Adjusting for covariates

Jeff Leek

@jtleek

www.jtleek.com

Adjusting in linear regression

- Equivalent to fitting multiple lines
- Can “account” for other effects
- All the same caveats apply
- Interpretation changes

http://apps.who.int/gho/athena/data/GHO/WHOSIS_000008.csv?profile=text&filter=COUNTRY:;SEX:

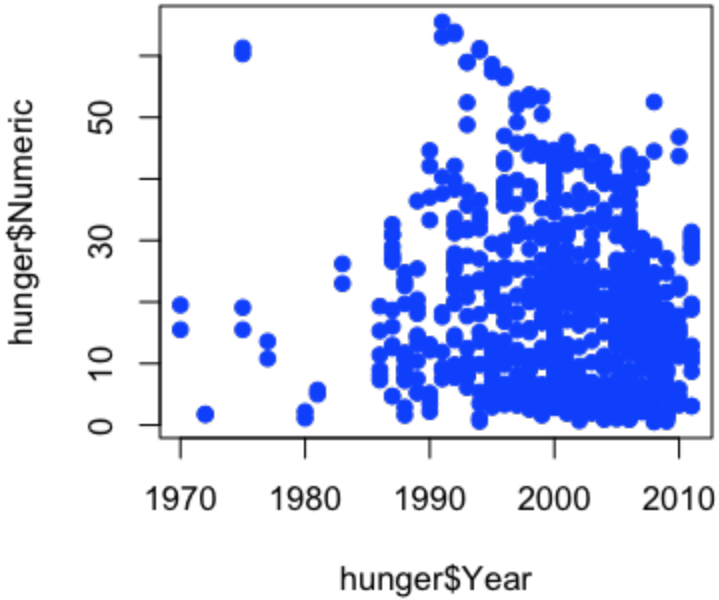
FACT SHEET

GOAL 1
Eradicate Extreme Poverty and Hunger

TARGETS

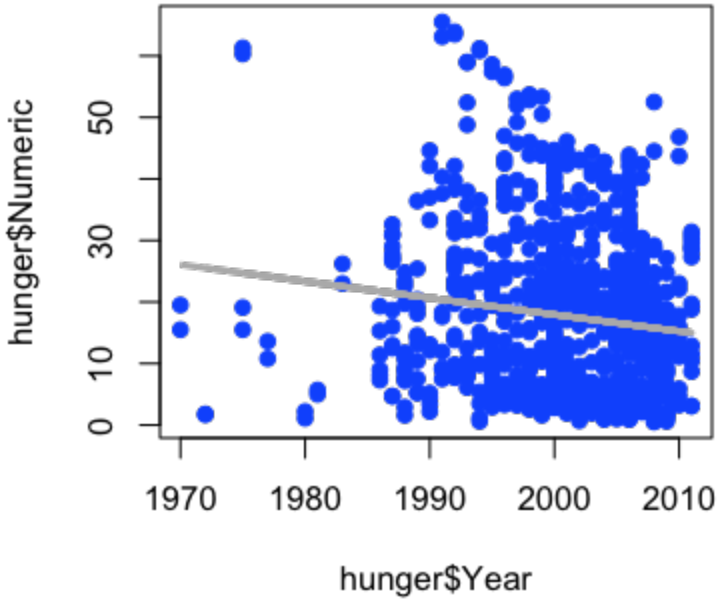
1. Halve, between 1990 and 2015, the proportion of people whose income is less than $1 a day
2. Achieve full and productive employment and decent work for all, including women and young people
3. Halve, between 1990 and 2015, the proportion of people who suffer from hunger

https://github.com/jtleek/dataanalysis/blob/master/week4/006multipleVariables/

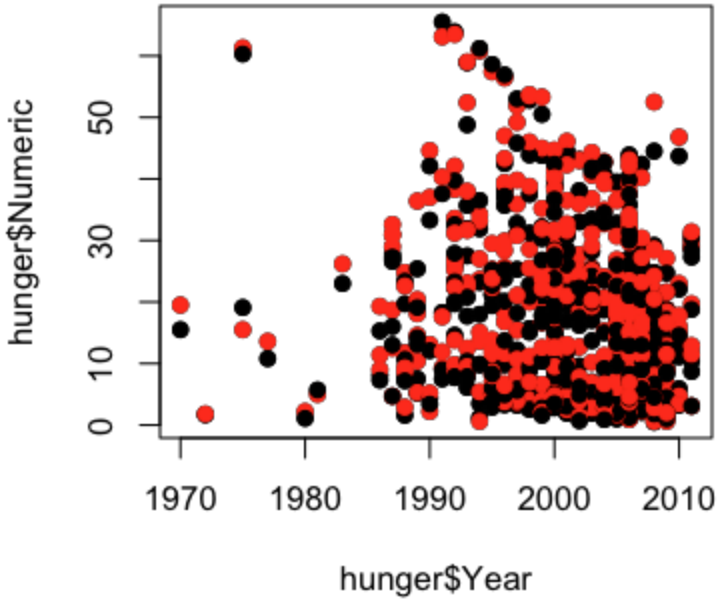


https://github.com/jtleek/dataanalysis/blob/master/week4/006multipleVariables/

$$Hu_i = b_0 + b_1 Y_i + e_i$$

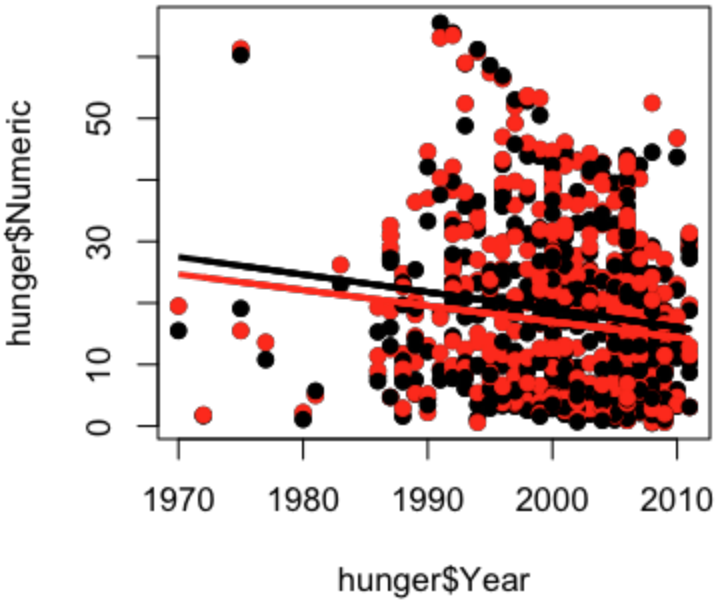


https://github.com/jtleek/dataanalysis/blob/master/week4/006multipleVariables/



https://github.com/jtleek/dataanalysis/blob/master/week4/006multipleVariables/

$$Hu_i = b_0 + b_1 Y_i + b_1 F_i + e_i$$



b_0 - percent hungry at year zero for females

$b_0 + b_1$ - percent hungry at year zero for males

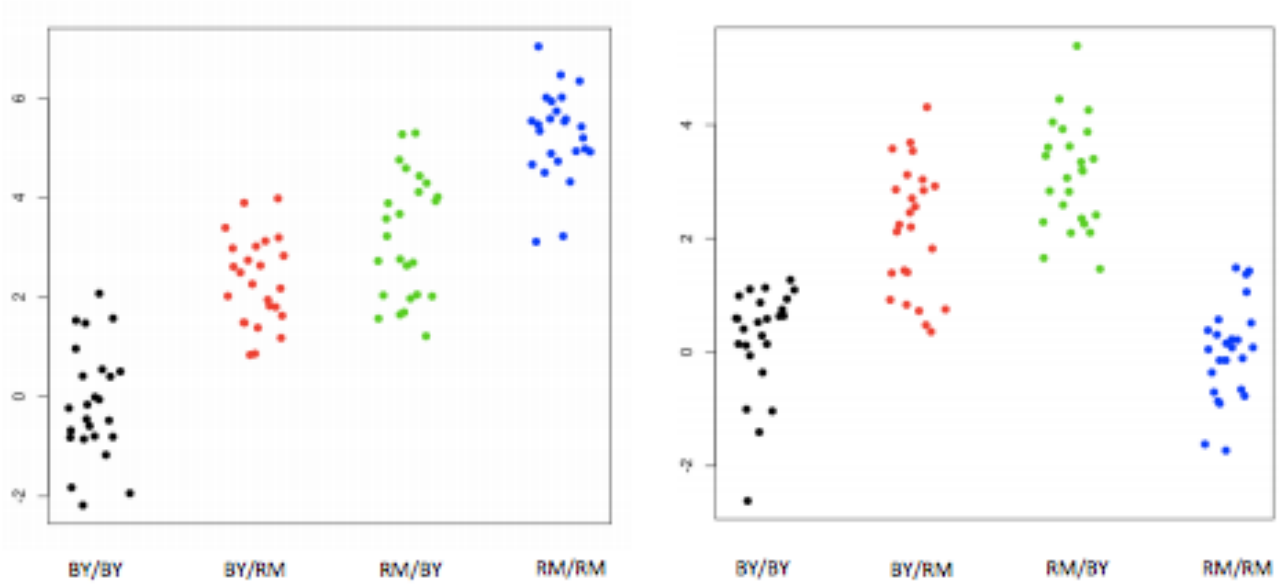
b_2 - change in percent hungry (for either males or females)

in one year

e_i - everything we didn't measure

Interaction terms

Slide courtesy: Ingo Ruczinski



Expression = Baseline + RM Effect + BY Effect + (RM Effect * BY Effect) + Noise

Notes and further reading

- Linear models is a whole class (no joke): https://www.coursera.org/course/regmods
- Basic thing to keep in mind is how many levels do you want to fit? What makes sense biologically?
- Great additional notes in Chapter 2 here: http://genomicsclass.github.io/book/