

Case study 19

cross sectional data: snapshot of a sample of the population at one moment in time

longitudinal data: following people along multiple time points

key fact: very hard to draw valid longitudinal conclusions from cross sectional data

why? PCFs, as usual: year of birth (main PCF)

w/cross sectional data, people in the snapshot differ from each other not only on x and y, but also on PCFs

gerontology (age-period cohort effect)

Why do so many people try to draw longitudinal conclusions from cross-sectional data?

- longitudinal data is very expensive and complicated to collect

p. 118 + 122 #2

association \leftrightarrow correlation

(causes)

$X \rightarrow Y$ or $Y \rightarrow X$ or

need r that is decently big

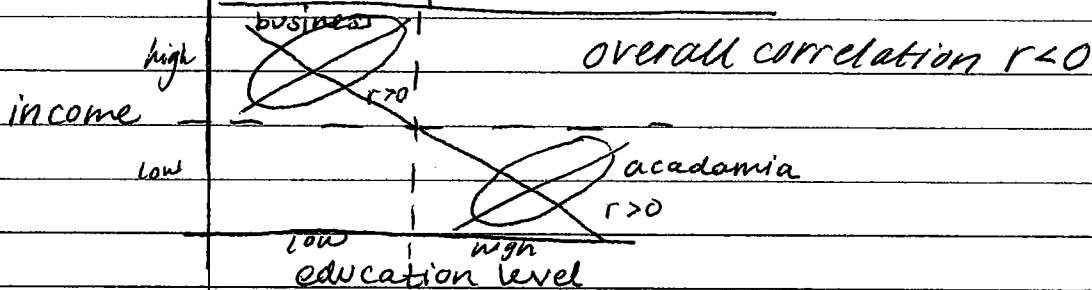
(related)

$X \rightarrow Y$

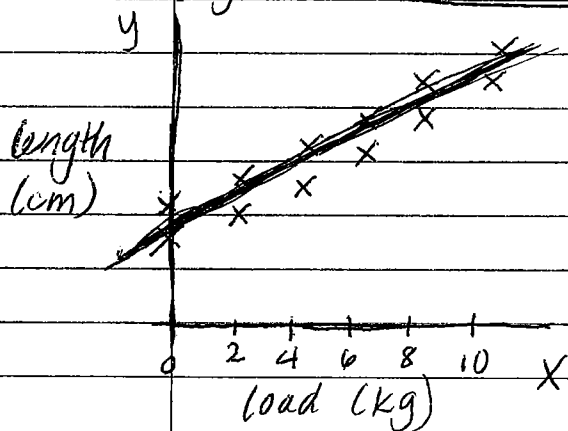
Z

(Z=PCF)

economics problem from HW



Regression in Controlled Experiments p. 120

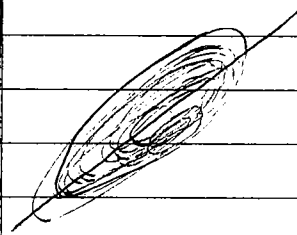


Hooke's Law $y = a + bx$

\uparrow \uparrow \uparrow \leftarrow load x in kg
length of string under load length under no load stretch per kg

When x is under experiment or control and all other factors have been held constant, the slope in regression equation has a valid cause and effect interpretation

y



cross sectional data on n persons
(measure x, y on each person)

Does the slope have a valid causal interpretation in this case?

- not necessarily, in fact often the slope has no causal meaning at all

x

p.122 #3

variable

education (x)

income (y)

mean

$\bar{x} = 14$ yrs

$\bar{y} = \$8000$

SD

$S_x = 3$ years

$S_y = \$3000$

$n = 144$

$r = +0.4$

5 number summary - only 5 numbers you need

y

income



reg. line

$$\text{slope} = r \cdot \left(\frac{S_y}{S_x} \right) = (.4) \left(\frac{\$3000}{3 \text{ yrs}} \right)$$

= \$400 income per year of education

Does this slope mean anything causally?

NO - causal conclusion on p.122 is an attempt to draw a longitudinal conclusion from cross sectional data - can't do

ed. level x

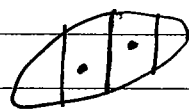
Q: If that's not what the slope means, what does it mean?

A: Associated with each year of education is an increase of \$400 in ~~the~~ income on average

(In other words, if I compare 2 groups of women whose education level differs on average by about 1 year, I expect their incomes to differ by \$400)

↑ "cross sectional conclusion"

y



shows difference of about \$400

x x+1 x