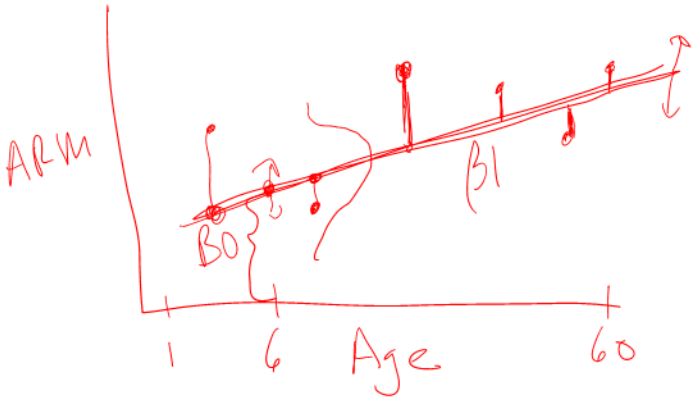
Lecture 5

The classical linear regression model

Review of key concepts from Lecture 3 and 4

- Simple linear regression model
  - $\underbrace{\text{ARM}} = \underbrace{B0 + B1\,(\text{age} - 6)}_{\text{systematic component}} + \underbrace{e}_{\text{random component}}$, $e \sim N(0,\sigma^2)$, independent



B0 = average ARM among 6 month old children

B1 = difference in average ARM comparing children who vary in a age by 1-month

Review of key concepts from Lecture 3 and 4

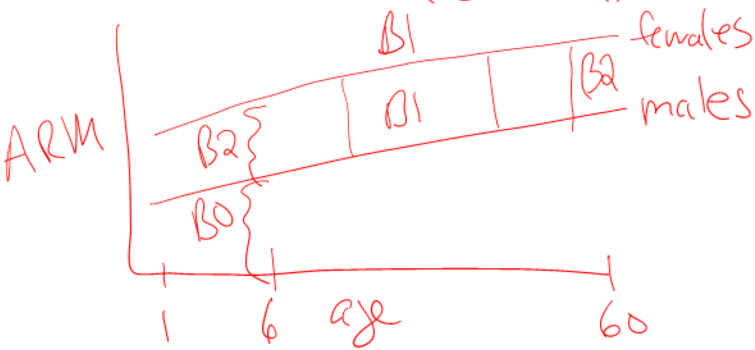
- Sex adjusted relationship between ARM and age
  - ARM = B0 + B1 (age – 6) + B2 Female + e, e~N(0,σ^2), independent

Controlling for
adjusting for

male: $ARM = B0 + B1(age-6) + e$

female: $ARM = B0 + B1(age-6) + B2 + e$

$= (B0 + B2) + B1(age+6) + e$



B0 = Average ARM among male children who are 6 months of age

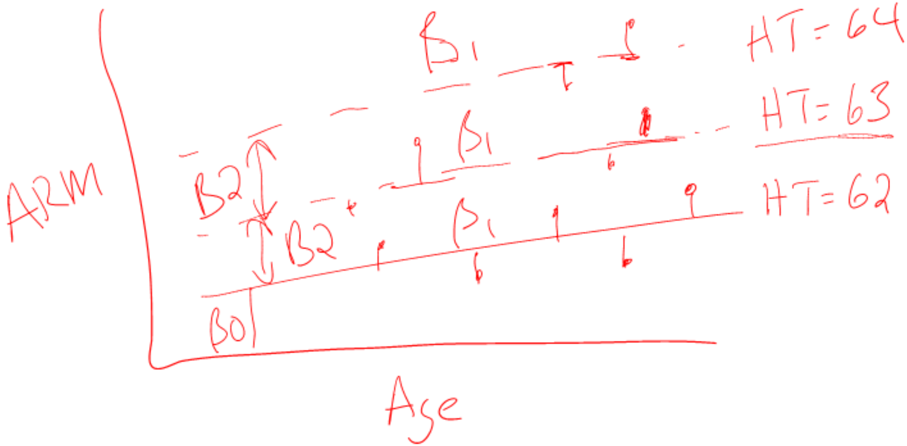
B1 = difference in mean ARM comparing two children who have the same sex but differ in age by 1 month

B2: Difference in mean ARM comparing females to males who are the same age

Review of key concepts from Lecture 3 and 4

- Height adjusted relationship between ARM and age
  - ARM = B0 + B1 (age – 6) + B2 (HT – 62) + e, e~N(0,σ^2), independent

↳ 62 cm = average HT among 6 month olds



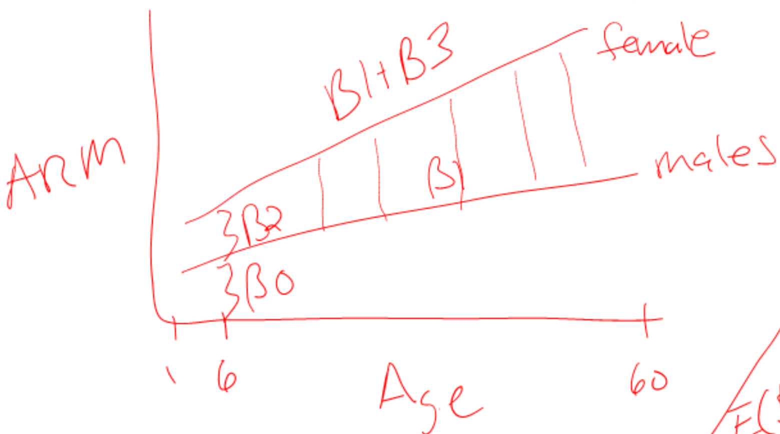
$Var(\hat{\varepsilon}_i)$

Review of key concepts from Lecture 3 and 4

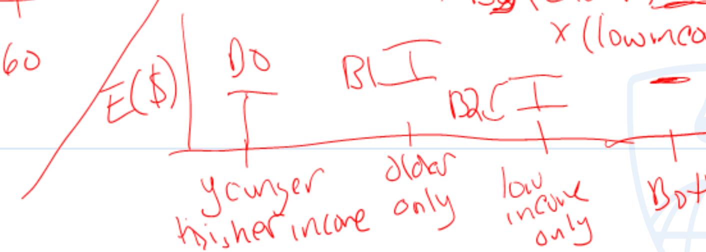
- Effect modification: Is the ARM vs. age relationship the same or different by sex
  - ARM = B0 + B1 (age – 6) + B2 Female + B3 (age – 6) Female + e, e~N(0,σ^2), independent

male: ARM = B0 + B1 (age − 6) + e

female: ARM = (B0 + B2) + (B1 + B3) (age − 6) + e



E($) = B0 + B1 (older) + B2 (lowincome) + B3 (older) x (lowincome)



Multiple Linear Regression Model

- Y is a random variable representing the outcome of interest in the population

y = observations

- The explanatory variables, $X_1, X_2, \ldots, X_p$ are fixed/known (not random or measured with error)

- Sample of size n is observed, data are: $\underbrace{(y_i, X_{1i}, X_{2i}, \ldots, X_{pi})}$

$$\rightarrow \quad Y_i = \underbrace{\mu_i(\beta, X_i)}_{\text{systematic}} + \varepsilon_i \; - \text{random component}$$

$E(Y_i \mid X_i)$

- X is the design matrix

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & & X_{p2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & & X_{pn} \end{bmatrix}$$

- X_i is the row of the design matrix corresponding to subject i

$X_n \leftarrow$

Multiple Linear Regression Model

$$Y_i = \mu_i(\underset{\sim}{\beta}, X_i) + \varepsilon_i \qquad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p+1 \times 1}$$

- Systematic component:
  - $\mu_i(\underset{\sim}{\beta}, \underline{X_i}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}$
- ε_i is the random components: $\varepsilon_i \sim N(0, \sigma^2)$, $Cov(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$
- The least squares solution finds the values of $\underset{\sim}{\beta}$ that minimize:

$$\sum_{i=1}^{n} \left(y_i - \underbrace{\mu_i(\underset{\sim}{\beta}, X_i)} \right)^2$$

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_p X_{pi} \right)^2$$

Least squares solution: simple linear regression

SLR = $p = 1$ $\qquad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$\frac{\text{covariance between } y \text{ and } X}{\text{variance in } X}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$$

Standardize Y and X to have mean 0 and SD = 1

$\hat{\beta}_1 = r$ Pearson ~~correla~~ correlation coefficient

Maximum likelihood inference in MLR

- Start with the MLR: $Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \varepsilon_i$

  $\varepsilon_i \sim N(0, \sigma^2)$ and independent

  Data: (y_i, X_i) for $i = 1, \dots, n$

- Other notation:

  Y_i = RV $\quad y_i$ = observation

  $\underset{\sim}{Y}$ = vector of RVs $\quad \underset{\sim}{y}$ = vector of observations $\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$

  X_i = ith row of design matrix X

  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta} X_{pi}$

  $\hat{R}_i = Y_i - \hat{Y}_i$

Likelihood function definition

- Model: systematic component $\mu_i(\beta, X_i)$ plus random component $\varepsilon_i \sim N(0, \sigma^2)$

$$Y_i \sim N(\mu_i(\beta, X_i), \sigma^2)$$

- Probability density function:

$$f(\underset{\sim}{y} \mid \underset{\sim}{\mu}(\beta, X), \sigma^2) = \prod_{i=1}^{n} f(y_i \mid \mu_i(\beta, X_i), \sigma^2)$$

↳ a function of $\underset{\sim}{y}$ given fixed mean $\mu_i(\beta, X_i)$ and variance σ^2

- Likelihood function:

$$L(\underset{\sim}{\mu}(\beta, X), \sigma^2 \mid \underset{\sim}{y}) = \prod_{i=1}^{n} L(\mu_i(\beta, X_i), \sigma^2 \mid y_i)$$

↳ a function of $\mu_i(\beta, X_i)$ and σ^2 given the observed data y_i

Maximum likelihood estimation under gaussian residuals

- Likelihood function $Y_i \sim \mathcal{N}\left(\mu_i(\beta, X_i), \sigma^2\right)$

$$L(\beta, \sigma^2 \mid y) = \prod_{i=1}^{n} L\left(\mu_i(\beta, X_i), \sigma^2 \mid y_i\right)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \underline{\mu_i(\beta, X_i)}\right)^2\right)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \beta_0 - \beta_1 X_{1i} - \dots - \beta_p X_{pi}\right)^2\right)$$

Maximum likelihood estimation under gaussian residuals

- Log Likelihood Function

$$\ell(\beta, \sigma^2 \mid y) = \log L(\beta, \sigma^2 \mid y)$$

$$= \sum_{i=1}^{n} \left(-\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2} \underbrace{\left(y_i - \mu_i(\beta, X_i)\right)}^2 \right)$$

= To find $\hat{\beta}$ and $\hat{\sigma}^2$ that maximize $\ell(\beta, \sigma^2 \mid y)$

w take the derivative w/t β and σ^2,

set the derivatives $= 0$ and solve for

β and σ^2

Maximum likelihood estimation under gaussian residuals

- Solution for β_j

$j = 0, \ldots, p$

$$\ell(\beta, \sigma^2 \mid y) = \sum_{i=1}^{n} \left(-\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}(y_i - \mu_i(\beta, X_i))^2 \right)$$

$$\ell(\beta, \sigma^2 \mid y) \propto \sum_{i=1}^{n} \frac{-1}{2\sigma^2}\left(y_i - \mu_i(\beta, X_i)\right)^2$$

as a function of β

Define a score equation for β_j

$$U_{\beta_j}(\beta \mid \sigma^2) = \frac{d}{d\beta_j} \ell(\beta, \sigma^2 \mid y)$$

$$= \frac{d}{d\beta_j} \sum_{i=1}^{n} \frac{-1}{2\sigma^2}\left(y_i - \beta_0 - \beta_1 X_{1i} - \ldots - \beta_p X_{pi}\right)^2$$

$$= \frac{-1}{2\sigma^2} \sum_{i=1}^{n} 2 \times \left(y_i - \mu_i(\beta, X_i)\right)(-X_{ij})$$

$= 0$ and solve for β_j

Maximum likelihood estimation under gaussian residuals

- Solution for β_j

Maximum likelihood estimation under gaussian residuals

- Solution for σ^2

MLEs for simple linear regression

MLEs for simple linear regression

MLEs for simple linear regression

Take away messages

Take away messages

Next time....

- Vector / Matrix representation of MLR
- Geometry of least squares
- Distribution of MLEs for regression parameters