Linear Regression Model (A review)

- Linear Regression Model (Review)
- Ordinary Least Squares (OLS)
- Maximum Likelihood estimation
- Distribution Theory

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

Regression Analysis

- **Classification of Variables**
- Response variable
  - ⋆ Dependent variable
  - ⋆ Outcome variable
- Predictor of interest (POI)
  - ⋆ Exposure variable
  - ⋆ Treatment assignment
- Confounding variables
  - ⋆ Associated with response and POI
  - ⋆ Not intermediate
- Precision variables
  - ⋆ Associated with response and not with POI
  - ⋆ Reduces response uncertainty

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:

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

$$y_{ij} = \beta_1 x_{ij1} + \quad \beta_p x_{ijp} + \epsilon_{ij}$$

- in the classical linear regression model

$$\epsilon_{ij} \sim N(0, \sigma^2),\ cor(\epsilon_{ij}, \epsilon_{ik}) = 0$$

- **ordinary least square estimation**

Review of linear model

$$\begin{aligned}
y_i &= \beta_1 x_{i1} + \quad + \beta_p x_{ip} + \epsilon_i \\
&= \boldsymbol{x}_i' \boldsymbol{\beta} + \boldsymbol{\epsilon}_i \\
\boldsymbol{x}_i &= (x_{i1}, \ldots, x_{ip}) \\
\boldsymbol{\beta} &= (\beta_1, \ldots, \beta_p) \\
\epsilon_i &\sim N(0, \sigma^2),\ i = 1, \ldots, m,\ iid \\
x_{i1} &= 1 \text{ then } \beta_1 \text{ is the intercept} \\
E(\epsilon_i) &= 0,\ V(\epsilon_i) = \sigma^2 \\
Y &= X\boldsymbol{\beta} + \boldsymbol{\epsilon} \\
\boldsymbol{\epsilon} &\sim MVN(0, \sigma^2 I) \\
\boldsymbol{\epsilon} &= (\epsilon_1, \ldots, \epsilon_m)
\end{aligned}$$

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

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

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

- *β 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 β

- Principle of maximum likelihood

  RA Fisher 1925

- *Estimate β by the values that make the observations maximally likely*
- Likelihood $P(\text{data} \mid \boldsymbol{\beta})$ - as a function of $\boldsymbol{\beta}$

Likelihood Inference

- Likelihood Inference is based on the specification of the probability density for the observed data

$$L(\boldsymbol{\theta} \mid \boldsymbol{y}) = f(\boldsymbol{y}; \boldsymbol{\theta})$$

- Maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is defined as

$$L(\boldsymbol{\theta} \mid \boldsymbol{y}) \leq L(\hat{\boldsymbol{\theta}} \mid \boldsymbol{y})$$

- $\hat{\boldsymbol{\theta}}$ is then regarded as the value of $\boldsymbol{\theta}$ which is most strongly supported by the observed data
- $\boldsymbol{\theta}$ is obtained by direct maximization of $L(\boldsymbol{\theta})$ or $\log L(\boldsymbol{\theta})$ by solving

$$S(\boldsymbol{\theta}) = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$$

- $S(\boldsymbol{\theta})$ is called score equation for $\boldsymbol{\theta}$

Linear Model – IID Normal Errors

$$\begin{aligned} L(\boldsymbol{\beta}; Y) &= \prod_{i=1}^{m} g(y_i; \boldsymbol{\beta}, \sigma^2) \\ &= \prod_{i=1}^{m} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2}{2\sigma^2} \right) \end{aligned}$$

Maximizing likelihood = maximizing log likelihood

$$\begin{aligned} \log L &= l(\boldsymbol{\beta}, \sigma^2, Y) = \\ &= \\ &= \sum_{i=1}^{m} \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2}{2\sigma^2} \right\} \\ &= \\ &= c(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 \\ &= \\ &= c(\sigma^2) - SS(\boldsymbol{\beta}) \end{aligned}$$

Maximizing normal likelihood = minimizing sum of squares

Geometry of least squares

$$\begin{aligned} \boldsymbol{x}_1 &= \begin{pmatrix} x_{11} \\ \vdots \\ x_{m1} \end{pmatrix}, \ldots, \boldsymbol{x}_p = \begin{pmatrix} x_{1p} \\ \vdots \\ x_{mp} \end{pmatrix} \\ X &= \left(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p \right), \text{dim}(X) = m \times p \\ X\boldsymbol{\beta} &= \boldsymbol{x}_1\beta_1 + \boldsymbol{x}_2\beta_2 + \ldots + \boldsymbol{x}_p\beta_p \\ SS(\boldsymbol{\beta}) &= \sum_{i=1}^{m} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 \\ &= (Y - X\boldsymbol{\beta})'(Y - X\boldsymbol{\beta}) \\ &= \mid Y - X\boldsymbol{\beta} \mid^2 \end{aligned}$$

$SS(\boldsymbol{\beta})$ is the LENGTH^2 of the residual vector $(Y - X\boldsymbol{\beta})$.

Choose $\hat{\boldsymbol{\beta}}$ so that the distance from Y to $X\boldsymbol{\beta}$ is as small as possible!

Methods of Least Squares

$\hat{\boldsymbol{\beta}}$ that minimizes the length of residual vector $Y - X\hat{\boldsymbol{\beta}}$ makes $Y - X\hat{\boldsymbol{\beta}}$ orthogonal to $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$

$$\begin{aligned}
\rightarrow X^{'}(Y - X\hat{\boldsymbol{\beta}}) &= 0 \\
\rightarrow X^{'}Y - X^{'}X\hat{\boldsymbol{\beta}} &= 0 \\
(X^{'}X)\hat{\boldsymbol{\beta}} &= X^{'}Y \\
\hat{\boldsymbol{\beta}} &= (X^{'}X)^{-1}X^{'}Y \\
\hat{Y} &= X\hat{\boldsymbol{\beta}} = X(X^{'}X)^{-1}X^{'}Y \\
&= HY \\
H \cdot H &= H \text{ (you check)}
\end{aligned}$$

H project Y onto space spanned by columns of x

Distribution Theory for Linear model

$$\begin{aligned}
Y &= X\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim MVN(0, \sigma^2 I) \\
\hat{\boldsymbol{\beta}} &= (X^{'}X)^{-1}X^{'}Y
\end{aligned}$$

$$\begin{aligned}
E\hat{\boldsymbol{\beta}} &= E\left\{(X^{'}X)^{-1}X^{'}Y\right\} = (X^{'}X)^{-1}X^{'}EY \\
&= (X^{'}X)^{-1}X^{'}X\boldsymbol{\beta} = \boldsymbol{\beta}
\end{aligned}$$

$\boldsymbol{\beta}$ is unbiased

$$\begin{aligned}
Var\hat{\boldsymbol{\beta}} &= Var\left((X^{'}X)^{-1}X^{'}Y\right) \\
&= (X^{'}X)^{-1}X^{'}VarY X(X^{'}X)^{-1} \\
&= (X^{'}X)^{-1}\sigma^2
\end{aligned}$$

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

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 $\boldsymbol{a}$ of known coefficients, $\phi = \boldsymbol{a}^{'}\beta$ then $\hat{\phi} = \boldsymbol{a}^{'}\hat{\beta}$ has the smallest possible variances amongst all unbiased estimators for ϕ which are linear combination of the Y_i. (*Gauss-Markov theorem*)