Linear regression review

Miaoyan Wang

Department of Statistics
UW Madison

Another view of T-test

- Recall the simple linear regression (SLR) model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \sim i.i.d.\ N(0, \sigma^2),$$

  for all $i = 1, \ldots, n$.

- Equivalently

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{MVN}(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

  where $\boldsymbol{X}_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$ denote the $n \times 2$ design matrix.

- One-sample test is a special case of SLR.
- Two-sample test is also a special case of SLR.

Equivalence to one-sample test

- Let
$$Y_i = \beta_0 + \varepsilon_i, \quad \varepsilon_i \sim i.i.d.\ N(0, \sigma^2),$$
for all $i = 1, \ldots, n$.
- Equivalently
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{MVN}(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$
where $\boldsymbol{X}_{n\times 1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ denote the $n \times 1$ design matrix, and $\boldsymbol{\beta} = \beta_0$.
- The one-sample mean test is equivalent to
$$H_0 : \beta_0 = \mu \text{ vs. } H_A : \beta_0 \neq \mu$$

Equivalence to two-sample test

- Let

$$Y_i = \beta_0 \mathbb{1}_{i \text{ is in group } 1} + \beta_1 \mathbb{1}_{i \text{ is in group } 2} + \varepsilon_i, \quad \varepsilon_i \sim i.i.d.\ N(0, \sigma^2),$$

  for all $i = 1, \ldots, n$.

- Equivalently

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{MVN}(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

  where $\boldsymbol{X}_{n\times 2} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}$ denote the $n \times 2$ design matrix, and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$.

- The unpaired two sample mean test is equivalent to

$$H_0 : \beta_0 - \beta_1 = 0 \text{ vs. } H_A : \beta_0 - \beta_1 \neq 0$$

Multiple Linear Regression Model

The multiple linear regression (MLR) model for the data $(x_{i1}, x_{i2}, \ldots, x_{i,p-1}, y_i)$ is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i,$$

for $i = 1, 2, \ldots, n$, where

- Y_i is the ith observation of the **response variable**.
- X_{ik} is the ith observation of the kth **explanatory variable** for $k = 1, \ldots, p-1$.
- ε_i is the ith **random error** term.
- The random errors follow a normal distribution with mean zero and variance σ^2 and are independent of each other.
- That is, $\varepsilon_i \sim \text{i.i.d. } N(0, \sigma^2)$.

Model Parameters

- The model parameters are $\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1}$, and σ^2 (population parameters).
- β_0 and $\beta_1, \beta_2, \ldots, \beta_{p-1}$: **regression coefficients**.
- β_0: **intercept**.
  β_0 interpreted as ____________________
- β_k: **slope** for $k = 1, \ldots, p-1$.
  β_k interpreted as____________________
- σ^2: **error variance**, sometimes written as σ_ε^2.

Q: How to estimate the model parameters based on data?

Example: $p = 3$

- Example: # of explanatory variables = 2.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \quad \varepsilon_i \sim \text{iid } N(0, \sigma^2),$$

  for $i = 1, \ldots, n$.

- Mean response:

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}.$$

- Interpretation:
  - β_0: Intercept. The mean response $\mathbb{E}(Y)$ at $X_1 = X_2 = 0$.
  - β_1: Slope. The change in the mean response $\mathbb{E}(Y)$ per unit increase in X_1, when X_2 is held constant.
  - β_2: Slope. The change in the mean response $\mathbb{E}(Y)$ per unit increase in X_2, when X_1 is held constant.

Models

- The relationship between the response variable Y and the explanatory variables $X_1, X_2, \ldots, X_{p-1}$ is

$$E(Y_i|\boldsymbol{X}_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} \qquad E(\varepsilon_i) = 0$$

- Equal variance:

$$Var(Y_i|\boldsymbol{X}_i) = Var(\varepsilon_i) = \sigma^2.$$

- Independence:

$$Cov(Y_i, Y_{i'}|\boldsymbol{X}_i, \boldsymbol{X}_{i'}) = Cov(\varepsilon_i, \varepsilon_{i'}) = 0 \quad \text{for} \quad i \neq i'.$$

- Normal distribution:

$$Y_i|\boldsymbol{X}_i \sim N(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1}, \sigma^2) \qquad \varepsilon_i \sim N(0, \sigma^2)$$

Models in matrix form

- Response variable: $\boldsymbol{Y}_{n\times 1} = (Y_1, Y_2, \ldots, Y_n)'$.
- Design matrix:

$$\boldsymbol{X}_{n\times p} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

- Random error: $\boldsymbol{\varepsilon}_{n\times 1} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$.
- Regression coefficients: $\boldsymbol{\beta}_{p\times 1} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$.
- The multiple linear regression model can be written as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}_{n\times 1}, \sigma^2 \boldsymbol{I}_{n\times n}).$$

Least Squares Estimation

- Consider the least-square cost:

$$Q(\beta) = (\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta).$$

- We have shown that the least squares estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y},$$

  assuming that the $p \times p$ matrix $\boldsymbol{X}'\boldsymbol{X}$ is invertible.

Fitted Values and Residuals

- Fitted values: $\hat{\boldsymbol{Y}} = (\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n)'$.
- Following the arguments for SLR in matrix terms, we have

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{Y}, \quad \text{where} \quad \boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.$$

  The matrix $\boldsymbol{H}$ is often referred to as the "hat matrix".
- Residuals: $\boldsymbol{e} = (e_1, e_2, \ldots, e_n)' \stackrel{\text{def}}{=} \boldsymbol{Y} - \hat{\boldsymbol{Y}}$. Sample quantities.
- To be distinguished from the model error $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n) \sim \mathcal{MVN}(0, \sigma^2 \mathbf{I})$, which are population quantities.
- Following the arguments for SLR in matrix terms, we have

$$\boldsymbol{e} \stackrel{\text{def}}{=} \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}.$$

Properties of the hat matrix $\boldsymbol{H}$

- $\boldsymbol{H}$ is symmetric and idempotent:
  $\boldsymbol{H}^2 = \boldsymbol{H}$, and $\text{Rank}(\boldsymbol{H}) = \text{Tr}(\boldsymbol{H}) = p$.
- $\boldsymbol{I} - \boldsymbol{H}$ is symmetric and idempotent:
  $(\boldsymbol{I} - \boldsymbol{H})^2 = \boldsymbol{I} - \boldsymbol{H}$, and $\text{Rank}(\boldsymbol{I} - \boldsymbol{H}) = \text{Tr}(\boldsymbol{I} - \boldsymbol{H}) = n - p$.

A geometric interpretation

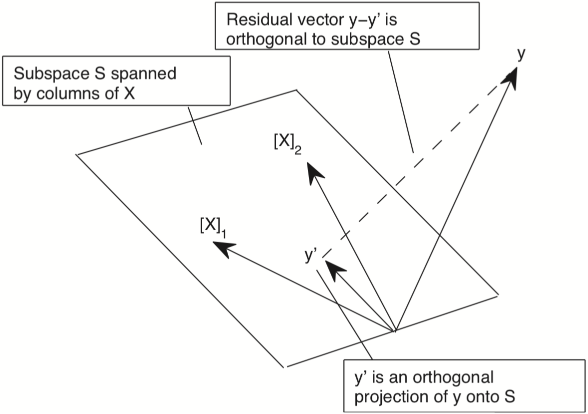
- Recall Least-square cost for linear regression:

$$Q(\beta) = (\boldsymbol{Y} - \beta \boldsymbol{X})'(\boldsymbol{Y} - \beta \boldsymbol{X})$$

- Normal equation (i.e. gradient):

$$\frac{\partial Q(\beta)}{\partial \beta} = 0 \to \boldsymbol{X}'(\boldsymbol{Y} - \beta \boldsymbol{X}) = 0$$

- Residual $\boldsymbol{e} = \boldsymbol{Y} - \hat{\beta}\boldsymbol{X}$ are orthogonal to columns of $\boldsymbol{X}$
- $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta}$ gives the "best" reconstruction of $\boldsymbol{Y}$ in the range of $\boldsymbol{X}$.
- Recall the range of a matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is the linear space $\subset \mathbb{R}^p$ spanned by the columns of $\boldsymbol{X}$.



- Recall "hat matrix": $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, and $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta} = \boldsymbol{H}\boldsymbol{Y}$
- $\boldsymbol{H}$ projects $\boldsymbol{Y}$ onto the span of $\boldsymbol{X}$.
- $\boldsymbol{I} - \boldsymbol{H}$ projects $\boldsymbol{Y}$ onto the space orthogonal to $\boldsymbol{X}$.

Estimation of Regression Coefficients

Distribution of regression coefficients estimates

$$\hat{\boldsymbol{\beta}} \sim \mathcal{MVN}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$$

- The LS estimate $\hat{\boldsymbol{\beta}}$ is an unbiased estimate of $\boldsymbol{\beta}$. That is,

$$\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}.$$

- The variance-covariance matrix is

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \in \mathbb{R}^{p \times p}$$

where

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \cdots & \text{Cov}(\hat{\beta}_0, \hat{\beta}_{p-1}) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) & \cdots & \text{Cov}(\hat{\beta}_1, \hat{\beta}_{p-1}) \\ \vdots & \vdots & \vdots & \\ \text{Cov}(\hat{\beta}_{p-1}, \hat{\beta}_0) & \text{Cov}(\hat{\beta}_{p-1}, \hat{\beta}_1) & \cdots & \text{Var}(\hat{\beta}_{p-1}) \end{bmatrix}$$

Inference of Regression Coefficients

- The estimated variance-covariance matrix.

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}})} = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- Marginally, we have

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\widehat{\mathrm{Var}(\hat{\beta}_k)}}} \sim T_{n-p}, \quad \text{for all } k = 0, 1, \ldots, p-1.$$

Inference of Regression Coefficients

- Thus the $(1-\alpha)$ confidence interval for β_k is

$$\hat{\beta}_k \pm t_{n-p,\alpha/2}\sqrt{\widehat{\mathrm{Var}(\hat{\beta}_k)}}.$$

- Hypothesis testing:

$$H_0 : \beta_k = \beta_k^0 \text{ versus } H_A : \beta_k \neq \beta_k^0.$$

- Under the H_0, we have

$$T^* = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\widehat{\mathrm{Var}(\hat{\beta}_k)}}} \sim T_{n-p}, \quad \text{Why } n-p?$$

Estimation of Mean Response

- Define a new observation with predictor $\boldsymbol{X}_h = (1, X_{h1}, \dots, X_{h,p-1})'$. Estimate $\mu_h = \mathbb{E}(\boldsymbol{X}_h'\boldsymbol{\beta} + \varepsilon_{\text{new}})$?
- The estimated mean response corresponding to $\boldsymbol{X}_h$:

$$\hat{\mu}_h = \boldsymbol{X}_h'\hat{\boldsymbol{\beta}}.$$

- Distribution of $\hat{\mu}_h$:

$$\hat{\mu}_h \sim N\left(\boldsymbol{X}_h'\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h\right)\right).$$

- Mean. ____________________
- Variance. ____________________

Confidence Intervals for Mean Response

- Estimated variance.

$$\widehat{\mathrm{SD}(\hat{\mu}_h)} = \hat{\sigma}\sqrt{\boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h}.$$

- The $(1-\alpha)$ confidence interval for $\hat{\mu}_h$ is

$$\hat{\mu}_h \pm t_{n-p,\alpha/2}\widehat{\mathrm{SD}(\hat{\mu}_h)}$$

- Hypothesis tests on μ_h can be carried out similarly.

Prediction of New Observation

- The predicted new observation corresponding to $\boldsymbol{X}_h$:

$$\widehat{Y}_h = \boldsymbol{X}_h' \hat{\beta}.$$

- What is the MSE of $\widehat{Y}_h$ for predicting $Y_{h(\mathrm{new})}$?
- Prediction error variance:

$$\mathrm{Var}(\widehat{Y}_h - Y_{h(\mathrm{new})}) = \sigma^2 \left(1 + \boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h\right).$$

- Distribution of $\widehat{Y}_h - Y_{h(\mathrm{new})}$:

$$\widehat{Y}_h - Y_{h(\mathrm{new})} \sim N\left(0, \sigma^2 \left(1 + \boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h\right)\right).$$

Prediction Intervals for New Observation

- The estimated prediction error variance is

$$\widehat{\sigma}_{\text{pred}} = \hat{\sigma}\sqrt{1 + \boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h}.$$

- The $(1-\alpha)$ prediction interval for $Y_{h(\text{new})}$ is

$$\widehat{Y}_h \pm t_{n-p,\alpha/2}\widehat{\sigma}_{\text{pred}}.$$

- Note that

$$\frac{\widehat{Y}_h - Y_{h(\text{new})}}{\widehat{\sigma}_{\text{pred}}} \sim T_{n-p}.$$