The Classical Linear Regression Model

The Classical Linear Regression

Least Square Estimation

Inferences About the Regression Model

Multivariate Multiple Regression

Multivariate Multiple Regression

Lecture 7 Multivariate Linear Regression

Shiwei Lan[1]

[1]School of Mathematical and Statistical Sciences
Arizona State University

STP533 Multivariate Analysis
Spring 2025

Table of Contents

1 The Classical Linear Regression Model
  - The Classical Linear Regression
  - Least Square Estimation
  - Inferences About the Regression Model

2 Multivariate Multiple Regression
  - Multivariate Multiple Regression

The Classical Linear Regression Model

The Classical Linear Regression

Least Square Estimation

Inferences About the Regression Model

Multivariate Multiple Regression

Multivariate Multiple Regression

Regression Analysis

- Regression analysis is the statistical methodology for predicting values of one or more *response* (dependent) variables from a collection of *predictor* (independent) variables.
- It can also be used for assessing the effects of the predictor variables on the responses.
- The name *regresion*, dated back to 1885 by F. Galton.
- We first review the classical linear regression model with a single response. Then we generalize to linear model fir several dependent variables.

The Classical Linear Regression

- Suppose we have p predictor variables $X_1, \cdots, X_p$ and a response variable Y.
- For example, Y=current market value of a house, X_1=square feet, X_2=location, X_3=appraised value of last year, and X_4=quality of construction.
- A classical linear regression relates the average value of Y with a linear combination of X_i's.

$$Y_i = \beta_0 + X_{i1}\beta_1 + \cdots + X_{ip}\beta_p + \epsilon_i, \quad i = 1, \cdots, n,$$

  where we assume $\epsilon_i \overset{iid}{\sim} (0, \sigma^2)$.
- If we denote $\mathbf{Y} = [Y_1, \cdots, Y_n]^T$, $\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$, and $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_p]^T$, then we can rewrite

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

One-Way ANOVA as A Regression

Example 7.2 (The design matrix for one-way ANOVA as a regression model) Determine the design matrix if the linear regression model is applied to the one-way ANOVA situation in Example 6.6.

We create so-called *dummy* variables to handle the three population means: $\mu_1 = \mu + \tau_1$, $\mu_2 = \mu + \tau_2$, and $\mu_3 = \mu + \tau_3$. We set

$$z_1 = \begin{cases} 1 & \text{if the observation is from population 1} \\ 0 & \text{otherwise} \end{cases} \qquad z_2 = \begin{cases} 1 & \text{if the observation is from population 2} \\ 0 & \text{otherwise} \end{cases}$$

$$z_3 = \begin{cases} 1 & \text{if the observation is from population 3} \\ 0 & \text{otherwise} \end{cases}$$

and $\beta_0 = \mu$, $\beta_1 = \tau_1$, $\beta_2 = \tau_2$, $\beta_3 = \tau_3$. Then

$$Y_j = \beta_0 + \beta_1 z_{j1} + \beta_2 z_{j2} + \beta_3 z_{j3} + \varepsilon_j, \qquad j = 1, 2, \ldots, 8$$

where we arrange the observations from the three populations in sequence. Thus, we obtain the observed response vector and design matrix

$$\underset{(8\times 1)}{\mathbf{Y}} = \begin{bmatrix} 9 \\ 6 \\ 9 \\ 0 \\ 2 \\ 3 \\ 1 \\ 2 \end{bmatrix}; \qquad \underset{(8\times 4)}{\mathbf{Z}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

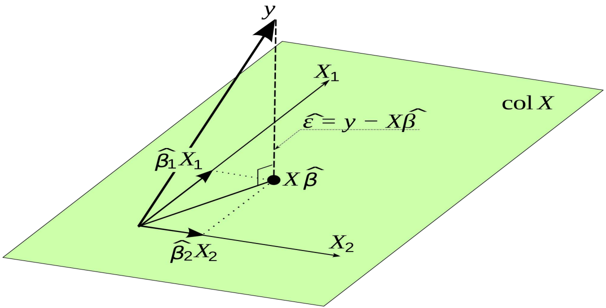
Least Square Estimation

- The least square estimation (LSE) minimizes the sum of square $S(\beta) = \|\mathbf{Y} - \mathbf{X}\beta\|_2^2$ with respect to β.
- Let $\mathbf{X}$ be full rank $p + 1 \leq n$. The LSE result of β is $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$.
- Let $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{H}\mathbf{y}$ be the *fitted values* of $\mathbf{y}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is called *hat matrix*.
- The residual vector can now be written

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

- The residual sum of squares becomes $S(\hat{\beta}) = \|\mathbf{e}\|_2^2 = \mathbf{y}^T(\mathbf{I} - \mathbf{H})\mathbf{y}$.

Sum-of-Squares Decomposition



- Note $\mathbf{X} \perp \mathbf{e}$ and $\hat{\mathbf{y}} \perp \mathbf{e}$. Why?
- Then we have $\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\mathbf{e}\|^2$.
- Further we have decomposition of the sum of squares about mean

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{SST} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{SSR} + \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{SSE}$$

Coefficient of Determination

- The decomposition of the sum of squares can also be written as $\mathbf{y}^T(\mathbf{I}-\mathbf{J})\mathbf{y} = \mathbf{y}^T(\mathbf{H}-\mathbf{J})\mathbf{y} + \mathbf{y}^T(\mathbf{I}-\mathbf{H})\mathbf{y}$.
- We define the *coefficient of determination* as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

- This quantity measure the proportion of the total variation in y's "explained" by the model with p predictors $\mathbf{X}$.
- If we plot $\hat{\mathbf{y}}$ against $\mathbf{y}$, what is the slope?

Sampling Properties of LSE

- We have the following property for LSE $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$

$$\mathrm{E}[\hat{\beta}] = \beta, \quad \mathrm{Cov}[\hat{\beta}] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

- The residual vector $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ has the following property

$$\mathrm{E}[\mathbf{e}] = \mathbf{0}, \quad \mathrm{Cov}[\mathbf{e}] = \sigma^2[\mathbf{I} - \mathbf{H}]$$

Inferences about the Regression Parameters

- Now we consider $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ with $\epsilon \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.
- Then the maximum likelihood estimator (MLE) of β is the same as LSE $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. Moreover, we have

$$\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$$

- The residual $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ is independent of $\hat{\beta}$ and $SSE/n = \|\mathbf{e}\|^2/n$ is the MLE of σ^2. Moreover,

$$\frac{\|\mathbf{e}\|^2}{\sigma^2} \sim \chi^2(n-p-1).$$

- $MSE = \frac{SSE}{n-p-1} = \frac{\|\mathbf{e}\|^2}{n-p-1} =: s^2$ is an unbiased estimator of σ^2.

Inferences about the Regression Parameters

- $100(1-\alpha)\%$ CR for $\boldsymbol{\beta}$ is determined by

$$(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X})^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) \leq (p+1)s^2F_{1-\alpha}(p+1, n-p-1).$$

- The $100(1-\alpha)\%$ SCI for β_j's are given by

$$\hat{\beta}_j \pm \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_j)}\sqrt{(p+1)F_{1-\alpha}(p+1, n-p-1)}, \quad j = 0, 1, \cdots, p.$$

where $\widehat{\mathrm{Var}}(\hat{\beta}_j)$ is the j-th diagonal element of $s^2(\mathbf{X}^T\mathbf{X})^{-1}$.

- For each β_j, the $100(1-\alpha)\%$ individual CI is

$$\hat{\beta}_j \pm t_{1-\alpha/2}(n-p-1)\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_j)}, \quad j = 0, 1, \cdots, p.$$

Likelihood Ratio Tests for the Regression Parameters

- Suppose you hypothesize that only the first $q \leq p$ predictors are significant in explaining the response variable.
- We want to test $H_0 : \beta_{1+1} = \beta_{1+2} = \cdots = \beta_p = 0$. Denote $\boldsymbol{\beta}_2 = [\beta_{q+1}, \cdots, \beta_p]^T$.
- We devide $\mathbf{X} = [(\mathbf{X}_1)_{n\times(q+1)} | (\mathbf{X}_2)_{n\times(p-q)}]$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T]^T$. Then

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \epsilon$$

- The LRT rejects $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ if

$$\frac{(SSE(\mathbf{X}_1) - SSE(\mathbf{X}))/(p-q)}{s^2} > F_{1-\alpha}(p-q, n-p-1).$$

Lecture 7

S.Lan

The Classical Linear Regression Model

The Classical Linear Regression

Least Square Estimation

Inferences About the Regression Model

Multivariate Multiple Regression

Multivariate Multiple Regression

Example 7.5 **(Testing the importance of additional predictors using the extra sum-of-squares approach)** Male and female patrons rated the service in three establishments (locations) of a large restaurant chain. The service ratings were converted into an index. Table 7.2 contains the data for $n = 18$ customers. Each data point in the table is categorized according to location (1, 2, or 3) and gender (male = 0 and female = 1). This categorization has the format of a two-way table with unequal numbers of observations per cell. For instance, the combination of location 1 and male has 5 responses, while the combination of location 2 and female has 2 responses. Introducing three dummy variables to account for location and two dummy variables to account for gender, we can develop a regression model linking the service index Y to location, gender, and their "interaction" using the design matrix

Table 7.2 Restaurant-Service Data

Location	Gender	Service (Y)
1	0	15.2
1	0	21.2
1	0	27.3
1	0	21.2
1	0	21.2
1	1	36.4
1	1	92.4
2	0	27.3
2	0	15.2
2	0	9.1
2	0	18.2
2	0	50.0
2	1	44.0
2	1	63.6
3	0	15.2
3	0	30.3
3	1	36.4
3	1	40.9

Table of Contents

1. The Classical Linear Regression Model
   - The Classical Linear Regression
   - Least Square Estimation
   - Inferences About the Regression Model
2. Multivariate Multiple Regression
   - Multivariate Multiple Regression

Multivariate Multiple Regression

- XXX