Regression Analysis

Q: What are the goals of regression analysis?

A: Estimation, Testing, and Prediction

- Estimation of the effect of one variable (exposure), called the predictor of interest, after adjusting, or controlling for other measured variables
  - Remove confounding variables
  - Remove bias

- Testing whether variables are associated with the response

- Prediction of a response variable given a collection of covariates

### Example

Impact of maternal smoking on low birth

- Measured variables
  - birth weight (g)
  - maternal smoking (yes/no)
  - maternal age (yrs)
  - maternal weight at last menses (kg)
  - race
  - history of premature labor (yes/no)
  - history of hypertension

Response:

Predictor of interest:

Confounder:

Precision:
\textbf{Linear regression model}

- $y_{ij}, \ j = 1, \ldots, n, \ i = 1, \ldots, m$
- $t_j$ corresponding times at which the measurements are taken on each unit
- $x_{ijk}, \ k = 1, \ldots, p$ explanatory variables
  \[ y_{ij} = \beta_1 x_{i1j} + \beta_p x_{ipj} + \epsilon_{ij} \]
- in the classical linear regression model
  \[
  \epsilon_{ij} \sim N(0, \sigma^2), \ \text{cor}(\epsilon_{ij}, \epsilon_{ik}) = 0
  \]
- ordinary least square estimation

\begin{align*}
  y_1 &= \beta_1 x_{11} + \beta_p x_{1p} + \epsilon_1 \\
  y_2 &= \beta_1 x_{21} + \beta_p x_{2p} + \epsilon_2 \\
  \vdots &= \vdots \\
  y_m &= \beta_1 x_{m1} + \beta_p x_{mp} + \epsilon_m \\
  Y &= X\beta + \epsilon
\end{align*}

- $\epsilon_i \text{ iid } N(0, \sigma^2)$ if and only if $\epsilon \sim MVN(0, \sigma^2 I)$
- $[\text{cov}(\epsilon)] = \text{cov}(\epsilon_i, \epsilon_j) = \sigma^2 I$
- $\sigma^2 I = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$

\textbf{Review of linear model}

\[
  y_i = \beta_1 x_{i1} + \beta_p x_{ip} + \epsilon_i = x_i^T \beta + \epsilon \\
  x_i = (x_{i1}, \ldots, x_{ip}) \\
  \beta = (\beta_1, \ldots, \beta_p) \\
  \epsilon_i \sim N(0, \sigma^2), \ i = 1, \ldots, m, \ iid \\
  x_{i1} = 1 \text{ then } \beta_1 \text{ is the intercept} \\
  E(\epsilon_i) = 0, \ V(\epsilon_i) = \sigma^2 \\
  Y = X\beta + \epsilon \\
  \epsilon \sim MVN(0, \sigma^2 I) \\
  \epsilon = (\epsilon_1, \ldots, \epsilon_m)
\]

\textbf{Review of linear model}

- Linear model includes as special cases:
  1. the analysis of variance, $x_{ij}$ are dummy variables
  2. multiple regression, $x_{ij}$ are continuous
  3. the analysis of covariance, $x_{ij}$ are dummy and continuous variables
- $\beta$ is the expected value of the response variable $Y$, per unit change of its corresponding explanatory variable $x$, all other variables held fixed.
Estimation of $\beta$

- Principle of maximum likelihood
  RA Fisher 1925
- Estimate $\beta$ by the values that make the observations maximally likely
- Likelihood $P(\text{data} \mid \beta)$ - as a function of $\beta$

Likelihood Inference

- Likelihood Inference is based on the specification of the probability density for the observed data
  \[ L(\theta \mid y) = f(y; \theta) \]
- Maximum likelihood estimate $\hat{\theta}$ is defined as
  \[ L(\theta \mid y) \leq L(\hat{\theta} \mid y) \]
- $\hat{\theta}$ is then regarded as the value of $\theta$ which is most strongly supported by the observed data
- $\theta$ is obtained by direct maximization of $L(\theta)$ or $\log L(\theta)$ by solving
  \[ S(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} = 0 \]
- $S(\theta)$ is called score equation for $\theta$

Linear Model – IID Normal Errors

\[
L(\beta; Y) = \prod_{i=1}^{m} g(y_i; \beta, \sigma^2)
= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y_i - x'_i \beta)^2}{2\sigma^2} \right)
\]

Maximizing likelihood = maximizing log likelihood

\[
\log L = l(\beta, \sigma^2, Y) = \quad \quad \quad = \quad \quad \quad = \sum_{i=1}^{m} \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y_i - x'_i \beta)^2}{2\sigma^2} \right\}
= c(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - x'_i \beta)^2
= c(\sigma^2) - SS(\beta)
\]

Maximizing normal likelihood = minimizing sum of squares

Geometry of least squares

\[
x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{m1} \end{pmatrix}, \ldots, x_p = \begin{pmatrix} x_{1p} \\ \vdots \\ x_{mp} \end{pmatrix}
\]

\[ X = (x_1, x_2, \ldots, x_p), \dim(X) = m \times p \]

\[ X\beta = x_1\beta_1 + x_2\beta_2 + \ldots + x_p\beta_p \]

\[ SS(\beta) = \sum_{i=1}^{m} (y_i - x'_i \beta)^2 = (Y - X\beta)'(Y - X\beta) = |Y - X\beta|^2 \]

$SS(\beta)$ is the LENGTH$^2$ of the residual vector $(Y - X\beta)$. Choose $\hat{\beta}$ so that the distance from $Y$ to $X\beta$ is as small as possible!
Methods of Least Squares

\( \hat{\beta} \) that minimizes the length of residual vector

\[ Y - X \hat{\beta} \text{ makes } Y - X \hat{\beta} \text{ orthogonal to } x_1, \ldots, x_p \]

\[
\begin{align*}
  \rightarrow X'(Y - X\hat{\beta}) &= 0 \\
  \rightarrow X'Y - X'X\hat{\beta} &= 0 \\
  (X'X)\hat{\beta} &= X'Y \\
  \hat{\beta} &= (X'X)^{-1}X'Y \\
  \hat{Y} &= X\hat{\beta} = X(X'X)^{-1}X'Y \\
  &= HY \\
  H \cdot H &= H \text{ (you check)} \\
\end{align*}
\]

\( H \) project \( Y \) onto space spanned by columns of \( x \)

Distribution Theory for Linear model

\[ Y = X\beta + \epsilon, \quad \epsilon \sim MVN(0, \sigma^2 I) \]

\[ \hat{\beta} = (X'X)^{-1}X'Y \]

\[
\begin{align*}
  E\hat{\beta} &= E \left\{ (X'X)^{-1}X'Y \right\} = (X'X)^{-1}X'EY \\
  &= (X'X)^{-1}X'X\beta = \beta \\
\end{align*}
\]

\( \beta \) is unbiased

\[
\begin{align*}
  Var\hat{\beta} &= Var \left( (X'X)^{-1}X'Y \right) \\
  &= (X'X)^{-1}X'VarYX(X'X)^{-1} \\
  &= (X'X)^{-1}\sigma^2 \\
\end{align*}
\]

- \( E\hat{Y} = EHY = HEY = HX\beta = X\beta \)
- \( Var\hat{Y} = HVarYH' = \sigma^2H^2 = \sigma^2H \)
- \( E\hat{\epsilon} = E(Y - \hat{Y}) = X\beta - X\beta = 0 \)
- \( Var\hat{\epsilon} = (I - H)\sigma^2I(I - H)' = \sigma^2(I - H) \)

Statistical Properties of \( \hat{\beta} \)

- It is an unbiased estimator: \( E(\hat{\beta}) = \beta \)
- \( Var(\hat{\beta}) = \sigma^2(X'X)^{-1} \)
- For any vector \( \alpha \) of known coefficients, \( \phi = \alpha'\beta \) then \( \hat{\phi} = \alpha'\hat{\beta} \) has the smallest possible variances amongst all unbiased estimators for \( \phi \) which are linear combination of the \( Y_i \). (Gauss-Markov theorem)