Confounding: Reducing it by Regression

Preamble

- Don't overlook classical, "non-regression" methods
- Regression methods are more "synthetic" (i.e. "artificial")
- Cf chapter 3 by Anderson et al. (c622; readings from aahovw)

Definitions ... / synonyms

Original (statistical, in design of experiments)

- inability to estimate higher order interactions
  (so typically assume they are zero)

- "mixed up with other effects" or "inextricable"

Epidemiological

- (osm)

Other terms

- "Lurking" (i.e. "hidden") variable

- "Simpson's Paradox" is the most extreme form

  *(see collection of Simpson's paradox examples under **Other Resources** on **c626**)*

Examples...

- Does using a Macintosh lead to sloppier writing? **a**
- Better Service from Canada Post after "Major Restructuring"**a**
- Salaries of Master's and PhD's **a**
- Outcomes of Pregnancy during Residency for women and wives of their male classmates• Admissions of Males & Females to Berkeley Graduate Schools **b**
- Percentage of White & Black Convicts Receiving Death Penalty **a**
- Intelligence Quotient (IQ) - Mother's Milk; Other Variables **a**
- Lung Function of Vanadium Factory Workers **Other resources, c697**

  - vs. reference group (matched for smoking and age) that was 3.4 cm different in ave. height

- Blood Pressure and Altitude - age; height; weight; country **b**
- Longevity - Sexual Activity; thorax size **c622**
- Fatalities & Speed Limit Change - Time **a**
- NEURODEVELOPMENT OF CHILDREN EXPOSED IN UTERO TO ANTIDEPRESSANT DRUGS **b**
- What Does It Take to Heat a New Room? **dataset, c697**

---

a notes on Ch 2, c607 **b** resources this course (678), session 5

Adjustment via regression ...

- "Outcome" **Y**

- Contrast with respect to **X** ("Exposure" variable)

  (for now, say X is binary **X=1** and **X=0**)

- Confounder **C**

CRUDE CONTRAST:

via $E[Y \mid X] = b_0 + b_X X$

b_X = crude difference = $\bar{Y}_{X=1} - \bar{Y}_{X=0}$

ADJUSTED CONTRAST:

$E[Y \mid X \underline{\underline{C}}] = b_0^* + b_X^* X + b_C C$

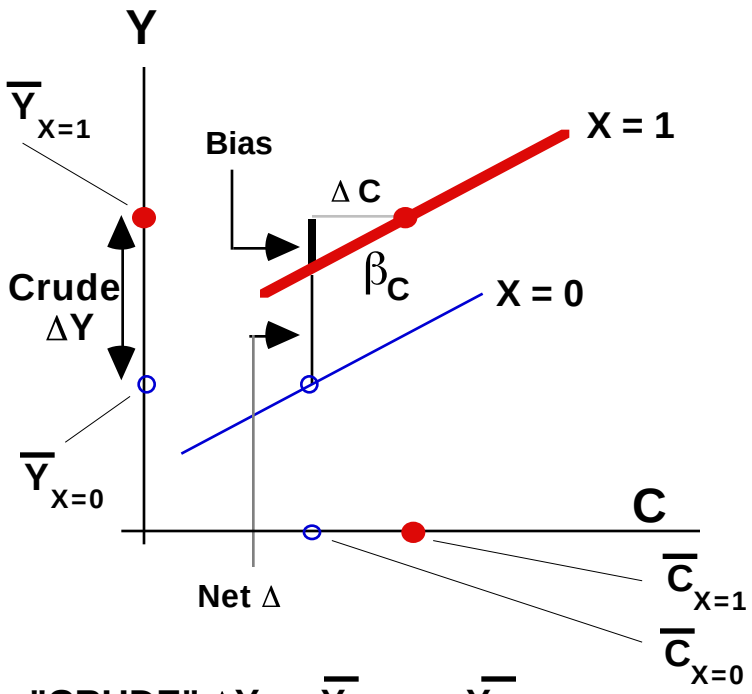
b_X^* = adjusted difference

= $\bar{Y}_{X=1} - \bar{Y}_{X=0}$ ***(CRUDE Δ)***

minus

$b_C (\bar{C}_{X=1} - \bar{C}_{X=0})$ ***(ADJUSTMENT)***

In Pictures... (cf Anderson et al. chapter)



"CRUDE" $\Delta Y = \bar{Y}_{X=1} - \bar{Y}_{X=0}$

$\Delta C = \bar{C}_{X=1} - \bar{C}_{X=0}$

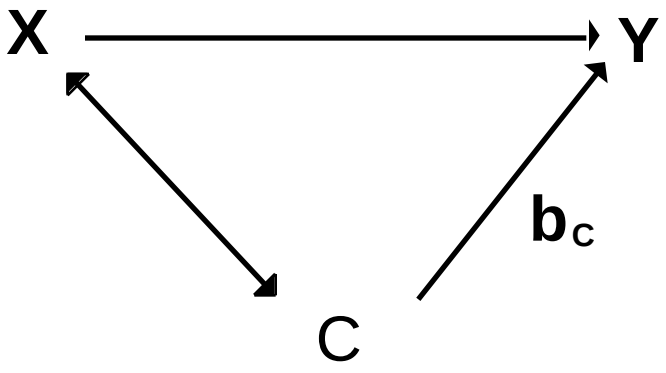
Bias $= \beta_C \times \Delta C$

"Net" $\Delta Y = \bar{Y}_{X=1} - \bar{Y}_{X=0} - \beta_C \times \Delta C$

cf Notes re "First Visit to Mars" (G&S Ch 1+) on c678 page

Anatomy of the "Adjustment"

$b_C \left(\bar{C}_{X=1} - \bar{C}_{X=0} \right)$



- for a NON-ZERO ADJUSTMENT...

b_C NON ZERO

AND

$\left(\bar{C}_{X=1} - \bar{C}_{X=0} \right)$ NON ZERO

Special issues

1.

- Adjustment uses a LINEAR relation Y <--> C

If Y <--> C relationship not linear, using a linear relation will not produce correct adjustment

e.g. Y= birthweight and C = Age in residents' study

2.

- If Y <--> C relationship not same at different levels of X

(ie if C is a modifier of X<->Y rel'n,
or X is a modifier of C<->Y rel'n
i.e. if X<-->C "interaction")

then cannot make a unique "adjustment"
(adjustment different at different levels of C)

e.g. gender D's in salary (C = # years experience)
c.f. Miettinen diagram (covariate as a modifier, confounder, or both)

3.

- Inappropriate Adjustment…

X ---> C ---> Y

X ---> Y ---> C