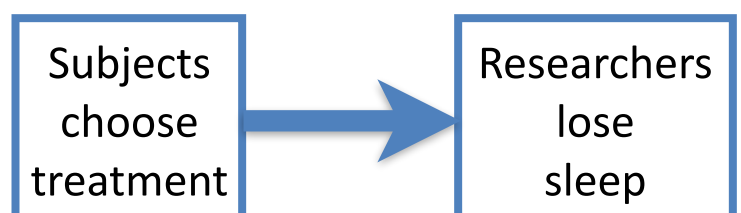
How do we identify confounders?

How do we **control** for them?

Best case: randomize treatment

In a randomized experiment, there should be no confounding variables.

In many social science settings, RCT is impossible: subjects (e.g. countries, individuals) choose their own treatment.



Next best: statistical control

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.189***	3.360	34
Internal conflict risk (1990–2004)	0.346**	2.097	32
Weighted domestic conflict index (1981–2009)	−105.0*	1.611	30
Weighted domestic conflict index (1990–2009)	−119.7**	2.177	33
Deaths from domestic terrorism (1985–2010)	−2.357**	1.728	33

* Statistically significant at the 10 percent level (one-tailed test)

** Statistically significant at the 5 percent level (one-tailed test)

*** Statistically significant at the 1 percent level (one-tailed test)

Source: Based on data in Kaufmann, Kraay, and Mastruzzi 2010; PRS Group 2004; Banks, 2010; and GTD Team 2010

Source: Lijphart (2012)

Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis

E. G. Wilmot · C. L. Edwardson · F. A. Achana ·
M. J. Davies · T. Gorely · L. J. Gray · K. Khunti ·
T. Yates · S. J. H. Biddle

Table 1 Characteristics of cross-sectional and prospective cohort studies included in meta-analysis

Author [ref.]	Design, sample size	Outcome, no. cases	Sedentary measure used in meta-analysis	Confounders measured	Quality
Dunstan et al 2004 [21]	Cross-sectional 8,299 Australian men and women	Diabetes 252 cases (3%)	TV viewing >14 vs <14 h/week	Adjusted for age, education, FHx DM, smoking, diet and PA	5
Dunstan et al 2010 [32]	Prospective 6.6 year f/u 8,800 Australian men and women	Cardiovascular mortality 87 cases (1%) All-cause mortality 284 cases (3.2%)	TV viewing ≥4 vs <2 h/day	Adjusted for age, sex, smoking, education, diet	6
Ford et al 2010 [24]	Prospective 7.8 year f/u 23,855 German men and women	Diabetes 927 cases (3.9%)	TV viewing <1 vs>4 h/day	Adjusted for age, sex, education, occupational activity, smoking, alcohol, PA, diet, systolic BP	3
Hawkes et al 2011 [29]	Prospective 3 year f/u 1,966 Australian	Diabetes 247 cases (12.6%)[a] Cardiovascular	TV viewing <2 vs >4 h/day	Sex, age, education, marital status Diabetes outcome[b]	4

When statistical control really matters

Classic example: Cochran (1968) on risk of **pipe smoking** vs **cigarette smoking**

THE EFFECTIVENESS OF ADJUSTMENT BY SUBCLASSIFICATION IN REMOVING BIAS IN OBSERVATIONAL STUDIES

W. G. COCHRAN

Harvard University, Cambridge, Mass., U. S. A.

SUMMARY

In some investigations, comparison of the means of a variate y in two study groups may be biased because y is related to a variable z whose distribution differs in the two groups. A frequently used device for trying to remove this bias is adjust-

Are pipes worse than cigarettes?

Source: Cochran (1968)

Are pipes worse than cigarettes?

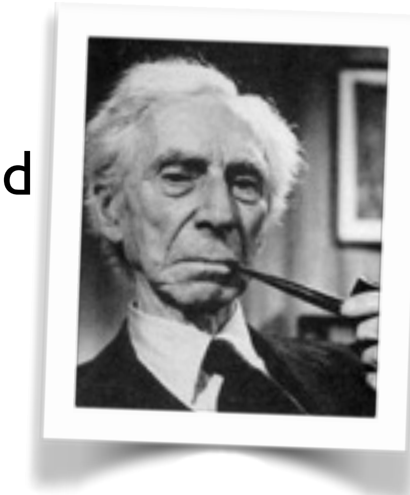
Outcome (y): Death rate

Source: Cochran (1968)

Are pipes worse than cigarettes?

Outcome (y): Death rate

Study groups (x): pipe smokers and cigarette smokers



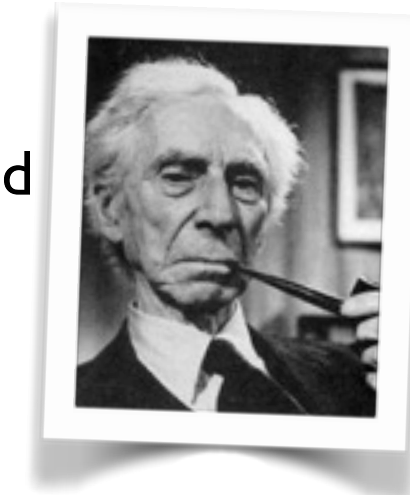
Source: Cochran (1968)

Are pipes worse than cigarettes?

Outcome (y): Death rate

Study groups (x): pipe smokers and cigarette smokers

Death rate is higher among pipe smokers.



Source: Cochran (1968)

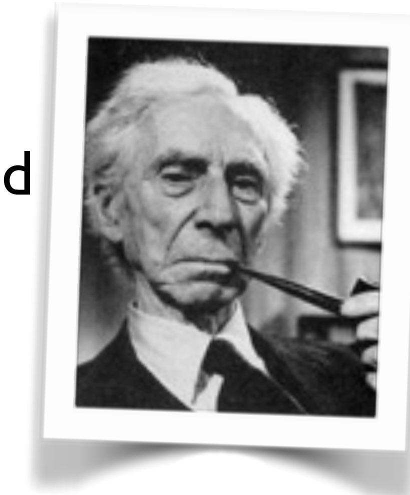
Are pipes worse than cigarettes?

Outcome (y): Death rate

Study groups (x): pipe smokers and cigarette smokers

Death rate is higher among pipe smokers.

But: pipe smokers are older (z).



Source: Cochran (1968)

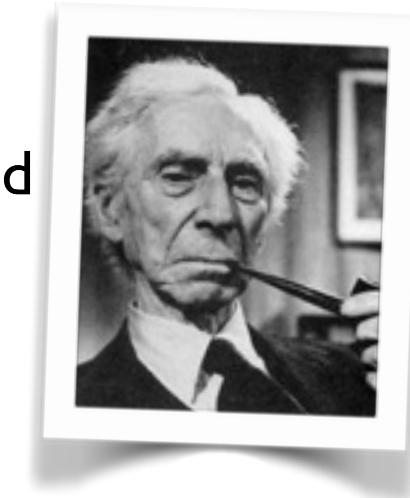
Are pipes worse than cigarettes?

Outcome (y): Death rate

Study groups (x): pipe smokers and cigarette smokers

Death rate is higher among pipe smokers.

But: pipe smokers are older (z).



Pipe smokers				
Cigarette smokers				

Source: Cochran (1968)

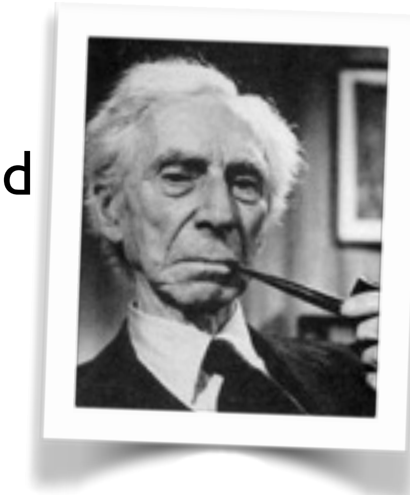
Are pipes worse than cigarettes?

Outcome (y): Death rate

Study groups (x): pipe smokers and cigarette smokers

Death rate is higher among pipe smokers.

But: pipe smokers are older (z).



	US		
	Raw death rates		
Pipe smokers	***17.4***		
Cigarette smokers	***13.5***		

Source: Cochran (1968)

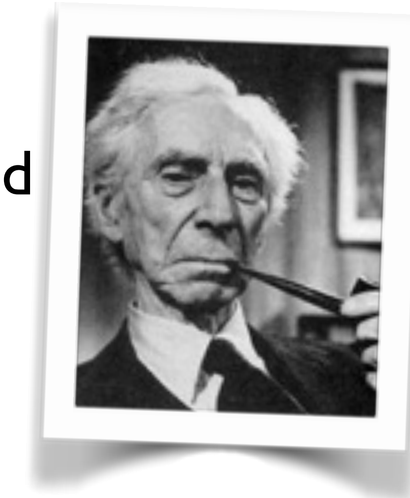
Are pipes worse than cigarettes?

Outcome (y): Death rate

Study groups (x): pipe smokers and cigarette smokers

Death rate is higher among pipe smokers.

But: pipe smokers are older (z).



	US	
	Raw death rates	***Adjusted for age***
Pipe smokers	***17.4***	***13.7***
Cigarette smokers	***13.5***	***21.2***

Source: Cochran (1968)

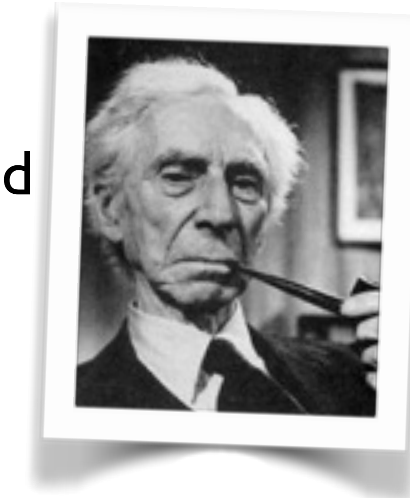
Are pipes worse than cigarettes?

Outcome (y): Death rate

Study groups (x): pipe smokers and cigarette smokers

Death rate is higher among pipe smokers.

But: pipe smokers are older (z).



	US		UK
	Raw death rates	*Adjusted for age*	*Raw death rates*
Pipe smokers	*17.4*	*13.7*	*20.7*
Cigarette smokers	*13.5*	*21.2*	*11.0*

Source: Cochran (1968)

Are pipes worse than cigarettes?

Outcome (y): Death rate

Study groups (x): pipe smokers and cigarette smokers

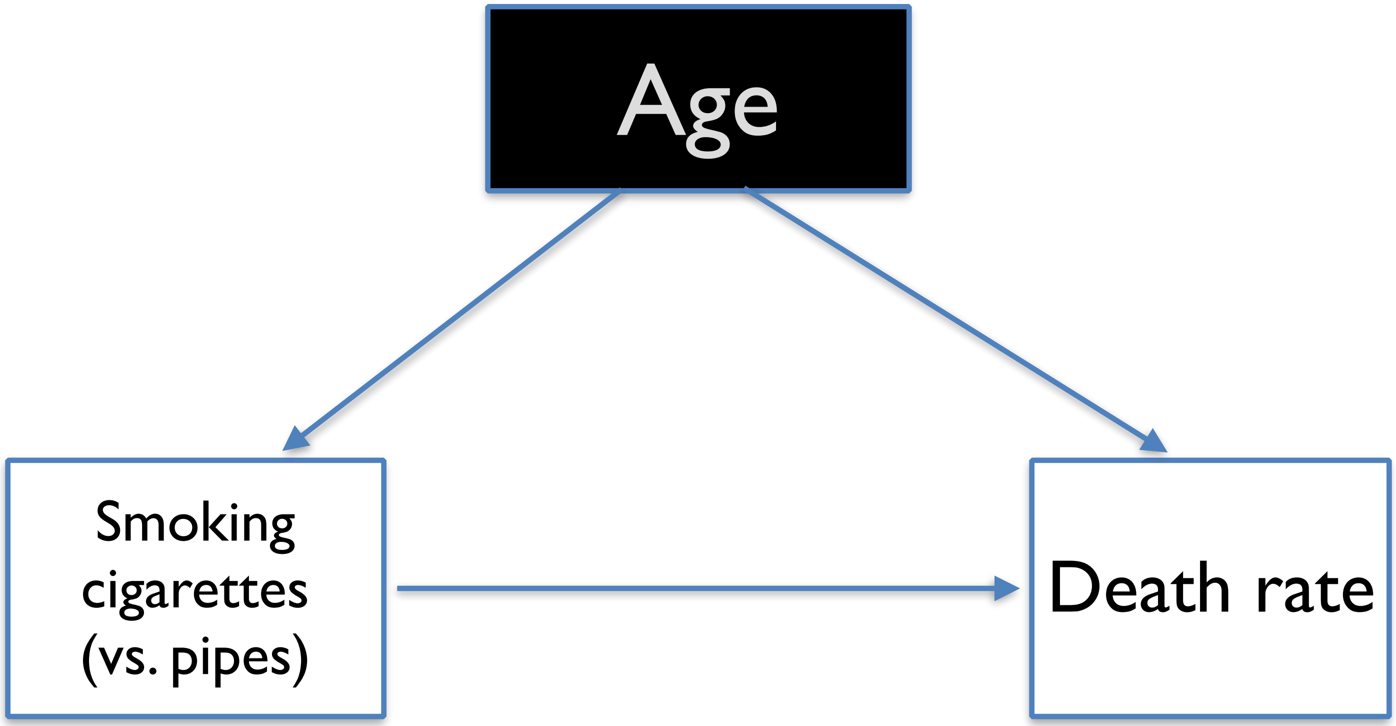
Death rate is higher among pipe smokers.

But: pipe smokers are older (z).

	US		*UK*	
	Raw death rates	***Adjusted for age***	***Raw death rates***	***Adjusted for age***
Pipe smokers	***17.4***	***13.7***	***20.7***	***11.0***
Cigarette smokers	***13.5***	***21.2***	***11.0***	***14.8***

Source: Cochran (1968)

Why does controlling for age **reverse** the conclusion?



When would controlling for a confounder **strengthen** the conclusion?

Is consensus democracy better than majoritarian democracy?

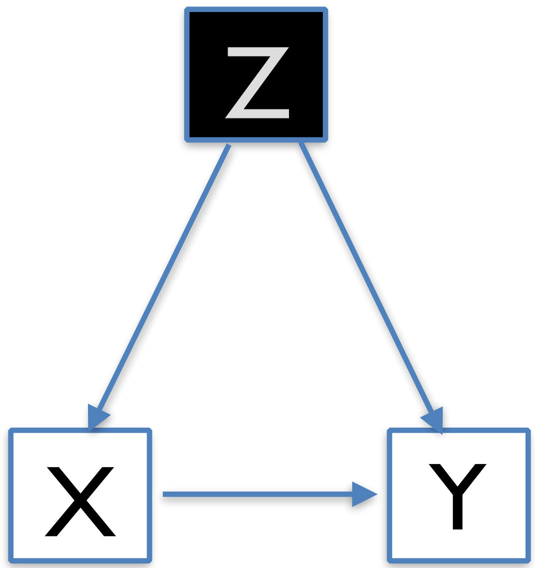
Outcome (y): e.g. political stability

Study groups (x): countries with consensus forms of democracy (e.g. Finland, Netherlands) and majoritarian forms (e.g. UK, Bahamas)

Political stability is higher in consensus democracies.

But: These countries differ in many other ways! (z). Which differences should we control for?

What do we need to control for?

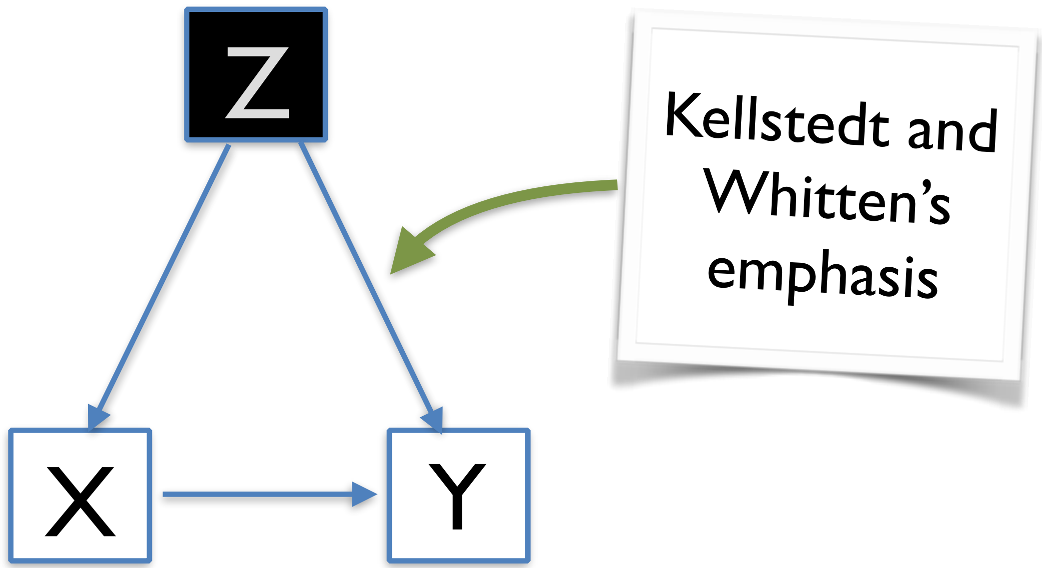


To study the effect of X on Y, we need to control for factors that affect both X and Y (directly or indirectly).

What should we control for when studying the effect of

- pipes vs. cigarettes on mortality?
- consensus democracy on political stability?

What do we need to control for?

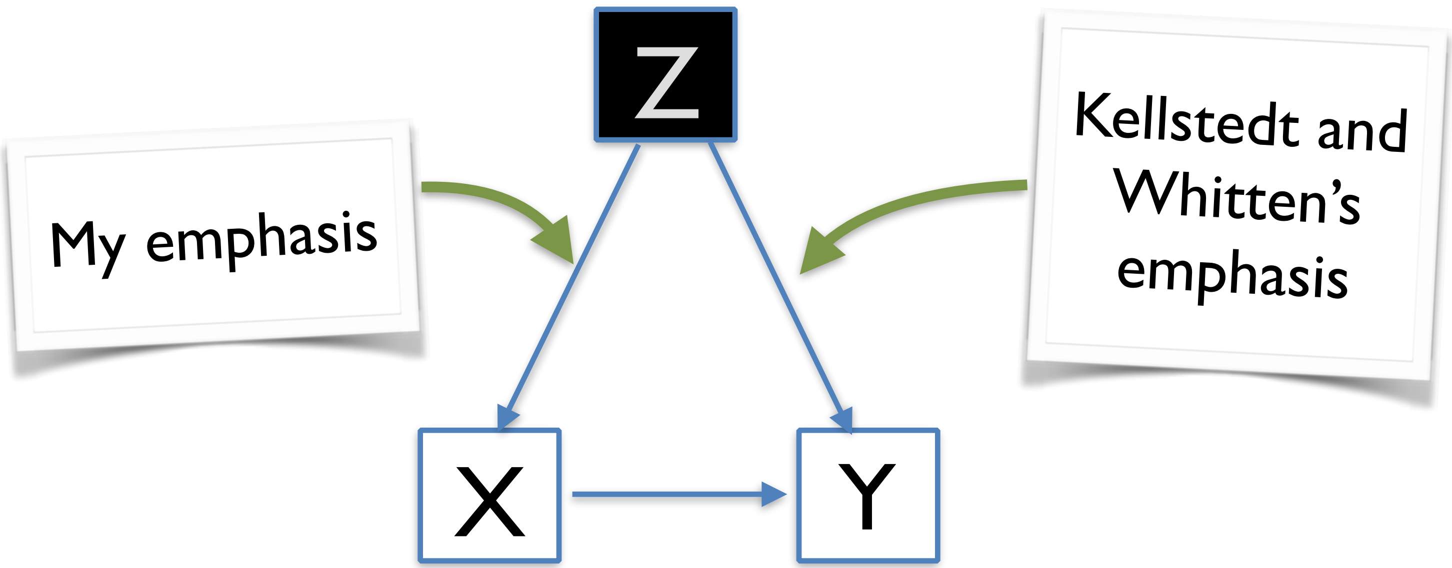


To study the effect of X on Y, we need to control for factors that affect both X and Y (directly or indirectly).

What should we control for when studying the effect of

- pipes vs. cigarettes on mortality?
- consensus democracy on political stability?

What do we need to control for?



To study the effect of X on Y, we need to control for factors that affect both X and Y (directly or indirectly).

What should we control for when studying the effect of

- pipes vs. cigarettes on mortality?
- consensus democracy on political stability?

How do we control?
Two intuitive approaches

How do we control?
Two intuitive approaches

Subclassification: compare outcomes for subjects within intervals of a covariate.

	Mortality	
	Cig. smoker	***Pipe smoker***
Age 55-60	***8.2***	***6.1***
Age 60-65	***10.4***	***8.7***

etc.

How do we control?
Two intuitive approaches

Subclassification: compare outcomes for subjects within intervals of a covariate.

	Mortality	
	Cig. smoker	***Pipe smoker***
Age 55-60	***8.2***	***6.1***
Age 60-65	***10.4***	***8.7***

etc.

Matching: for every "treated" unit, find a similar "untreated" unit. Compare the two groups.

How do we control?
Two intuitive approaches

Subclassification: compare outcomes for subjects within intervals of a covariate.

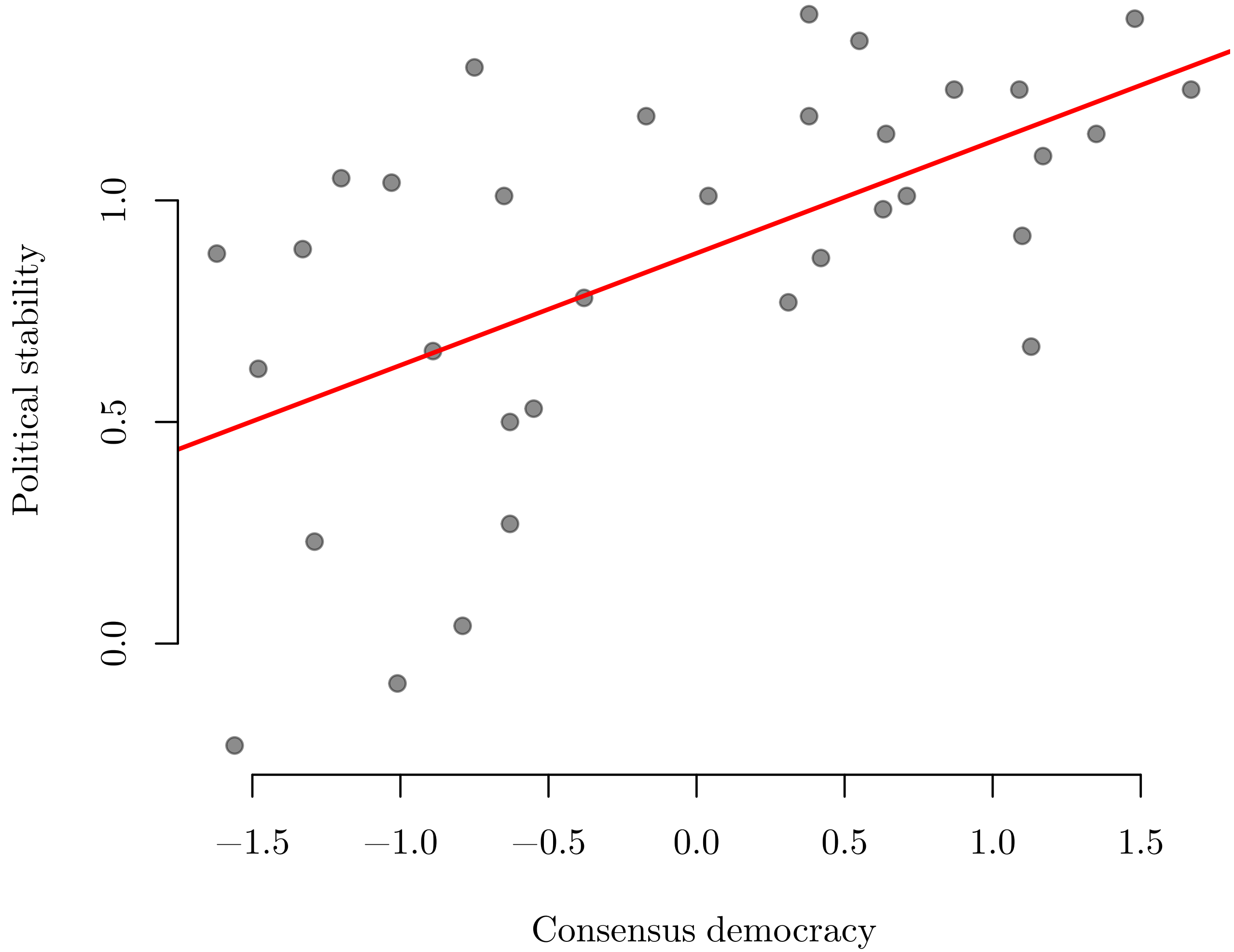
	Mortality	
	Cig. smoker	*Pipe smoker*
Age 55-60	*8.2*	*6.1*
Age 60-65	*10.4*	*8.7*

etc.

Matching: for every "treated" unit, find a similar "untreated" unit. Compare the two groups.

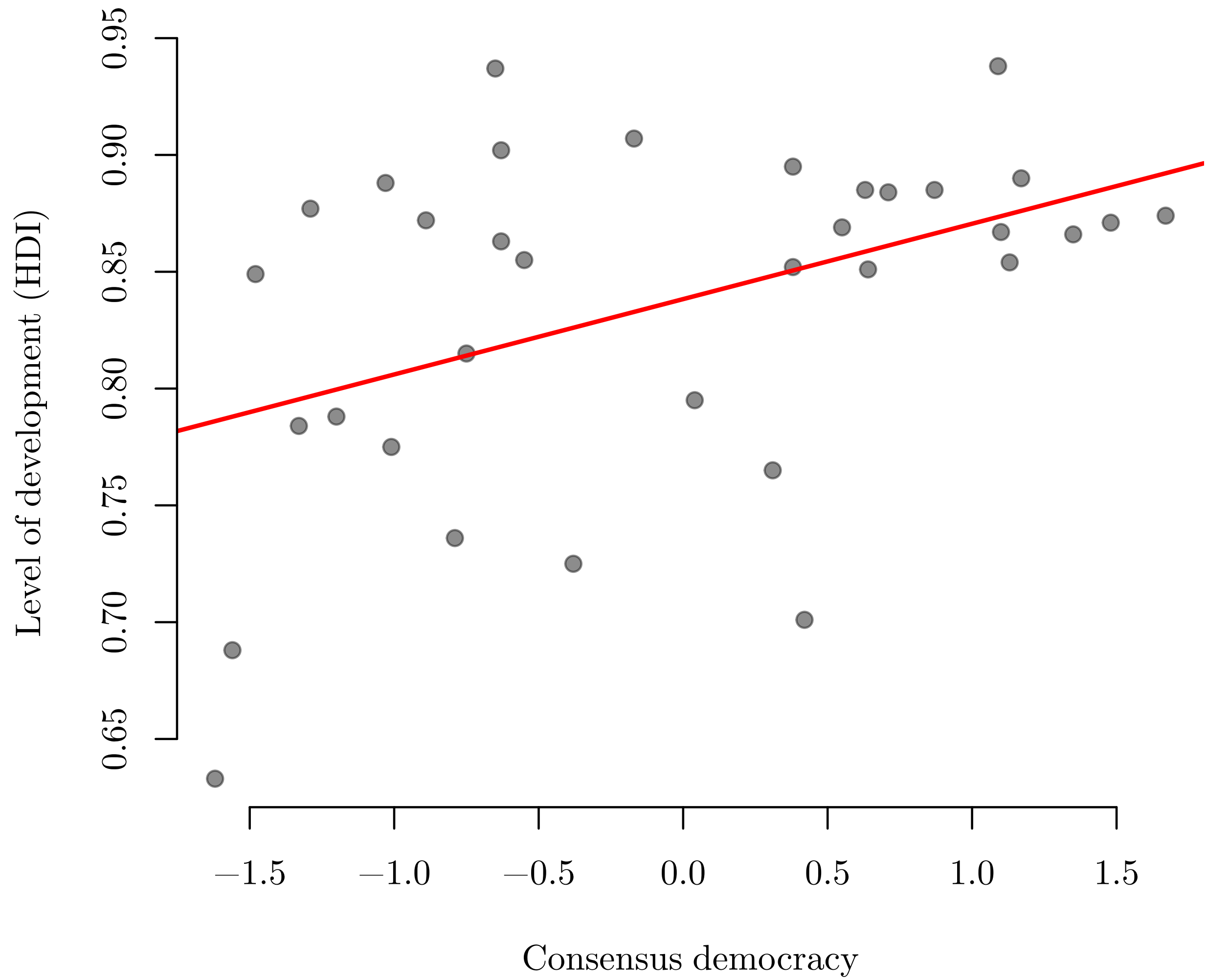
Both reasonable, but less common than regression, partly because less flexible.

Bivariate relationship



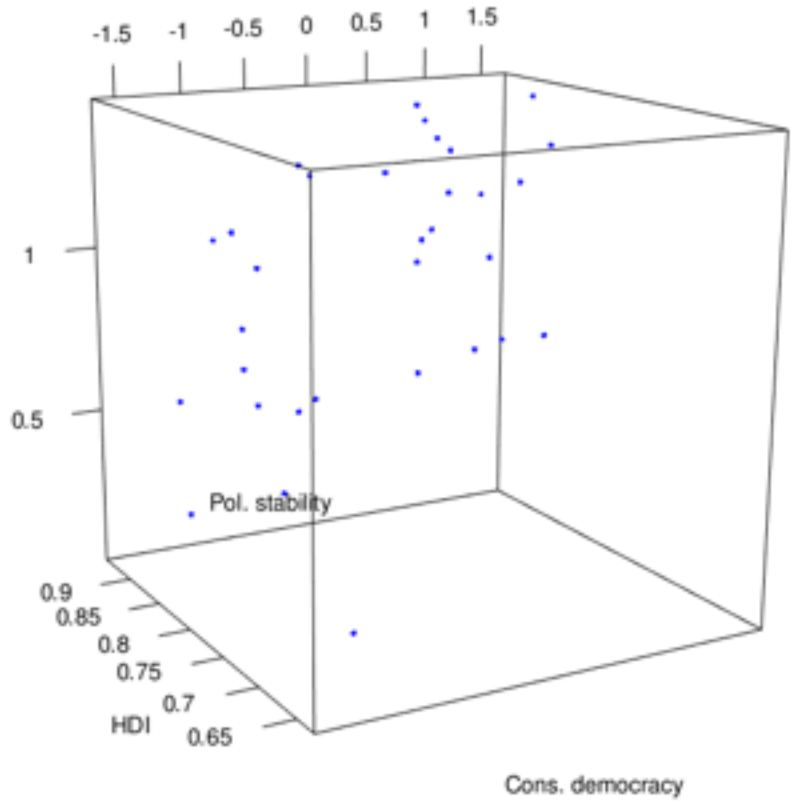
Consensus democracy is positively related to political stability.

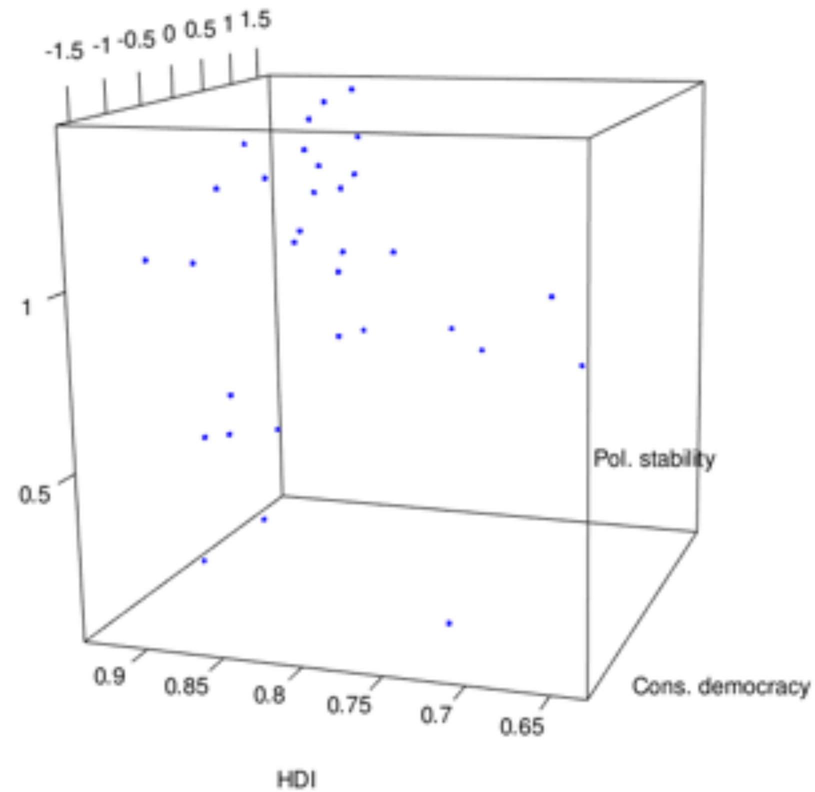
Another bivariate relationship



Consensus democracy is positively related to development.

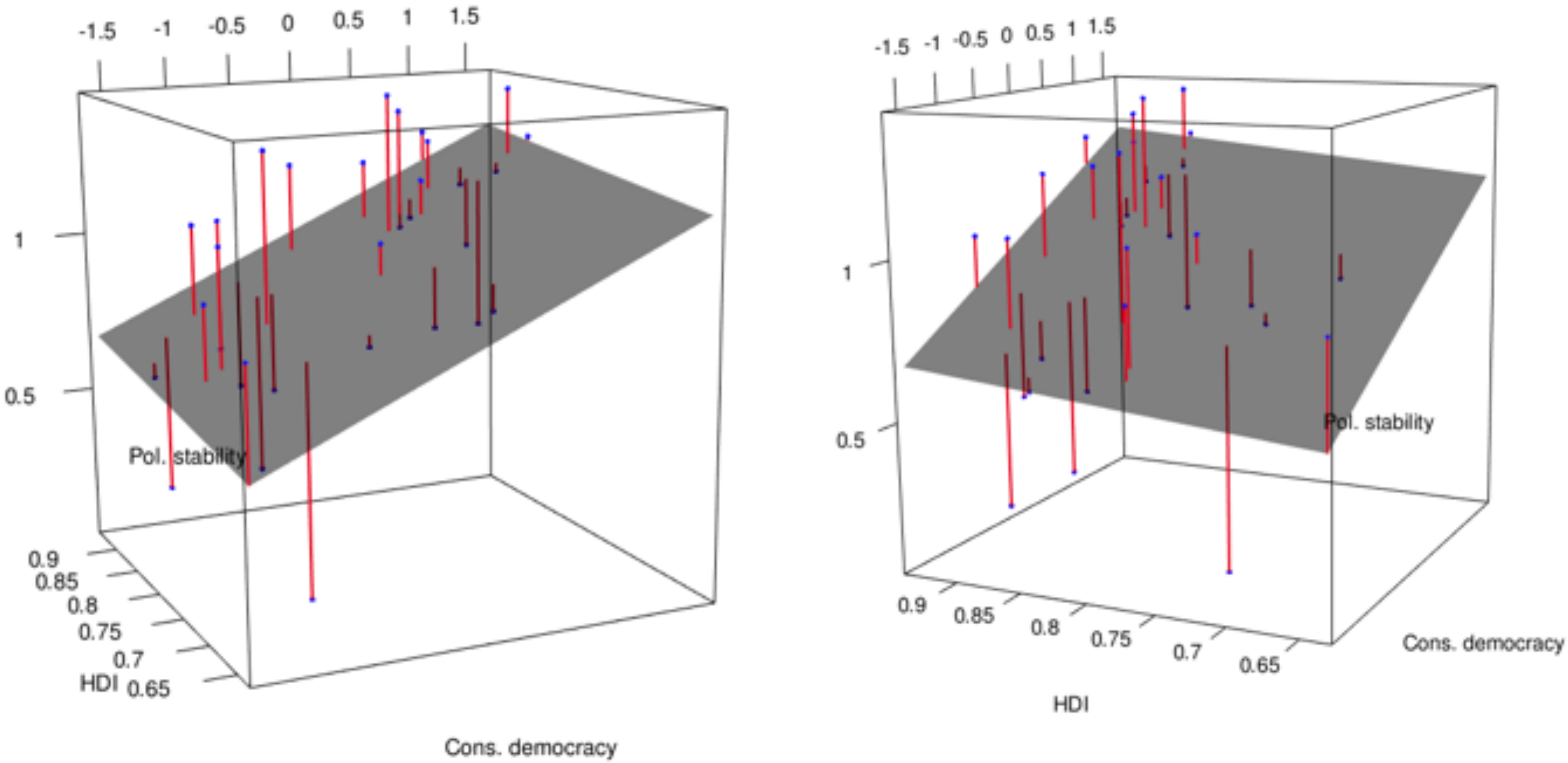
Multivariate relationship





See package rgl, persp() command

Multivariate regression

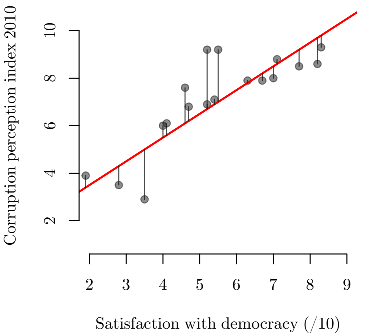


With two predictors, our prediction is a *plane*.

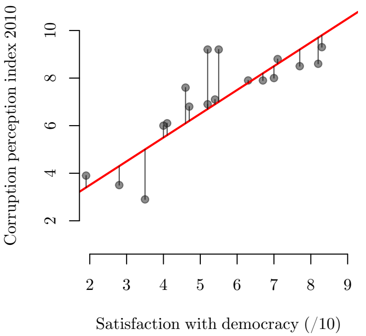
Residuals are given by line from the point to the plane.

OLS regression picks plane that minimizes sum of squared residuals.

Bivariate regression (recap)



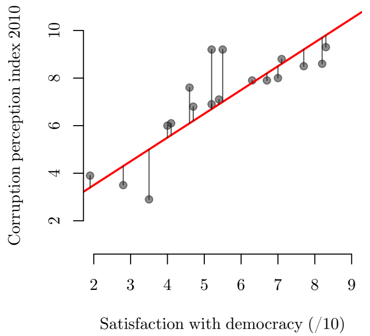
Bivariate regression (recap)



With one predictor (bivariate regression), the regression equation is:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i$$

Bivariate regression (recap)



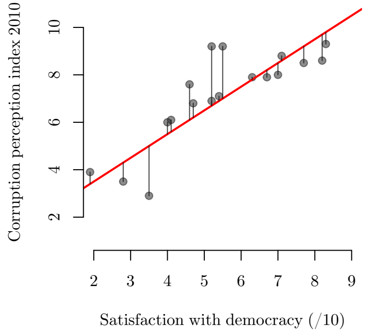
With one predictor (bivariate regression), the regression equation is:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i$$

In R, the command is (excluding outliers as Lijphart does):

```
> lm(pol_stab ~ cons_democ, data = d[!d$country %in% c("IND", "ISR"),])
```

Bivariate regression (recap)



With one predictor (bivariate regression), the regression equation is:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i$$

In R, the command is (excluding outliers as Lijphart does):

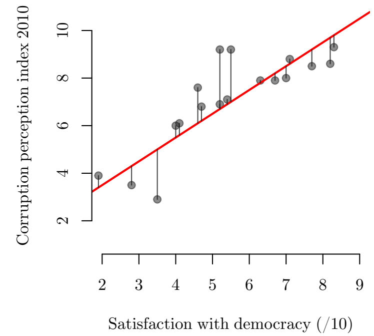
```
> lm(pol_stab ~ cons_democ, data = d[!d$country %in% c("IND", "ISR"),])
```

And the output is:

```
Call:
lm(formula = pol_stab ~ cons_democ, data = d[!d$country %in%
    c("IND", "ISR"), ])

Coefficients:
(Intercept)   cons_democ
     0.8808       0.2528
```

Bivariate regression (recap)



With one predictor (bivariate regression), the regression equation is:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i$$

In R, the command is (excluding outliers as Lijphart does):

```
> lm(pol_stab ~ cons_democ, data = d[!d$country %in% c("IND", "ISR"),])
```

And the output is:

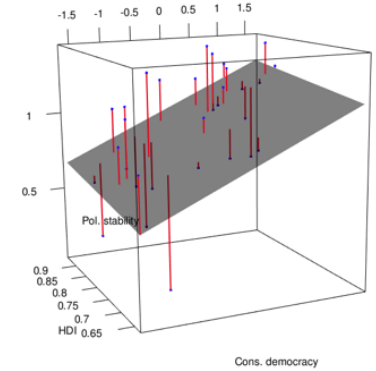
```
Call:
lm(formula = pol_stab ~ cons_democ, data = d[!d$country %in%
    c("IND", "ISR"), ])

Coefficients:
(Intercept)   cons_democ
     0.8808       0.2528
```

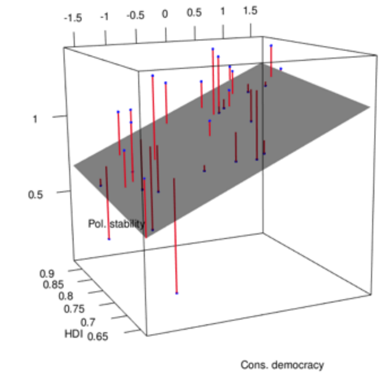
Intercept, i.e. β_0

Slope, i.e. β_1

Multivariate regression



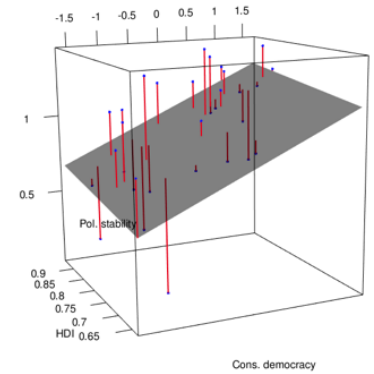
Multivariate regression



With two predictors (multivariate regression), the regression equation is:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

Multivariate regression



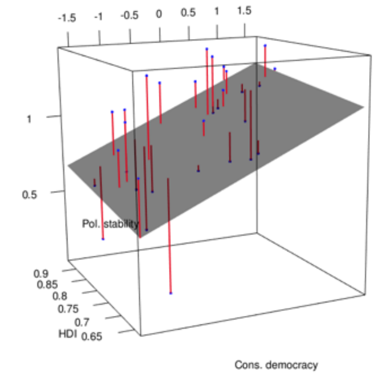
With two predictors (multivariate regression), the regression equation is:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

In R, the command is (excluding outliers as Lijphart does):

```
> lm(pol_stab ~ cons_democ + development, data = d[!d$country %in% c("IND", "ISR"),])
```

Multivariate regression



With two predictors (multivariate regression), the regression equation is:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

In R, the command is (excluding outliers as Lijphart does):

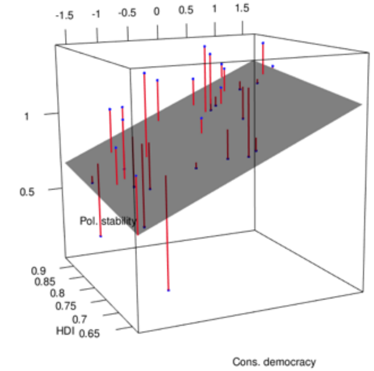
```
> lm(pol_stab ~ cons_democ + development, data = d[!d$country %in% c("IND", "ISR"),])
```

And the output is:

```
Call:
lm(formula = pol_stab ~ cons_democ + development, data = d[!d$country %in%
    c("IND", "ISR"), ])

Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

Multivariate regression



With two predictors (multivariate regression), the regression equation is:

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

In R, the command is (excluding outliers as Lijphart does):

```
> lm(pol_stab ~ cons_democ + development, data = d[!d$country %in% c("IND", "ISR"),])
```

And the output is:

```
Call:
lm(formula = pol_stab ~ cons_democ + development, data = d[!d$country %in%
    c("IND", "ISR"), ])

Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

Intercept, i.e. β_0

First slope, i.e. β_1

Second slope, i.e. β_2

Multivariate regression (2)

Multivariate regression (2)

You can use more than two predictors.

How many is Lijphart using in Table 15.2?

270 EFFECTIVE GOVERNMENT AND POLICY-MAKING

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

Multivariate regression (2)

You can use more than two predictors.

How many is Lijphart using in Table 15.2?

270 EFFECTIVE GOVERNMENT AND POLICY-MAKING

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

One predictor (no control variables):

```
> lm(pol_stab ~ cons_democ, data = d[!d$country %in% c("IND", "ISR"),])
```

Multivariate regression (2)

You can use more than two predictors.

How many is Lijphart using in Table 15.2?

270 EFFECTIVE GOVERNMENT AND POLICY-MAKING

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

One predictor (no control variables):

```
> lm(pol_stab ~ cons_democ, data = d[!d$country %in% c("IND", "ISR"),])
```

Two predictors (one control variable):

```
> lm(pol_stab ~ cons_democ + development, data = d[!d$country %in% c("IND", "ISR"),])
```

Multivariate regression (2)

You can use more than two predictors.

How many is Lijphart using in Table 15.2?

270 EFFECTIVE GOVERNMENT AND POLICY-MAKING

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

One predictor (no control variables):

```
> lm(pol_stab ~ cons_democ, data = d[!d$country %in% c("IND", "ISR"),])
```

Two predictors (one control variable):

```
> lm(pol_stab ~ cons_democ + development, data = d[!d$country %in% c("IND", "ISR"),])
```

Four predictors (three control variables):

```
> lm(pol_stab ~ cons_democ + development + logpop + socdiv, data =
d[!d$country %in% c("IND", "ISR"),])
```

How to think about (multivariate) regression

How to think about (multivariate) regression

Regression produces **conditional predictions**: our best guess of the outcome as a linear function of the predictors.

How to think about (multivariate) regression

Regression produces **conditional predictions**: our best guess of the outcome as a linear function of the predictors.

This regression output…

```
Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

How to think about (multivariate) regression

Regression produces **conditional predictions**: our best guess of the outcome as a linear function of the predictors.

This regression output…

```
Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

…implies this prediction equation:

$$\text{PolStab}_i = -0.07 + 0.22 \times \text{ConsDemoc}_i + 1.13 \times \text{Development}_i$$

How to think about (multivariate) regression

Regression produces **conditional predictions**: our best guess of the outcome as a linear function of the predictors.

This regression output…

```
Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

…implies this prediction equation:

$$\text{PolStab}_i = -0.07 + 0.22 \times \text{ConsDemoc}_i + 1.13 \times \text{Development}_i$$

So what would we predict for a country with ConsDemoc = 1 and Development = .9?

How to think about (multivariate) regression

Regression produces **conditional predictions**: our best guess of the outcome as a linear function of the predictors.

This regression output…

```
Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

…implies this prediction equation:

$$\text{PolStab}_i = -0.07 + 0.22 \times \text{ConsDemoc}_i + 1.13 \times \text{Development}_i$$

So what would we predict for a country with ConsDemoc = 1 and Development = .9?

-.07 + 0.22×1 + 1.13×0.9 = 1.17

How to think about (multivariate) regression coefficients

How to think about (multivariate) regression coefficients

A regression **coefficient** tells us how our prediction of the outcome changes with a one-unit change in the associated predictor *(holding other predictors fixed)*.

How to think about (multivariate) regression coefficients

A regression **coefficient** tells us how our prediction of the outcome changes with a one-unit change in the associated predictor *(holding other predictors fixed)*.

This regression output…

```
Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

How to think about (multivariate) regression coefficients

A regression **coefficient** tells us how our prediction of the outcome changes with a one-unit change in the associated predictor *(holding other predictors fixed)*.

This regression output…

```
Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

…implies this prediction equation:

$$\text{PolStab}_i = -0.07 + 0.22 \times \text{ConsDemoc}_i + 1.13 \times \text{Development}_i$$

How to think about (multivariate) regression coefficients

A regression **coefficient** tells us how our prediction of the outcome changes with a one-unit change in the associated predictor *(holding other predictors fixed)*.

This regression output…

```
Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

…implies this prediction equation:

$$\text{PolStab}_i = -0.07 + 0.22 \times \text{ConsDemoc}_i + 1.13 \times \text{Development}_i$$

How much does the predicted political stability change when the consensus democracy measure increases by one unit?

How to think about (multivariate) regression coefficients

A regression **coefficient** tells us how our prediction of the outcome changes with a one-unit change in the associated predictor *(holding other predictors fixed)*.

This regression output…

```
Coefficients:
(Intercept)   cons_democ  development
   -0.06594      0.21635      1.12932
```

…implies this prediction equation:

$$\text{PolStab}_i = -0.07 + 0.22 \times \text{ConsDemoc}_i + 1.13 \times \text{Development}_i$$

How much does the predicted political stability change when the consensus democracy measure increases by one unit?

.22

Another way to think about (multivariate) regression coefficients

Another way to think about (multivariate) regression coefficients

Another way to get the **coefficient** β_1 in this regression…

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

Another way to think about (multivariate) regression coefficients

Another way to get the **coefficient** β_1 in this regression…

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

…is to first estimate the **residuals** from this regression…

$$\text{ConsDemoc}_i = \alpha_0 + \alpha_1 \text{Development}_i$$

Another way to think about (multivariate) regression coefficients

Another way to get the **coefficient** β_1 in this regression…

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

…is to first estimate the **residuals** from this regression…

$$\text{ConsDemoc}_i = \alpha_0 + \alpha_1 \text{Development}_i$$

… and then run this regression:

$$\text{PolStab}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{ResidualsFromRegressionAbove}_i$$

Another way to think about (multivariate) regression coefficients

Another way to get the **coefficient** β_1 in this regression…

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

…is to first estimate the **residuals** from this regression…

$$\text{ConsDemoc}_i = \alpha_0 + \alpha_1 \text{Development}_i$$

… and then run this regression:

$$\text{PolStab}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{ResidualsFromRegressionAbove}_i$$

$\tilde{\beta}_1 = \beta_1$.

Another way to think about (multivariate) regression coefficients

Another way to get the **coefficient** β_1 in this regression…

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

…is to first estimate the **residuals** from this regression…

$$\text{ConsDemoc}_i = \alpha_0 + \alpha_1 \text{Development}_i$$

… and then run this regression:

$$\text{PolStab}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{ResidualsFromRegressionAbove}_i$$

$\tilde{\beta}_1 = \beta_1$.

More generally: to get the regression coefficient on a variable X_1 in the regression of Y on X_1 and X_2, X_3, etc, you can regress Y on the residuals from a regression of X_1 on X_2, X_3, etc.

Another way to think about (multivariate) regression coefficients

Another way to get the **coefficient** β_1 in this regression…

$$\text{PolStab}_i = \beta_0 + \beta_1 \text{ConsDemoc}_i + \beta_2 \text{Development}_i$$

…is to first estimate the **residuals** from this regression…

$$\text{ConsDemoc}_i = \alpha_0 + \alpha_1 \text{Development}_i$$

… and then run this regression:

$$\text{PolStab}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{ResidualsFromRegressionAbove}_i$$

$\tilde{\beta}_1 = \beta_1$.

More generally: to get the regression coefficient on a variable X_1 in the regression of Y on X_1 and X_2, X_3, etc, you can regress Y on the residuals from a regression of X_1 on X_2, X_3, etc.

This means: a coefficient on a variable in a multivariate regression tells us about the relationship between the outcome (Y) and the part of the variable that is not "explained" by the other variables.

Back to chocolate!

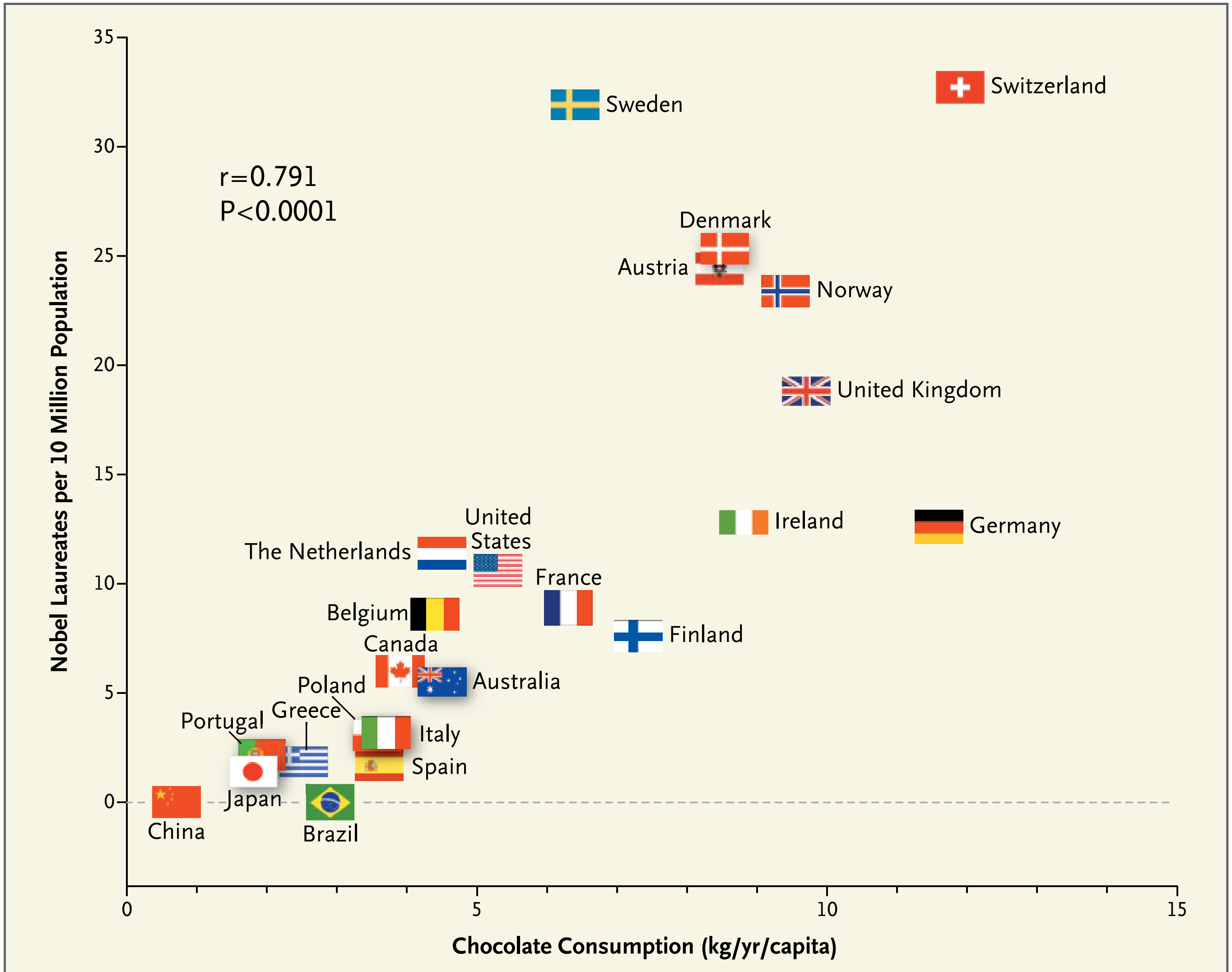


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Adding control variables: R session

Adding control variables: R session

```
> lm(lnobel_rate ~ lchocolate, data = cc)

Call:
lm(formula = lnobel_rate ~ lchocolate, data = cc)

Coefficients:
(Intercept)   lchocolate
     -1.629        2.092
```

Bivariate regression: no controls

Adding control variables: R session

```
> lm(lnobel_rate ~ lchocolate, data = cc)

Call:
lm(formula = lnobel_rate ~ lchocolate, data = cc)

Coefficients:
(Intercept)   lchocolate
     -1.629        2.092
```

Bivariate regression: no controls

```
> lm(lnobel_rate ~ lchocolate + GDP_capk, data = cc)

Call:
lm(formula = lnobel_rate ~ lchocolate + GDP_capk, data = cc)

Coefficients:
(Intercept)   lchocolate     GDP_capk
    -3.1664       1.0262       0.1049
```

Controlling for GDP per capita

Adding control variables: R session

```
> lm(lnobel_rate ~ lchocolate, data = cc)

Call:
lm(formula = lnobel_rate ~ lchocolate, data = cc)

Coefficients:
(Intercept)   lchocolate
     -1.629        2.092
```

Bivariate regression: no controls

```
> lm(lnobel_rate ~ lchocolate + GDP_capk, data = cc)

Call:
lm(formula = lnobel_rate ~ lchocolate + GDP_capk, data = cc)

Coefficients:
(Intercept)   lchocolate     GDP_capk
    -3.1664       1.0262       0.1049
```

Controlling for GDP per capita

```
> lm(lnobel_rate ~ lchocolate + GDP_capk + nw.europe, data = cc)

Call:
lm(formula = lnobel_rate ~ lchocolate + GDP_capk + nw.europe,
    data = cc)

Coefficients:
  (Intercept)     lchocolate       GDP_capk  nw.europeTRUE
      -2.9818         0.7090         0.1057         0.5488
```

Controlling for GDP per capita and NW Europe

Adding control variables: prediction equations

$$\text{NobelRate}_i = -1.63 + 2.09 \times \text{Chocolate}_i$$

$$\text{NobelRate}_i = -3.17 + 1.03 \times \text{Chocolate}_i + 0.10 \times \text{GDP}_i$$

$$\text{NobelRate}_i = -2.98 + 0.71 \times \text{Chocolate}_i + 0.11 \times \text{GDP}_i + 0.55 \times \text{NWEurope}_i$$

Adding control variables: prediction equations

$$\text{NobelRate}_i = -1.63 + 2.09 \times \text{Chocolate}_i$$

Bivariate regression: no controls

$$\text{NobelRate}_i = -3.17 + 1.03 \times \text{Chocolate}_i + 0.10 \times \text{GDP}_i$$

Controlling for GDP per capita

$$\text{NobelRate}_i = -2.98 + 0.71 \times \text{Chocolate}_i + 0.11 \times \text{GDP}_i + 0.55 \times \text{NWEurope}_i$$

Controlling for GDP per capita and NW Europe

Adding control variables: regression table

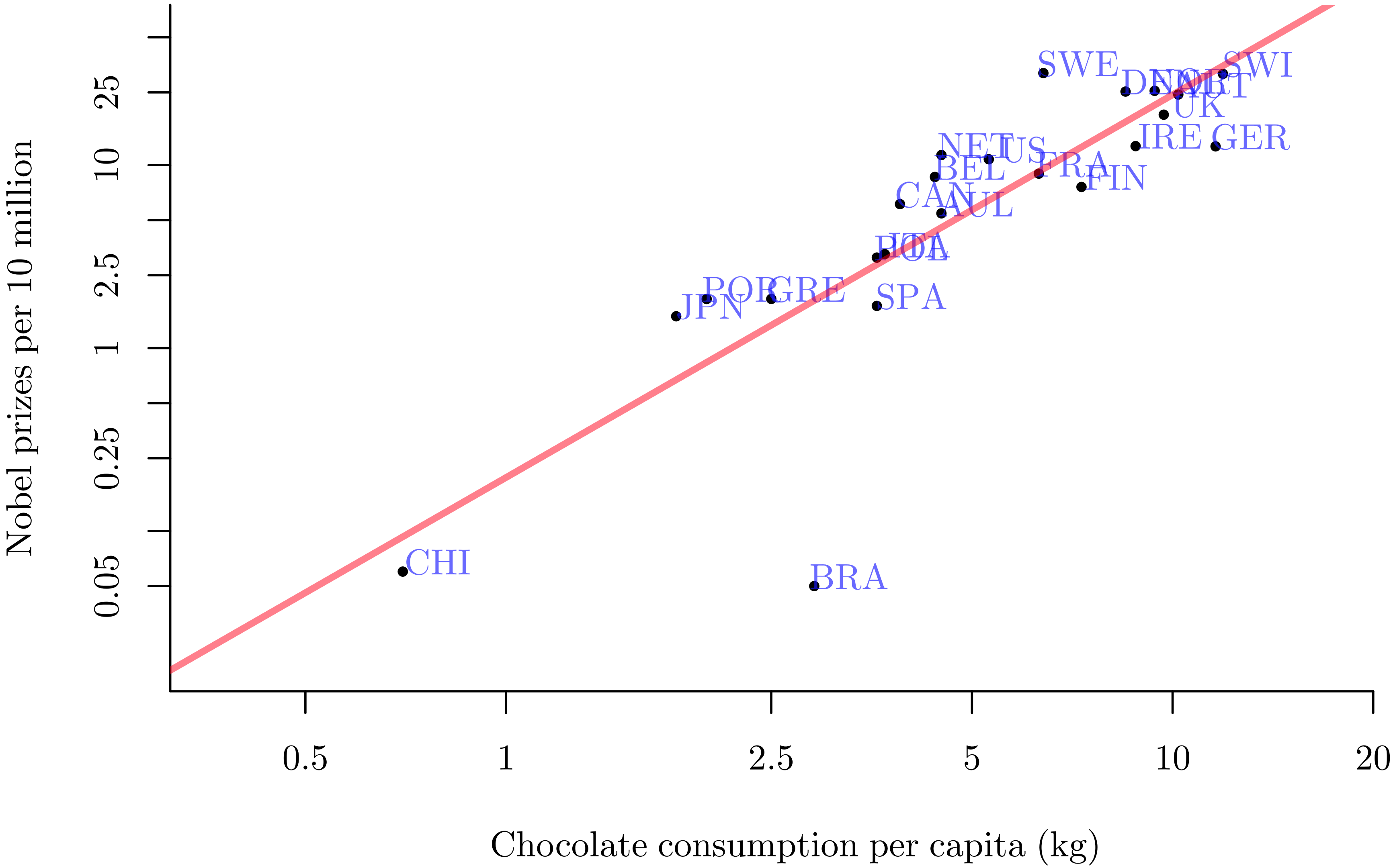
Dependent variable: Nobel Prizes awarded per capita (in log scale)

	(1)	(2)	(3)
Intercept	-1.629* (0.509)	-3.166* (0.511)	-2.982* (0.527)
Chocolate consumption per capita (log scale)	2.092* (0.298)	1.026* (0.326)	0.709 (0.415)
GDP/capita (thousands of USD)		0.105* (0.024)	0.106* (0.024)
NW Europe			0.549 (0.452)
R^2	0.70	0.85	0.86
N	34	34	34

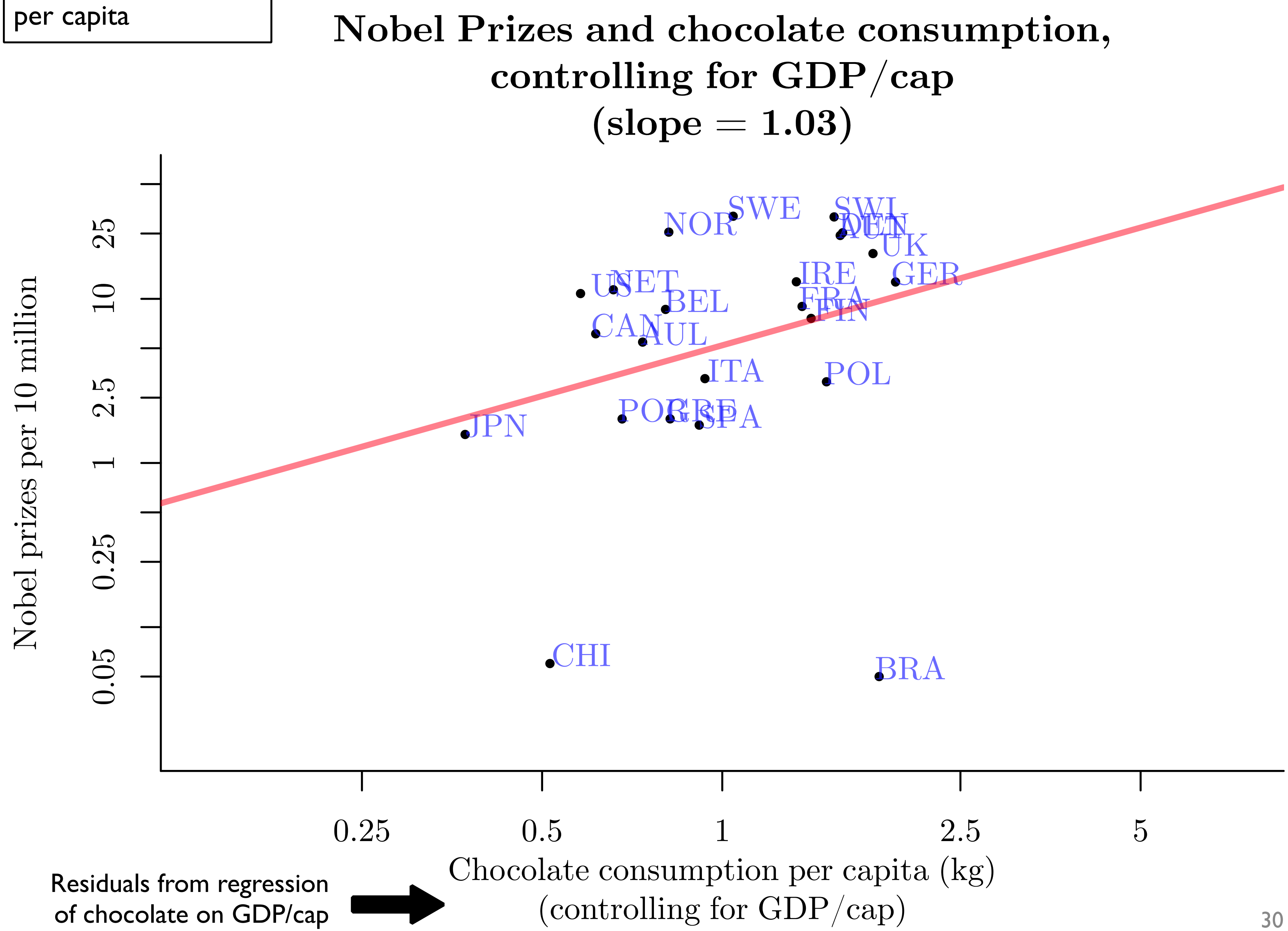
* Indicates p<0.05

Bivariate regression:
no controls

Nobel Prizes and chocolate consumption (slope = 2.09)

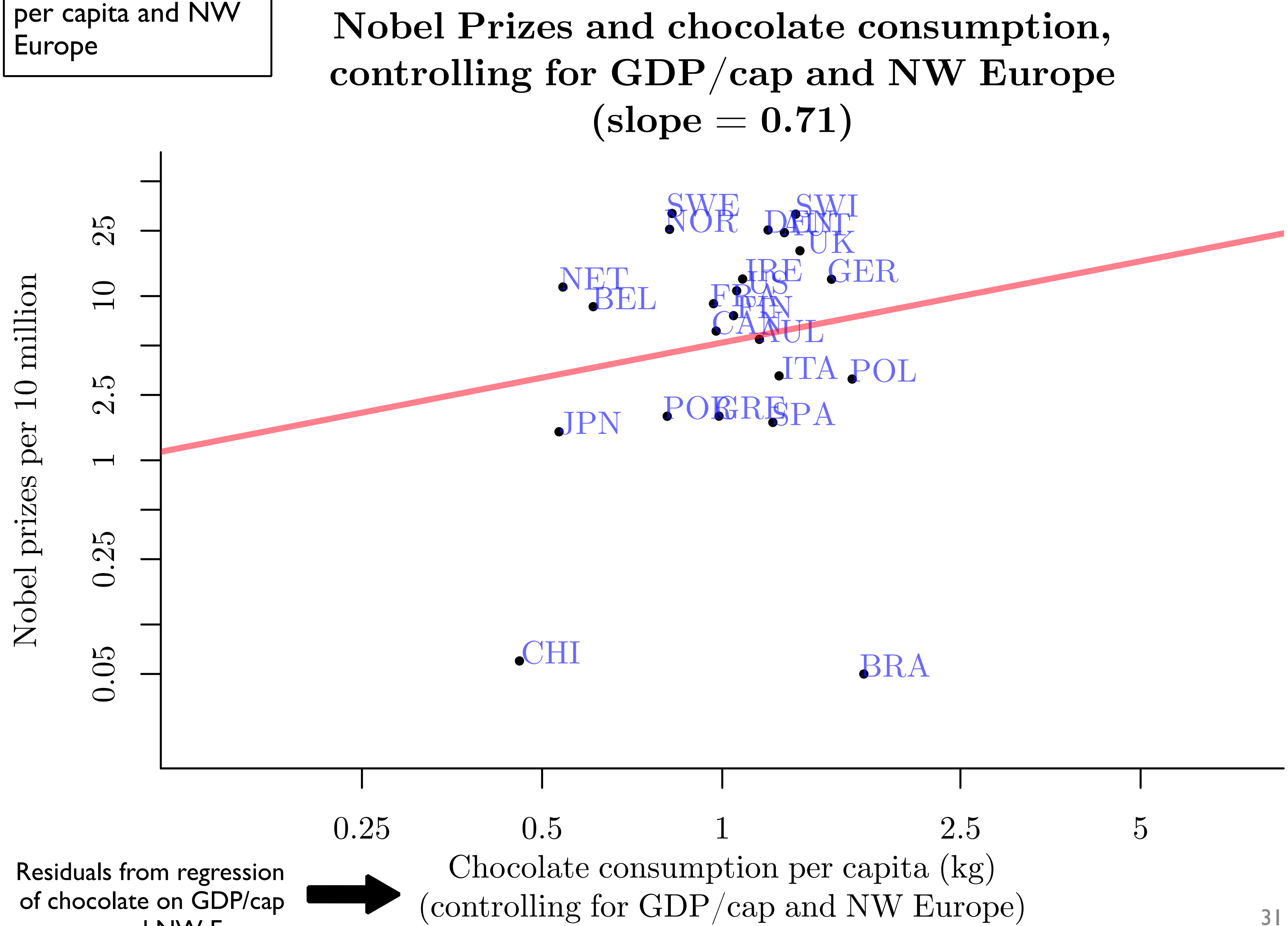


Controlling for GDP per capita



Residuals from regression of chocolate on GDP/cap

Controlling for GDP per capita and NW Europe

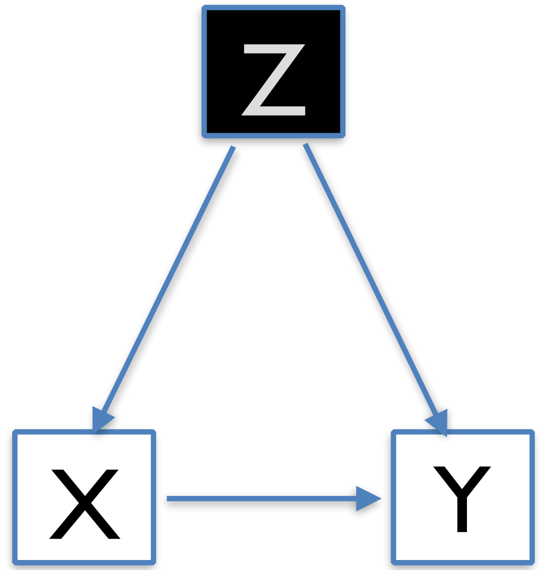


Residuals from regression of chocolate on GDP/cap and NW Europe

How this relates to your essay

Some questions you can ask about one of Lijphart's findings:

- What are some difference between consensus democracies and majoritarian democracies that Lijphart doesn't control for?
- What **should** Lijphart control for, given his questions and claims?
- Are the regression results the same when you control for an additional variable?
- Are the regression results the same when you include or exclude outliers?



This week's labs: regressions!

Upcoming lectures:

- Next week: Inference, i.e. assessing our confidence in an estimate.
- Week 8: Applying what you've learned to analyzing research in political science