Week 6 – Tuesday Session

Linear Regression

EPI202 – Epidemiologic Methods II
Murray A. Mittleman, MD, DrPH
Department of Epidemiology, Harvard TH Chan School of Public Health

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Advantages of Regression Models

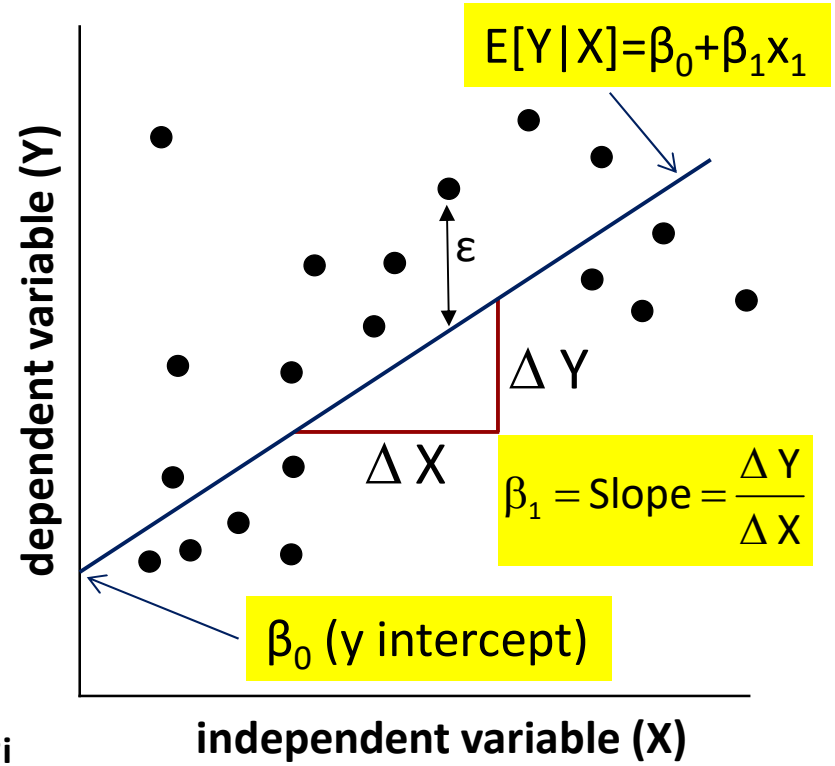
- Generally more efficient than stratification-based methods when data are sparse
- Modelling of a continuous outcomes
- Specify continuous and categorical exposures, confounders and modifiers
- Specify interactions to model effect modification
- Model nonlinear relations associations hips between exposure and outcome and other covariates

Regression Models

y	=	$\beta_0+\beta_1x_1+\ldots\beta_ix_i+\varepsilon$
Dependent		Independent
Predicted		Predictor variables
Response variable		Explanatory variables
Outcome variable		Covariables

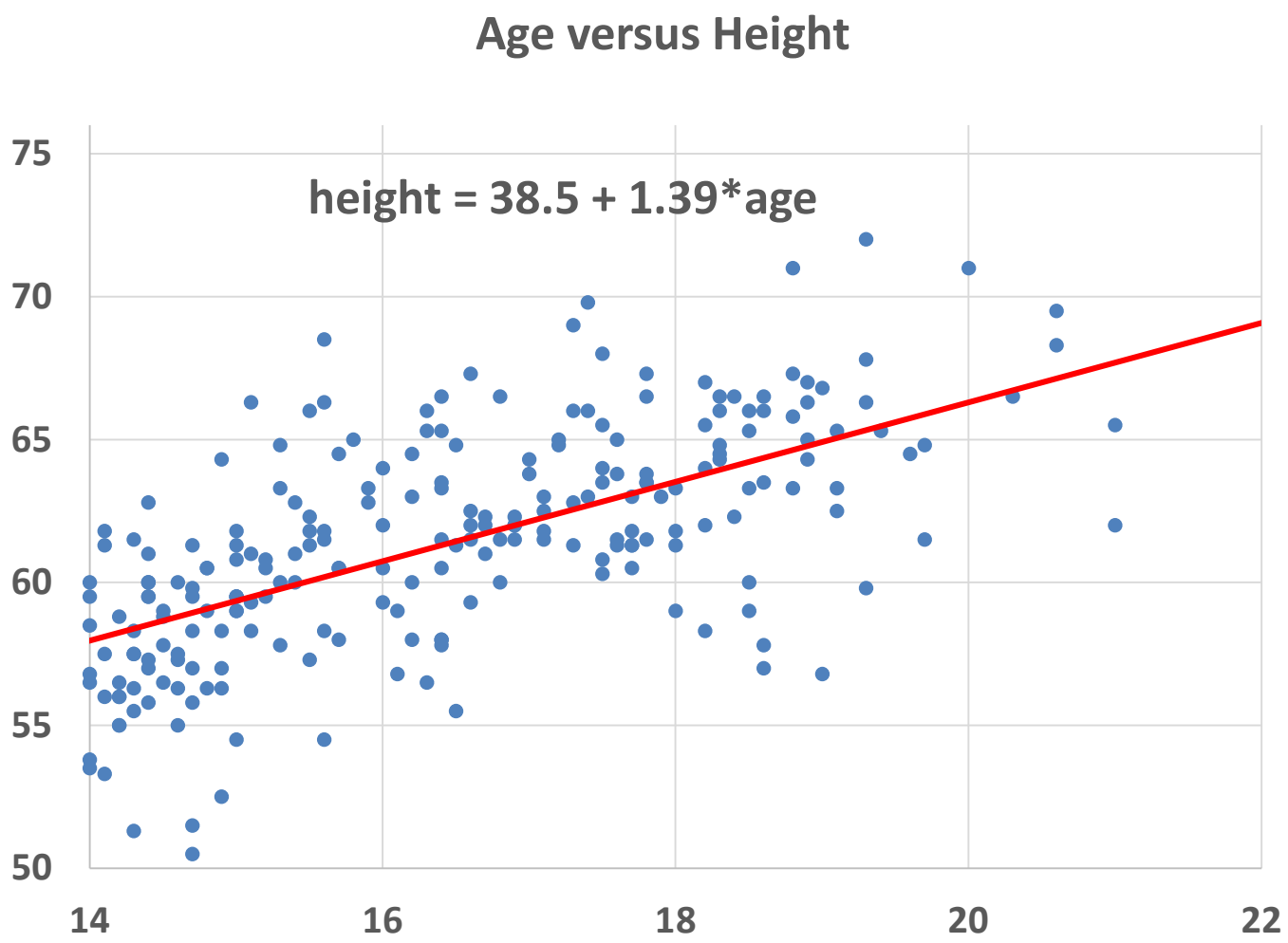
Simple Linear Regression

- Predict a continuous dependent (outcome) variable y from a continuous independent (exposure) variable x
- Simple linear regression fits a straight line to the data using the least squares method.
- Regression line: $\mathbf{E[Y|X]=\beta_0+\beta_1 x_1}$
  - Often presented as y = mx + b where
    - b=y-intercept
    - m=slope=$\Delta y/\Delta x$ (rise/run)



- Individual predicted value: $\mathbf{Y_i=\beta_0+\beta_1 x_{1i}+ \varepsilon_i}$
  - β_0 =y-intercept (where the line crosses the Y-axis)
  - β_1 =slope=$\Delta y/\Delta x$ =average change in y when x changes by one unit
  - X_1 is a known constant
  - ε , the error, is an observation's deviation from the conditional mean, N(0, $\sigma 2$)

Linear Regression



Week 6 Linear Regression Exercise

HAVE A GOOD WEEK