JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

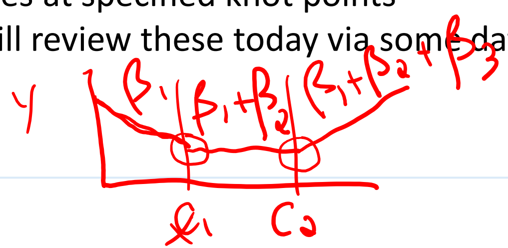
Lecture 3

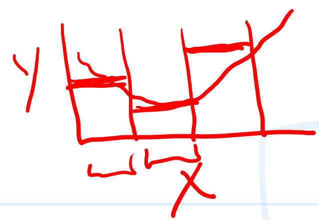
Basic functions for building regression models: indicator variables; splines; interactions WITH applications

Review of key concepts from Lecture 2

- Two main uses / purposes:
  - Etiology: creating useful models to describe how Y is caused or associated with a set of Xs
  - Prediction: predicting Y using X

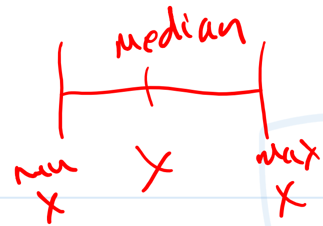
- Tools for building models to described how Y changes with X Single variable
  - Step function: Mean of Y as a function of X changes with abrupt jumps (up and down)
  - Linear spline: Mean of Y as a function of X changes linearly with a slope that changes at specified knot points
  - Cubic spline: Mean of Y as a function of X is locally cubic with curvature that is changes at specified knot points
  - We will review these today via some data analysis





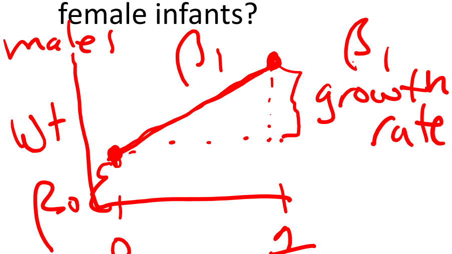
Review of key concepts from Lecture 2

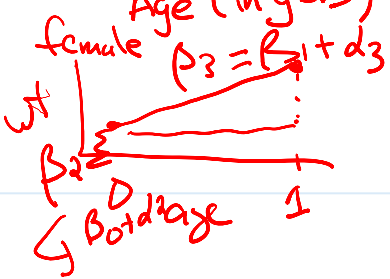
- Additional smoothers from Lab 1
  - "loess": locally estimated scatterplot smoothing
  - "lowess": locally weighted estimated scatterplot smoothing
    - Estimates mean Y at a given X using a polynomial function (degree 1 or 2) based on some fraction of the observed data (span, e.g. 0.3 or 30% of the data) using weights: $(1- |d|^3)^3$
  - "natural spline" or "natural cubic spline"
    - Cubic splines that assume a linear function beyond the boundary knots
    - You can provide the knots
    - Alternatively, you specify the "degrees of freedom": defines the number of interior knots (df – 1 – intercept) set at appropriate quantile, where the default boundary knots are the min/max value of X 1



Interactions of simple functions

- Interactions allow for E(Y|X) = f(x) to vary across subsets of the population of interest
- Effect modification
- During the first year of life, is the average "growth rate" in weight for male infants the same as for female infants?





$$E(wt \mid age) = \beta_0 + \beta_1 age \Rightarrow male$$

$$E(wt \mid age) = \beta_2 + \beta_3 age \Rightarrow females$$

$$F = \begin{cases} 1 & F \\ 0 & m \end{cases}$$

$$E(wt \mid age) = \beta_0 + \beta_1 age + \alpha_2 F + \alpha_3 F age$$

$H_0: \alpha_3 = 0$

$H_A: \alpha_3 \neq 0$

Male : $\beta_0 + \beta_1 age$

females : $F = 1 \quad (\beta_0 + \alpha_2) + \underbrace{(\beta_1 + \alpha_3)}\, age$

Interactions of simple functions

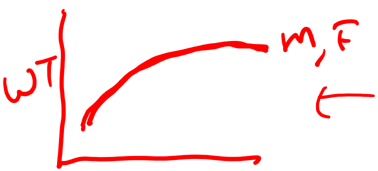
- During the first year of life, is the average "growth rate" in weight for male infants the same as for female infants?

Interactions of simple functions

- During the first year of life, is the average "growth curve" for weight for male infants the same as for female infants?

growth is quadratic

$$E(wt \mid age, F) = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 F + \beta_4 F age + \beta_5 F \times age^2$$

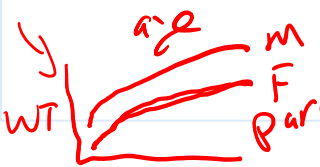


model

m: $\beta_0 + \beta_1 age + \beta_2 age^2$

F: set F=1

$(\beta_0 + \beta_3) + (\beta_1 + \beta_4) age + (\beta_2 + \beta_5) age^2$



parallel lines

$\boxed{\beta_4, \beta_5 = 0}$

$\beta_3, \beta_4, \beta_5 = 0$

Interactions of simple functions

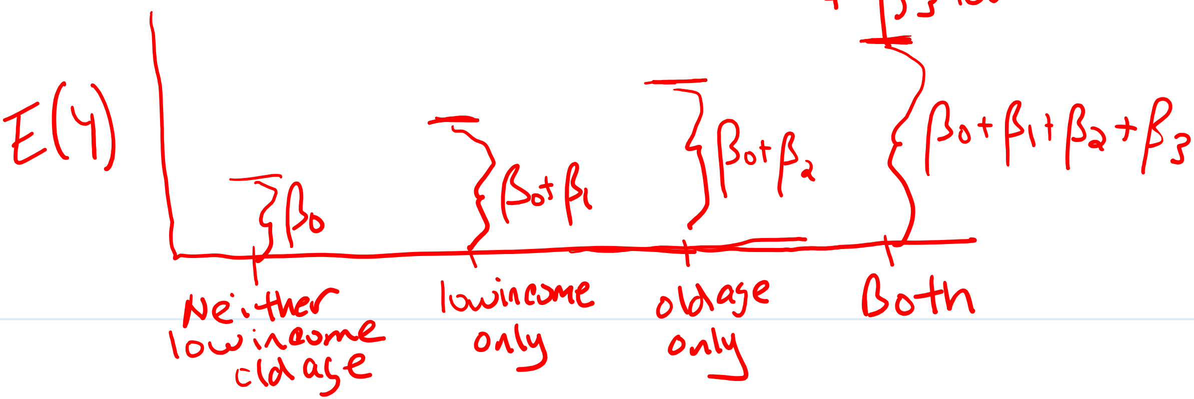
- Is the effect on average medical expenditures of being both poor and older greater than would be expected given the independent effects of poverty and old age alone

Y = medical expenditures

$$\text{lowincome} = \begin{cases} 1 & \text{low income} \\ 0 & \text{o/w} \end{cases}$$

$$\text{oldage} = \begin{cases} 1 & \text{age} \geq 65 \\ 0 & \text{age} < 65 \end{cases}$$

$$E(Y \mid \text{lowincome}, \text{oldage}) = \beta_0 + \beta_1 \text{lowincome} + \beta_2 \text{oldage} + \beta_3 \text{lowincome} \times \text{oldage}$$



And now some ANALYSIS!

- For the rest of today, we are going to work through several analyses for the Nepali Anthropometry study.