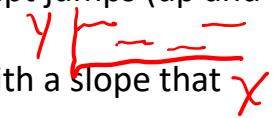
Lecture 3

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 — a single X, continuous
  - 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

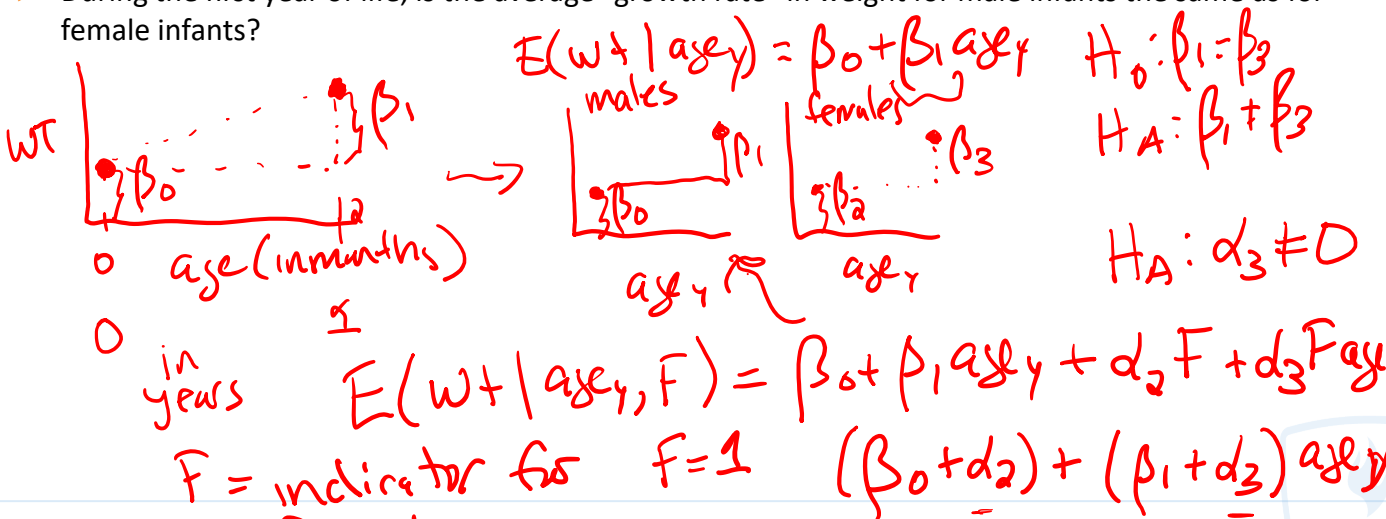


Review of key concepts from Lecture 2

- 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

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?

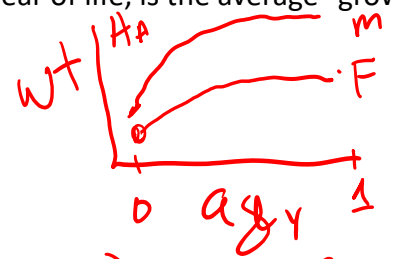


Interactions of simple functions

- 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?



quadratic function

$$E(wt \mid age_Y, F) = \underbrace{\beta_0 + \beta_1 age_Y + \beta_2 age_Y^2}_{\text{quadratic function for males}} + \beta_3 F + \beta_4 F age_Y + \beta_5 F age_Y^2$$

Females: $F = 1$

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

$H_0: \beta_4 = 0, \beta_5 = 0$ $\quad$ H_A: at least one $\beta_i \neq 0$, $i = 4, 5$

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 expenditure

Income = 1 → poor
0 → ow

Age = 1 age > 65
0 ow

null model

$E(Y \mid Income, Age) = \beta_0 + \beta_1 Income + \beta_2 Age$

interaction model

$= \beta_0 + \beta_1 Income + \beta_2 Age + \beta_3 Income\ Age$

And now some ANALYSIS!

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